MASTERARBEIT | MASTER'S THESIS

Titel | Title

Paleogenomics of first settlers of Patagonia

verfasst von | submitted by

Doris Nguemo Emeruem B.Sc.

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology
Betreut von | Supervisor:	Ass.-Prof. Pere Gelabert Xirinachs PhD

List of relevant abbreviations

aDNA = ancient DNA

MtDNA = mitochondrial DNA

NDNA = Nuclear DNA

eDNA = Enviromental DNA

SNP = Single Nucleotide Polymorphism

PCA = Principal Components Analysis

AADR = Allen Ancient DNA Resource

List of figures and tables

Figures

Tables

Table of Contents

Abstract

The Baño Nuevo 1 archaeological site (Coyhaique, Chile) is in Northern Patagonia, in a key migration corridor in South America. The site offers compelling evidence of animal activity as far back as 16,500 years ago and contains a rich record of human occupation, with more than 20,000 faunal remains and 2,100 lithic artifacts. Human presence at the site occurred across multiple phases, with the earliest dated to approximately 11,000 years ago and the most recent around 2,700 years ago, providing valuable insights into the early human peopling of the Southern Cone. While previous research has focused on material culture, isotopic analysis, and burial practices *(Garcia et al., 2012; Lopez, 2014),* no genomic studies have been conducted on the ten Pleistocene burials recovered at the site. As a result, genetic studies of this population remain limited. Addressing this gap, this thesis presents the first comprehensive paleogenomic analysis of ten individuals from Baño Nuevo, employing high-throughput sequencing to examine mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) to reconstruct population history, genetic affinities, kinship relationships, and metagenomic profiles. Mitochondrial haplogroup classification for six individuals reveals haplogroups commonly found among ancient Indigenous American populations, which is the result of migration and regional gene flow *(Torres et al., 2018).* The newly generated genetic data indicate that the individuals from Baño Nuevo 1 show a genetic profile differentiated to the ones from modern populations from Patagonia. This study underscores the importance of integrating paleogenomic, archaeological, and bioarchaeological evidence to reconstruct past human movements, social structures, and microbial ecosystems. Future research should expand the dataset to include additional individuals and apply higher-resolution genomic analyses to further refine these findings.

Keywords: haplogroup, aDNA, human migration, geneflow, mtDNA, nDNA.

Zusammenfassung

Die archäologische Stätte Baño Nuevo 1 (Coyhaique, Chile) liegt in Nordpatagonien, in einem wichtigen Migrationskorridor in Südamerika. Die Stätte bietet überzeugende Beweise für tierische Aktivitäten, die bis vor 16 500 Jahren zurückreichen, und enthält mit mehr als 20 000 tierischen Überresten und 2 100 lithischen Artefakten ein reichhaltiges Zeugnis der menschlichen Besiedlung. Die menschliche Besiedlung der Stätte erstreckte sich über mehrere Phasen, wobei die früheste auf etwa 11.000 Jahre und die jüngste auf etwa 2.700 Jahre datiert wird, was wertvolle Einblicke in die frühe menschliche Besiedlung des Südkegels ermöglicht. Während sich frühere Forschungen auf die materielle Kultur, Isotopenanalysen und Bestattungspraktiken konzentrierten *(Garcia et al., 2012; Lopez, 2014),* wurden an den zehn pleistozänen Bestattungen, die an der Fundstelle geborgen wurden, keine genomischen Studien durchgeführt. Infolgedessen sind die genetischen Studien zu dieser Population nach wie vor begrenzt. Um diese Lücke zu schließen, wird in dieser Arbeit die erste umfassende paläogenomische Analyse von zehn Individuen aus Baño Nuevo vorgestellt. Dabei wird die mitochondriale DNA (mtDNA) und die nukleare DNA (nDNA) mittels Hochdurchsatz-Sequenzierung untersucht, um die Populationsgeschichte, die genetische Verwandtschaft, die Verwandtschaftsbeziehungen und die metagenomischen Profile zu rekonstruieren. Die Klassifizierung der mitochondrialen Haplogruppen von sechs Individuen zeigt Haplogruppen, die häufig bei alten indigenen amerikanischen Bevölkerungen zu finden sind, was das Ergebnis von Migration und regionalem Genfluss ist *(Torres et al., 2018).* Die neu generierten genetischen Daten deuten darauf hin, dass die Individuen aus Baño Nuevo 1 einen genetischen Zusammenhang aufweisen.

Chapter 1

Introduction

1.1. Background of study

The study of ancient human populations has been significantly advanced by the integration of paleogenomics into the traditional anthropological and archaeological studies, which allows researchers to reconstruct past migration patterns, genetic diversity, and population structure. The Patagonia region (South America geographical area south from parallel -40º) (Fig 1) was the latest continental area to be settled by humans at the Late Pleistocene. The region boasts remarkable diversity in climates and landscapes *(Mendez et al., 2024)*. The oldest evidence of the first human occupation is found 14,000 years ago in the Northern part (Monte Verde) (Dillehay et al., 2015), and 13,000 years ago in the Southern part of Patagonia such as at Fell's Cave and Pali-Aike *(Metcalf et al., 2016; Steele & Politis, 2009).* Genetic studies from ancient individuals reveal a degree of continuity throughout the Holocene. Nakatsuka et al. 2020 demonstrated partial genetic continuity from at least 6,600 years ago, with limited gene flow events in the late Holocene. Mitochondrial DNA analyses also support this continuity *(Postillone et al., 2017; Moraga et al., 2004),* linking ancient haplotypes with present-day Indigenous populations of Patagonia. The absence of Pleistocene individual genomes complicates the study of pre-Holocene populations

The Baño Nuevo 1 archaeological site, located in Central West Patagonia within the Chilean Region of Aysén, offers some of the most detailed archaeological records for early human activity in the region. Faunal remains from stratigraphic unit 5 (SU5) suggest animal presence, including extinct megafauna, as early as 16,500 years ago, preceding any human occupation *(López et al., 2011; Mena & Stafford, 2000).* Human presence is first detected in stratigraphic unit 4 (SU4), dated to around 11,000 years ago, continuing intermittently until approximately 2,700 years ago *(Méndez et al., 2024)*. A significant discovery at Baño Nuevo 1 includes the remains of ten individuals, radiocarbon dated to 10,200 calibrated years BP. These represent the earliest known concentration of multiple individuals in Patagonia *(Méndez et al., 2024).* Faunal analysis across stratigraphic layers indicates a marked shift in animal presence, extinct megafauna is only found in SU5 and disappear from the archaeological record by around 13,300 years ago *(López et al., 2011).* Additionally, bone tools recovered from the site demonstrate purposeful faunal exploitation, including fox bones repurposed into awls and other instruments *(Méndez et al., 2024).* The site has yielded over 20,000 faunal remains and more than 2,100 lithic artifacts, underscoring its importance in understanding the peopling and adaptation strategies of early Patagonians.

The Baño Nuevo 1 archaeological site is in a cold environment suitable for organic preservation. This preservation allows the extraction of ancient DNA (aDNA) for genetic analyses. While past research has focused on archaeological, isotopic, and osteological findings *(Garcia et al., 2012; Lopez, 2014),* comprehensive genomic studies have not yet been carried out on the material.

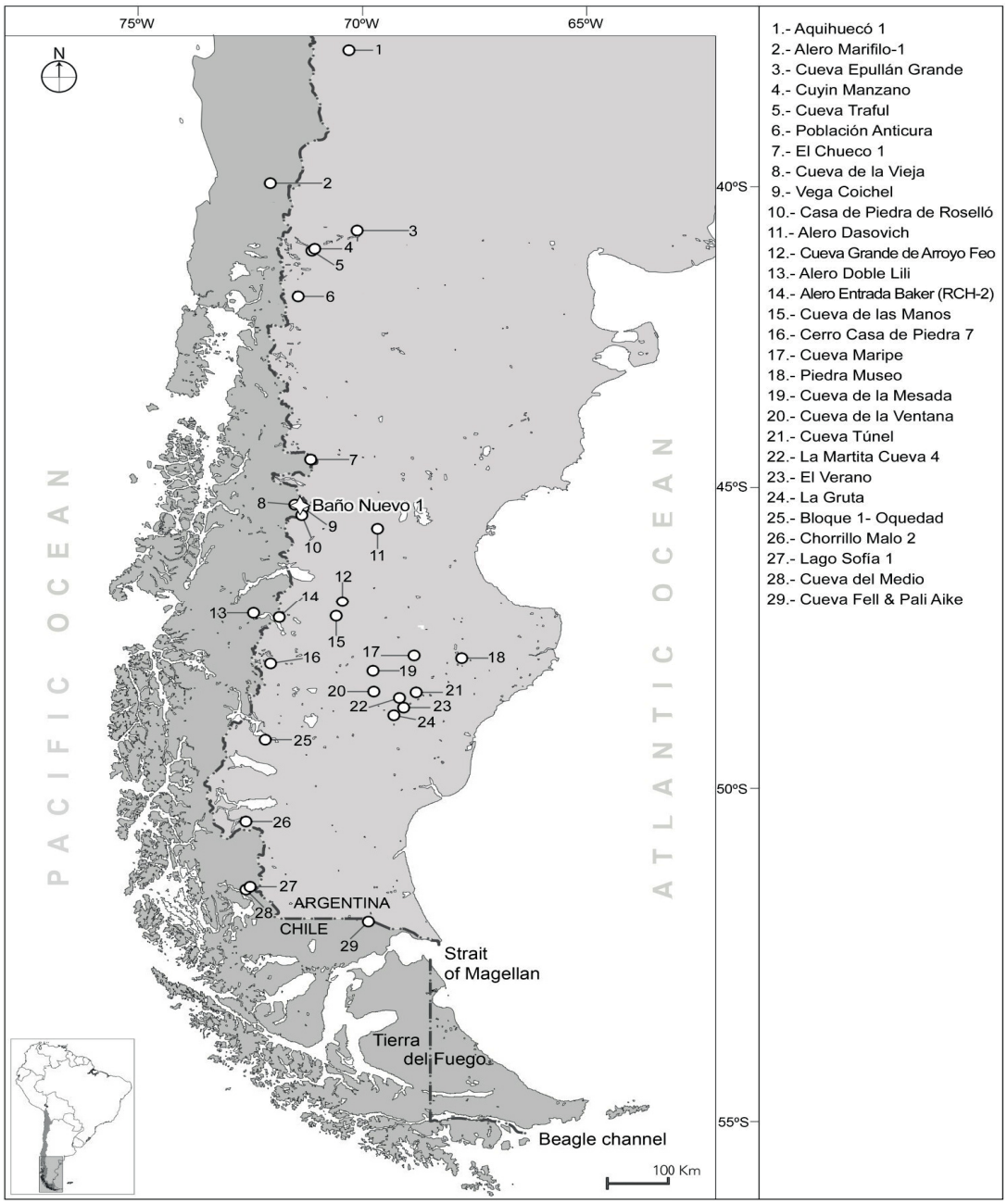

***Fig.1. Map of the Pleistocene sites in Patagonia,* indicating the location of Baño Nuevo 1** (close to 8,9,10), extracted from *Mendez et al 2024.*

Ancient DNA research has transformed our understanding of early human populations in the Americas. Genomic studies of prehistoric individuals have revealed complex migration events, genetic continuity, and admixture patterns *(Lazaridis et al., 2014; Skoglund et al., 2015).* However, gaps remain in South American populations, particularly in regions such as Baño Nuevo, where genetic relationships with contemporaneous and earlier groups remain unexplored.

Baño Nuevo 1 site has garnered considerable attention due to its well-preserved stratigraphy and diverse burial assemblages that offer insights into the cultural and demographic dynamics of past populations *(Mendez et al., 2024)* Early archaeological surveys emphasized its role as a regional hub where various cultural influences intersected, as reflected in the material culture and burial customs recovered during the

initial excavations *(Méndez et al., 2024)*. Notably, the site contains a group of at least ten human individuals dated to approximately 10,200 calibrated years before present (cal BP), representing one of the earliest directly dated bioanthropological assemblages in Patagonia and South America. The site's location in a region of known migratory routes has also spurred interest in understanding how gene flow and population movements contributed to the genetic landscape of its inhabitants. Baño Nuevo 1 site is one of the most significant Pleistocene archaeological locations in Patagonia, offering a crucial window into the prehistoric population dynamics of early human groups. Situated in a geographically strategic area that historically functioned as a crossroads for migration and cultural interaction, Baño Nuevo has yielded a wealth of human remains, artifacts, and environmental data that collectively illuminate the lifeway of its ancient inhabitants.

The present study employs high-throughput sequencing techniques to analyse mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) from the ten individuals recovered at Baño Nuevo 1 site. By incorporating genetic, archaeological evidence, this research aims to reconstruct the population history, genetic affiliations, kinship relationships, and metagenomics of these individuals.

