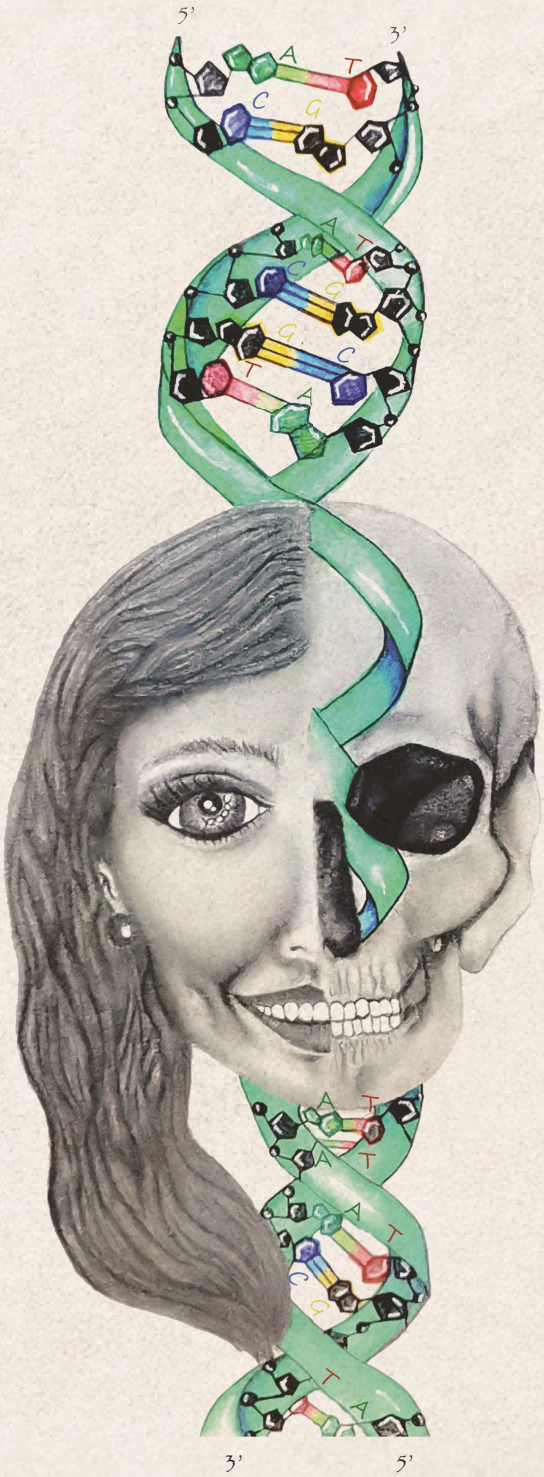
Universitat Autònoma de Barcelona

Methodological Challenges in Ancient and Forensic DNA Analysis: Improvements in DNA Extracti on and Geneti c Characterizati on from Human Skeletal Remains

DIANA C. VINUEZA-ESPINOSA



UNIVERSITAT AUTÒNOMA
DE BARCELONA

Departament de Biologia Animal,
Biologia Vegetal i Ecologia
Unitat d'Antropologia Biològica
Laboratori de DNA antic

Methodological Challenges in Ancient and Forensic DNA Analysis: Improvements in DNA Extraction and Genetic Characterization from Human Skeletal Remains

DIANA C. VINUEZA-ESPINOSA

PhD Thesis

Supervisors:
Assumpció Malgosa, PhD
Cristina Santos, PhD

Barcelona, 2022

Universitat Autònoma de Barcelona

Departament de Biologia Animal, Biologia Vegetal i d'Ecologia
Unitat d'Antropologia Biològica
Laboratori d'DNA antic

Ph.D. Dissertation

Methodological Challenges in Ancient and Forensic DNA Analysis: Improvements in DNA Extraction and Genetic Characterization from Human Skeletal Remains

By
Diana C. Vinueza Espinosa

Barcelona, 2022

Methodological Challenges in Ancient and Forensic DNA Analysis:
Improvements in DNA Extraction and Genetic Characterization from Human Skeletal Remains

Dissertation presented by Diana Carolina Vinueza Espinosa in fulfillment of the requirements for the Doctorate in Biodiversity of Department de Biologia Animal, de Biologia Vegetal I d'Ecologia, Universitat Autònoma de Barcelona, directed by:

Dr. Assumpció Malgosa, Chair Professor at Unitat d'Antropologia Biològica, Departament de Biológica Animal, de Biologia Vegetal i d'Ecologia, Universitat Autònoma de Barcelona.

Dr. Cristina Santos, Associate Professor at Unitat d'Antropologia Biològica, Departament de Biològica Animal, de Biològia Vegetal i Ecologia, Universitat Autònoma de Barcelona.

Dr. Assumpció Malgosa **Dr. Cristina Santos**

Diana C. Vinueza Espinosa

To my husband, best friend, life partner, angel, weakness, and strength,
thanks for being the support in this arduous journey.
We can do it! I cannot believe it.

A mi madre, la luz de mis ojos me diste las alas y me enseñaste a volar,
me empoderaste con tú amor infinito y con la fé que siempre has tenido en mí
incluso cuando pensé que no lo lograría, soy lo que soy gracias a ti.

A mi hermano Luis, sobrina Brittany, suegros Manuel y Mélida
y cuñados Javier y Cami mis mayores refugios lleno de brazos dispuestos
para arroparme en los momentos más difíciles.

Y a todas aquellas personas que migran de su país en
búsqueda de nuevas y mejores oportunidades.

ACKNOWLEDGMENTS

For all of us who have been through this long journey, the doctoral thesis is the culmination of many years of hard work and effort, so the last station stop of this journey ends here.

However, this dissertation would not have been possible without the help and support of many beautiful people. First and foremost, I would like to thank my supervisors, Assumpció Malgosa and Cristina Santos, for the much-needed guidance and support. You have done that my work is valuable in the lab and have allowed me to leave a footprint in the department. Thanks to those who have passed for the laboratory and are still in this gripping world of research, Daniel Ruiz and Xavier Roca Rada; thanks for the aid and support; this would have been impossible without you. Let us go for much more publications in the future!

I am profoundly grateful to Cristina Martínez-Labarga, Professore Associato- Centro di Antropologia molecolare per lo studio del DNA antico, Laboratorio di Biologia dello scheletro e Antropologia forense, Dipartimento di Biología Università degli Studi di Roma “Tor Vergata” for the great welcome and knowledge acquired during the stay. I am proud of being part of your research group and being in the great ancient DNA lab; you are a great researcher. To Hannes Schroeder, Associate Professor, Globe Institute, Faculty of Health and Medical Sciences, University of Copenhagen, for this stay full of new knowledge, language and culture enrichment in COVID time. A dream come true to have been in one of the best ancient DNA laboratories in the world. To Anna Knudsen for your help in all the lab work. Thanks so much, both of you.

I am thankful for the scholarship and the financial support from Interchange Erasmus Project 2019-2020, with which I was able to do my international stay. Also, for all the projects in which I had an excellent opportunity to work and give all of me to identify victims: Proyecto de Memoria Democrática (Baleares y Catalunya), Cementeri de Montjuic, Proyecto d’Enterrament Islamics (Born) y Víctima de la guerra del Pacífico-Chile.

My sincere thanks to those who made this less complicated than what could have been, Nuria and Jessy, thanks for giving me invaluable help in lab and for the long

conversations such pleasant in the lunch. I will carry it in my heart forever. To my girls, Maritxu, Aidaxtu, and María have been a great support at this stage; you deserve more thanks than I can express here. Thanks for the love and adventures over the past few years, which may have been crazy but have kept me sane all the same. I love you so much. ¡Nos merecemos un Goya! Merci, Eskerrik asko. To my friends, who I knew through the desire of fitting in a doctoral society: Lino, Moni, mi bichita, Jelly, and Sergio, thanks for great moments together and because the only thing I learned in Català, "es que esmorzar un bikini no es malo." Los quiero.

To all whom I have known in the process and of one way or another I have worked: María Pilar Aluja, Cristina Afonso, Amanda Ramos, Nika Nociarová, Nuria Armentano, ATICS, Iltirta, Andrea, Nuria, Marta, Raquel Rasal and Ferran Casals.

Finally, to my tiny family (Mom Amparito, Brother Luis, Niece Brittany, Grandpa Bolivar, Grandma Zoila, Cousin Dani, Aunt Tere, Mother-in-law Mélida, Father-in-law Manuel, Brothers-in-law Javier and Cami), thanks for having been at the forefront of that effort, for constant encouragement and for always being there for me, taking interest in our steps. To the unyielding support of my life, my husband, Andrés, thanks for being the main character of this journey that we have chosen and for flying together to come true our aims and dreams, for sharing the love of science and for the endless days in front of the computer. I love you eternally.

Thank you, God, for helping me not lose heart and because you gave me the force necessary when I was about to give up.

Thanks so much to all of you

Diana C. Vinueza-Espinosa

Todo es muy difícil antes de ser fácil. La perseverancia hace que todo lo que se creía imposible sea posible

-Goethe

ABSTRACT

In the last 40 years, ancient and forensic DNA fields have experienced an impressive development of methods and procedures. These advances have been from a simple Polymerase Chain Reaction (PCR) to Next Generation Sequencing (NGS), reducing costs and time invested and increasing the genetic information obtained. The goals of both disciplines are different forensic genetics focuses on identification for legal purposes, while paleogenetics has broader goals, such as population reconstruction or the study of the evolutionary past. This divergence leads to the use of different methodologies. Forensic genetics needs the standardization and validation of procedures between laboratories. The field of ancient DNA has always been more purposeful than forensic genetics and has developed more rapidly procedures and new platforms and methodologies. In contrast, forensic DNA is conservative and leads to consensus methods, which are often difficult to change. Although ancient DNA (aDNA) analysis and forensic DNA profiling have traditionally required different sets of DNA markers to achieve their respective goals, some commonalities have been evidenced. Both fields face various technical problems, mainly the recovery of degraded DNA from skeletal remains. In this type of specimen, harsh environmental conditions such as long-term UV irradiation, extreme temperature gradients, low/high pH, and humidity have triggered DNA degradation, resulting in various biochemical nucleotide modifications and making DNA recovery difficult. Given the great success of the aDNA community in recovering degraded human DNA, the critical forensic samples could be treated as ancient genomes laboratory-wise because their DNA is also highly fragmented in minute amounts and accumulates molecular damage.

This dissertation aims to connect current standard practices in ancient DNA research and forensic genetics to increase the recovery of genetic information from highly degraded skeletal remains. The tested techniques could be an alternative route to achieving the genetic identification of victims of armed conflict, missing persons, or victims of massive disasters. Different factors that could affect the recovery of degraded DNA were evaluated.

Chapter 1 tackles a systematic comparison between types of skeletal remains (petrous bone, different parts of the tooth, and postcranial bones) and four different extraction methods, based on the silica-in suspension and the silica-in column protocols and the traditional

organic protocol (Ph-Cl), used commonly in forensic genetics until now. Autosomal STR typing, a fragment of nuclear DNA, and the mtDNA control region (HVRI and HVRII) were molecular markers used to evaluate if the success of recovery depends on the type of skeletal remains or the extraction method-specific used, or the combination of both factors. The results showed that methods based on silica-DNA binding, specifically the silica in-suspension method, recovered the highest amount of short fragments characteristic of highly degraded DNA, typical of forensic and ancient samples. The method of silica-in suspension obtained the highest number of positive amplifications of the mtDNA fragments (HVRI and HVRII). In contrast, the Phenol-Chloroform (Ph-Cl) was associated with the poorest results. Regarding the type of bone, petrous bones, and pulp cavity displayed the best results in the concentration of amplified mtDNA products (HVRI and HVRII) using the methods of silica-in suspension and silica in the HE-membrane column. At the same time, the rib did not show amplification of HVRI using the methods of silica in the XS plasma-column and Phenol-Chloroform. The autosomal STRs profiles from the petrous bone obtained with the method of silica in-suspension and silica in the HE-membrane column showed the highest percentage of reportable alleles with the highest values of the Peak Height Ratio (PHR).

Chapter 2 compares the recovery of DNA, evaluated by NGS parameters, from different types of skeletal remains from newborn individuals. Two specific goals were defined based on intra and inter-individual approaches: 1) to investigate if there are differences between petrous bone and vertebra when both remains belong to the same infant individual, and 2) to investigate if petrous and other skeletal remains present differences and if there are differences between performance according to the age of the individuals (perinatal, infantile, juvenile and adult). Petrous bones from newborns gave superior results in comparison with other bones and they behaved equally or even better than petrous bones from adult individuals.

Chapter 3 describes the typing of the mtDNA control region (mtDNA-CR) by NGS from highly degraded skeletal remains coming from different chronologies by using the new forensic kit *PowerSeq™CRM Nested System* (Promega Corporation). Two NGS protocols alternative to the manufacturer's recommendation were evaluated. In addition, two analysis tools were compared to the NGS data analysis (an in-house pipeline and the GeneMarker HTS software). The results showed that the M3 protocol (35 cycles and longer times in some steps: 96ºC for 15 min, 35 cycles of 96ºC for 15 seconds, 60ºC for 35 seconds, 72ºC for 30 seconds, and 60ºC for 2 min) successfully recovered the whole mtDNA-CR from highly degraded skeletal samples of different chronologies. Moreover, our freely available in-house pipeline can provide variants concordant with the GeneMarker HTS software. It also can confirm the molecular damage of degraded DNA that is impossible to determine in the GeneMarker HTS software.

RESUMEN

En los últimos 40 años, los campos del ADN antiguo y forense han experimentado un impresionante desarrollo de sus métodos y procedimientos. Estos avances han sido desde una simple Reacción en Cadena de la Polimerasa (PCR) hasta la Secuenciación Masiva o de Nueva Generación (NGS), reduciendo los costes y tiempo invertido, he incrementando la información genética obtenida. Los objetivos de ambas disciplinas son diferentes, la genética forense se enfoca en la identificación con fines legales, mientras que la Paleogenética tiene objetivos más amplios, tales como la reconstrucción de las poblaciones o el estudio del pasado evolutivo. Esta divergencia lleva al uso de diferentes metodologías. La genética forense necesita la estandarización y validación de procedimientos entre laboratorios. El campo del ADN antiguo siempre ha tenido más propósitos que la genética forense y ha desarrollado más rápidamente procedimientos y nuevas plataformas y metodologías. Por el contrario, el ADN forense es conservador y conduce a métodos de consenso, que a menudo son difíciles de cambiar. Aunque el análisis de ADN antiguo (aDNA) y el genotipado del ADN forense tradicionalmente han requerido diferentes conjuntos de marcadores moleculares de ADN para lograr sus respectivos objetivos, hay varios puntos en común que se han evidenciado.. Ambos campos se enfrentan a varios problemas técnicos, principalmente la recuperación de ADN degradado de restos óseos. En este tipo de muestra, las condiciones ambientales hostiles, como la larga exposición a radiación UV, los gradientes de temperatura extremos, el pH bajo/alto y la humedad, han desencadenado la degradación del ADN, lo que ha dado lugar a diversas modificaciones bioquímicas de los nucleótidos dificultando la recuperación de material genético. Dado el gran éxito de la comunidad de aDNA en la recuperación de ADN humano degradado, las muestras forenses críticas podrían tratarse como genomas antiguos en el laboratorio porque su ADN también está muy fragmentado en cantidades diminutas y acumula daño molecular. Esta tesis doctoral tiene como objetivo conectar las actuales prácticas metodológicas usadas en la investigación del ADN antiguo y la genética forense para aumentar la recuperación de información genética de restos óseos altamente degradados. Las técnicas probadas podrían ser una ruta alternativa para lograr la identificación genética de víctimas de conflictos armados, personas desaparecidas o víctimas de desastres masivos. Se evaluaron diferentes factores que podrían afectar la recuperación del ADN degradado.