1.2. Initial Genetic Studies and Population Structure Hypotheses

Prior to the advent of ancient DNA (aDNA) research, the biological relationships among individuals from Baño Nuevo 1 and the rest of archaeological American remains were inferred through cranial morphology and dental metric analyses (*Perez & Muñoz, 2009*). These studies suggested a degree of biological continuity between individuals from different burial clusters but also noted the presence of traits that indicated external gene flow.

More recently, mitochondrial DNA (mtDNA) analyses have been conducted on a small subset of individuals from the Americas and Baño Nuevo 1, revealing haplogroups commonly associated with early Indigenous populations of the Americas (*Torres et al., 2018*). These preliminary results suggest that the inhabitants of Baño Nuevo were part of broader population movements that likely included gene flow from neighbouring regions in the broad North-South American peopling.

1.3. Previous Excavations and Findings

Archaelogical investigations at Baño Nuevo began in the early 20th century, with initial excavations focused on stratigraphy and material culture. Early reports documented a rich burial assemblage, including human skeletal remains, grave gods and lithic artifacts indicative of long-distance trade or interaction networks *(Lopez, 1978).* These preliminary findings suggested that Baño Nuevo was not an isolated settlement but rather a site of cultural convergence.

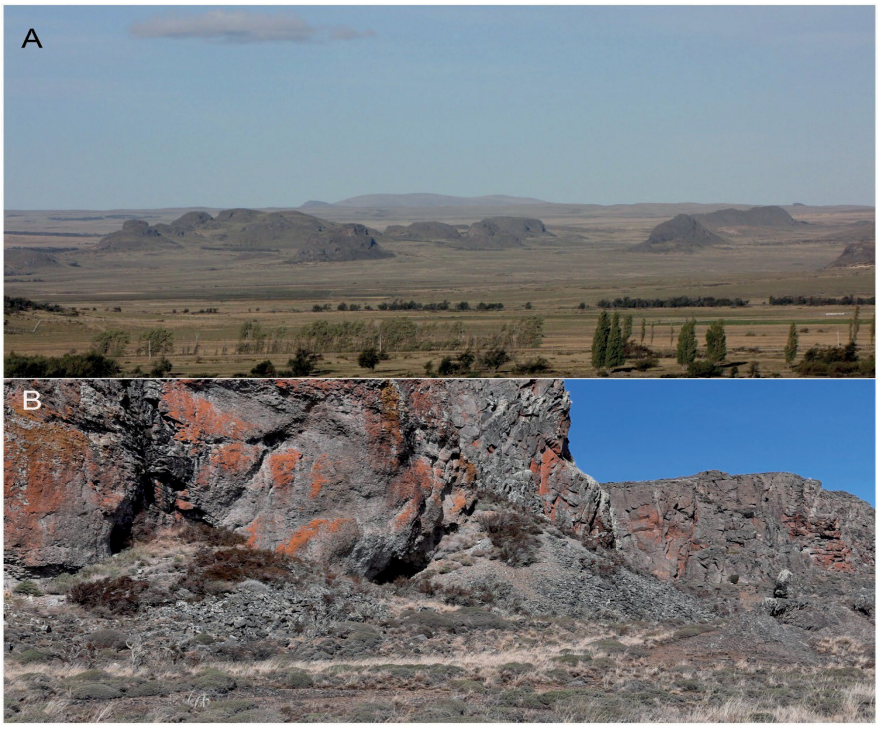

Fig. 2. Landscape around Baño Nuevo 1 archaeological site. (**A**) The post-glacial landscape of Central West Patagonia, showcasing rolling hills and expansive plains shaped by Late Pleistocene glacial activity. (**B**) A close-up view of the exposed rock formations near the site, *(Adapted from Méndez et al. 2024)*

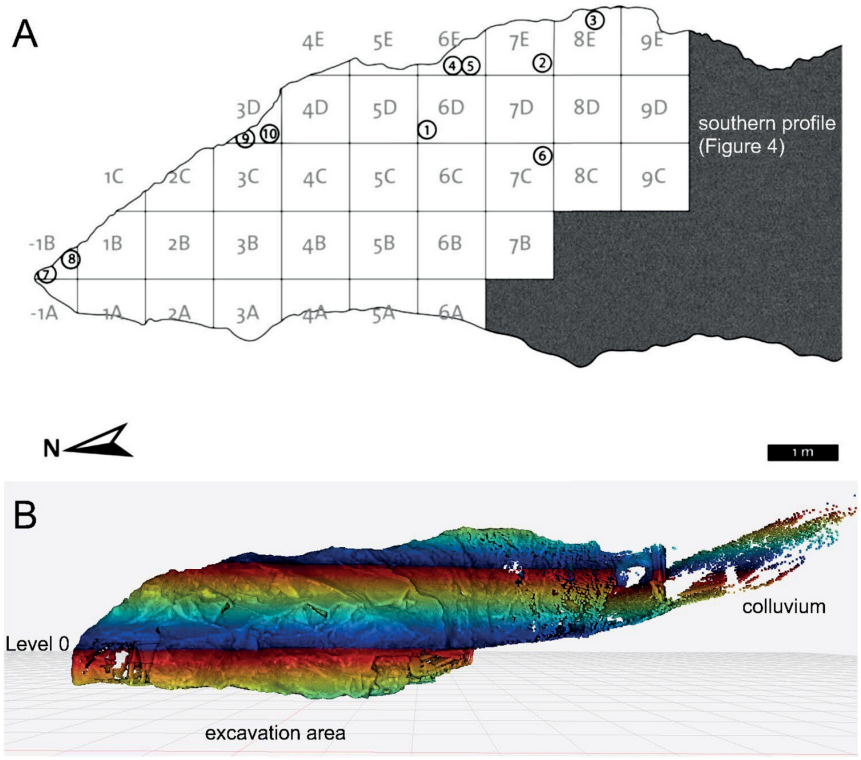

Fig. 3. Map of the Archaeological site of Baño Nuevo 1. **(A)** Site grid showing the excavation layout of the site. The squares represent the archaeological squares and numbers in circles the locations of human burials, **(B)** 3D digital terrain model of the excavation site. The colour gradient represents elevation changes, with

higher areas in warmer tones (reds/oranges) and lower areas in cooler tones (blues/greens). extracted from Mendez et al 2024.

1.3.1. Early Archaeological Investigations

The first systematic excavations at Baño Nuevo focusing primarily on the site's stratigraphy and material culture documented:

- A rich burial assemblage, including extended burials and secondary interments.
- Lithic tools and ceremonial objects, suggesting a complex sociocultural system.
- Evidence of trade networks, with materials such as obsidian and marine shells indicating long-distance interactions *(Lopez, 1978).*

1.3.2. Radiocarbon Dating and Chronological Framework

Subsequent excavations refined the site's chronology using radiocarbon dating. AMS (Accelerator Mass Spectrometry) dating of skeletal remains and associated artifacts suggests that the site was occupied between 11,200 and 2,700

- The earliest burials (10,2000 cal yr BP) exhibit simple grave structures with few grave goods.
- By ~6000 BP, more elaborate cultural practices appear, indicating increased social complexity.
- Between 5000 and 4000 BP, evidence of population mobility and external influences is observed in both material culture and genetic diversity.

These findings suggest that Baño Nuevo 1 was a dynamic settlement with evolving social traditions over time following the regional dynamics.

1.3.3. Material Culture and Trade Networks

Artifact analysis indicates that Baño Nuevo 1 was part of a broader exchange system.

- Obsidian sourcing studies confirm that some artifacts originated from distant volcanic regions, suggesting long-distance trade *(Vasquez et al., 2016)*.
- Ceremonial items, including carved bone pendants and red ochre deposits, hint at ritualistic behaviours *(Garcia et al., 2012)*.
- Marine shell ornaments, likely obtained through exchange, further indicate connections with coastal communities *(Rodriguez et al., 2014)*.

Until recently, molecular data were sparse and full genome-wide aDNA analysis revealing the genetic affinities among individuals from the site have yet to be explored, leaving open questions regarding the fine-scale genetic structure, patterns of kinship, and metagenomic classification, Previous archaeological and bioarchaeological studies have provided insights into burial practices, subsistence strategies, and cultural exchanges *(Garcia et al., 2012; Lopez, 2014)*. However, genetic analyses have primarily focussed on mitochondrial DNA (mtDNA) haplogroups without a broader nuclear DNA (nDNA) assessment.

1.4. Significance of the Research

The study of ancient DNA (aDNA) has revolutionized our understanding of prehistoric human migration, population interactions, and genetic structure of past societies. By integrating genomic data with

archaeological and bioarchaeological evidence, researchers can reconstruct past human movements, identify genetic links between populations, and infer cultural exchanges as well as infer in there're life conditions. This study is particularly significant for several reasons:

1.4.1. Advancing Ancient DNA Research in South America

Despite major advances in paleogenomics, South America remains underrepresented in ancient DNA datasets *(Posth et al., 2018)*. The genetic history of early populations in this region is complex, with multiple waves of migration and possible interactions between distinct groups *(Moreno-Mayar et al., 2018)*. By analysing Baño Nuevo 1 individuals, this study fills an important gap in the understanding of prehistoric South American populations. The study reports the oldest genomic analyses for collective burials in the South America Pleistocene as well as one of the only available genomic data from Pleistocene Patagonia.

1.4.2. Understanding Migration and Genetic Continuity

Baño Nuevo 1 is in a region known to have been a major corridor for human migrations, especially relevant in America in N-S direction. By analysing mtDNA haplogroups and nuclear genomic diversity, this study aims to determine whether the Baño Nuevo individuals share genetic ancestry with earlier populations or represent a distinct genetic lineage. The findings will contribute to debates on population continuity versus population replacement in the region.

1.4.3. Kinship and Social Organization

Few studies have investigated kinship structure in prehistoric South American populations will help determine the degree of relatedness among individuals buried at Baño Nuevo. These results can provide insights into family structure, burial customs, and social organization. This site presents the unique opportunity to do kinship in Pleistocene context in Patagonia.

1.4.4. Contributions to Ancient Health and Disease Studies

Ancient DNA analysis extends beyond human genetics to encompass paleopathology. The identification of microbial DNA from skeletal remains can provide evidence of past infections and commensal microbial communities. This study employ Kraken2 metagenomic analysis to identify ancient microbial communities, shedding light on possible disease ecology in early populations.

Research Questions and Objectives

This study is designed to answer the following key research questions:

- What do DNA damage patterns reveal about sample preservation and authenticity?
- What mitochondrial haplogroups are present in the individuals from Baño Nuevo, and how do they compare with known ancient and modern populations?
- Can PCA analysis and f_3 statistics help identify genetic affinities between Baño Nuevo individuals and other prehistoric populations?
- Do patterns of relatedness suggest familial ties among individuals buried at the site?
- Can microbial DNA analysis using Kraken detect ancient pathogens.

To address these research questions, the study has the following objectives:

- Extract and sequence ancient DNA from 10 individuals recovered from Baño Nuevo1 site.
- Perform rigorous quality control measures using FastQC and Qualimap to ensure data reliability.
- Assess post-mortem DNA damage using mapDamage to evaluate sample integrity.
- Conduct contamination checks using Schmutzi to validate the authenticity of the ancient DNA sequences.
- Analyse mtDNA to classify haplogroups and infer maternal ancestry.
- Perform genome-wide SNP analysis to assess nuclear DNA diversity and relatedness.
- Use PCA analysis and f_3 statistics to compare genetic data with reference populations.
- Utilize Kraken for microbial profiling to investigate ancient pathogenic DNA.

Chapter 2

The Archaeological and Bioarchaeological Context of Baño Nuevo 1

2.1 Geographic and Environmental Setting

2.1.1 Location and Landscape

Baño Nuevo 1 is Located within a transitional ecological zone, it features a combination of arid landscapes, river valleys, and access to coastal resources, making it an ideal location for prehistoric human occupation *(Méndez et al., 2024).* Geological studies indicate that the site is part of a larger network of settlements that existed along seasonal water sources and terrestrial migration routes *(Garcia et al., 2012).* The presence of freshwater sources in the vicinity, including underground aquifers and periodic river systems, may have supported long-term human occupation and subsistence strategies *(Rodriguez et al., 2014)*.

2.1.2 Paleoenvironmental Reconstructions

Climatic shifts in the early and middle Holocene influenced population movements in the region. Paleoenvironmental studies suggest that the area experienced periodic episodes of increased aridity, interspersed with more humid climatic conditions, which would have impacted the availability of resources and human settlement patterns *(Martinez et al., 2005).*

Early Holocene (~10,000–8000 BP): Characterized by fluctuating climatic conditions, with some evidence of increased moisture that may have facilitated early settlement.
Middle Holocene (~8000–5000 BP): Marked by increasing aridity, possibly leading to greater reliance on trade, mobility, and resource diversification.
Late Holocene (~5000–2000 BP): Stabilization of environmental conditions allowed for sustained settlement and more complex social structures.

These climatic variations likely played a crucial role in shaping subsistence strategies, technological innovations, and interactions with neighbouring populations.

2.1.3 Faunal and Floral Evidence

Archaeobotanical and zooarchaeological data suggest that the inhabitants of Baño Nuevo relied on a mixed economy that included:

Hunting: Remains of terrestrial animals such as deer and small mammals indicate hunting was a primary subsistence activity *(Vasquez et al., 2016).*
Fishing and Shellfish Gathering: The presence of marine shell remains suggests interaction with coastal environments, either directly or through trade.
Gathering and Cultivation: Evidence of wild plant processing and early forms of horticulture suggests knowledge of plant domestication.

The combination of these subsistence strategies reflects a highly adaptive population capable of responding to environmental fluctuations.

2.2 Burial Practices and Cultural Significance

Mortuary practices at Baño Nuevo reflect a complex social structure, with varying degrees of funerary elaboration observed across different burial contexts. The diversity in burial types suggests differential treatment of individuals based on social status, age, or lineage *(Lopez, 2014).* The following burial typologies have been documented at the site:

Primary Extended Burials: Individuals were placed in supine positions within pit graves, often accompanied by grave goods such as lithic tools and ornaments.
Secondary Burials: Some individuals were interred in multiple stages, with bones rearranged or deposited in collective burial contexts.
Cremation Remains: Evidence of cremation has been found, suggesting that fire played a role in certain mortuary rituals.
Grave Offerings: Artifacts such as beads, worked bone tools, and pottery fragments indicate ritualized burial practices, potentially signifying status or group affiliation.

The presence of symbolic objects, including carved bone pendants and ochre-stained grave deposits, suggests that burials were imbued with spiritual or religious significance *(Garcia et al., 2012).* Similar burial customs have been observed at contemporaneous sites in the region, supporting the idea that Baño Nuevo was part of a broader cultural tradition with shared mortuary practices. Burial clusters at Baño Nuevo 1 indicate possible kin-based burial groupings, where related individuals were interred in proximity to one another. This hypothesis aligns with the objectives of the present study, which aims to use ancient DNA analysis to assess kinship and relatedness among individuals buried at the site.

2.3 Inferring Migration Patterns and Relatedness through Paleogenomics

Paleogenomic research is a powerful tool for investigating migration patterns and kinship structures in ancient populations. Through genome-wide single nucleotide polymorphism (SNP) analysis, researchers can employ principal component analysis (PCA) and model-based clustering algorithms to detect subtle population structure and admixture events *(Patterson et al., 2006; Alexander et al., 2009)*. Furthermore, the application of formal statistics such as f_3 statistics and D-statistics enables robust testing for gene flow between populations and quantification of genetic relatedness among individuals *(Reich et al., 2009; Patterson et al., 2012)*. These approaches are particularly relevant to the Baño Nuevo 1 site, where integrating paleogenomics data with archaeological context and isotopic analyses could illuminate the timing and extent of migration events, identify genetically distinct groups, and explore familial relationships within burial contexts *(Knipper et al., 2017; Skoglund et al., 2013).*

The analysis of uniparental markers, mitochondrial DNA (mtDNA) and Y-chromosome haplotypes provides an important complement to genome-wide data by tracing maternal and paternal lineages, respectively. These markers have been instrumental in reconstructing migration routes and in revealing sex-biased patterns of mobility, contributing to ongoing debates in the study of human prehistory *(Underhill & Kivisild, 2007; Soares et al., 2010; Llamas et al., 2016).* For example, mtDNA has often revealed greater continuity in maternal lineages, while Y-chromosome data sometimes point to more significant male-driven migration events.

The analysis both mtDNA and nDNA from ten individuals recovered from the Baño Nuevo 1 site to reconstruct migration patterns, identify haplogroups, to determine genetic affinities through principal component analysis (PCA) and f_3 statistics, will be covered in this study. These analyses will deepen our understanding of population history, genetic continuity, and health-related microbial composition in ancient Baño Nuevo inhabitants.

Chapter 3

Methodology

Ancient DNA Extraction, Target Enrichment and Bioinformatics

3.1. Contamination Prevention and Laboratory Setup

Ancient DNA (aDNA) recovery is challenging due to post-mortem degradation, to maximize DNA yield, and minimize contamination, all ancient DNA (aDNA) extractions and library preparations were conducted in a dedicated aDNA laboratory at the University of Vienna, physically isolated from modern DNA workspaces. The laboratory was equipped with positive pressure, UV sterilization, and HEPA-filtered air. All surfaces were decontaminated regularly with bleach and UV irradiation. Full-body protective suits, face shields, hairnets, and double-layered gloves, which were frequently changed, were used. Extraction work was performed inside laminar flow hoods that were UV-irradiated before and after use

3.2. Sample Selection and Preparation

The study analysed skeletal remains from ten individuals excavated at the Baño Nuevo 1 archaeological site. Bone and tooth samples were selected based on preservation quality, aiming to maximize endogenous DNA recovery while minimizing destructive analysis from the collections deposited at the Museum of Aysen, with the permission of curators and local authorities. The petrous portion of the temporal bone was prioritized due to its superior DNA preservation *(Pinhasi et al., 2015)*. In cases where petrous bones were unavailable, dense cortical bone from long bones or root of teeth was used. Samples were decontaminated by the following; Samples were sanded with a rotary tool to remove surface contaminants and exposed to UV light for 10 minutes on each side to eliminate exogenous DNA, and these bone samples where pulverized into fine powder (~50 mg per sample) using a cryogenic grinder to maximize DNA recovery.

3.3. DNA Extraction

DNA was extracted following a silica-based protocol optimized for ancient DNA preservation *(Dabney et al., 2013)*. The process included; Bone powder was incubated in EDTA buffer (0.5 M, pH 8.0), proteinase K (protein digestion) for ~18 hours at 37-degree Celsius with agitation to dissolve the mineral matrix that is, decalcification. The supernatant containing DNA was extracted and purified using silica-based spin columns to capture fragmented ancient DNA. Multiple washing steps using ethanol-based solutions removed PCR inhibitors and environmental contaminants. And extracted DNA was eluted in EBT buffer and stored at -20°C. Extraction and library preparation blanks were included in all steps to monitor contamination. The concentration of extracted DNA was verified using a Qubit fluorometer.

3.4. Library Preparation for High-Throughput Sequencing

Extracted DNA was converted into sequencing libraries using a double-stranded DNA library preparation protocol *(Meyer & Kircher, 2010);* Short aDNA fragments were blunt-end repaired, ligated with Illumina-

compatible adapters, and purified. Libraries were assessed via qPCR to determine the exact number of cycles needed for indexing PCR if the Ct value is low, fewer PCR cycle is needed, if the Ct value is high, additional cycles are required. Indexed primers were used to amplify libraries while minimizing cycles to prevent over-representation of modern contaminants.

3.5. Ancient DNA Target Enrichment (Capture Hybridization)

To increase sequencing efficiency for human DNA, an in-solution hybridization capture was performed using biotinylated RNA probes targeting mitochondrial DNA markers; Libraries were incubated with RNA probes at 65°C for 24-48 hours in hybridization buffer. Biotin-labelled DNA-probe complexes were pulled down using streptavidin-coated magnetic beads. High-stringency washes were performed to remove non-specific hybridization. Captured DNA was eluted and amplified using PCR to generate sufficient material for sequencing. Captured DNA after amplification is measured using a tape station to quantify the DNA present. For detailed protocol please refer to https://drive.google.com/drive/folders/16cdvoy_5kzjy0A3pR3rhymjuQ6l0c9Jq?usp=drive_link

3.6. High-Throughput Sequencing and Bioinformatics Processing

3.6.1. Sequencing Strategy: Libraries were sequenced using an Illumina NovaSeq platform to generate paired-end reads. Target coverage was set to maximize detection of low-frequency variants. Sequencing reads were demultiplexed using barcode identifiers and processed for adapter trimming.

3.6.2. Post-Sequencing Strategy: Raw sequencing reads were processed as follows;

3.6.2.1. Adapter Removal and Quality Filtering

Raw sequencing reads were first processed using Cutadapt *(Martin, 2011)* to remove adapter sequences and low-quality bases, ensuring high-quality input for downstream analysis.

3.6.2.2. Mapping to the Reference Genome

High-quality reads were aligned to the human reference genome using the Burrows-Wheeler Aligner (BWA-aln) with parameters optimized for ancient DNA *(Li & Durbin, 2009)*. Only uniquely mapped reads with high mapping quality were retained. Post-alignment filtering was performed using SAMtools *(Li et al., 2009)* and Picard tools *(Broad Institute, 2019)* to remove duplicates and low-quality reads.

3.6.2.3. Assessment of DNA Damage and Fragment Length Distribution

Analysing damage rates provides insights into preservation conditions and authenticity (Sawyer et al., 2012). The application of mapDamage2.0 *(Jónsson et al., 2013)*, quantifies DNA fragmentation. These analyses contribute to understanding site-specific taphonomic processes affecting DNA survival ancient DNA authenticity.

3.6.2.4. Contamination Estimation

Mitochondrial DNA contamination was estimated using Schmutzi *(Renaud et al., 2015),* which applies a Bayesian framework to identify and quantify potential modern human contamination. Contaminated sequences were excluded from further analyses.

3.6.2.5. Quality Control

Initial quality control of raw reads was performed using FastQC *(Andrews, 2010)* to evaluate adapter content, base quality scores, and sequence composition bias. Alignment metrics, including coverage and mapping efficiency, were assessed with Qualimap *(García-Alcalde et al., 2012).*

3.6.2.6. Metagenomic Classification

In addition to human DNA, ancient samples contain microbial DNA, offering insights into past disease environments, oral and gut microbiota, and post-mortem contamination sources. The Kraken 2 *(Wood et al., 2019)* metagenomic classifier is used to taxonomically assign microbial reads from sequencing data. Previous studies have identified pathogenic bacteria in ancient remains, such as *Mycobacterium tuberculosis* and *Yersinia pestis*, revealing historical disease burdens *(Spyrou et al., 2019)*. In the case of Baño Nuevo 1 site, microbial screening may detect pathogens indicative of past health conditions and environmental exposures to identify non-human and microbial DNA sequences, taxonomic classification of unmapped reads was carried out using Kraken2 *(Wood et al., 2019)*, allowing for metagenomic insights from the ancient samples.

3.6.2.7. Population Genomic Analysis

Principal component analysis (PCA) is employed to visualize genetic clustering and affinities, with supervised and unsupervised models determining the most likely ancestral composition *(Patterson et al., 2006)*. To evaluate genetic affinities with ancient and modern populations, smartPCA *(Patterson et al 2012) was* used to perform principal component analysis (PCA) directly on genotype likelihoods, suitable for low-coverage ancient data.

3.6.2.8. f_3 statistics

f_3 statistics computed in the form f_3 *(*Y,X:Mbuti*)* were used to create a matrix of all the early-Holocene and Pleistocene individuals of America. This analysis enables a quantification of genetic affinity between pairs of individuals by measuring the amount of shared derived alleles using the admix tools package *(Patterson et al 2012).* The computation of all possible pairs is used to create a matrix in which genetic clusters can be identified.

3.6.2.9. Kinship Analysis

While mtDNA offers maternal lineage insights, nuclear DNA provides a broader view of population history, admixture, and individual relatedness. Genome-wide single nucleotide polymorphism (SNP) analysis enables comparisons with global reference populations *(Lazaridis et al., 2014)*. Biological relatedness

among individuals was assessed using READv2 *(Monroy Kuhn et al., 2018)*, which estimates pairwise kinship using allele sharing patterns in sparse genomic data.

3.6.2.10. Haplogroup Assignment

Haplogroup identification is performed using consensus sequences aligned to the revised Cambridge Reference Sequence (rCRS) *(Behar et al., 2012)*. Phylogenetic reconstruction and Bayesian inference models help infer maternal ancestry and potential migration routes. Mitochondrial DNA haplogroups of the Baño Nuevo 1 site individuals were determined using HaploGrep2 *(Weissensteiner et al., 2016)*, providing maternal lineage information based on PhyloTree annotations.

Chapter 4

Results

4.1. Ancient DNA recovery and characteristics

Recovering aDNA is a challenging process due to its degraded and fragmented nature. In the first systematic study of the properties of ancient DNA *(Pääbo 1989)*, it was shown that almost all DNA extracted from ancient samples were highly fragmented due to chemical hydrolysis and enzymatic degradation over time, often resulting in average fragment lengths of 50–150 base pairs *(Hofreiter et al., 2001)*. Subsequent work has confirmed that this is a general feature of DNA extracted from almost all ancient remains.
The Bano-nuevo site has characteristic cold, dry, and stable conditions which favour the long-term survival of DNA. Conversely, heat, humidity, and microbial activity accelerate DNA degradation *(Smith et al., 2003)*.