El Capítulo 1 aborda una comparación sistemática entre tipos de restos óseos (hueso petroso, diferentes partes del diente y huesos poscraneales) y cuatro métodos de extracción basados en protocolos de sílica en suspensión y en columna, y el protocolo orgánico tradicional (Ph-Cl), utilizado comúnmente en genética forense hasta ahora. El tipado de STR autosómicos, un fragmento de ADN nuclear y la región control del ADNmt (HVRI y HVRII) fueron marcadores moleculares utilizados para evaluar si el éxito de la recuperación depende del tipo de restos óseos o del método de extracción específico utilizado, o la combinación de ambos factores. Los resultados mostraron que los métodos basados en la unión sílica-ADN, específicamente el método de sílica en suspensión, recuperaron la mayor cantidad de fragmentos cortos, característico del ADN altamente degradado, propio de muestras forenses y antiguas. El método de sílica en suspensión obtuvo el mayor número de amplificaciones positivas de los fragmentos de ADNmt (HVRI y HVRII). Por el contrario, el método de fenol-cloroformo (Ph-Cl) se asoció con los peores resultados. Considerando el tipo de hueso, los huesos petrosos y la cavidad pulpar mostraron los mejores resultados en la concentración de productos amplificados de ADNmt (HVRI y HVRII) usando los métodos de sílica en suspensión y sílica en columna de membrana (HE). En paralelo, la costilla no mostró amplificación de HVRI usando los métodos de sílica en columna de membrana (XS plasma) y fenol cloroformo. Los perfiles de STR autosómicos obtenidos a partir de los métodos de sílica en suspensión y en columna de membrana (HE) mostraron el mayor porcentaje de alelos reportables con los valores más altos del radio de la altura del pico medida en RFUs (PHR).

El **Capítulo 2** compara la recuperación de ADN, evaluado a través de parámetros de NGS, a partir de diferentes tipos de restos óseos de individuos recién nacidos. Se definieron dos objetivos específicos basados en enfoques intra e interindividuales: 1) investigar si hay diferencias entre el hueso petroso y la vértebra cuando ambos restos pertenecen al mismo individuo infantil, y 2) investigar si el hueso petroso y otros restos óseos presentan diferencias y si existen diferencias entre el rendimiento según la edad de los individuos (perinatal, infantil, juvenil y adulto). El hueso petroso de individuos recién nacidos mostró resultados superiores en comparación con otros huesos y se comportaron igual o mejor que los huesos petrosos de individuos adultos.

El **Capítulo 3** describe el tipado de la región control del ADNmt por NGS, a partir de restos esqueléticos altamente degradados de diferentes cronologías utilizando el nuevo kit forense *PowerSeq™CRM Nested System* (Promega Corporation). Se evaluaron dos protocolos de NGS alternativos al protocolo recomendado por el fabricante. Además dos herramientas de análisis fueron comparadas para analizar datos generados por NGS (un pipeline elaborado en nuestro grupo de investigación, y el software GeneMarker HTS). Los resultados mostraron que el protocolo M3 (incluye 35 ciclos y tiempos más largos en algunos pasos: 96ºC por 15 min, 35 ciclos de 96ºC por 15 segundos, 60ºC por 35 segundos, 72ºC por 30 segundos y 60ºC por 2 min) recuperó exitosamente la región control completa del ADNmt a partir de restos óseos

altamente degradados de diferentes cronologías. Además, nuestro pipeline disponible on line puede proporcionar variantes concordantes con el software GeneMarker HTS. También puede confirmar el daño molecular de ADN degradado que no es posible determinar en el software GeneMarker HTS.

INDEX

LIST OF TABLES

LIST OF FIGURES

1

Introduction

"It has long been an axiom of mine that the little things are infinitely the most important."
-Sherlock Holmes (A case of Identity)

1.1 A brief overview of Ancient and Forensic DNA

Every life form on Earth is written in a unique and singular language: THE DNA. It contains the information necessary for species and life to exist, evolve, adapt, and be sustained over time. The discovery of DNA structure (Klug, 1968), the invention of PCR (Templeton, 1992), the sequencing of the human genome (Venter et al., 2001), and the ultimate advances in molecular biology techniques have led to remarkable changes in life science and enabled the development of a wide range of diverse research. Today, through paleoanthropology and molecular anthropology, a deeper understanding can be gained of when and where our existence as modern humans and species originated and how we differed from and interacted with other forms of humans that are now extinct.

From the mid-1980s to the present, molecular anthropology has enabled the recovery of genetic information from a wide range of archaeological remains or subfossil materials, including hard tissue—bones, teeth (Drancourt et al., 1998; Habelberg et al., 1989) — soft tissue—mummified skin and hair (Gilbert et al., 2004)—or sediments (Willerslev et al., 2003). The genetic material preserved in these samples is an ultra-short and degraded DNA known as ancient DNA (aDNA). aDNA research combines paleoanthropology and molecular anthropology, and its application to ancient humans and archaic hominins has flourished in the last decade (Ermini et al., 2015). Genetic patterns of variation in modern populations can help reconstruct this roughly 200,000-year-old population history. However, by merging these data with genetic data from ancient remains, we can now directly access our evolutionary past and learn about our population history in much greater detail. For example, the discovery of new relationships between present-day species and other human forms that existed until about 30,000 years ago can be analyzed and incorporated into the human phylogenetic tree

of extinct species (Pääbo, 2014) —e.g., Denisovans and Neanderthals—(Meyer et al., 2012; Sawyer et al., 2015). In addition, it is possible to learn more about the paths our ancestors took in colonizing the planet, who they interbreed (Nielsen et al., 2017), how they domesticated plant and animal species (Frantz et al., 2020; Orlando et al., 2015; Pedersen et al., 2014), how they responded to lifestyle and environmental changes, and what diseases or pathogens devastated their populations (Spyrou et al., 2019). All these strategies have the potential to resolve many pending controversies about their origins that cannot be traced with modern patterns of genetic variation alone, providing a formidable toolkit for a new generation of molecular anthropologists.

DNA technology has opened a new front in forensic science almost simultaneously with aDNA research. Non-human and human genetic variations are known as alleles, polymorphisms, or mutations and allow us to differentiate within and between populations and as unique individuals. An example of this was given in 1984 when Sir Alec Jeffreys stated that individuals could be distinguished based on easily detectable differences (alleles) in their DNA. As proposed by Jeffreys, DNA profiling consists of molecular analysis of multiple polymorphic sites for genetic identification (Jeffreys et al., 1985a). In this way, no two people are genetically identical; even identical twins (developing from the same zygote) have unusual genetic variation due to mutations during development and variation in gene copy-number (Bruder et al., 2008; Weber-Lehmann et al., 2014). Moreover, these variants are inherited from generation to generation, making it possible to study kinship relationships.

The first criminal case in which DNA profiling was used occurred between 1983 and 1986 in the United Kingdom. In this case, DNA analysis allowed Richard Buckland to be exonerated and Colin Pitchfork to be subsequently convicted of the rapes and murders of Lynda Mann and Dawn Ashworth. Since then, the field of molecular identification appears to have undergone an almost limitless technological improvement, allowing forensic experts to improve the performance of analysis and genetic identification. When sufficient DNA is available, the analysis of a DNA profile from a single person's sample is straightforward and provides meaningful scientific evidence to exclude or include a person as a possible

source of that DNA. This DNA analysis includes the match probability calculation; this probability indicates how rare a matching DNA profile is in a specific population.

Currently, a wide variety of evidence can be used as sources of DNA. However, minute amounts of degraded DNA have been found in samples such as cigarette butts, fingerprints, or bone remains. DNA profiling in such samples yields no or incomplete DNA profiles with a low success rate. This represents one of the greatest challenges of molecular biology in anthropological analysis, both forensically and archaeologically. For this reason, this dissertation will analyze DNA from bones and teeth. Until recently, molecular biology techniques commonly used in forensic laboratories were applied to bones or tooth samples collected from archeological sites to establish population or kinship relationships (Afonso et al., 2019). Nowadays, aDNA research has greatly improved its analysis through Next Generation Sequencing (NGS) techniques, which allows for determining some prehistoric kinships at the archaeological level (Monroy Kuhn et al., 2018).

In contrast to the aDNA community, the application of NGS in forensics has been slower, mainly due to the lack of accredited sequencers, kits, and relatively higher sequencing error rates compared to standard Sanger sequencing. Nowadays, most of these problems have been solved, and forensic scientists increasingly use NGS.

Although aDNA analysis and forensic DNA profiling have traditionally required different sets of DNA markers and treatment to achieve their respective goals, a great number of commonalities exist. For instance, mitochondrial DNA (mtDNA) analysis and Single Nucleotide Polymorphisms (SNPs) are markers shared by ancient and forensic DNA. In addition, both fields face various technical problems, mainly the recovery of degraded DNA from skeletal remains. In this type of specimen, harsh environmental conditions such as long-term UV irradiation, extreme temperature gradients, low/high pH, and humidity have triggered DNA degradation, resulting in various biochemical nucleotide modifications and making DNA recovery difficult. Given the great success of the aDNA community in recovering degraded human DNA, the critical forensic samples could be treated as ancient

genomes laboratory-wise because their DNA is also highly fragmented in minute amounts and accumulates molecular damage.

1.2 The Human Genome

With the advent of the genomic era, genetics has shifted from proteins to sequence polymorphisms as study tools. The number of polymorphisms discovered in recent years is enormous due partly to the development of the Human Genome Project (HGP), a collaborative international project launched in the 1990s to describe all human genomic variability. The goal of **the HGP** was to **identify more than 20,000 human genes**, make them accessible for further biological study, determine the sequences of the 3 billion chemical base pairs that make up human DNA, and store all this information in databases. The much smaller **mitochondrial genome** was also **sequenced in the early 1980s** (Anderson et al., 1981). Both projects completed a draft of the human genome sequence in 2001, and the results were published in the scientific journals *Nature* and *Science* in February of that year (Brown, 2002; Venter et al., 2001). Although the HGP was completed in 2003, the data analysis will continue for many years. Just a few weeks ago, on 1 April, Science published a special issue "Completing the human genome" in each Telomere-to-Telomere (T2T) Consortium presents a complete 3.055 billion-base pair sequence of a human genome, that provides for the first time the sequence of heterchromatic regions of the human genome (Nurk et al., 2022).

1.2.1. Nuclear DNA

Nuclear DNA (nDNA) is the genetic material contained in each nucleus of a eukaryotic organism (Figure 1). It encodes most of the **genome in eukaryotes**. In Humans, its sequence is packaged in **23 distinct chromosome pairs** in each diploid cell **(2n).** Of the 23 chromosome pairs, 22 are autosomal chromosomes; one pair of sexual chromosomes determines gender —**XX** for females and **XY** for males—. Half of a child's DNA comes from the mother and half from the father and full siblings share, on average, half of their DNA. The **haploid genome (n)**, a single copy per chromosome, comprises a total length of about 3.4 Gb base pairs and 20465-24849 genes (Howe et al., 2021). The human genome is commonly divided into **coding** or **exome** and **non-coding** or **"junk" DNA sequences**.

1.2.1.1 Coding and non-coding regions

Exome or **coding DNA** is the sequence that can be transcribed into mRNA and translated into proteins during the human life cycle; these sequences make up only a small portion of the genome **(<2%)**. These regions, also known as protein-coding sequences, are the most studied and understood component of the human genome. All human proteins are produced from these sequences by various biological processes such as alternative pre-mRNA splicing or DNA rearrangements, etc. The production of unique proteins may even exceed the number of protein-coding genes (Pierce, 2003).

All DNA sequences within a genome that are not found in exomes are **non-coding DNA** (ncDNA). These regions sometimes called "junk DNA" are interspersed throughout the genome. The non-coding regions are transcribed but are neither translated nor directly involved in the translation process and therefore are not included in the amino acid sequence of functional proteins. More than **98% of the human genome** consists of **ncDNA**, and some of these regions have a demonstrated biological function. Certain regions produce transcripts directly involved in RNA processing and translation rather than expressed in messenger RNAs encoding proteins. These regions include transfer RNAs (tRNA), ribosomal RNAs (rRNA), small nuclear RNAs (snRNA), small nucleolar RNAs (snoRNA), etc. There are other classes of non-coding sequences such as pseudogenes, introns, **repetitive DNA sequences**, untranslated regions of mRNA, and regulatory DNA sequences (Alexander et al., 2010; Shanmugam et al., 2017). Bellow, some of these regions will be described.

- ***Pseudogenes*** are inactive copies of protein-coding genes created by gene duplication. These regions are nonfunctional and accumulate inactivating mutations (insertions and deletions). The number of pseudogenes in the human genome is about 15217 (Howe et al., 2021), and in some chromosomes, it almost equals the number of functional protein-coding genes. Gene duplication is an important mechanism for generating new genetic material during molecular evolution (Pei et al., 2012).

- ***Introns*** are detached by the process of "splicing" leaving only the "exons" which are translated further into proteins. Introns consist of large stretches of DNA whose biological functions are in the process of being elucidated.

- ***Repetitive DNA sequences*** can be divided into tandem repetitive sequences (known as satellite DNA) and interspersed repeats. These regions represent approximately **50% of the human genome**, of which only **8% are tandem sequences**. The term satellite describes DNA sequences that comprise short head-to-tail tandem repeats incorporating specific motifs. These make up one-third of DNA repeats and are exemplified by the minisatellites and microsatellites. The latter is the most relevant because is highly variable, even among closely related individuals, being widely used for genealogical DNA testing, forensic DNA analysis, and medicine. A summary of the satellite DNA repeats is given in Table 1.

Table 1 — Types of repetitive DNA regions located in the genome. Adapted from Trend, 2012.

Type of variation	Description
Minisatellite	It is a type of VNTR (variable number of tandem repeats) with 9-100 bp units repeated in tandem thousands of times. This large polymorphism is found mostly in centromeres and telomeres.
Microsatellite or Short Tandem Repeats (STRs)	They are called SSR (simple sequence repeat). They involve small tandem repeats, e.g., 2–6 bp in size (Figure 7). For identification purposes, these microsatellites are used in gene discovery by linkage analysis, paternity testing or forensic DNA testing (section 1.4.4.1).
Copy number variations (CNVs)	Deletions and duplications in the Kb to Mb range are carried out by these structural variants, resulting in the change in the copy number for a specific genome region. The size CNVs contribute more than SNPs in the variation of the genome as well as functioning as polymorphisms. These changes can cause genetic disease by interfering with gene function or dosage (gene copy number) effects. (Alexander et al., 2010; Mills et al., 2007)
Interspersed DNA repeats	There are a variety of interspersed DNA repeats. Long interspersed elements (LINEs) occupy about 15% of the human genome. They have been inserted randomly into eukaryotes during evolution, e.g., retrotransposons that encode reverse transcriptase and can function as polymorphisms depending on their presence or absence in the genome. Active LINEs enlarge the genome by copying themselves to new sites. Short interspersed repeats (SINEs) comprise about 10% of the human genome. A type of SINE is an Alu element of 300 bp in size of genes transcribed by RNA polymerase III that depend on LINE elements for their replication. Long terminal repeat (LTR) retrotransposons are flanked at both ends by direct LTRs. The latter can be generated by insertion, deletion, reciprocal translocation, or inversion.

1.2.2 Mitochondrial DNA

Mitochondria are cellular organelles located in the **cytoplasm** of eukaryote cells containing a small circular chromosome separated from the nuclear genome (Figure 1). They originated from endosymbiotic bacteria in proto-eukaryotic cells about 1.5 billion years ago and were used for energy production. Many features of mitochondria resemble those of bacteria: the circular rather than linear genome, the absence of histones, the discrete origins of replication (the absence of introns), the absence of scattered repeats, and the very small amount of inter-genic DNA (Archibald, 2015).