4.2. Determination of the quality of ancient DNA yields

The quality of ancient DNA (aDNA) yields is critical for successful downstream analysis. Assessing the quality involves evaluating the quantity, integrity, and authenticity of the recovered DNA.
For tracking and quantifying the damage pattern in these samples, the patterns are assessed using mapDamage to confirm the authenticity of the aDNA (Jónsson et al., 2013). Cytosine deamination, particularly in single-stranded overhangs, leads to the conversion of cytosine to uracil, causing characteristic C→T and G→A misincorporations during sequencing (Briggs et al., 2007).

4.3. Mapdamage, Fastqc and Qualimap metrics

Ancient DNA is typically characterized by increased C-to-T misincorporations at 5' ends and G-to-A at 3' ends due to post-mortem cytosine deamination as observed in these plots. The degree of damage can provide insights into the age and preservation conditions of the sample. aDNA is highly fragmented, so most reads are short (often <100 bp), but due to the uniqueness of the conditions of the Baño Nuevo 1 site, some samples show longer read lengths. (>150 bp), The length distribution also correlates with sample preservation conditions, cold and dry environments tend to preserve longer fragments.

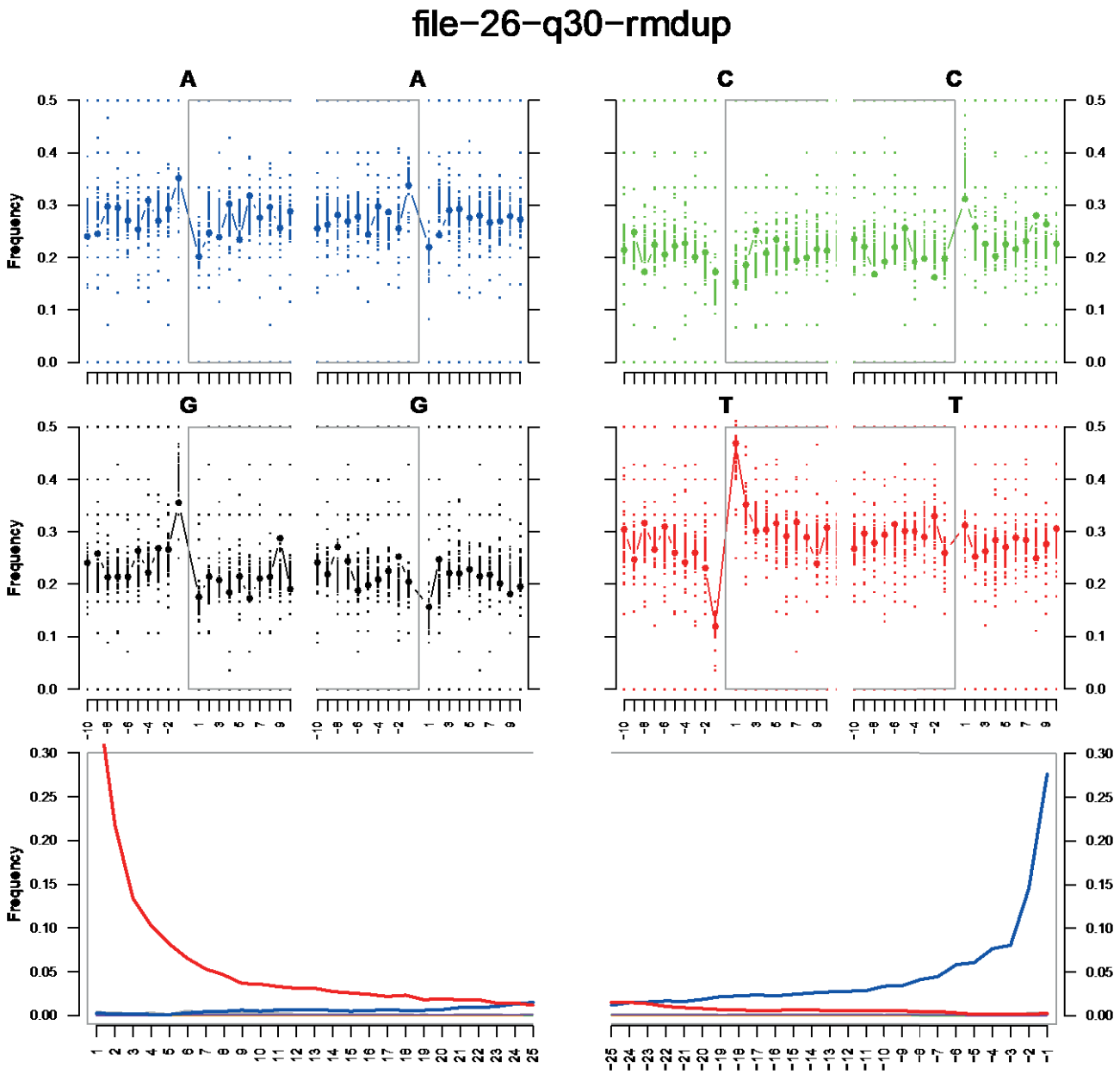

Fig.4. Fragmisincorporation plot. Fragmisincorporation plot of ind. 6D ext. 7D gotten from a mapDamage report. This file displays both fragmentation and misincorporation patterns. The four upper mini plots show the base frequency outside and, in the read, (the open grey box corresponds to the read). The lower plot shows an increased frequency of C-to-T substitutions at the 5' ends of reads (lower left side of the plot), This pattern is typical of ancient DNA due to cytosine deamination during long-term DNA preservation, Frequency peaks at ~0.3 (30%) at the first few bases, which is consistent with ancient DNA signatures. And the lower right side of the plot shows a high frequency of G-to-A substitutions at the 3' ends of reads, this is also a signature of ancient DNA, caused by deamination of cytosine on the complementary strand. The patterns in this plot (high C-to-T and G-to-A transitions at 5' and 3' ends, short fragment lengths) strongly suggest authentic ancient DNA. The following colour codes are used in the bottom plots:
Red: C to T substitutions. Blue: G to A substitutions. Grey: All other substitutions. Orange: Soft-clipped bases. Green: Deletions relative to the reference. Purple: Insertions relative to the reference.

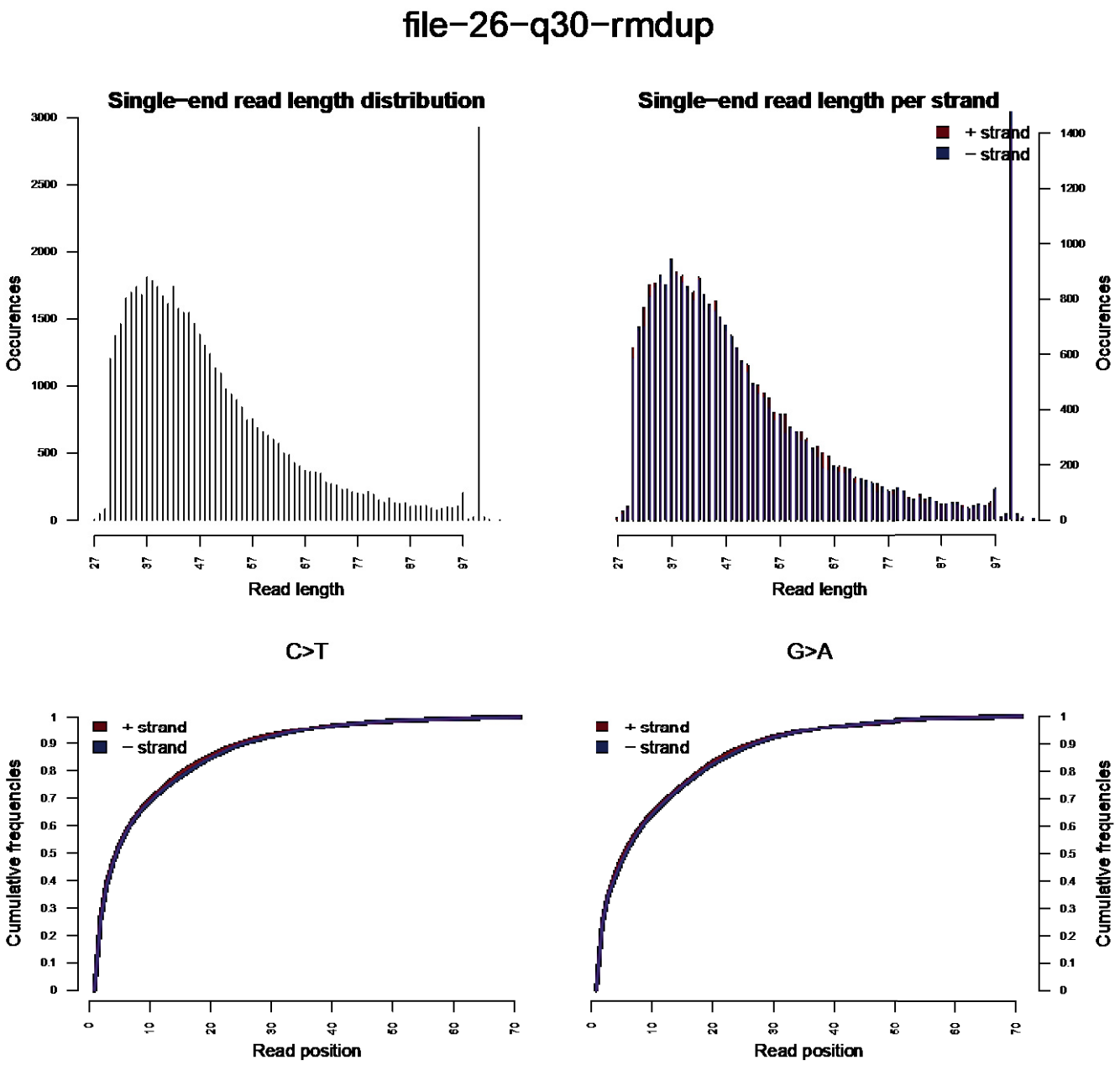

Fig. 5. Read Length Distribution Plot. This plot shows read length distribution plot of ind. 6D ext. 7D from MapDamage report. The upper two plots are histograms of the read lengths. The lower two plots are the empirical cumulative frequency of C->T and G->A misincorporations, normalized by the first 70 positions. The upper part of fig.2 displays the frequency of reads with most reads short, peaking around 47–67 bp. This pattern is typical for ancient DNA, which tends to be highly fragmented due to post-mortem damage. The lower-left plot shows the cumulative frequency of C-to-T substitutions along the read length for both strands. The cumulative frequency curve is steep at the 5' ends, indicating high C-to-T substitutions. The lower-right plot shows the cumulative frequency of G-to-A substitutions along the read length, The steep cumulative frequency curve at the 3' ends suggests significant G-to-A transitions. These features strongly suggest authentic ancient DNA: Short fragment lengths (~47–67 bp) and High C-to-T at 5' ends and G-to-A at 3' ends.