Mitochondria play a critical role in cellular energy provision for numerous cellular functions, including **ATP production**, cellular homeostasis, and apoptosis. Human mitochondrial DNA (mtDNA) is a five mm histone-free circular double-stranded DNA molecule with around **16569 base pairs in length and weighing 10^7 Daltons**. The human mitochondrial genome was first completely **sequenced in 1981** (Anderson et al., 1981); the Cambridge Reference Sequence (CRS), often referred to as the Anderson sequence, was later reanalyzed to confirm and correct the sequence, named the revised Cambridge Reference Sequence (rCRS) (Andrews et al., 1999).

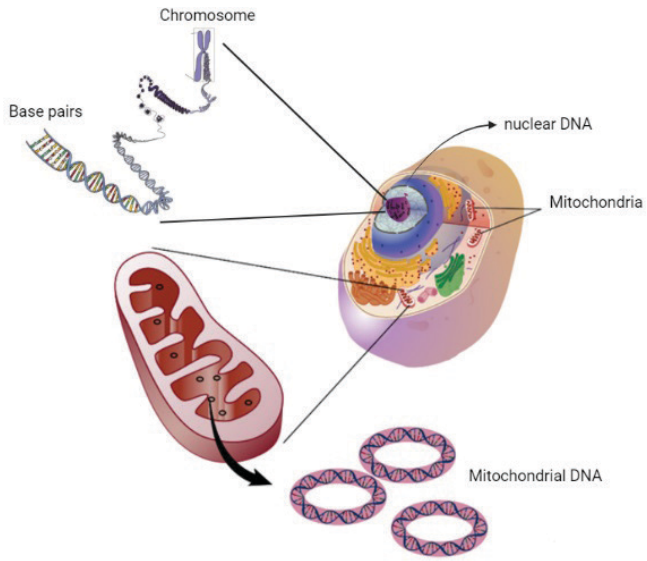


Figure 1 — Basic representation of the human genome of a eukaryote nucleated cell showing nuclear DNA and mitochondrial DNA. Adapted from William Gahl, 2021.

MtDNA consists of **two strands**, which have different densities depending on the G+T base composition in each strand. The **heavy (H) strand** encodes more information, with genes for two rRNAs (12S and 16S), twelve polypeptides, and fourteen tRNAs, while the **light (L) strand** encodes eight tRNAs and one polypeptide. All the 13 protein products are part of the enzyme complexes that constitute the oxidative phosphorylation system (Taanman, 1999). The origin of replication is positioned in the **non-coding region**, an **1121 base pairs** segment located between positions **16024 and 576** (Figure 2). According to Anderson (Anderson et al., 1981), nucleotide positions in the mtDNA genome are numbered, later slightly modified by Andrew (rCRS) (Andrews et al., 1999). The mtDNA sequence is published in GenBank NC012920 (Lott, 2019).

The **mutation rate** in the mitochondrial genome is significantly higher than in the **nuclear genome (0.5×10^{-9}/bp/year)** (Scally, 2016) due to the lack of repair mechanisms and the low fidelity of the mtDNA polymerase. The mutation rate for the **non-coding region is 4.1878×10^{-6} mutations/site/generation**. If 675 bp of **HVRI** and **HVRII** are considered, the mutation rate becomes **6.0367×10^{-6} mutations/site/generation** (Santos et al., 2005). Most of the sequence variability between individuals is found in two specific regions of the control region, called the hypervariable 1 (HVR1), which extends from 16024 to 16365, and the hypervariable 2 (HVR2) region lies from 73 to 340. A third hypervariable region (HVR3) located in positions 438 to 574 has additional polymorphic positions that can be useful when mtDNA sequences do not differ in the HV1 and HV2 regions (Court, 2021; Lutz et al., 2000). The high variability in the non-coding regions explains why these regions are most frequently sequenced in forensic analysis and archaeological research.

In addition to rCRS and after refining the human mtDNA phylogeny, Behar and collaborators propose switching the reference to a **Reconstructed Sapiens Reference Sequence (RSRS)** identified by considering all available mitogenomes from *Homo neanderthalensis*. This "Copernican evidence," as the authors call it, is a reassessment of the human mtDNA tree from its deepest root; according to the authors, it will be able to resolve problems of public perception of human evolution by clarifying the core

principles of common ancestry for extant descendants (Behar et al., 2012), however, its use still being very limited.

1.2.2.1 Mitochondrial DNA Inheritance

The number of mitochondria per cell depends on the energic requirement of that cell. For instance, nerve and muscle cells require high energy to perform their biological functions, therefore, containing thousands of mitochondria with **2-10 copies of mtDNA**. In contrast, other cell types may contain only a few hundred. **Human eggs contain** more than **100,000 copies of mtDNA,** whereas **sperm** contain approximately **100 copies** (Hecht et al., 1984). Hence the paternal contribution compared to the mtDNA pool of the zygote is expected to be relatively small. Some studies show that its contribution is null (Birky, 1995; Greenberg et al., 1983).

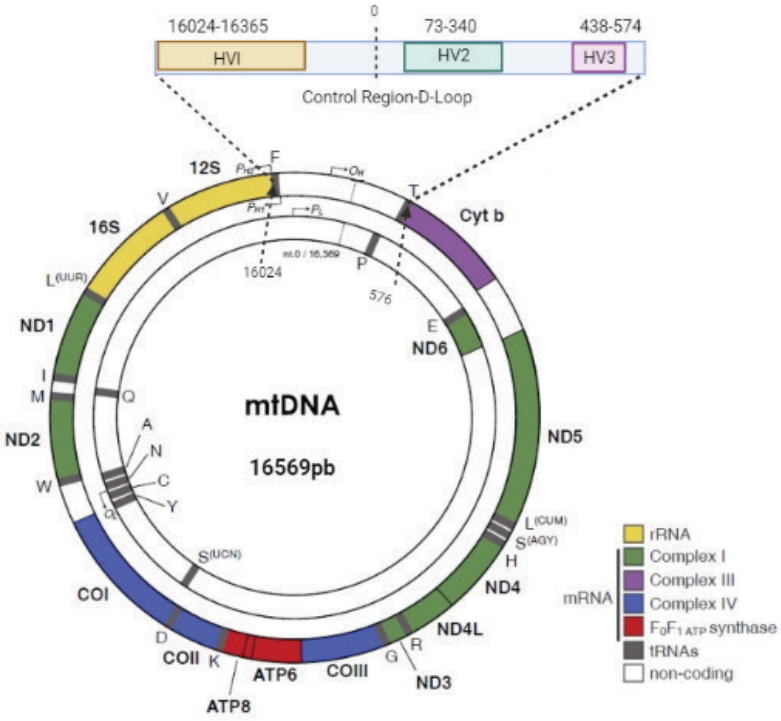


Figure 2 — Diagram of the basic structure of the mitochondrial genome to highlight the hypervariable regions (HV1, HV2, and HV3) that lie within the 'control region' or D-loop area. Adapted from Picard et al., 2016.

For a long time, it was considered that mtDNA is inherited in a non-Mendelian manner without recombination (Giles et al., 1980). Unlike the mammalian mitochondrial genome, mtDNA is recombined in other organisms such as plants, fungi, mollusks (Ladoukakis & Zouros, 2001), and fish. Moreover, the dogma of strict maternal mtDNA inheritance in humans was questioned when Luo and collaborators compelling evidence that, in rare cases, the father might pass on his mtDNA to the offspring (Luo et al., 2018).

The analysis showed that some individuals with heteroplasmy (the presence of differences in the mtDNA copies of the same individual) had inherited mtDNA from both parents, breaking the usual pattern of exclusive maternal inheritance of mtDNA. While some authors conclude that paternal mitochondrial transmission is unlikely to be a common occurrence and, therefore, at this point, they would not recommend changes in clinical practice (Rius et al., 2019) and, consequently, in evolution and forensic research. Therefore, and while there is not a probative study, the inheritance of mtDNA is considered strictly maternal (Figure 3) and individuals harbor only one mtDNA: that of the mother.

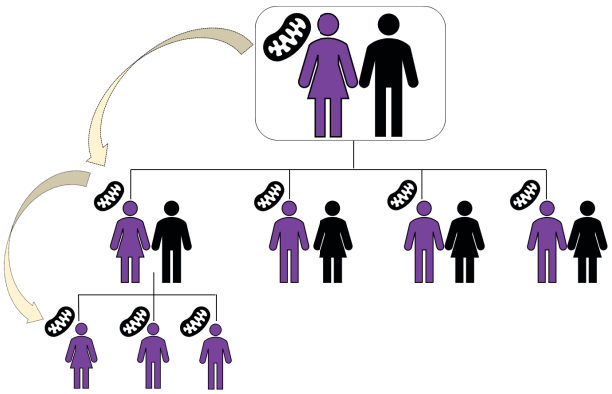


Figure 3 — Illustration of mtDNA inheritance over two generations indicates in purple the individuals that share the same mtDNA haplotype from mother to offspring.

1.2.2.2 Heteroplasmy and Homoplasmy

The strict maternal transmission of mtDNA results in **homoplasmic individuals**, who typically have a **single mtDNA haplotype**, the maternal one. However, **heteroplasmy** is

the simultaneous presence of **two or more types of mtDNA** in the same individual and has often been reported in some species, including humans (Ramos et al., 2012; Santos et al., 2008). There are two types of heteroplasmy: length and point or sequence (Bendall et al., 1996; Bendall & Sykes, 1995; Melton, 2004; Parakatselaki & Ladoukakis, 2021). Length heteroplasmy often occurs around the homopolymeric C-stretches in HV1 at positions 16184 to 16193 and HV2 at 303 to 310 and consists of the insertion/deletion of cytosine. A heteroplasmy results from a point mutation where a single base is substituted in a portion of the genomes, demonstrating two different bases in the same position (Figure 4). Recent advances in sequencing techniques have increased the ability to detect and quantify heteroplasmy using Next Generation Sequencing (González et al., 2020).

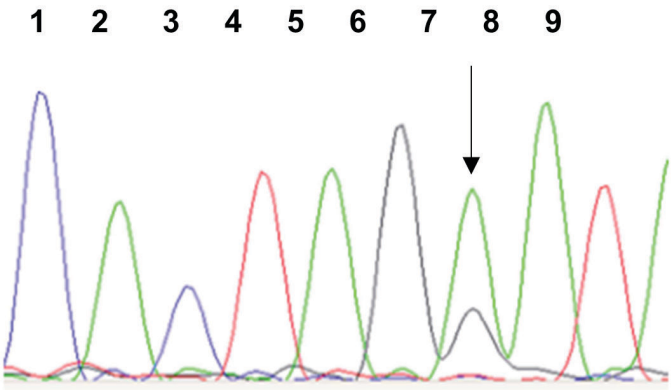


Figure 4 — Sanger sequencing electropherogram of a sample showing point heteroplasmy in the 7th base shows two peaks representing bases A (in green) and G (in black). C and T bases are represented in blue and red, respectively. Adapted from Court, 2021.

Heteroplasmy commonly occurs through somatic mutations during an individual's lifetime. Somatically, the level of heteroplasmy can be different in the different cells from the same tissue or organ, between organs from the same person, or even between individuals of the same family (Wallace & Chalkia, 2013).

In forensic human identification, only point heteroplasmy is important; the length heteroplasmy or overall length polymorphisms are not mandatory information because they do not influence the impact in haplogroup definition (Parson et al., 2014).

Reported heteroplasmy at two sites in the same individual may occur; that condition is known as *triplasmy* and has been reported by Tully et al., 2000. This genetic event occurs at lower frequencies than single-site heteroplasmy, and it is normally rare to find more than one heteroplasmic position in the 610 nucleotides sequenced for HV1 and HV2 (Tully et al., 2000).

1.2.2.3 MtDNA sequence nomenclature

A **haplotype** is a **set of single nucleotide polymorphisms (SNPs)** strictly specific to populations because of uniparental inheritance. Polymorphism or mutations are terms used for variants in a population with a **frequency of 1%** or higher naturally occurring with a neutral or beneficial effect (Brookes, 1999). Thanks to these spontaneous variants are possible, for instance, to track the human matrilineal inheritance back to its origins in Africa and describe the migration or spread across the world.

The haplotype of an individual is obtained from the mtDNA sequence and **is reported as SNPs differences found in front of the revised Cambridge Reference Sequence (rCRS**). The haplotypes are clustered by the combination similar to single nucleotide polymorphisms (SNPs) in mtDNA inherited from a common ancestor, giving place to the generation of haplogroups that originated as a result of the sequential accumulation of mutations by maternal linage (Mitchell et al., 2014). The fast accumulation of nucleotide substitutions makes it possible to distinguish populations that had been separate in relatively recent times. In other words, if different populations share one or more mitochondrial haplogroups, it is possible to indicate that their divergence may be relatively recent.

The process of naming mtDNA sequences seems simple and obvious; however, it is crucial because it can lead to ambiguities resulting in different analysts describing the same samples differently although they agree on the nucleotide. Most ambiguous alignment/nomenclature arises due to insertions and deletions (indels). Phylogenetic-based nomenclature guidelines have been proposed to avoid errors by insertions/deletions,

facilitate haplotype identification, and well-assign this haplotype to existing haplogroups or new haplogroups (Bodner et al., 2016).

In the forensic community, there are some rules published that advise the naming of homopolymeric C-stretches regions and provide nomenclature recommendations on substitutions and indels, transitions, and transversions. They recommend a blended application of rule-based and phylogenetic approaches using the EMPOP database. As a result, they introduce guidelines regarding sequence generation, and control measures based on the known worldwide mtDNA phylogeny that can be applied to ensure the highest population data quality. These annotations are used for storing haplotypes in the EMPOP and have been adopted by the Scientific Working Group on DNA Methods (SWGDAM) in the United States (Parson et al., 2014).

Regarding the categorization, the **haplogroups lineages** have been named in letter order, **A to Z**, according to their discovery. The most recent common ancestor from which all mtDNA originated belonged to haplogroup L, known as 'Mitochondrial Eve' and dated to ~120 000 to 156 000 years ago (van Oven & Kayser, 2009). African and especially sub-Saharan African populations are clustered into one of the main haplogroup lineages that diverge from **macro-haplogroup L: L0, L1, L2, L3, L4, L5, and L6** (Allard et al., 2005; Bandelt et al., 2001; Behar et al., 2008; Chen et al., 2000; Gonder et al., 2007; Pakendorf & Stoneking, 2005; Rosa et al., 2004; van Oven & Kayser, 2009). More than 90% of the individuals of the European and USA Caucasian populations are categorized into ten main haplogroup lineages**: H, I, J, K, M, T, U, V, W, and X** (Behar et al., 2008; Budowle et al., 2002; Pakendorf & Stoneking, 2005). Concerning African-American populations, the most observed haplogroups are L2a, L1c, L1b, and L3b (Allard et al., 2005). The main haplogroups found in Asian populations are haplogroups M and N (Allard et al., 2004; Kivisild, 2015). Likely migration routes of the main haplogroups are illustrated in Figure 5.

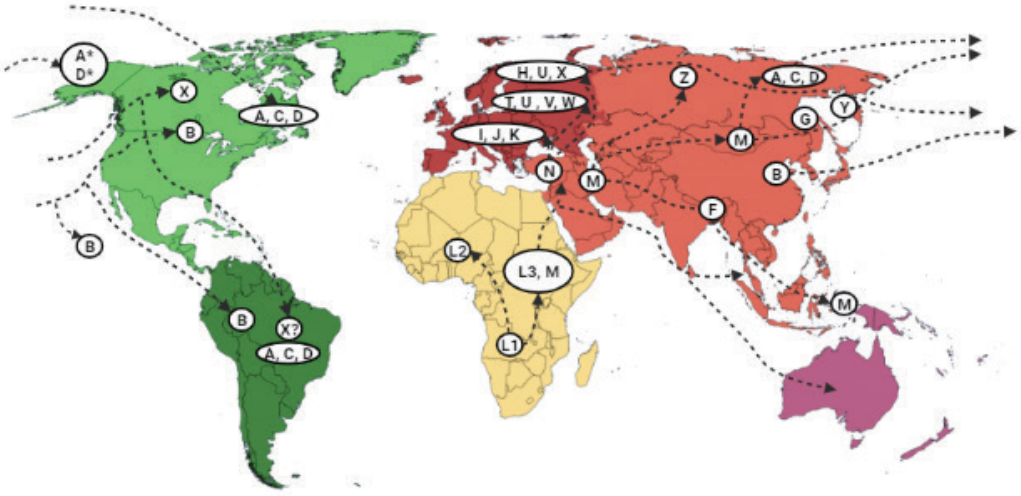


Figure 5 — Representation of the geographical origin of major mtDNA haplogroups across the world and probable migration routes based on Court, 2021.

1.2.2.4 MtDNA Population Data and Database

Implementing mtDNA databases can entail privacy questions when genetic information of any nature is published (Guillén et al., 2000; Wallace et al., 2014). For instance, it is known that mitochondrial diseases affect between 1 in 4,000 and 1 in 5,000 people, and this genetic condition can be inherited; this information about the mitochondrial genome composition may therefore enable the identification of the current or future state of health of an individual. In other cases, such as forensic identification and population, databases play an important role in estimating the expected frequency of mtDNA haplotypes when the suspect's mtDNA sequence match that of an evidentiary sample. Some informatics tools provide genetic information confidentially to solve different scientific questions either with forensic applicability or population. **EMPOP**, **Mitomap**, **HmtDB**, and **AmtDB** are human mtDNA databases while **PhyloTree** and **Haplogrep** are classification tools of mtDNA haplogroups.

EMPOP is the most comprehensive mtDNA database worldwide and the most useful forensically. **This database was developed through the European DNA Profiling Group (EDNAP)** and is freely accessible from the website (www.empop.online). EMPOP uses a

string-based search algorithm (SAM2) that converts sequences into alignment-free strings to ensure that haplotypes are properly defined regardless of alignment differences. It was designed to serve as a reference population database to be used in the evaluation of mtDNA forensic evidence around the world. The high mtDNA quality data lead to EMPOP serving as a reference population database and quality-control tool for scientists in forensic genetics and other disciplines (Parson & Dür, 2007). At EMPOP, the tool haplogroup browser represents all the established Phylotree haplogroups and provides the number of EMPOP sequences assigned to the respective haplogroups by estimating mitochondrial DNA haplogroups using the maximum likelihood approach (Röck et al., 2013). In addition, the most recent common ancestor (MRCA) haplogroups are provided for multiple possible haplogroups. The geographical haplogroup patterns can be visualized via maps to well-understand their geographical distribution.

Mitomap is another important human mtDNA database (www.mitomap.org). This online database developed in 1996 published human mtDNA variation and geographic and disease-specific variants. Currently, **mitomap is frequently updated and high-quality human mtDNA data is avaliable for clinicians, investigators, and geneticists**. A wide list of mtDNA variants from both healthy individuals and patients is stored, and the frequencies of the variants are calculated from human mitogenomes retrieved from the GenBank (Lott, 2019; Ruiz-Pesini et al., 2007).

HmtDB (Human mtDNA Database) (https://www.hmtdb.uniba.it/) is an open resource created to support population genetics and mitochondrial disease studies. The **database hosts human mitochondrial genome sequences** annotated with population and variability data. Classifier tools implemented in HmtDB allow the prediction of the haplogroup for any human mitochondrial genome currently stored in HmtDB or externally submitted de novo by an end-user. Haplogroup definition is based on the Phylotree system.

AmtDB (Ancient mtDNA DataBase) is the first database of ancient human mitochondrial genomes. This database **contains 1107 mtDNA ancient genomes** (data found on the webpage on 03 April 2022) that are freely accessible for download, together

with the individual descriptors, including geographic location, radiocarbon dating, and archaeological culture affiliation. Also comprised of an interactive map for sample location visualization. AmtDB is a key platform for ancient population genetic studies and is available at **(https://www.amtdb.org).**

PhyloTree (www.phylotree.org) is an available online tool that comprises the human mtDNA phylogeny, based on whole mtDNA sequences (van Oven & Kayser, 2009). Provides a **theoretical phylogenetic continuously updated with new information on haplogroups or variants**, offering information on how these newly identified haplogroups have evolved from a series of common ancestors.

Haplogrep is another fast and free haplogroup classification tool **(http://haplogrep.uibk.ac.at)** where mtDNA profiles can be uploaded and aligned to rCRS or RSRS and receive mitochondrial haplogroups in return, subsequently are classified into phylogenetic clusters (haplogroups). This tool is highly relevant for evolutionary, forensic, and medical genetics. Today, **HaploGrep 2 (https://haplogrep.i-med.ac.at/)** offers several advanced features, including a system of quality control (QC) that allows detecting artificial recombinants, missing variants, annotating rare and phantom mutations. The handling of high-throughput data in VCF files is now directly supported; moreover, it generates a publication-ready phylogenetic tree of all input samples encoded relative to the revised Cambridge Reference Sequence (Weissensteiner et al., 2016).

1.2.3 Comparison between nuclear and mtDNA

Comparing nuclear and mitochondrial DNA, some fundamental differences stand out, coming from their respective functions. They affect the size, the number of copies per cell, heritability, the repair mechanisms and the mutation rate (Table 2). The result of mtDNA typing is more successful than nuclear DNA because of the higher number of copies per cell. A remarkable similarity in both genomes is the presence of coding and non-coding regions, being the non-coding regions highly variable and greatly studied for identification purposes in forensic and archaeological research.

Table 2 — Comparison of some basic characteristics of nuclear DNA and mitochondrial DNA.

Characteristic	Nuclear DNA	Mitochondrial DNA (mtDNA)
Size genome	~3.2 billion bp	~16,569pb
Copies per cell	2 (1 allele from each parent)	Can be >1000
Percent of total DNA	99.75%	0.25% content per cell
Structure	Linear; packaged in chromosomes	Circular
Inherited from	Father and mother (except Y-chromosome)	Mother
Chromosomal pairing	Diploid (2n)	Haploid (n)
Generational recombination	Yes	No
Replication repair	Yes	No
Unique	Unique to the individual (except identical twins*)	Not unique to the individual (same as maternal relatives)
Mutation rate	Low (although variable between regions)	At least 5-10 times more than nuclear DNA
Reference sequence	Described in 2001 by the Human Genome Project	Described in 1981 by Anderson and co-workers

1.3 Ancient DNA

1.3.1 History of ancient DNA: from paleogenetics to paleogenomics

Evolutionary records, adaptative mechanisms, or species migration are phenomena that have questions and concerns yet to be resolved. Researchers' hypotheses to resolve these doubts have usually been based on morphological and archaeological data until the standing out of molecular biology as a powerful tool to understand the genetic code. Thus, molecular biology applied to the archaeological context gave rise to a branch of knowledge known as paleogenetics. This discipline focuses on the study of the past through the analysis of old biological remains.

Almost four decades ago, the birth of ancient DNA was the first attempt using bacterial cloning to amplify DNA sequences from the Quagga, an extinct member of the horse family (Higuchi et al., 1984). The short fragments retrieved from the dried muscle of this species of Zebra were limited and showed that the genetic material surviving in ancient remains belonged principally to the microbial or fungal origin and that the endogenous DNA was limited to **very low concentrations of short and damaged fragments** (Willerslev & Cooper, 2005).

With the advent of the **Polymerase Chain Reaction (PCR)**, Svante Pääbo, Russell Higuchi, and Allan Wilson (Higuchi et al., 1984; Pääbo, 1989) were able to target specific genomic fragments of interest from archaeological and fossilized samples. Since then, an innovative way to extract and characterize the DNA was developed managing to elucidate the genetic information of extinct species, the relationship to contemporary species, or solving phylogenetic relationships within the genus *Homo*, throwing light on the origin of modern humans (Stringer & Andrews, 1988).

The invention of PCR made it possible to recover information even from a single surviving molecule with molecular damage. However, in ancient samples contaminated with modern DNA, the amplification of **endogenous DNA** will be masked with exogenous DNA providing **erroneous genetic information**. This fact generated some inconsistent and unreliable data in some aDNA studies (Gutiérrez & Marín, 1998; Zischler et al., 1995). Consequently, the aDNA community targeted much of its work and effort to understand the tempo and mode of DNA damage and repair, the longevity of DNA molecules depending on environmental conditions, and the dynamic of contamination of biological samples.

In this context Cooper and Poinar (2000) (Cooper & Poinar, 2000) proposed the improvement of practices in traditional lab-based methods, including PCR and Sanger sequencing, and some principles for quality ancient DNA research and confirmation of the authenticity of endogenous results among which we can highlight: i) the work with ancient DNA must occur in dedicated **"clean" lab facilities**, in which PCR is performed in a separate area, ii) independent extractions from separate sample materials should be performed, iii) replicability with additional sample material from the same individual, iv) results should show characteristics of ancient DNA behavior **(fragmentation and deamination)**.

Besides the improvement and advance in the knowledge of these important factors, the arrival and development of **massive parallel sequencing (MPS) or next-generation**

sequencing (NGS) have meant by far the most transformative and revolutionary technology in the history of aDNA community research (Orlando et al., 2021). PCR-based analysis and Sanger sequencing have been replaced almost entirely by genome-scale studies and NGS. NGS sequencing has shown to **be superior to Sanger** because of the capability for sequencing **smaller fragment sizes between 30-400 base pairs**, whereas the shortest possible fragment sizes with **Sanger are around 60-100 base pairs**. This fact represents an advantage for working with degraded samples (Hofreiter et al., 2015), which is very common in ancient DNA research. Therefore, **NGS technology** provides an **ideal platform for ancient DNA** because it favors the **recovery** of small fragment sizes **(30bp or above)** through library preparation techniques that allow for the capture of degraded DNA (Dabney et al., 2013; Hofreiter et al., 2015).

The ability to sequence DNA and analyze the entire genome from ancient specimens or even extinct species has allowed the emergence of paleogenomics. Nowadays, the time range of amenable DNA analysis extends to more than half a million years, including the whole mitochondrial genome from too degraded human specimens (Meyer et al., 2016; Orlando et al., 2015; Slon et al., 2017) and also complete genomes of extinct species such as woolly mammoths (Palkopoulou et al., 2015) or Neanderthals (Mafessoni et al., 2020). Modern sequencing technologies enable the sequence of any protein-encoding gene in any extant species. Ancestral sequence reconstruction allows researchers to determine the protein's sequences at various ancestral points in a phylogenetic tree. This approach allows us to understand how the functions encoded in ancient genomes have changed compared to those observed in present-day genomes (Selberg et al., 2021). Thus, the publication of partial and complete paleogenomes within the last few years has reinvigorated research in ancient DNA (Shapiro & Hofreiter, 2014).

1.3.2 Ancient DNA characteristics

While an organism is alive, damage to the DNA strands, such as UV radiation following exposure to the sun, is repaired via a suite of host repair mechanisms. DNA damage continues after death, but the repair pathways, which are energy-requiring, no longer

function. As a result, DNA strands accumulate various forms of damage after death. This damaged genetic material is known as aDNA, and few copies of this endogenous DNA tend to survive, containing a variety of chemical modifications. In addition, these DNA strands are short — fewer than 100 base pairs (bp)— in old specimens (Poinar et al., 2006). The difficulty in the recovery of this **damaged DNA** is a consequence of **post-mortem DNA degradation processes**, which can cause miscoding lesions, potentially leading to sequence errors, or physical destruction of the DNA molecule, thus increasing the risk for preferential amplification of exogenous contaminant sequences. However, there are authenticity criteria that demonstrate the reliability and authenticity of the aDNA recovery (Montiel et al., 2007).

In this sense, it is important to know the types of molecular damage that have been found in aDNA, since this damage is a signature of authenticity.

1.3.2.1 Ancient DNA damage

The chemical modifications such as hydrolysis and oxidation have been largely found in ancient samples (Dabney et al., 2013; Gilbert et al., 2003; Skoglund et al., 2014), resulting in different types of DNA damage. Table 3 shows the types of ancient DNA damage analyzed routinely to demonstrate that the results are obtained from authentic endogenous DNA instead of a contaminating DNA.

DNA fragmentation is owing to **hydrolytic depurination** and subsequent **β elimination** resulting in single-strand breaks (Figure 6A). Hydrolysis can also induce to deamination of Cytosine to Uracil, causing C to T transitions which miscoding lesions are commonly generated (Lindahl, 1993) (Figure 6B). Oxidation reaction can induce lesions that block polymerases and either stop amplification and sequencing or lead to '*jumping PCR*' generating chimeric sequences. The blocking lesions are presented as nucleotide modifications and cross-links, formed between DNA strands, fragments, or between DNA and other molecules (Dabney et al., 2013; Pääbo, 1989).

Table 3 — Type of ancient DNA damages. Adapted from Fulton & Shapiro, 2019.

Damage Type	Type of process	Effects on the DNA molecule	Solutions in aDNA classical sequencing methodologies
Strand Breaks	Nuclease activity	Low quantity of surviving DNA, short fragment length	Amplify short (<100-300pb) overlapping fragments
	Microorganism degradation		
	Desiccation, heat chemicals, etc.		
	Direct cleavage (Hydrolysis)		
	Depurination causes abasic site (hydrolysis)		
Miscoding lesions (hydrolysis)	Deamination causes miscoding lesions:	Base misincorporations	Multiple extractions and amplifications; cloning; UDG (uracil DNA glycosylase) to remove uracil
	1. Adenine to hypoxanthine	A to G	
	2. Cytosine to uracil*	C to T	
	3. 5-methylcytosine to thymine*	C to T	
	4. Guanine to xanthine	G to A	
Blocking and miscoding lesion (oxidation)	Base modifications	No amplification; jumping PCR	Special polymerases; cloning; multiple amplifications
	5-OH-5- methylhydantoin (blocking)		
	5-OH-hydantoin (blocking)		
	8-Oxoguanosine (miscoding G to T)	Base misincorporation	
Cross-links	DNA-DNA cross-links via alkylation	No amplification	PTB (N-phenacylthiazolium bromide) to cleave cross-links
	DNA-protein cross-link		

**Most observed ancient DNA damage manifests as C to T or complementary strand G to A transitions*

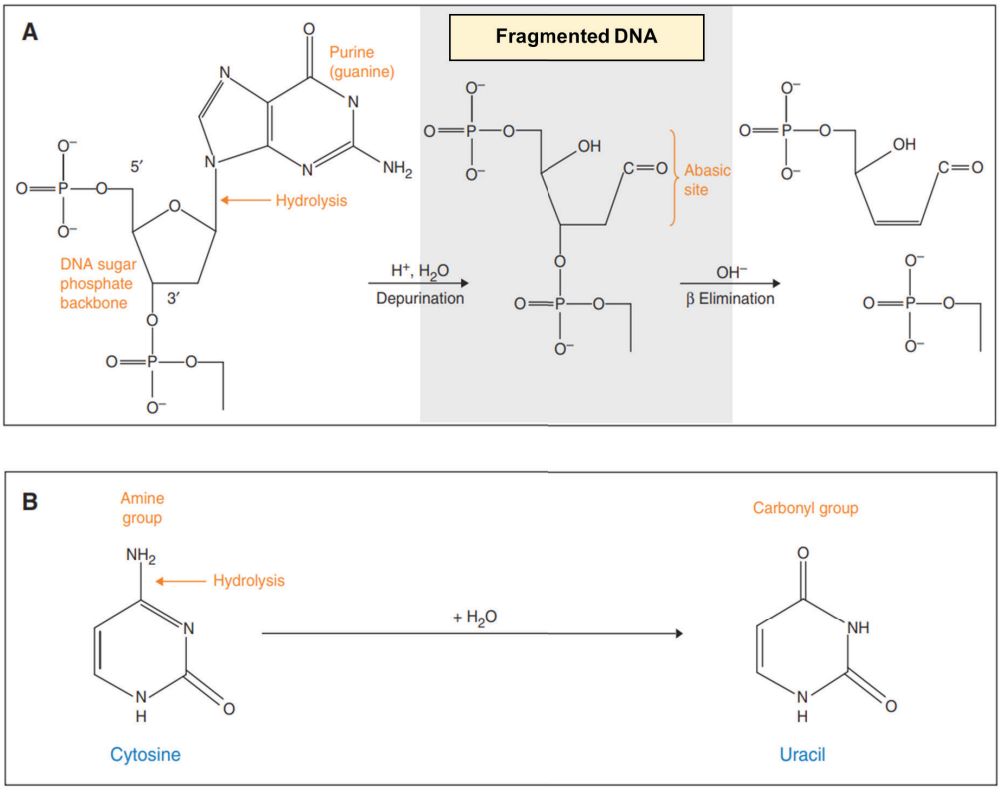


Figure 6 — Hydrolysis reactions induce fragmentation and deamination of ancient DNA (A) Depurination. The N-glycosyl bond between a sugar and an adenine or guanine residue is cleaved, resulting in an abasic site. The DNA strand is then fragmented through β elimination, leaving 3'-aldehydic and 5'-phosphate ends. (B) Deamination of cytosine to uracil leads to miscoding lesions. DNA polymerases will incorporate an A across from the U, and in turn, a T across from the A, causing apparent G to A and C to T substitutions. Adapted from Dabney et al., 2013.