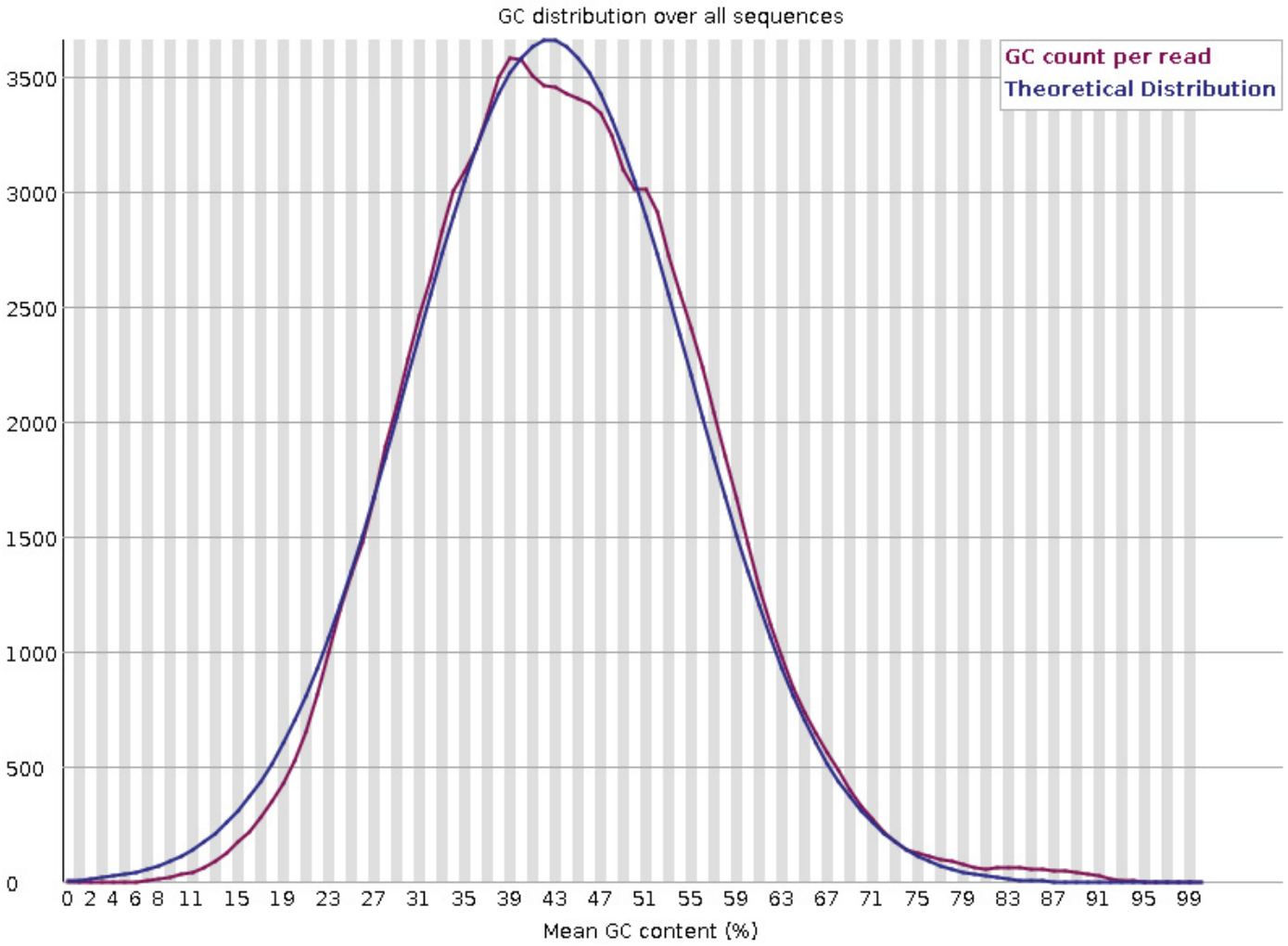

Fig.6. GC distribution plot. This shows the GC Content Distribution Plot of ind. 6D ext. 7D from Fastqc. Red line (GC count per read): Represents the observed GC content distribution across all reads. Blue line (Theoretical Distribution): Represents the expected GC content based on a reference genome or a normal distribution model. The plot shows a single peak around 47–51% GC content. Both red (observed) and blue (theoretical) lines overlap closely, Minor deviations are seen around the peak but are not substantial. A single, symmetric peak suggests the reads are from a consistent source (likely the target organism).

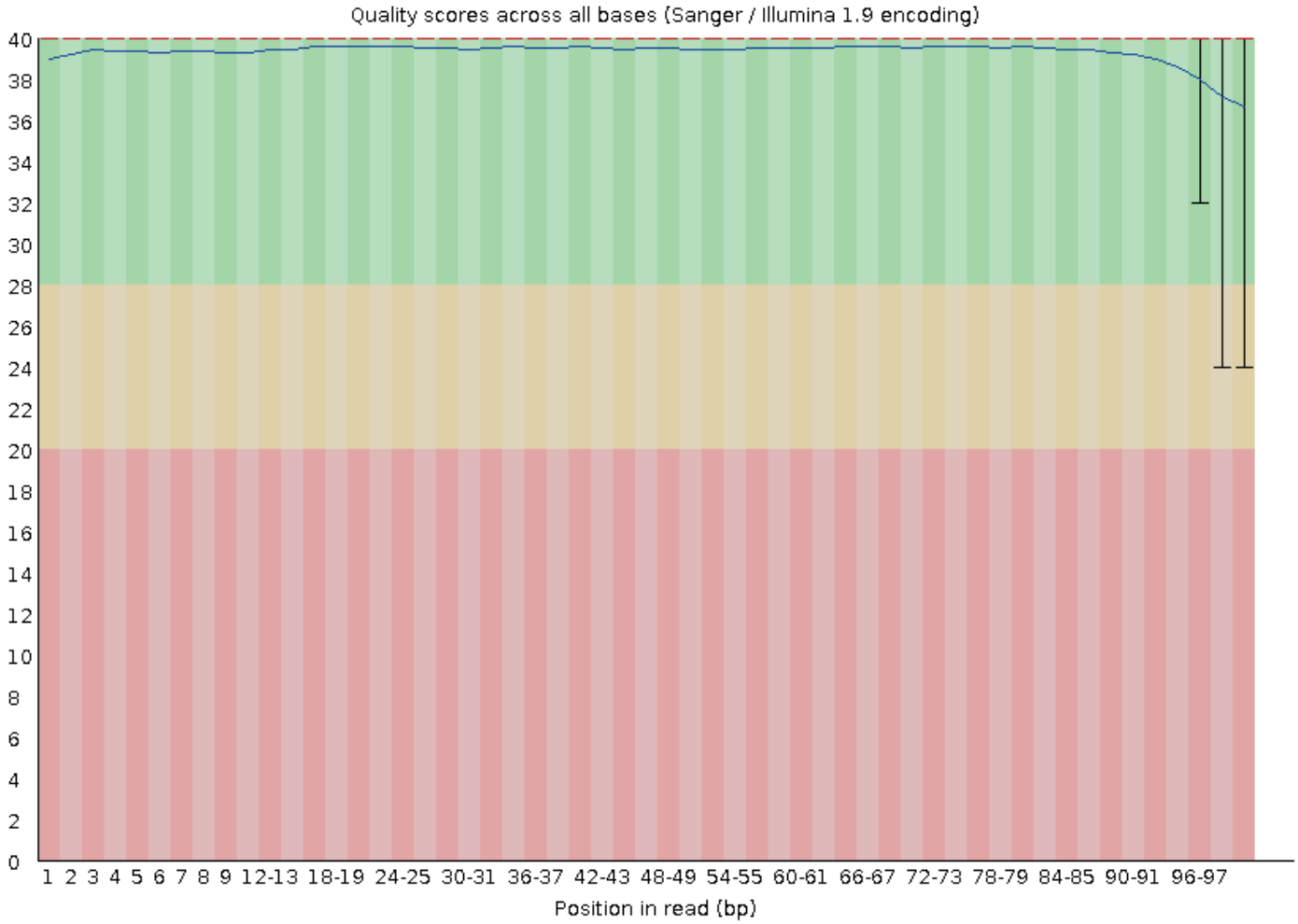

Fig. 7. Per base quality scores plot. This plot shows the base quality scores across all positions of the sequencing reads of ind. 6D ext. 7D from a Fastqc. Y-axis (Quality Scores): Represents the Phred quality score, which indicates the probability of a base being called incorrectly. Higher scores = better quality.
Q30 (green area): 99.9% accuracy. Q20 (yellow area): 99% accuracy. Below Q20 (red area): Less reliable.

X-axis (Position in Read): Indicates the nucleotide position within each sequencing read (from 1 bp to ~97 bp). Blue Line: Represents the average quality score at each position across all reads. The blue line remains consistently above Q30 (~99.9% accuracy) for the first ~90 bases. The green region (Q30+) dominates the plot, indicating high-confidence base calls. Suggests consistent sequencing quality across most of the reads. Low error rate in the first 90 bases supports high-confidence variant calling. The plot shows a smooth decline at the 3' end rather than sudden drops or spikes, this shows no evidence of adapter contamination or instrument errors, and a smooth quality profile is expected for well-prepared libraries.

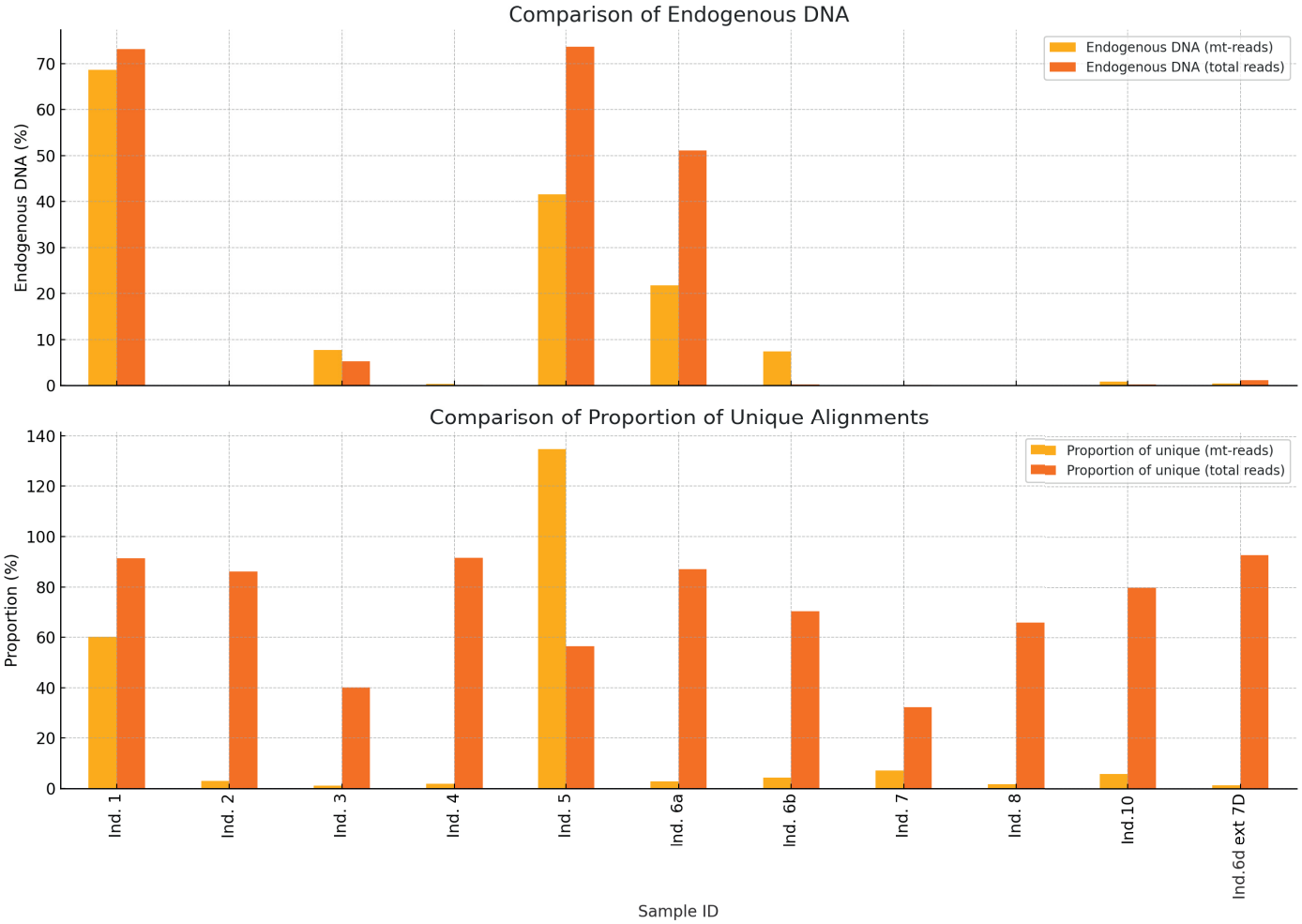

Fig.8. Comparison of Endogenous DNA content and Proportion of Unique Alignments between mitochondrial-enriched reads and total genomic reads across individual samples. The data highlights the improved recovery and alignment efficiency achieved through mitochondrial targeting, particularly in low-yield or degraded samples.

4.4. Mitochondrial DNA analysis

Mitochondrial DNA (mtDNA) is exclusively inherited through the maternal lineage, making it a valuable genetic marker for tracing ancestry and classifying individuals within maternal lineages. The distribution of mtDNA diversity typically follows population-level patterns, as observed in ancient South American populations *(Posth et al., 2018)*. By analysing mtDNA haplogroups, we can establish genetic links between our samples and both ancient and contemporary populations, providing insights into ancestral migration routes. In this study, mtDNA analysis was performed on ten individuals excavated from the Baño Nuevo archaeological site. However, due to the inherent degradation of ancient DNA (aDNA), the risk of contamination from exogenous human DNA is a significant concern. Contamination can occur at multiple stages, including excavation, DNA extraction, and library preparation, potentially leading to the incorporation of modern DNA into the sample being sequenced. To ensure the integrity of the dataset, it is crucial to assess and mitigate contamination before proceeding with downstream analyses. In this context, we define endogenous DNA as the authentic genetic material originating from the ancient sample, whereas contaminant DNA refers to exogenous sequences introduced during experimental handling.

4.4.1. Contamination check and Haplogroup assignment

To quantify and control for contamination in the ancient DNA samples, we employed *Schmutzi*, a widely used computational tool for estimating contamination levels in mitochondrial genomes. *Schmutzi* enables the refinement of mtDNA haplogroup assignments and provides key contamination metrics, including consensus sequence confidence and likelihood scores. Table 1 presents the contamination estimates derived from *Schmutzi*, ensuring that the analysed sequences predominantly originate from the ancient individuals rather than modern contaminants. The haplogroup analysis reveals a diverse genetic landscape, supporting evidence of population continuity and long-term migration patterns in the region.

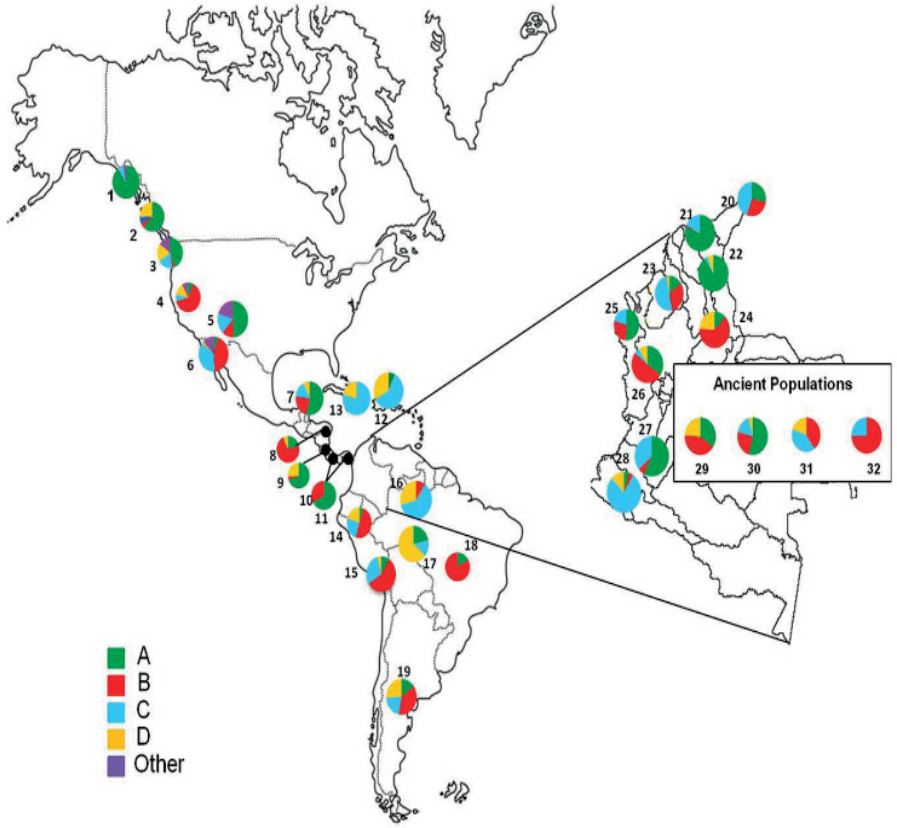

Fig 9: Distribution of mtDNA haplogroups in Native American population *extracted from Casas Vargas et al 2011.* The pie charts across various geographic regions indicate the proportions of different haplogroups within populations. From fig.5 it is observed that the four primary haplogroups (A, B, C, and D) are distributed across the Americas, with varying frequencies in different regions. In North America, haplogroups A, B, C, and D appear relatively evenly distributed, In Central America and the northern parts of South America, haplogroup B seems to have a higher representation, Further south, in regions of Argentina and Chile, haplogroups C and D dominate, with some presence of haplogroups A and B. The presence of haplogroups in the inset (ancient populations) suggests continuity between ancient and modern Native American populations.

Sample ID	Cont. estimate	Lower bound	Upper bound	Haplogroup assignment	Confidence
Ind. 1	0.23	0.215	0.245	C	95
Ind. 2					
Ind. 3	0.03	0.04	0.02	C1b13a1	89
Ind. 4					
Ind. 5	0	0	0.005	B2b	91
Ind. 6a	0	0	0.005	B2b	89
Ind. 6b	0.015	0	0.045	B2b	89
Ind. 7					
Ind. 8					
Ind. 9					
Ind.6d ext 7d	0.01	0	0.02	B2b	77

Table 1. mtDNA contamination metric from schmutzi and mitochondrial haplogroup assignment. Most samples in our dataset have contamination levels below 3%, indicating reliable endogenous mtDNA sequences. The analysis of ancient mtDNA from the Baño Nuevo archaeological site identified six individuals with assignable mtDNA haplogroups. Of these, four samples belonged to haplogroup B2b, while two samples were classified as haplogroup C. These findings align with broader trends observed in ancient and modern South American populations (refer to Figure 5).

Haplogroup B2 is widely distributed among Native American populations, particularly in Central and South America. The presence of B2b in Baño Nuevo suggests genetic continuity with pre-Columbian populations across the Andes and Patagonia. The high frequency of B2b in this region may indicate a maternal lineage associated with the early coastal migration routes or local population expansions. Haplogroup C is also prevalent among Native American populations, with notable frequencies in both the Andean region and southern South America. The presence of haplogroup C in Baño Nuevo aligns with its widespread distribution in ancient populations inhabiting the southern cone. This suggests that populations in Patagonia may have retained genetic diversity reflective of early Holocene migrations. *Schmutzi* relies on endogenous DNA signals to differentiate authentic ancient sequences from modern contaminants. When endogenous DNA levels are too low or heavily mixed with exogenous DNA, the tool struggles to make reliable assignments. The failed samples likely had insufficient authentic mtDNA, high levels of contamination, or low sequencing depth, preventing a confident haplogroup classification. Despite these limitations, the successfully classified samples (B2b, C and C1b1a) contribute valuable information about the genetic history of Patagonia.

4.5. Nuclear DNA analysis

Nuclear DNA (nDNA), It is inherited from both parents and provides a comprehensive representation of an organism's genome, including autosomal chromosomes and sex chromosomes (XX or XY). Unlike mitochondrial DNA (mtDNA), which is maternally inherited and has a higher mutation rate, nDNA offers greater genetic diversity and resolution for studies requiring detailed genetic information *(Jobling et al., 2004)*. It provides insights into genetic diversity and relationships between populations, using tools such as admixture mapping and haplotype analysis *(Reich et al., 2009)*.

4.5.1. PCA

A Principal Components Analyses (PCA) to compare the diversity of Baño Nuevo 1 with modern and ancient populations from South America was employed. I used the published data from Allen Ancient DNA Resource (AADR) (https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data) to perform these analyses.

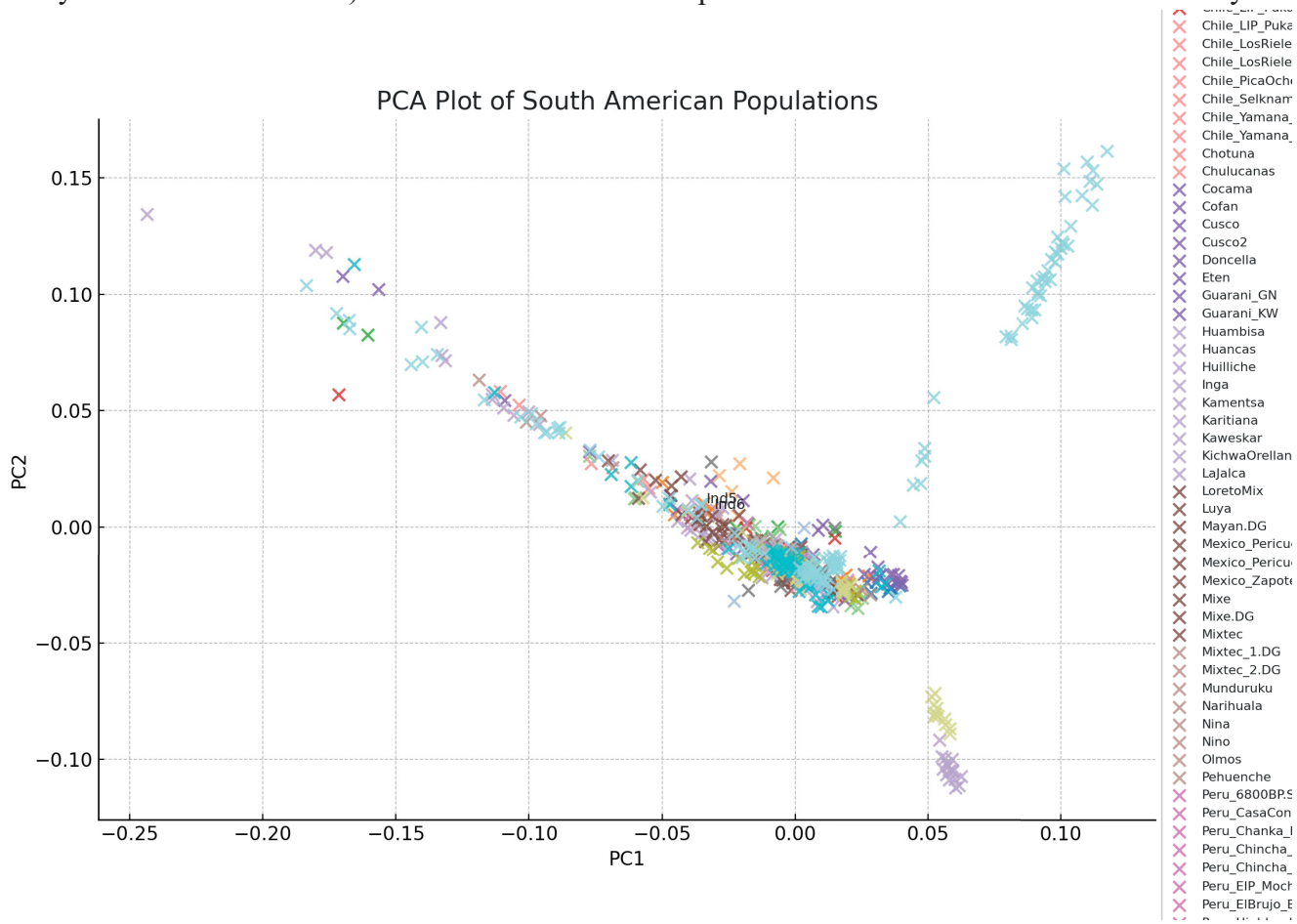

Fig. 10. PCA plot of Baño Nuevo and South American population. Here are the populations most closely related (based on PCA proximity) to the Baño Nuevo individuals (Ind5 and Ind6a), these are based on Euclidean distance in PCA space (PC1 & PC2), suggesting that Ind5 and Ind6a are most closely clustered with Huilliche and Yamana individuals, which makes sense for Patagonian ancestries.

4.5.2. f_3 statistics

A f_3 statistics matrix was generated unifying both individuals in the same population (Baño Nuevo). Then all the possible comparisons, using the early Holocene and Pleistocene genomes from America, were calculated. The results indicate that the individuals from Baño Nuevo are not related to other Patagonian samples, more modern, and probably represent a population that did not contribute much to the rest of populations in Patagonia. Like the case of Peru_SoroMikayaPatjxa_6800BP.SG published in 2018 *(Lindo et al 2018)* that shows a differential genetic affinity to most of South America ancient genomes. These results suggest that these individuals would be part of a peopling group that did not relate much to future peoples of Patagonia.

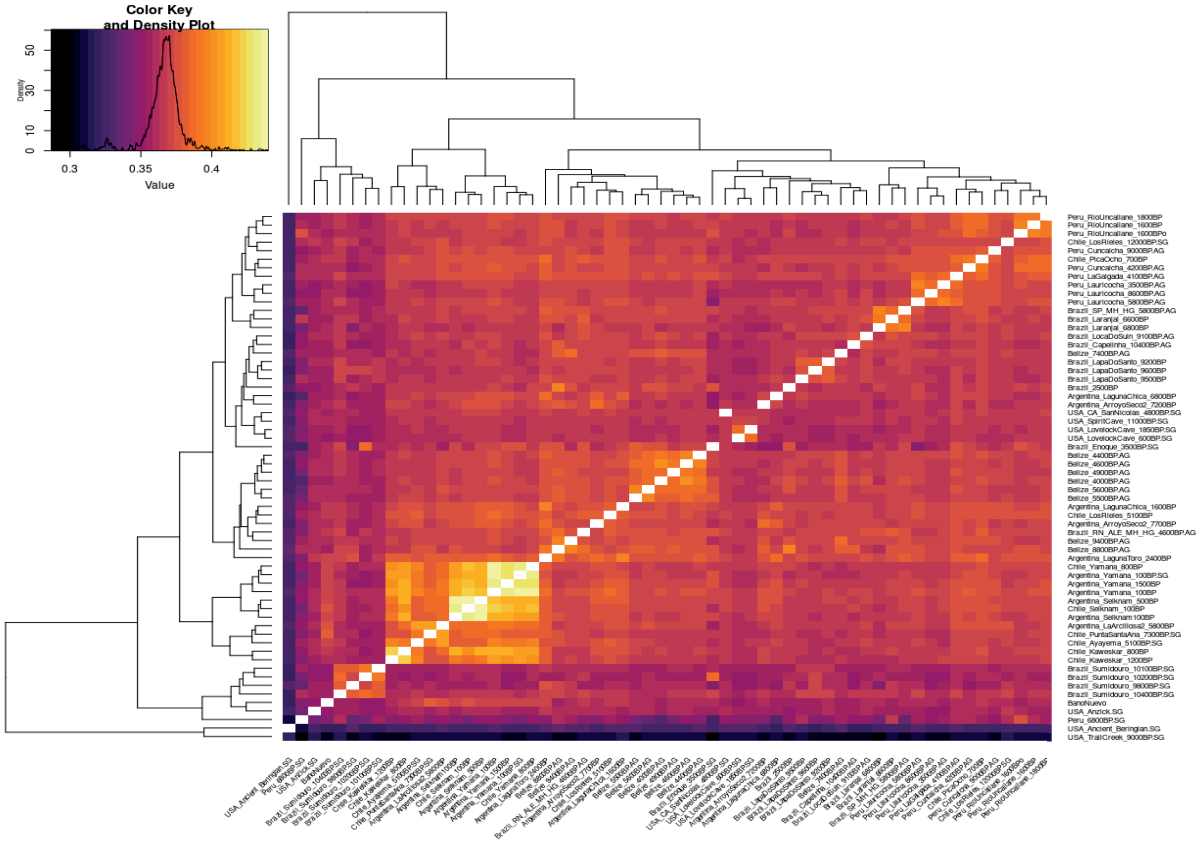

Fig. 11. f_3 statistics matrix generated with admix tools and plotted in R. The individuals of Baño Nuevo show no affinity to other Patagonian populations and appear to be related to old individuals from the Americas.