The extent of **damage** is sample-dependent and **linked to preservation conditions**. Cold, dry, temperature-stable environments such as permafrost regions and caves are among the best places for well-preserved specimens and have permitted large-scale population studies (Bower & Jones, 2010). From these environments, reasonably well-preserved specimens with low contamination levels have yielded exceptional amounts of data when NGS techniques are applied. As techniques for recovering and authenticating ancient DNA improve, large-scale studies are progressively more attainable (Hofreiter, 2008).

1.3.2.2 Contamination

The high sensitivity of PCR allows amplification from only one or a few starting copies of the target sequence and allows contaminating DNA to be amplified. PCR will preferentially amplify modern DNA over damaged ancient molecules even if the level of contamination is extremely low. There are different sources of contamination that may be described.

Contamination by the sample itself: The sample itself may be contaminated, for instance, bones and teeth are porous, and contamination may occur by contact or uptake of exogenous DNA from microorganisms located in the environment and own microorganism of the decomposition of the skeletal remains.

Contamination during downstream experimental processes: Laboratory personnel may introduce their DNA or any DNA deposited into the lab, through their shoes or clothing, in any processes including DNA extraction, sequencing library preparation, or PCR set-up. Reagents may also be contaminated with human or animal DNA (Leonard et al., 2007).

Contamination by DNA aerosolized: A previously amplified DNA may be present in the laboratory environment (airborne particulates), being this a potential source of contaminating DNA. Even the tiny amount of DNA aerosolized when a tube is opened may contain over a million template copies. Strict separation should be maintained between pre and post-PCR labs, where samples are processed after amplification to prevent this problem.

1.3.3 Authentication criteria

For more than a decade ago, twelve authenticity criteria were established by Cooper and Poinar (Cooper & Poinar, 2000) (Table 4). Criteria relevant to ancient DNA research in the next-generation sequencing era will be discussed afterwards. However, because not all ancient DNA research uses NGS, several criteria are more applicable to traditional PCR-based analysis. The criteria 3, 4, 5, and 10 are focused on PCR-based methodology, the criteria 7, 11, and 12 could be applied to NGS-technique while the criteria 1,2, 6, 8, and 9 are for both approaches.

Table 4 — Twelve authenticity criteria of the aDNA recovery that should be applied to fossils remain proposed by Cooper & Poinar (2000).

Nº	Criteria of authenticity	Description	Important points to note
1	Physical isolation of Pre-PCR facility- strict maintenance of a "one way."	Rule of movement up the concentration gradient: Reagents and equipment must only move in the direction of the Pre-PCR to the post-PCR facility. In many labs, additional precautions are taken into account once laboratory personnel has entered any building where PCR is performed; they can only reenter the pre-PCR facility after fully showering and changing clothes.	Nightly UV irradiation of surfaces, isolated ventilation, flow hoods or glove boxes, daily cleaning with bleach
2	Negative extraction and PCR controls	Extraction and PCR controls containing no DNA should be carried out alongside the samples at every step.	Extraction and PCR controls should be performed at a ratio of at least 1:5.
3	Appropriate molecular behavior	An inverse relationship should be observed between the targeted PCR fragment's length and the amplification strength in studies PCR-based. If shorter fragments are amplified such readily as very long fragments, the product is likely a modern contaminant.	Analyses of the length distribution of fragments sequenced using high-throughput approaches should provide supporting data, where ancient DNA extracts tend to be dominating for short rather than long fragments.
4	Reproducibility	Overlapping PCR products and amplification replicate from multiple DNA extractions must be consistent. If differences occur, more replicates should be performed. For NGS techniques, sequencing libraries or DNA extracts should be consistent, although the variation in endogenous content between extracts from the same sample is common.	Replication of results by different technicians or operators.
5	Cloning and sequencing	A subset of PCR amplification should be cloned to assess the damage and detect nuclear mitochondrial insertions (NUMTs), chimeric sequences from jumping PCR, and contaminants. This point is applicable for PCR-based results, while high-throughput screening of sequencing libraries is a tool more powerful to accomplish the same goal.	To PCR strategies: cloning product amplified by a vector, e.g., TOPO vector. To NGS: library elaboration based on single-strand or double-strand)
6	Independent replication	A second laboratory should replicate the results. This criterion is important when results are evolutionarily unexpected, or data are recovered from a taxon or DNA source for which no data have been isolated.	Independent replication should be performed in another laboratory to rule out intra-laboratory contamination.
7	Biochemical preservation	An evaluation of the likelihood of DNA preservation may be performed. The classical method may be a test of amino acid racemization (Poinar et al., 1996) or bone porosity/density (Pääbo et al., 2004). To NGS methods: the observation of post-mortem nucleotide misincorporation arising from cytosine deamination	Post-mortem DNA damage patterns can be quantified in non-USER-treated or partial USER-treated libraries using the tool map Damage2 (Jónsson et al., 2013). Cytosine deamination can also be estimated using the tool PMDtools (Skoglund et al., 2014).
8	Quantification of starting material	If few surviving DNA molecules are present, the amount of damage in the starting molecules PCR may produce sequence errors that appear in most clones. In the NGS era, quantification is also useful when determining the optimal number of cycles to which a sequencing library should be amplified before sequencing, a low amplification will result in insufficient data, but over-amplification will reduce the sequence complexity of a library (high duplicates)	Techniques of qPCR are an adequate tool to know the number of cycles necessary to obtain a good sequence complexity.
9	DNA from associated remains	It can provide strong circumstantial support in ancient human and microbial studies.	

Table 4 — Continuation.

Nº	Criteria of authenticity	Description	Important points to note
10	Use of "carrier DNA negative" in PCR-based assays	The presence of any other DNA may carry the contaminating DNA through the reaction, resulting in amplification of the contaminant molecules of a target DNA that have not been detected previously and misassigned as a clean negative control.	Carrier DNA should be selected as a type of DNA that will not be amplified in the PCR, e.g., Lamba or vector. DNA is easily accessible.
11	Patterns of DNA damage and sequence diversity by time and preservation-dependent	Sequences isolated from increasingly old or highly degraded samples should be more damaged than younger and better-preserved samples.	
12	Phylogenetic sense	Critical assessment of the sensibility of the results obtained from ancient DNA is an important aspect taken into account. Although the recovered and assembled sequence may differ from any known sequence, the results should be reasonable based on the targeted genomic regions. If the sequence is highly divergent from any known sequence, that result should make sense given the taxon being studied and the data available for comparison.	Tests such as BLAST should be used to ensure that the sequences are not human or environmental contaminants when other species are expected. Using MEGAN is possible to explore the diversity of sequences preserved and provide additional information about preservation, contamination, and authenticity.

1.3.4 Genetic markers in ancient DNA

A wide variety of studies have revealed the true potential of aDNA research to provide methods and processes to reconstruct patterns of evolution, population genetics, and palaeoecological change. It is known that modern humans have widely migrated throughout history, providing different scientific evidence that the entire human population descended from just several thousand African migrants. Various molecular markers have been used to elucidate this and all the population migrations; moreover, revealing the past social, nutritional, and environmental alterations and their influence on people's health. Bellow will detail some studies that have used the main molecular markers to reveal important human aDNA findings.

1.3.4.1 MtDNA -A tool for determining the origin of populations and maternal phylogenetic relationships.

The evolution of modern humans has been a long and hard process that began with their first appearance and continues today. The study of population genetic origins can aid in determining population kinship and better understanding the progressive changes in the gene pool through time. Mitochondrial DNA (mtDNA) has been a proper tool for determining the origin of populations. The extraordinarily high number of copies, the mutation rate, the lack of recombination, and maternal inheritance, made this DNA the easiest to obtain and interpret. Because of the poor rate of DNA recovery from ancient remains, mitochondrial DNA is always the best choice to target because sequences from multi-copy loci have a higher likelihood of being found in ancient remains than single-copy loci. The mtDNA variation data can be used for creating genealogical trees that contain information about the order of the evolutionary processes in space and time.

The non-coding regions (the hypervariable segment I and hypervariable segment II) were the first regions of mtDNA studied based on PCR-target by Sanger (Salas et al., 2000; Sharma et al., 2005). During the last 10 years, the entire mtDNA has been used to trace population origins by direct sequencing with no targeted amplification steps based on shotgun or capture by NGS (Diroma et al., 2021; Loreille et al., 2018; Marciniak et al., 2015). Ancient mitochondrial DNA retrieved from a museum specimen, archaeological findings, and fossil

remains can provide direct evidence for population origins and migration processes through analyses and comparisons between ancient and modern mtDNA (Yao & Zhang, 2003). The first mtDNA tree was created by Vigilant et al. 1991 (Vigilant et al., 1991), and the first five branching points of the mtDNA tree were from people living in sub-Saharan Africa. Understanding the development of the mtDNA pedigree aided population geneticists in tracing the ancestors of contemporary humans back to their African origins and subsequent distribution in the world. In the last 20 years, ancient mtDNA studies strongly imply creating chronology linking modern humans with their ancestors, determining features of the prehistoric migrations, demographic expansion, and identifying individuals' genealogy with a high level of authenticity.

Some findings showed that the evolutionary history of Neanderthals and modern humans was characterized by similar demographic parameters (Lalueza-Fox et al., 2005) and by a positive selection of introgressed variants that occurred either soon after interbreeding with archaic humans and/or after modern humans expanded throughout Eurasia (Jagoda et al., 2022; Yair et al., 2021). The hypothesis of interbreeding between Neanderthals and modern humans has also been proved by the existence of similar anatomically features between them. The examination of the mandible found in a Middle Paleolithic rock shelter in the Monti Lessini, Verona-Italy, showed that it was anatomically typical of modern humans but genetically and morphologically typical of late Neanderthals (Condemi et al., 2014). The studies confirmed that this change in the mandibular morphology could be the result of a small degree of interbreeding with anatomically modern humans or genetic admixtures between *Homo sapiens* and *Homo neanderthalensis* (Anders & Andrea, 2012).

In 1995, a study of archaeological and paleoecological data of ancient samples showed that genetic differences observed in Basques are compatible with migration during the last glacial period of around 18,000 years (Bertranpetit et al., 1995). Particularly in the Neolithic period, after the spread of modern humans in Europe, high levels of migration were performed, probably contributing to the homogenization of the gene pool in Europe.

MtDNA outcomes from skeletons of preNeolithic hunter-gatherers and early Neolithic farmers were analyzed and compared with modern data. The results showed that mtDNA belonging to haplogroup H was indicative of population expansions and the spread of farming culture while mtDNA belonging to haplogroup U showed that the earlier hunters-gatherers adopted farming practices and admixed with the immigrant farming populations. This evidence demonstrated that the spread of agriculture in Europe pushed the expansion of farming populations into the continent followed by the admixture of resident hunters-gatherers (Fu et al., 2012; Sampietro et al., 2007).

Until now, the recovery of genetic information was solely possible when ancient skeletal remains were available in the archaeological site. However, in a present study of exploration of Neanderthal populations in western Europe (Burgos-Spain) was possible to obtain nuclear and mtDNA information using sediment samples (~20 mg of soil) from Pleistocene caves: Galería de las Estatuas, a site in northern Spain with Neanderthal occupation but that is genetically unexplored. The result showed high-coverage genomes of two Neanderthals and one Denisovan hominin. Namely, the results determined the existence of three individuals of different ages who lived 100.000 years ago in the Estatuas cave, and for unknown reasons, they were substituted by another group of the same species, but with a different genetic profile.

In addition, to elucidate our phylogenetic origin, mtDNA has been occasionally used to answer questions of historic interest related to the genetic identification by maternal kinship. Louis XVI inherited France's throne from 1793 to 1795 but died so young. In 1998, a German clockmaker claimed a direct kinship, however, the mtDNA analysis showed no match (Jehaes et al., 1998). Contrary, maternally related living Habsburg descendants showed a consensus sequence in non-coding regions (HVI and HVII) that matches the Louis' heart. Another known case is the identity of the Romanov family. Nine skeletons found in a shallow grave in Ekaterinburg, Russia were tentatively identified as the remains of the last Tsar. Analysis of mtDNA revealed a match between some skeletal remains and a living maternal relative (Coble et al., 2009; Gill et al., 1994). Both historical events have been performed by a conventional technique based on PCR following all the recommendation criteria proposed by the ancient DNA community.

1.3.4.2 Nuclear DNA- A tool for sex determination, phylogenetic analysis and inferring phenotypic features.

Determining the **biological sex** of ancient remains is an important feature for critically testing hypotheses about social structure in prehistoric societies. However, **morphological techniques are not enough** or inadequate for diagnosing the sex of juvenile individuals and fragmentary remains. Techniques based on the targeted region in the amelogenin gene or flanking regions in SRY or alphoides are studied to determinate biological sex (Afonso et al., 2019). Same as happened with mtDNA, the advent of shotgun sequencing or capture by NGS that allows the reliable analysis to identify chromosomal sex simply by considering the ratio of sequences aligning to the X and Y chromosomes. This analysis can be performed in the presence of substantial amounts of present-day contamination by using the signature of cytosine deamination, a characteristic feature of ancient DNA (Skoglund et al., 2014).

Human characteristics such as skin color and behavioral features can be obtained when geneticists access old nuclear DNA. A question in aDNA about if Neanderthals spoke and, if so, how well they did it is one interesting subject that was analyzed by Krause and coauthors (Krause et al., 2007). They analyzed the *FOXP2* gene, intimately connected with the ability to speak in humans. They characterized two SNP in the gene and found that Neanderthal and modern humans share the same polymorphisms at the *FOXP2* locus, which differ from those found in all other mammals, suggesting *Homo sapiens neanderthalensis* had likely the ability to speak like modern humans. Also, it was possible to determine that the Neanderthal Y chromosome sequence falls outside the range of variation observed in modern humans, demonstrating that the *FOXP2* sequences are authentic and that the Neanderthal paternal contribution to the human nuclear genome is low.