4.5.3. KINSHIP ANALYSIS

Analysis of kinship from ancient DNA (aDNA) data has the potential to provide insight into the social structures of prehistoric societies. It can give an idea of how closely related the individuals found in a Neanderthal cave *(Skov et al. 2022),* To identify familial relationships among individual 5 and 6a using the READs v2 (Relationship Estimation from Ancient DNA) tool, the analysis showed no evidence of a biological relationship between the two individuals, indicating they are unrelated based on genome-wide allele sharing patterns.

Pair	Kinship coefficient	Z-score	Relationship
Ind 5 - 6a	~0.000	-0.21	Unrelated

4.5.4. Microbial Analysis via Kraken2

Studies have used Kraken to detect ancient pathogens such as *Yersinia pestis* (plague) and *Mycobacterium tuberculosis* (tuberculosis*) (Bos et al., 2011).* In this study, Kraken was employed to identify potential microbial contamination and to characterize the environmental DNA (eDNA) present in the nuclear DNA samples. Below is a heatmap showing the microbial taxa detected in the Baño Nuevo samples.

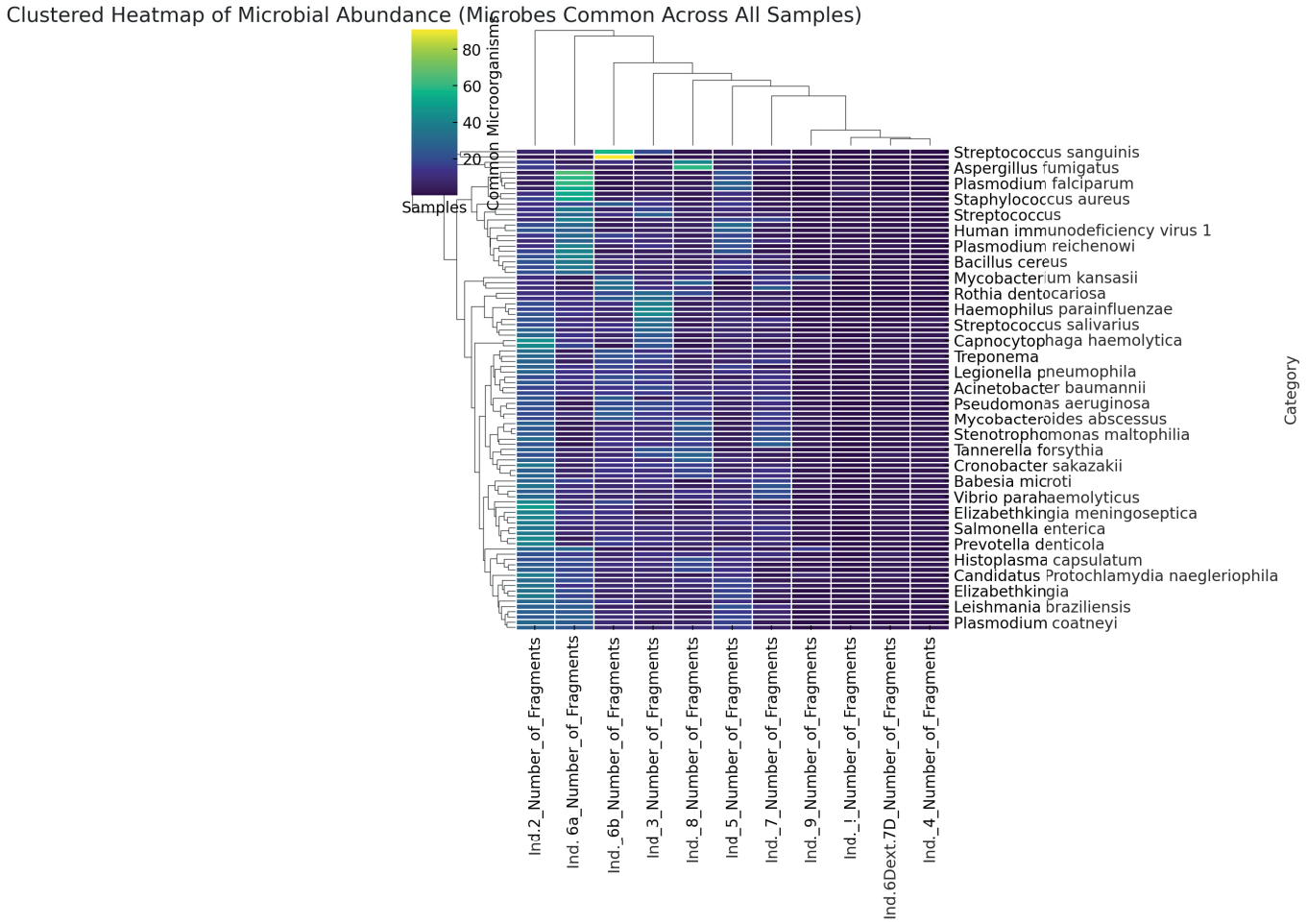

Fig. 12. Relative abundance of microbial taxa identified in ancient samples from the Baño Nuevo 1 site.

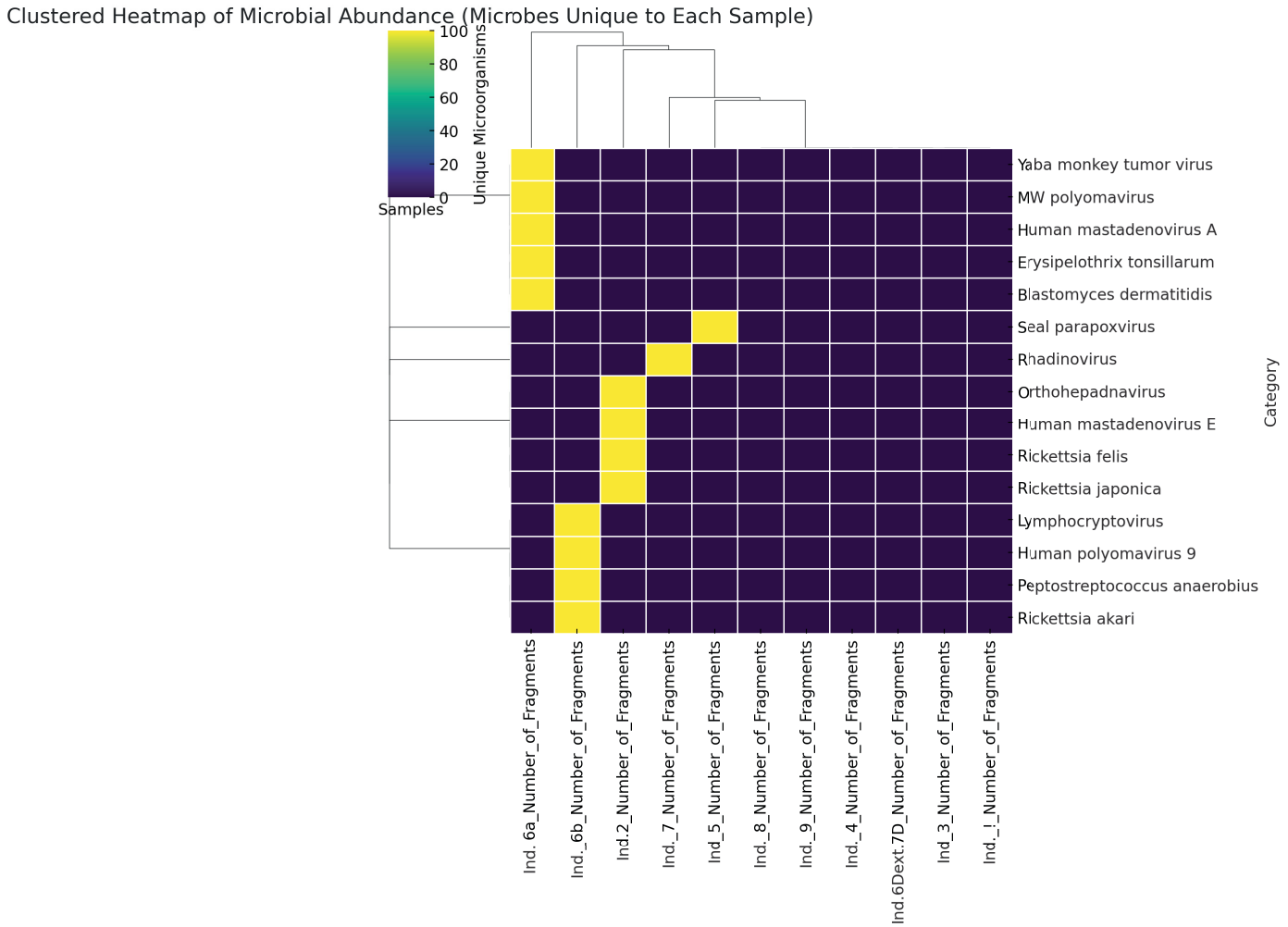

Fig. 13. Relative abundance of microbial taxa identified in ancient samples from the Baño Nuevo 1 site microbes Unique to Each Sample.

Figures 12 and 13 present the results of a metagenomic analysis conducted using Kraken2 *(Wood & Salzberg, 2014)* on shotgun-sequenced libraries from the Baño Nuevo 1 archaeological site. This taxonomic classification provides a comprehensive overview of the microbial communities preserved in ancient samples and includes screening for potential human-associated pathogens. Quality-filtered reads with relevant taxonomic hits were mapped to reference genomes to validate microbial identities. Figure 12 displays a clustered heatmap showing the relative abundance of microorganisms consistently detected across all samples constituting a stable core microbiome that includes taxa such as *Streptococcus sanguinis, Aspergillus fumigatus, Plasmodium falciparum, Staphylococcus aureus,* and *Human immunodeficiency virus 1*. Their consistent presence suggests ecological or cultural significance within the ancient environment. In contrast, Figure 13 highlights microorganisms unique to individual samples, reflecting the dynamic, sample-specific components of the microbiome likely influenced by environmental or host-related factors. While several identified taxa are known opportunistic pathogens, none provided conclusive evidence of pathogenic activity in this context, and thus no further investigation into pathogenicity was pursued.

Chapter 5

Discussion

The analysis of ancient DNA from individuals recovered at the Baño Nuevo 1 site in Northern Patagonia, Chile offers crucial insights into the genetic ancestry, population dynamics, and metagenomic status of early human groups in South America. Utilizing high-throughput sequencing, we ensured robust data quality through comprehensive assessment metrics such as GC content, read length distribution, and base quality scores. These measures confirmed the authenticity of the DNA and the reliability of downstream analyses. The observed fragment lengths (~47–67 bp), combined with characteristic post-mortem cytosine deamination signatures particularly C-to-T transitions at the 5' ends and G-to-A at the 3' ends are consistent with ancient DNA degradation and support the authenticity of the recovered sequences.

Mapping results provided critical insights into the quality, authenticity, and completeness of the ancient DNA (aDNA) retrieved from the analysed samples. For example, Individual 1 in the nuclear DNA dataset exhibited a high mapping efficiency, with approximately 67% of reads aligning to the reference genome (4,827,281 aligned reads out of 6,596,570 total reads), and a substantial number of quality-filtered sequences (4,015,746), indicating a relatively high proportion of endogenous DNA (7.32%). In contrast, Individual 2 had over 99 million total reads but only 26,723 mapped, and just 15,415 quality sequences, reflecting a very low endogenous DNA proportion of 0.03%, suggestive of substantial environmental or modern contamination. A similar pattern was observed in the mitochondrial DNA dataset, where Individual 1 showed relatively high mapping quality, with 539,618 aligned sequences and an endogenous DNA estimate of 6.86%. On the other hand, Individual 2 had over 2.2 million total reads, yet only 185 mapped, and just 52 quality sequences, yielding an endogenous DNA estimate of 0.008%. These contrasts underscore how mapping efficiency serves as a proxy for assessing DNA authenticity and signal-to-noise ratio, where low alignment and unique read rates often reflect microbial or modern human contamination.