Lalueza-Fox and colleagues sequenced a fragment of the melanocortin, one receptor (*MC1R*) gene from two Neanderthal remains (Lalueza-Fox et al., 2007). The *MCR1R* gene is one of the genes responsible for regulating pigmentation in humans and other mammals. Variants in this gene that leads to a reduced function in melanocortin receptor are associated with

light skin color and red hair in humans of European origin. This study showed that both Neanderthals specimens carry a mutation in the *MCR1* gene that is completely absent in the 3700 samples analyzed from modern humans, suggesting that the pigmentation levels in Neanderthals varied potentially on the same scale as observed in modern humans. Another example of molecular evolutionary history based on nuclear markers is the study of a subset of autosomal recessive primary microcephaly genes (*MCPH*) —*CEP135*, *ZNF335*, *PHC1*, *SASS6*, *CDK6*, *MFSD2A*, *CIT*, and *KIF14*— from the complete genomic sequences of archaic humans, the Neandertals and Denisovans. The result showed that the increase protein-coding of *MCPH* genes might not be the sole determinant of the increase in relative brain size during primate evolutionary history (Pervaiz et al., 2021). The ability of some human adults to digest lactose—the sugar in milk—is evidence of recent human evolution. All mammalian babies can digest lactose, using an enzyme called lactase. By adulthood, however, most mammals stop producing this enzyme. Only a small percentage of human adults maintain a functioning lactase enzyme. These "lactase-persistent" individuals have a mutation in their DNA that keeps the lactase gene turned on into adulthood, and they are therefore able to digest lactose. These lactase-persistence mutations arose in the last 10,000 years in different populations and increased in frequency by natural selection. Thus, some findings have explored this nuclear region of lactase-persistence observing changes between ancient and modern populations (Burger et al., 2007; Gamba et al., 2014; Patterson et al., 2022; Tishkoff et al., 2007).

1.4 Forensic Genetics

1.4.1 History of Forensic genetics: from RFLPs to NGS

Human identification has evolved along with recent human history. Conventional methods related to physical evidence such as fingerprints or dentition have been used frequently with relative success. However, in some cases, the state of preservation is insufficient to identify individuals. In this context, DNA typing for human identification has become a powerful tool developed for forensic purposes in recent years.

The first experience to investigate genetic differences between individuals was performed in 1985 with the use of biochemical markers such as conventional blood group and enzyme

analysis (e.g., typing of ABO typing, Km, Gm, EsD, PGM1, AcP). However, this technique had a limited applicability to identify biological samples in criminal cases. Botstein and colleagues were the first in using RFLP (restriction fragments length polymorphisms) to map genes of the human genome (Botstein et al., 1980). Since then the advent of the DNA analysis by Restriction Fragment Length Polymorphism (RFLP) provided an incomparable higher discrimination power than the biochemical tool (Gill et al., 1985; Jeffreys et al., 1985b). A few years later, Alex Jeffreys, an English geneticist, was the first to describe an identification technique, "DNA fingerprinting" or "DNA typing," based on a set of genetic markers named variable number of tandem repeats (VNTRs). Jeffreys identified that these markers are highly polymorphic regions of DNA and that an individual has a unique pattern of minisatellites (Jeffreys et al., 1985b). This pioneering methodology used RFLPs analysis, digesting DNA with specific restriction enzymes, and separating these fragments by electrophoresis. The first forensic cases in which this technique was used successfully involved an immigration dispute and a double rape/homicide of teenage girls (Jeffreys et al., 1985b). In 1985 Kary Mullis marked the beginning of a new era in DNA fingerprinting and forensic genetics, developing a revolutionary technique, PCR or Polymerase Chain Reaction (Mullis, 1990). Forensic samples usually range from high quality to extremely degraded samples; PCR became an important analytical method because of its specificity.

STRs typing became the primary method used in forensic laboratories. In 1997, the Federal Bureau of Investigation established a set of CORE STRs loci for use within a national USA DNA database known as CODIS (Combined DNA Index System) (Budowle et al., 1999). In a forensic context, the nature of some samples is a challenge for the scientific community. The **insufficient template of DNA** quality and quantity for STR typing has led to mitochondrial DNA (mtDNA) as a useful source of DNA (Budowle et al., 2003). Despite mtDNA having less discriminatory power than STRs, it is useful for familial relationships (Coble et al., 2009). Since, as previosuly stated, it is better preserved in degraded samples. At the same time, mtDNA maternal inheritance is useful for kinship testing. The differentiation between individuals is possible by sequencing the two hypervariable segments (HV1 and HV2) of the non-coding region (Parson & Dür, 2007).

However, the arrival of novel sequencing technologies, named **NGS-next generation sequencing** o **MPS-massive parallel sequencing**, could be considered the last milestone in forensic research. NGS technology, including library preparation and targeted captures, allow the sequencing of autosomal/sexual chromosomes STRs, single-nucleotide polymorphisms (SNPs), or mtDNA. Improvements in sequencing technology, library preparation, and targeted captures have promoted the development of novel forensic markers with the ability to differentiate between individuals, interrogate biogeographical ancestry, and deconvolute mixtures (Bulbul et al., 2018).

NGS platforms such as the Illumina MiSeq FGx in combination with the Verogen ForenSeq DNA Signature Prep kit, combine traditional forensic autosomal STRs with ancestry, phenotypic, and identity SNPs markers. The mtDNA can be also analyzed by the ForenSeq mtDNA Whole Genome Kit or mtDNA Control Region Kit; all mentioned kits belong to the same company. On the other hand, MiSeq in combination with PowerSeq® 46GY System from Promega combines traditional forensic autosomal STRs and Y-STRs in the same panel while the PowerSeq® CRM Nested System analyzes the control region of mitochondrial DNA. Other platforms, such as the Life Technologies Ion S5, are also for NGS applications for forensic samples. Remarkable advantages of the NGS approach in front of early conventional methods are mentioned in Table 5.

1.4.2 Forensic DNA characteristics

Forensic DNA analysis only examines a small subset of genetic variation within the human genome to differentiate among individuals. The resulting DNA profile for a sample, a combination of individual STR genotypes, is compared to other samples. All these genetic investigations utilize the same underlying approach: the generation of DNA profiles from the unknown evidentiary sample or questioned samples, referred to as a 'Q,' and a known reference sample referred to as a 'K' followed by a comparison. If the DNA profiles are not consistent, then a person can be eliminated from contributing their transfer DNA to an item, or the identification hypothesis must be rejected. If the DNA profiles are consistent, statistical calculations are performed to provide information regarding the confidence in

Table 5 — Comparison of NGS, RFLP, and PCR-based DNA typing methods.

Characteristic	RFLP Methods	PCR Methods	NGS Methods
Time required to obtain results	6-8 weeks with radioactive probes; ~ 1 week with chemiluminescent probes	1-2 days	Less two days
Amount of DNA needed	50-500ng	0.1-1ng	0.065 ng
Molecular status of DNA	High-molecular- weight, intact DNA	It may be highly degraded	It may be highly degraded
Capable of handling samples mixtures	Yes (single-locus probes)	Yes	Yes
Allele identification	Binning is required since the distribution of sizes is observed	Discrete alleles observed	Yes, the number of repetitions and the sequence
Form used in the analysis	DNA must be double-strand for restriction enzymes to work	DNA can be either single-stranded or double-stranded	DNA can be either single-stranded or double-stranded
Power discrimination	~1 in 1 billion with six loci	~1 in 1 billion with 8 to 24 loci	Depending on the sequencer but could arise 200,000 reads of 110 base pairs (bp) in length (the current maximum read length is 1000 bp)
Automatable and capable of high-volume sample processing	No	Yes	Yes
Tools informatic to the analysis of results	No	Yes, basic tools	Yes, advanced bioinformatic analysis

the match. A forensic DNA laboratory often must deal with less ideal DNA samples. The biological material serving as evidence of a crime may have been left exposed to a harsh environment for a long time, such as in the case of an investigation into a missing person. The bodies of homicide victims are typically found in places where the biological material may have been exposed to a harsh environments like direct sunlight or damp woods. Therefore, at the same aDNA, the DNA molecules from forensic samples may be highly degraded and provide Low Copy Number (LCN) DNA testing, which typically refers to examining less than 100pg of input DNA.

1.4.3 Quality control in forensic genetics

Unlike ancient DNA laboratories, standardized protocols for DNA extraction, DNA quantification, DNA amplification based on PCR using commercial kits and DNA sequencing of short tandem repeats in autosomal (auSTRs) or sexual chromosomes (X-STRs and Y-STRs) and the Control Region of mitochondrial DNA (CR-mtDNA) are

required to validate and to be part of the workflow in forensic laboratories all over the world (Butler, 2012). The necessity for validation of methods is understandable because the genetic results obtained to be used in medico-legal contexts must be admissible in court. In contrast, ancient DNA laboratories are not subject to such limitations and have more leeway regarding method development and modifications to laboratory protocols.

In forensic laboratories, two topics are commonly referred to when discussing the importance of maintaining good laboratory practices to obtain accurate scientific results: Quality assurance (QA) and Quality control (QC). QA refers to those planned or systematic actions necessary to provide adequate confidence that a product or service will satisfy given requirements for quality. QC refers to the day-to-day operational techniques and the activities used to fulfill requirements of quality which include the obligation of having a laboratory with separated areas of work pre-PCR of post-PCR.

1.4.4 Genetic markers in forensic genetics

1.4.4.1 Microsatellites-Nuclear autosomal STRs: The gold standard of contemporary forensic genetics.

Short Tandem Repeats (STRs) are the microsatellites most used in forensic genetics to identify individuals. These markers are comprised of a core motif of repeated sequences in tandem of 2-7pb (Figure 7), are highly polymorphic, are localized in non-coding regions which are on different chromosomes (loci) and are much shorter than minisatellites (VNTRs). Therefore, even trace quantities of degraded material can be analyzed. Nowadays, STRs typing involves simultaneous amplifying many loci by multiplex polymerase chain reaction (PCR) to obtain a panel microsatellite. These multi-locus genotyping are separated by capillary electrophoresis (CE), giving an STRs electropherogram or DNA profile. Figure 7 illustrates an STRs profile that refers to the genotype (the number of repeats found in each allele of the analyzed STR marker). In this example, the amplification of two molecular markers is shown. This technique of STRs multi-locus analysis is considered the gold standard of contemporary forensic genetics (Kowalczyk et al., 2018).

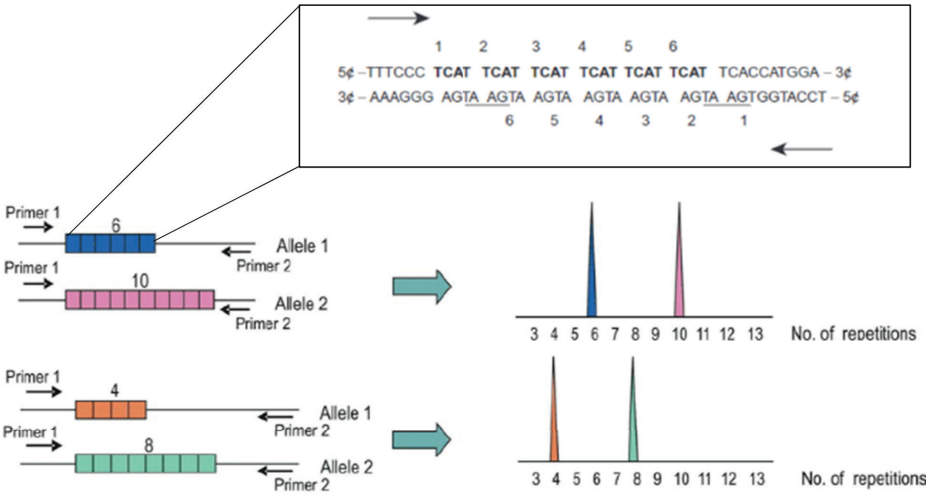


Figure 7 — DNA profile of two STRs markers shows heterozygous alleles of 6 and 10 and 4 and 8 repetitions, respectively. In the example of allele 6, the repeat region is shown. The top strand has 6 [TCAT] units, while the bottom has 6 [TGAA] repeat units. According to ISFG recommendations, the top strand from GenBank should be used. Therefore, this example would be described as having [TCAT] as the repeat motif. The number of repeat units is indicated above, and below the sequence, the arrows stand for the 5'-to-3' direction. PCR amplifies the STRs with fluorescent primers flanking the interest region. PCR products are separated according to size by electrophoresis in an automated sequencer Adapted from Butler, 2012; Sitnik et al., 2006.

The kits of STRs markers traditionally included the analysis of 13-16 STR loci ranging from 90 to 500 base pairs (bp) using various sets of primers. This analysis at least must have the 13 common core STR markers by CODIS (Combined DNA Index System), which have been used for criminal casework and to create the National DNA database of the United States (Katsanis & Wagner, 2013) and the DNA database of European by DNA Working Group of the ENFSI (European Network of Forensic Science Institutes) (ENFSI, 2012). In recent years, there has been a tendency to increase the number of loci amplified to improve the power of discrimination and avoid adventitious matches as databases grow. Nowadays, the standard number of STRs markers used in many cases is extended up to 24 STR loci, and they are included in the available commercial kit (GlobalFiler STR kit and Power Plex 24 kit) (Figure 8). There are a variety of STR multiplex kits commercially available for DNA testing. The majority contains the 13 core CODIS markers, Amelogenin for sex determination, and a few STR markers unique to each provider's kit (Table 6). The nomenclature of the STR loci and the allelic variants was established in 1993 by the DNA Commission of the International Society of Forensic Genetics (ISFG).

Table 6 — Characteristic of autosomal STR Loci present in 31 commercially available STR kits. Adapted from Butler, 2012.

Autosomal STR Loci				Autosomal STR Kits																														
Chr	STR Locus	Repeat	Allele Range	CODIS 13 (US 1997-present)	CODIS 20 (US future)	ESS12 (EU 2009-present)	**PowerPlex 16 (HS)**	PowerPlex 18D	PowerPlex ESI16	PowerPlex ESX16	PowerPlex ESI17Pro	PowerPlex ESIX17	**PowerPlex 21**	PowerPlex CS7	PowerPlex S5	**PowerPlex Fusion**	ProFiler Plus	COFiler	Profiler	CGM Plus	SEFiler Plus	SinoFiler	MiniFiler	**Identifiler**	VeriFiler	NGM	NGM SElect	**GlobalFiler**	ESSPlex	ESSPlex SE	Hexaplex ESS	Nonaplex ESS	Decaplex SE	IDplex
			(Buttler et al. 2012)	Required			Promega STR kits										Life Technologies- Thermo Fisher (ABI) STRs Kits												Qiagen STRs Kits					
1q31	F13B	AAAT	6 to 11											2																				
1q42	D1S1656	TAGA	10 to 19.3						3	2	3	2	3			3									2	3	3	2	2	2	3	2		
2p25.3	**TPOX**	AATG	5 to 13				4	4					6			6		3	3					3				5						1
2p14	D2S441	TCWA	8 to 17						5	1	5	1				4									1	1	1	1	1	1	1	1		
2q35	D2S1338	TKCC	15 to 27					2	4	3	4	3	3			3				4	4	5	2	5	2	4	4	4	5	5			4	5
3p21.31	**D3S1358**	TCTR	11 to 24				1	1	2	2	2	2	2			2	1	1	1	1	1	1		1		2	2	1	3	3		3	3	3
4q31.3	**FGA**	YTYY	16.2 to 43.2				5	5	2	4	2	4	4		2	4	3		3	3	3	3	2	3		4	4	4	4	3		4	2	4
5q23.2	**D5S818**	AGAT	7 to 15				1	1					5			5	1		1			2		2				2						4
5q33.1	**CSF1PO**	AGAT	7 to 15				5	5					4			4		4	4			4	1	4				4						2
6p24	F13A01	AAAG	13.2 to 17											4																				
6q14	SE33	AAAG	6.3 to 36								3	4									3						5	5		4		3	1	
6q15	D6S1043	AGAY	8 to 26										4									2			2									
7q21.11	**D7S820**	GATA	6 to 14				3	3					4			4	3	1	3			3	2	3				4						2
8p22	LPL	AAAT	7 to 15											1																				
8q24.13	**D8S1179**	TCTR	8 to 18				3	3	1	3	1	3	1		3	1	2			2	2	1		1		2	2	3	4	4		2	3	2
9p13	PENTA C	AAAAC	5 to 16											1																				
10q26.3	D10S1248	GGAA	8 to 19						4	1	4	1				5									1	1	1	1	1	1	2	1		
11p15.5	**TH01**	TCAT	5 to 11				2	2	1	3	1	3	1		1	1		2	2	2	2			2	3	3	3	3	2	2	1	2	2	2
12p13.31	**vWA**	TCTR	11 to 21				2	2	2	2	2	2	2			2	2		2	2	2	2		2		2	2	2	4	4		4	4	4
12p13.2	D12S391	AGAY	14 to 27						4	2	4	2	2			2						3			3	4	4	3	3	3	2	3		
13q31.1	**D13S317**	TATC	8 to 15				2	2					5			6	2		2			3	1	3				3						3
15q25	FESFPS	ATTT	5 to 14											3																				
15q26.2	PENTA E	AAAGA	5 to 25				5	5					6	1		7																		
16q24.1	**D16S539**	GATA	5 to 15				4	4	1	4	1	4	1			1		2		3	3	4	1	4		3	3	3	1	1		1	1	

Table 6 — Cotinuation.

Autosomal STR Loci				Autosomal STR Kits																																				
Chr	STR Locus	Repeat	Allele Range	CODIS 13 (US 1997-present)	CODIS 20 (US future)	ESS12 (EU 2009-present)	**PowerPlex 16 (HS)**	PowerPlex 18D	PowerPlex ESI16	PowerPlex ESX16	PowerPlex ESI17Pro	PowerPlex ESIX17	**PowerPlex 21**	PowerPlex CS7	PowerPlex S5	**PowerPlex Fusion**	ProFiler Plus	COFiler	Profiler	CGM Plus	SEFiler Plus	SinoFiler	MiniFiler	**Identifiler**	VeriFiler	NGM	NGM SElect	**GlobalFiler**	ESSPlex	ESSPlex SE	Hexaplex ESS	Nonaplex ESS	Decaplex SE	IDplex
			(Buttler et al. 2012)	Required			Promega STR kits										Life Technologies- Thermo Fisher (ABI) STRs Kits												Qiagen STRs Kits						
18q21.33	**D18S51**	AGAA	9 to 28				4	4	2	5	2	5	2		2	2	4			4	2	4	2	4		4	4	5	2	2		1	1	1
19q12	D19S433	WAGG	9 to 18.2					1	3	3	3	3	3			3				1	1	1		1	2	2	2	2	3	3			2	3
21q21.1	**D21S11**	TCTR	24.2 to 39				3	3	3	4	3	4	3			3	3			3	1	2	3	2		3	3	4	5	5		5	5	5
21q22.3	PENTA D	AAAGA	2.2 to 17				6	6					5	5		5																		
22q12.3	D22S1045	ATT	8 to 19						5	1	5	1				5									1	1	1	1	2	2	2	2		
XP, YP	Amelogenin	-	-				1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1
Yq11-21	DYS391	TCTA	7 to 13													7												6						
Yq11.221	Y-Indel	TTCTC/-	"1" or "2"																									1						
autosomal STRs amplified							15	17	15	15	16	16	20	7	4	22	9	6	9	10	11	15	8	15	9	15	16	22	15	16	6	13	11	15

The allele range is from NIST 1036 data set. Numbers inside the colored boxes are indicative of the relative size position for that locus within a dye channel for the specific STR kit (Fig 8)

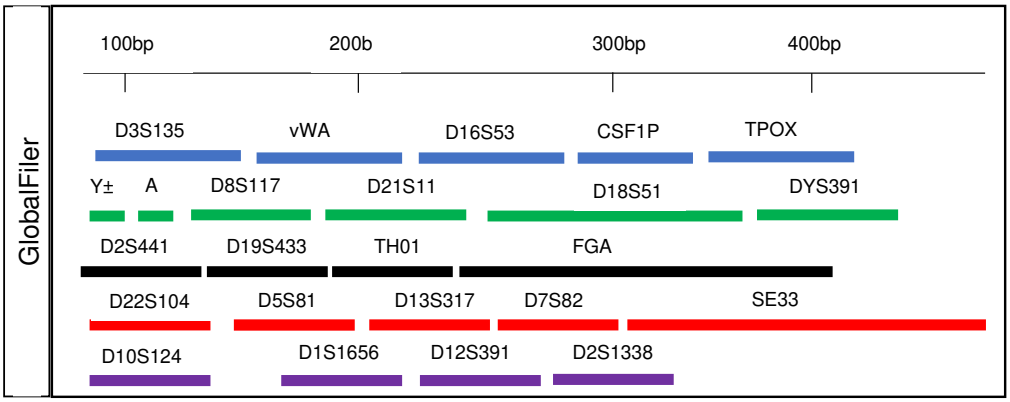


Figure 8 — The layout of loci by dye channel and relative size in selected ThermoFisher (Applied Biosystems, ABI) GlobalFiler STR kit. Adapted from Butler, 2012.

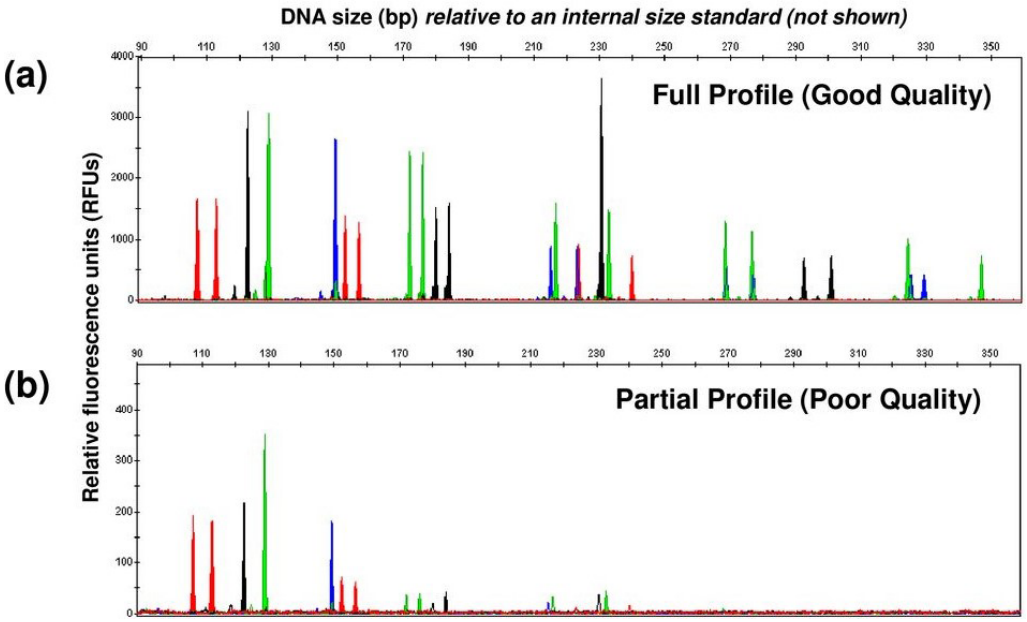


Figure 9 — Comparison of DNA profiles originating from the same biological source but of different qualities. (a) Intact, good-quality DNA yields a full profile. (b) Degraded, poor-quality DNA yields a partial profile with only the lower-size PCR products producing a detectable signal. With the degraded DNA sample shown in (b), information is lost at the larger-sized STR loci. Also, note the lower relative fluorescence units (RFUs) with the poor-quality partial profile in (b) Adapted from Butler, 2012.

Several experiments have shown an inverse relationship between the size of the locus and successful PCR amplification from degraded DNA samples, such as those obtained from a crime scene or a mass disaster (Whitaker et al., 1995). The STR loci with larger-sized amplicons in a multiplex amplification, such as D18S51 and FGA or CSF1PO, SE33 and Penta D, are the first to drop out or suffer from allelic loss when amplifying extremely degraded DNA samples. As a result, only a few of the lower molecular weight markers (<200pb) may be possible to recover and therefore obtain a partial or incomplete DNA profile (Figure 9).

Data used for short tandem repeat (STR) analysis come from Genetic Analyzer, which is a laser-induced fluorescence (LIF) capillary electrophoresis (CE) instrument from Applied Biosystems. A key component of the data detection process is the charge-coupled device, or CCD camera, which collects visible light from the laser-induced fluorescence of the dyes attached during the PCR process that copies STR markers. The dye-labelled STR PCR products are separated by mobility or size via CE (Butler, 2012). Data collection is shown as peaks, and the methods for distinguishing true signal from background noise have been studied for years (Gilder et al., 2007).

In an electropherogram of a DNA profile, some parameters are calculated. Among them is the limit of detection (LOD) which reflects the point at which the signal can be reliably distinguished from noise. Thus, a peak with a height just above the LOD has a 95% or 99.7% chance of not being noisy. However, a signal just under the LOD is likely contaminated with some noise. Therefore, the limit of quantitation (LOQ) should also be analyzed. This parameter reflects a point at which a signal can be reliably quantified, so a peak should not be determined until its height is at least above the LOD and LOQ. Having established this analytical threshold, a sample electropherogram is ready to decipher whether detected peaks are alleles or artifacts. The SWGDAM, 2010, has to settle downed interpretation guidelines to distinguish alleles from artifacts (SWGDAM, 2010). The parameters such as height peak in RFU (relative fluorescence unit), the artifacts or stutter, alleles drop-out, alleles drop-in, and peak height ratio (PHR) between two heterozygous alleles are also evaluated.

Autosomal nuclear DNA is highly preferred for the genetic identification from highly challenging samples such as skeletal remains because it is individual-specific and it provides bi-parental kinship information, namely, each DNA profile will consist of two copies or alleles of a locus (Figure 7), one from the mother and one from the father. The inherited alleles are called the maternal and parental alleles, respectively. There is a 50% chance for each of the two alleles to pass on to a child for any autosomal marker. In cases of missing persons or armed conflict, when skeletal remains are the only sample for genetic identification, the DNA profile is obtained and compared to the DNA profile from a possible relative. The relationship could be confirmed through a statistical calculation, known as (LR) likelihood ratio, in which conditional probabilities between two potential relatives are calculated. This value demonstrates how much more likely it is that the DNA evidence would be observed under a hypothesis that the evidence came from people with a specific relationship as opposed to seeing the DNA evidence given a hypothesis that the observed data came from two presumably unrelated people (Butler, 2012):

$$LR = \frac{\Pr(E|H1)}{\Pr(E|H2)}$$

E=the DNA profiles for the tested individuals
H1= the hypothesis that an alleged father is the biological father of the tested child
H2=the hypothesis that a random, unrelated man is the biological father of the tested child

Therefore, if statistical analyses were based only on the 13 routinely used Combined DNA Index, the probability of a random match between unrelated individuals would increase so the incorporation of more STR markers into the common forensic typing assay currently used has been recommended. However, simultaneous detection of more STR markers would be very difficult due to the technical limitations of fluorescent-based CE sequencers currently in use.

As mentioned before, traditional STR typing using CE detects DNA fragment size. Therefore, alleles of identical or similar lengths but different sequences cannot be distinguished. The advent of NGS can distinguish alleles with similar lengths and obtain the sequence by digital read count, which facilitates the identification of mixed samples and analysis of complex paternity cases. Some researchers have recently started using NGS

technology for STR testing. A pioneering study applied to forensic genetics was performed by Irwin et al. 2011, who analyzed 13 CODIS STR loci by comparing the Illumina and the 454 Genome Sequencing System from common samples used in cases of missing persons such as swab buccal, bloodstains, and skeletal remains. The preliminary results suggest that the genetic information obtained was feasible in both the Illumina and 454 platforms with degraded samples (Irwin et al., 2011). In addition, it was expected to generate the same high-quality raw and consensus data using NGS technologies when Sanger sequencing was used, resulting in a high level of comfort and confidence in the use of NGS for forensics.

The recovery of a large number of markers, for instance, more than 50 STRs or SNPs in a single assay, has been restricted by technical limitations in sequencing based on capillarity, particularly when considering the typical damaged and degraded forensic specimens regularly encountered in missing person casework. However, these limitations do not apply to NGS. Instead, the simultaneous recovery of the standard autosomal DNA STRs, mitochondrial DNA, as well as SNPs, X and Y chromosomal markers regularly assayed in forensic genetics may well be possible with these new technologies.

Some commercially available kits based on amplicon-NGS have been launched in recent years. Precision ID GlobalFiler kit includes the same 21 autosomal STRs, Y markers, and amelogenin sex markers found in GlobalFiler™ PCR Amplification Kit. In place of SE33, this kit includes additional multiallelic STR Penta D and Penta E to aid in mixture interpretation for complex casework samples. ForenSeq DNA Signature Prep Kit of Verogen is the only next-generation sequencing (NGS)-based short-tandem repeat (STR) sequencing chemistry approved for upload to the National DNA Index System (NDIS) for casework. The kit eliminates the need to run multiple STR tests because combining over 200 markers, including autosomal, X- and Y-STRs, identity-informative, phenotype-informative, and biogeographical ancestry-informative SNPs into a single streamlined workflow. This kit is part of an integrated workflow that sequences libraries on the MiSeq FGx Sequencing System, and finally PowerSeq® 46GY System kit contains reagents to amplify autosomal and Y-STR loci as small amplicons (140–300bp) that can be used to

prepare MPS libraries and generate sequencing data compatible with MiSeq® technologies. Recent studies have tested these STRs-NGS kits from highly degraded skeletal remains, and the results show that the use of this technology increases the retrieval of genetic information which by conventional panels is not possible (Gettings et al., 2015; Jäger et al., 2017; Zupanič Pajnič & Fattorini, 2021). Although in some cases is difficult to recover long STRs from degraded samples, it is possible to analyze other sources of genetic variations that have been demonstrated to present more specialized uses in forensic identification as autosomal SNPs, the markers on the Y chromosome, and mitochondrial DNA (mtDNA).

1.4.4.2 MtDNA-As molecular marker in genetic identification

Just like ancient DNA, the analysis of mitochondrial DNA is an alternative tool when nuclear DNA profiling has not had enough success from bones, teeth, or hair samples. The mitochondrial genome has a uniparental inheritance pattern, that means an individual's mtDNA will be identical to all consanguineous maternal relatives. As a result, mtDNA profiles can only offer an approximate identification. When materials are deteriorated and do not provide nuclear profiles, when additional genetic information is needed, or when sufficient autosomal STR references are not available, mtDNA profiles are relevant in forensic investigations.

In a forensic DNA laboratory, the genetic information from each marker is referred to as a haplotype rather than a genotype because there is only a single allele per individual. Therefore, haplotypes obtained from lineage markers can never be as effective in differentiating between two individuals as genotypes from autosomal markers that are unlinked and segregate separately from generation to generation.

The haplotype or mitochondrial DNA sequence is typically reported in terms of variation compared to rCRS. Thus, the observation of a C nucleotide at position 16126, which contains a T in the reference sequence, would be reported as 16126C. If no other nucleotide variants are reported, then it is assumed that the remaining sequence contains the same sequence as the rCRs. MtDNA sequence results for forensic identification purposes are based on comparing

two or more samples (questioned and known samples), showing one of three possible conclusions: exclusion, inconclusive, or failure to exclude according to SWGDAM guidelines.

- *Exclusion: if there are two or more nucleotide differences between the samples compared so they can be excluded as originating from the same person or maternal lineage.*
- *Inconclusive: The results will be inconclusive if there is one nucleotide difference between the samples compared.*
- *Failure to exclude: If the sequences from the samples under comparison have a common base at each position, they cannot be excluded as originating from the same person or maternal lineage.*

A statistical estimate of the significance of a match is needed when *failure to exclude* between the samples is concluded. The current practice is to use the counting method, which evaluates the rarity of an mtDNA haplotype among unrelated individuals. In other words, it is based on counting the number of times an mtDNA haplotype (sequence) is seen in a database. This approach depends entirely on the number of samples present in the database that is searched.

$$p = \frac{X}{N}$$

p= frequency of the haplotype in database
X= number of times where an mtDNA profile is observed in a datab
N=Number of total profiles in a database

Several forensic laboratories, including the FBI Laboratory (Washington, DC, USA), the Armed Forces DNA Identification Laboratory (Rockville, MD, USA), and the Forensic Science Service (Birmingham, UK), are actively involved in using mtDNA for forensic identifications. For instance, human skeletal remains from the Vietnam War and other wars are routinely analyzed and identified by mtDNA typing. A repository for blood samples from all those entering the US Armed Forces is performed in the Armed Forces DNA Identification Laboratory so that, if necessary, in the future, the DNA in the stored blood samples can be used to identify war casualties resulting in no more *unknown soldiers*. However, no blood repository existed for past soldiers, so family reference samples from siblings or maternal relatives are used for confirming the human remains' identity. Skeletons stored from

Guatemalan mass graves and the Spanish civil war have also been successfully analyzed by mtDNA testing (Ginther et al., 1992; Malgosa et al., 2021; Palomo-Díez et al., 2019).