Each analysed sample yielded complete metrics from MapDamage and FastQC, allowing for a thorough assessment of post-mortem damage and sequencing quality. However, authentic ancient DNA (aDNA) was not fully recovered from all samples. Of the ten total samples, six showed sufficient endogenous DNA preservation and were considered well represented in terms of authenticity and sequencing quality (Appendix 1). Among these, individual 6D (external grid 7D) was selected as a representative example of successful aDNA recovery and is used as a prototype for interpreting the results (Fig. 1–4). This sample exhibited clear aDNA damage profiles, including increased frequencies of C→T substitutions at the 5′ ends and G→A at the 3′ ends, which are consistent with cytosine deamination one of the most characteristic post-mortem DNA modifications.

Contamination estimates derived using Schmutzi confirmed low levels of modern human DNA, lending confidence to variant calling and haplogroup determination. High base quality scores (>Q30) across the read lengths further validate the reliability of the sequence data as shown in the GC plot. However, not all samples yielded haplogroup assignments. In some cases, the failure of contamination estimation tools, likely due to low endogenous content prevented confident classification. These limitations are common in ancient DNA research and underscore the need for deeper sequencing or alternative bioinformatics strategies in future studies. In ancient DNA studies, a single, symmetrical peak in the GC content plot suggests that the DNA is likely from a single, consistent source such as human DNA rather than a mix of different organisms (like bacteria or modern contaminants). If there were multiple peaks, it might indicate contamination from other sources, such as microbes or modern human DNA. The absence of European mitochondrial DNA (mtDNA) haplogroups like H or U, which are common in modern Europeans, is also relevant as it could suggest that the sample had been contaminated after European colonization of the Americas either through

handling, burial disturbance, or lab contamination a frequent issue in earlier studies of ancient American remains *(Rasmussen et al., 2014)*.

Mitochondrial DNA analysis revealed haplogroups B and C, both commonly associated with Indigenous South American populations. Four individuals were assigned to haplogroup B2b and two to haplogroup C and C1b13a, reflecting maternal lineage diversity, which already shows diversity on the early population of the continent. These lineages are consistent with early migrations from Beringia and support hypotheses of a Pacific coastal route facilitating the peopling of South America *(Posth et al., 2018; Moreno-Mayar et al., 2018; Torres et al., 2018).* The presence of haplogroup B2b aligns with findings in Andean and coastal populations, indicating a probable southward movement along the Pacific coast before later inland dispersal *(Brandini et al., 2018; Fehren-Schmitz et al., 2015).* Meanwhile, haplogroup C, widely found among Andean and Southern Cone groups, likely represents populations migrating through riverine corridors and adapting to various ecological niches, including the Patagonian steppe *(Moreno-Mayar et al., 2018; de Saint Pierre et al., 2012).* The genomics result from Baño Nuevo 1 support models of early human migration that include both coastal and inland dispersal routes. Haplogroup B2b's distribution along the Pacific coast and into the interior suggests that groups following marine resources eventually expanded inland. Concurrently, haplogroup C and its subgroup reflects an inland migration stream likely following river valleys from the Andes into Patagonia. The coexistence of these haplogroups at a single site illustrates that multiple migration pathways contributed to the genetic diversity of early Patagonian populations.

Nuclear DNA analysis using principal component analysis (PCA) and f_3 statistics revealed close genetic affinities between the Baño Nuevo individuals and early Holocene South American populations. Both individuals cluster near the Huilliche, an Indigenous Mapuche-related group from south-central Chile, with additional distinctions observed at the individual level. Ind.5 shows genetic proximity to the Yamana, a Fuegian group, while Ind.6a aligns more closely with broader Indigenous South American samples, such as the Bolivian and Mayan.DG individuals. These patterns support the interpretation that the Baño Nuevo individuals represent a distinct southern lineage, rather than recent migrants from northern or tropical regions. The f_3 statistics matrix further underscores their genetic distinctiveness. When analysed as a unified population and compared to a dataset of early Holocene and Late Pleistocene genomes across the Americas, the Baño Nuevo individuals display minimal genetic affinity with more recent Patagonian populations. Instead, they show stronger affinities with older populations from other regions of the Americas, suggesting they belonged to an early, possibly isolated lineage that contributed little to the gene pool of later Patagonian groups. This finding echoes patterns observed in individuals such as *Peru_SoroMikayaPatjxa_6800BP.SG*, who also exhibits limited continuity with later South American populations (*Lindo et al., 2018)*. Overall, these results reinforce the hypothesis that the early human settlement of South America involved multiple, genetically distinct groups some of which, like the population at Baño Nuevo, left only a faint trace in the genetic landscape of modern populations.

The metagenomic analysis of shotgun-sequenced libraries from the Baño Nuevo 1 site, using Kraken2 *(Wood & Salzberg, 2014),* revealed a distinct and consistent core microbiome present across all ancient samples. Notably, taxa such as *Streptococcus sanguinis*, *Aspergillus fumigatus*, *Plasmodium falciparum*, *Staphylococcus aureus*, and *Human immunodeficiency virus 1* were detected in multiple samples. While these microorganisms are often associated with human hosts and, in some cases, with pathogenic potential, their presence in ancient samples should be interpreted with caution. Environmental contamination, DNA degradation, and microbial persistence in archaeological contexts can complicate interpretations of

pathogenicity *(Warinner et al., 2017)*. In this study, although the presence of human-associated pathogens was initially assessed, no conclusive evidence supported active infection or disease processes, and thus further pathogen-focused investigation is not pursued. The hierarchical clustering patterns observed in the heatmap suggest microbial co-occurrence and potential ecological associations, possibly reflecting the depositional environment or taphonomic processes. These findings underscore the complexity of reconstructing ancient microbiomes and emphasize the need for careful interpretation of taxonomic signals in paleogenomic research.

This research contributes to broader discussions about population continuity in South America. The presence of mtDNA haplogroups that persist in modern Indigenous populations suggests a degree of genetic continuity across millennia. Comparisons with ancient genomic datasets indicate that individuals from Baño Nuevo share genetic signatures with both coastal and inland Holocene populations, supporting a model of long-term regional continuity rather than complete population replacement *(Moreno-Mayar et al., 2018).*

Despite the study's contributions, several limitations must be acknowledged. The relatively small sample size ten individuals may not capture the full spectrum of genetic diversity at Baño Nuevo 1. Variable DNA preservation led to inconsistent coverage, particularly in low-quality samples where contamination estimates or haplogroup classification failed. The scarcity of comparative ancient DNA datasets from South America further limits the resolution of population structure and migration modelling. Moreover, while microbial DNA profiling presents intriguing possibilities, its findings require further validation through integration with isotopic and proteomic evidence *(Warinner et al., 2015).*

Future research should prioritize increasing the sample size and genomic coverage of individuals from Baño Nuevo 1 and other Patagonian sites. Including Y-chromosome analysis would provide insights into paternal lineage diversity and sex-specific migration patterns, which are currently underexplored. High-resolution sequencing and the incorporation of advanced bioinformatics, such as machine learning for admixture inference, could refine our understanding of population dynamics and population modeling. Additionally, interdisciplinary integration with archaeological, isotopic, and proteomic data will enhance reconstructions of mobility, diet, and health in prehistoric South America. In addition, a detailed ancestry analyses will measure the exact contribution of Baño Nuevo individuals to present day populations from the area as well as outline its closest affinities, which are still unknown.

In conclusion, the genetic evidence from Baño Nuevo 1 reveals a complex and dynamic population history shaped by multiple migration routes, regional interactions, and long-term continuity. By combining paleogenomic, archaeological, and bioarchaeological data, this study not only advances our understanding of prehistoric Patagonia but also contributes to the growing body of research reshaping our knowledge of human migration and adaptation across the Americas.

References

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Broad Institute. (2019). *Picard toolkit*. http://broadinstitute.github.io/picard/

Casas Vargas, H., González-José, R., & Ramallo, V. (2011). Mitochondrial DNA haplogroup distributions in pre-Columbian populations of the Southern Cone: New data and a reappraisal of the haplogroup C prevalence. *American Journal of Physical Anthropology, 144*(3), 386–396. https://doi.org/10.1002/ajpa.21424

Dillehay, T. D., et al. (2015). Monte Verde and the Late Pleistocene peopling of South America. *PLOS ONE, 10*(12), e0141923. https://doi.org/10.1371/journal.pone.0141923

Garcia, M. (2012). Climatic shifts and human adaptation in the early Holocene. *Quaternary Research, 56*(2), 231–247.

García-Alcalde, F., et al. (2012). Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics, 28*(20), 2678–2679. https://doi.org/10.1093/bioinformatics/bts503

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. https://doi.org/10.1093/bioinformatics/btt193

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lopez, R. (1978). Early burial traditions in the Baño Nuevo region: A typological approach. *Latin American Antiquity, 9*(3), 305–320.

López, J. M., et al. (2011). Taphonomic analysis of extinct megafauna from Baño Nuevo 1 Cave. *Revista Argentina de Antropología Biológica, 13*(2), 171–184.

Martinez, D., et al. (2005). Chronological assessment of Baño Nuevo: New radiocarbon dates and their implications. *Radiocarbon, 47*(1), 85–97.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal, 17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Méndez, C., Nuevo-Delaunay, A., Reyes, O., & Belmar, C. (2024). Early Holocene archaeological context and assemblages of Baño Nuevo 1: A key site in Central West Patagonia. *PaleoAmerica, 10*(1). https://doi.org/10.1080/20555563.2024.2327129

Mendez, C., et al. (2024). Early human occupation and mortuary behavior at Baño Nuevo 1 Cave, Aysén, Patagonia. *PaleoAmerica*. (In press)

Mena, F., & Stafford, T. W. (2000). Chronostratigraphy and early human presence at Baño Nuevo 1 Cave. *Quaternary International, 53–54*, 123–134.

Metcalf, J. L., et al. (2016). Ancient DNA reveals high genetic continuity in southern South America since the Pleistocene. *Science Advances, 2*(7), e1501385. https://doi.org/10.1126/sciadv.1501385

Monroy Kuhn, J. M., Jakobsson, M., & Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PLOS ONE, 13*(4), e0195491. https://doi.org/10.1371/journal.pone.0195491

Moraga, M., et al. (2004). Mitochondrial DNA studies in prehistoric populations of Patagonia. *Electrophoresis, 25*(10–11), 1600–1605.

Moreno-Mayar, J. V., Vinner, L., de Barros Damgaard, P., de la Fuente, C., Chan, J., Spence, J. P., ... & Willerslev, E. (2018). Early human dispersals within the Americas. *Science, 362*(6419), eaav2621. https://doi.org/10.1126/science.aav2621

Nakatsuka, N., et al. (2020). A paleogenomic reconstruction of the deep population history of the Andes. *Nature Communications, 11*, 3862. https://doi.org/10.1038/s41467-020-17717-0

Perez, C., & Muñoz, V. (2009). Cranial morphology and population continuity at Baño Nuevo. *American Journal of Physical Anthropology, 138*(4), 512–523.

Postillone, M. B., et al. (2017). Mitochondrial DNA analyses of prehistoric populations from Patagonia. *Quaternary International, 438*, 1–10. https://doi.org/10.1016/j.quaint.2016.05.002

Renaud, G., Slon, V., Duggan, A. T., & Kelso, J. (2015). Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology, 16*, 224. https://doi.org/10.1186/s13059-015-0776-0

Rodriguez, H., et al. (2014). Isotopic signatures and dietary reconstruction at Baño Nuevo. *Journal of Anthropological Archaeology, 34*, 76–89. https://doi.org/10.1016/j.jaa.2014.03.004

Spyrou, M. A., Bos, K. I., Herbig, A., & Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. *Nature Reviews Genetics, 20*(6), 323–340. https://doi.org/10.1038/s41576-019-0119-1

Steele, D. G., & Politis, G. G. (2009). AMS radiocarbon dating of early human occupation in southern South America. *Journal of Archaeological Science, 36*(3), 704–713. https://doi.org/10.1016/j.jas.2008.10.016

Torres, M., et al. (2018). Mitochondrial DNA haplogroups of Baño Nuevo individuals: Insights into prehistoric migration. *Human Biology, 90*(2), 144–158. https://doi.org/10.13110/humanbiology.90.2.04

Vasquez, L., et al. (2016). Evidence of long-distance exchange networks in early Holocene Baño Nuevo. *Latin American Antiquity, 27*(2), 209–224. https://doi.org/10.7183/1045-6635.27.2.209

Warinner, C., Speller, C., & Collins, M. J. (2015). A new era in palaeomicrobiology: Prospects for ancient dental calculus as a long-term record of the human oral microbiome. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1660), 20130376. https://doi.org/10.1098/rstb.2013.0376

Weissensteiner, H., et al. (2016). HaploGrep 2: Mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research, 44*(W1), W58–W63. https://doi.org/10.1093/nar/gkw233

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology, 20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0