MtDNA analyses have also been used from hair shafts to connect individuals to crime scenes or exclude a suspect when the suspect was not available for testing (Butler & Levin, 1998). When skeletal remains are used to obtain genetic information, the great limitation is molecular damage caused by DNA degradation, producing consequently false sequences. This event is the main common point with ancient DNA. The same molecular damages of ancient DNA mentioned in section 1.3.2.1 are manifested when mtDNA sequences are analyze from compromised forensic samples such as skeletal remains.

Holland and collaborators (2021) (Holland et al., 2021) analyzed the positions with hot spots in the control region from forensic samples (Table 7). They can be taken into account as molecular damage in this position when a mixture of bases with an alternative underlying nucleotide is shown. However, the study by Nelson and Melton (2007) showed that the levels of heteroplasmy had been found from skeletal remains at a rate similar to that in hair (~10%) (Nelson & Melton, 2007). Therefore, variation or mutations in mitochondrial sequences continuously affect the germline on occasions, resulting in a heteroplasmy transmitted through the proposed genetic bottleneck. Over time, through genetic drift, the wild variant balance can be altered, and when the sequence change is beneficial to the individual can become fixed as a homoplasmy in the population. From a forensic approach, variabilities in the level of heteroplasmy can be seen as discordance between the haplotypes from a single hair root and blood in the same individual; however, the analysis of additional hairs should be undertaken in these very unusual circumstances. From skeletal remains, there can be three possible events of bases mixture: heteroplasmy variants, mixture profiles, and hot spots of molecular damage that can occur simultaneously, being important to identify these events. The criteria by Nelson and Nelson (2007)(Nelson & Melton, 2007) determines as **heteroplasmy variants** for a single nucleotide position within the analyzed region the presence of two nucleotides at a single position, **mixture profiles** for profiles where two or more sites have two or more nucleotides present, and **hot spots**

Table 7 — Potential damage hotspots (n = 35) were observed three or more times across all samples in the study by Holland et al. 2021.

Base position	Total observations	Base change/# observations
67	3	G-T(2); G-C-1
81	3	G-T(2); G-A-1
107	3	G-T(2); G-A-1
113	3	C-T-3
143	3	G-A(2); G-T-1
162	4	C-A(3);C-T-1
172	4	T-C-4
187	4	G-T-4
189	3	A-G(2);A-C-1
195	3	T-C-3
389	4	G-T-4
412	3	G-T-3
545	4	G-T(2); G-A-2
16026	3	C-A-3
16027	3	T-C-3
16065	3	G-T(2); G-A-1
16118	4	G-T-4
16179	4	C-T(3);C-A-1
16218	3	C-T(2);C-A-1
16222	3	C-T-3
16261	3	C-T-3
16265	4	A-G(3); A-C-1
16291	3	C-T-3
16328	3	C-T(2);C-A-1
16348	3	C-T(1);C-A-2
16390	7	G-A(6); G-T-1
16401	3	C-T-3
16414	3	G-T-3
16460	3	C-T-3
16496	3	G-T(2); G-A-1
16506	3	T-C-3
16519	4	T-C-4
16520	4	C-A(2);C-T-2
16558	3	G-T(2); G-A-1
16566	4	G-T(3); G-A-1

of molecular damage for one or more nucleotide substitutions with no evidence of an alternative underlying nucleotide.

1.5 Factors influencing DNA retrieval in skeletal remains

The retrieval and analysis of DNA from human skeletal tissues—teeth and bones—have become challenging for scientific fields ranging from ancient/archaeological DNA to forensic context. This fact is due to the damage that DNA suffers from when an organism dies.

In living cells, DNA molecules frequently suffer chemical modifications, repaired by enzymatic mechanisms that maintain the integrity of the genome (Lindahl, 1993). On death, these cellular repair mechanisms decay and cease their function. As a result, DNA strands become exposed to numerous factors that threaten their stability and generate chemical reactions to modify and break the DNA strand producing patterns of molecular damage. These extrinsic factors include external environmental conditions such as UV high radiation, extreme temperature, humidity, and low pH, which speed up the growth of microbes present inside and around the body showing. As a result, the increase of enzymatic reactions could lead to the loss of all DNA (Alaeddini et al., 2010b). In tissues frozen or become desiccated quickly after death, the enzymatic degradation procedures may be reduced or inhibited, and the destruction of all endogenous DNA may be avoided because the tissues are under favorable environmental conditions for DNA preservation. Although DNA preservation in cold environments may reduce nuclease activity and decrease some of the damage molecular, non-enzymatic DNA degradation reactions are carried out. In addition to external factors the type of skeletal remain and DNA extraction methods may be also a limiting factor for the successful recovery of degraded DNA.

1.5.1 Type of skeletal remains used in ancient DNA and forensic genetics

In addition to the many environmental factors, intrinsic characteristics such as bone type can play a role in DNA degradation and consequently in the successful retrieval of genetic information. In skeletal remains, the preservation of DNA is influenced by the size and architecture of the bones. The composition of bone tissue is made up of an extracellular matrix

split into i) an inorganic phase constituted by 65-70% of carbonated hydroxyapatite (HA) $Ca_{10}(PO_4)_6(OH)_2$, and by mostly collagen but also 25-30% of glycoproteins, proteoglycans and sialoproteins (Higgins & Austin, 2013), and ii) cellular bodies such as osteogenic cells, osteoblasts, osteocytes and osteoclasts (Figure 10). Bone has a complex material and can provide information at many different levels (isotopic, molecular, biochemical, and structural), which are of great importance in research. Diagenetic studies involve the processes that change the nature of bone during burial and how these processes are environmentally determined, and how specific types of information are altered or recovered.

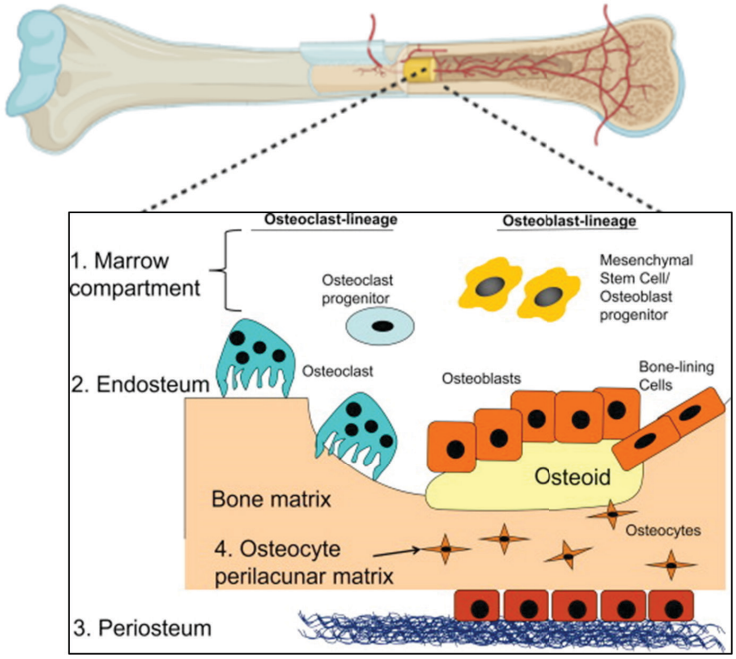


Figure 10 — Structure of compact bones. Histological structure of compact bone-Femur. Adapted from Alford et al., 2015.

Some studies have reported that DNA has a strong affinity for hydroxyapatite, and its degradation is related to the crystallinity loss in HA and the loss of collagen (Götherström et al., 2002; Scorrano et al., 2015). Some bones are more compact than others; therefore, crystallinity is related to the type of bone and the porosity grade. Most of the prevailing knowledge and assumptions about what skeletal parts are superior to others in preserving DNA from highly degraded skeletal remains are based on ancient DNA and forensic

researches (Barta et al., 2014; Hansen et al., 2017c; Pilli et al., 2018). The general conclusions in both research fields are that the density of a bone is positively correlated with DNA preservation and that sampling should be carried out in sites as dense as possible. Bellow will be describing the main samples of skeletal remains used in aDNA and forensic genetics to recover genetic information.

1.5.1.1 Postcranial bones, tooth, and temporal bones

Molecular human identification has conventionally focused on DNA sampling from dense, weight-bearing cortical bone tissue, typically femora or tibiae. However, there was a huge debate regarding what skeletal elements provide the best results to recover genetic information that allows DNA typing for forensic purposes or obtaining the whole genome for studies in aDNA research. Some studies that evaluate the DNA yield rates for different skeletal elements have been reported. In forensic research, Mundorff and collaborators (2009) showed that smaller elements, such as metatarsals and metacarpals, provide some of the greatest success for STR analysis (Mundorff et al., 2009). This finding is supported by Andronowski and colleagues (Andronowski et al., 2017), who conclude that differences in bone microstructure can explain differential nuclear DNA yield among bone tissue. Osteocytes and other bone cells house DNA in bone tissue; thus, examining the density of their lacunae may explain why nuclear DNA yield rates differ among bone tissue types. Barta and collaborators (2014) (Barta et al., 2014) also suggest that non-weight-bearing bones (postcranial bones) may provide improved results for mitochondrial DNA (mtDNA) testing, recommending ribs, as they are simple to remove; however, this work was largely based on nonhuman remains (i.e., seals). Edson (2019) evaluated the extraction of DNA from skeletonized postcranial (clavicle, femur, fibula, fragments, humerus, jaw, metacarpal, metatarsal, pelvis phalanx, radius, rib, scapula, talus, tibia, ulna, vertebra) by NGS and conventional techniques and in general, the recovery of mitochondrial DNA and STR testing through complete demineralization coupled with an organic purification was the optimal extraction protocol regardless of the osseous element tested (Edson, 2019). All these findings have shown a great variety of results that need to be supported by new systematic studies that allow for confirming which postcranial bones could show advantages in the recovery of degraded DNA.

A tooth comprises four major tissues: enamel, cementum, dentin, and pulp (Figure 11). Enamel is the hardest tissue in the human body, being 96% mineral (Nanci, 2008; Vaissier Welborn, 2020), is acellular, and does not contain DNA. In contrast to enamel, the dentin and pulp complex are highly cellular and make up the bulk of the tooth. Dentine comprises 65% mineral (hydroxyapatite), organic macromolecules (mostly collagen), and water. It usually does not contain any nucleated cell bodies; however, mitochondrial DNA may be accumulated from the odontoblastic process (Figure 11A) (Yu & Abbott, 2007). The dental pulp is a highly vascularized and innervated connective tissue that is consisted of numerous cellular types: cells that form dentine (odontoblasts), fibroblasts, defense cells (histocytes and macrophages), plasma cells, and nerve cells, namely is rich in DNA (Chiego, 1994).

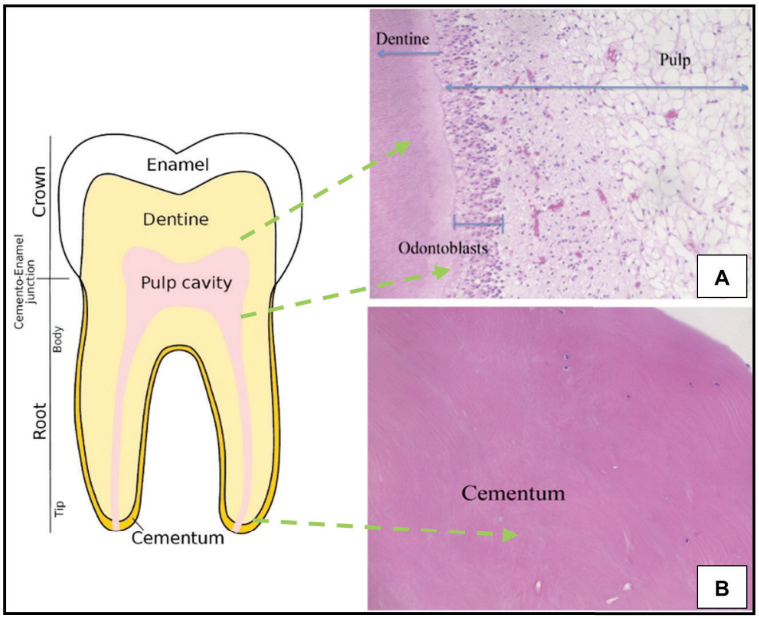


Figure 11 — Tooth structure of a human molar. Tooth regions: root body and tip. Cellular tooth tissues: dentine and cementum Adapted from Adler et al., 2011b; Higgins et al., 2013. **A.** Histological view of the pulp, dentine (displaying odontoblasts) **B.** and cementum (displaying cellular cementum with cementoblasts) of a sectioned human molar.

Since dentine and pulp are protected with enamel, this tissue provides a physical barrier protecting DNA against external factors like heat, sunlight, moisture, and microbial attack (Adler et al., 2011a) and, as a result, DNA is well preserved. Cementum covers the roots of the teeth and is interlocked firmly with the dentin of the root; it is a mineralized connective

tissue like bone except that it is avascular; the mineral is also apatite, and the organic matrix is largely collagen. The cells that form cementum are called cementoblasts. Although cementum has provided better results in DNA preservation than dentine, it could also display a lower endogenous DNA proportion because cementum is a layer exposed at the root surface and could potentially be more affected by microbial colonization.

In the DNA forensic laboratory routine, the most typical substrates used are the femur and tooth. However, teeth are superior in obtaining DNA profiles to identify an individual. There are differences in the amount of DNA recovered in the different tissue regions of the tooth. Knowing the composition of a tooth is necessary for well sampling and to provide good quality results because it has been shown that the nuclear DNA concentrations decline drastically in the inner dentine layer throughout the life of an individual (Amory et al., 2012), whereas levels of nucleated cells in the apical cementum layer are unaffected by age (Higgins et al., 2013).

On the lateral sides of the skull, on the temporal lobes of the cerebral cortex, are located the temporal bones. They house the structures of the ear and seven nerves and cranial vessels pass through them. It is anatomically divided into four regions: the scaly, mastoid, temporal and petrous. The petrous part of the temporal bone is pyramid-shaped and is wedged in at the base of the skull between the sphenoid and occipital bones (Figure 12). Petrous comes from the Latin word petrosus, meaning "stone-like, hard." It is one of the densest bones in the body. The petrous bone is important for studies of ancient DNA from skeletal remains, as it contains extremely well-preserved DNA.

A study performed by Pinhasi and collaborators found that the petrous bone has noticeable well-preserved DNA (Pinhasi et al., 2015). Two years later, another study compared DNA from different skeletal sites, and the results showed that the inner part of petrous bones (part of cortical bone encircling the osseous inner ear or otic capsule) is the densest, and the cementum layer in teeth roots are currently recognized as the best substrates for ancient DNA research. Both substrates display significantly higher endogenous DNA content (average of 16.4% and 40.0% for teeth and petrous bones, respectively) than parietal skull bone (average

of 2.2%) (Hansen et al., 2017a). Consequently, petrous bones are now the most widely-used skeletal site for studying ancient DNA. In recent years, the high quality of results obtained in the retrieval of DNA from petrous bones in the aDNA community has motivated the forensic DNA community to explore the use of petrous bone as part of the specimens for sampling and consequently to obtain the genetic identification STRs and mtDNA (Gonzalez et al., 2020)

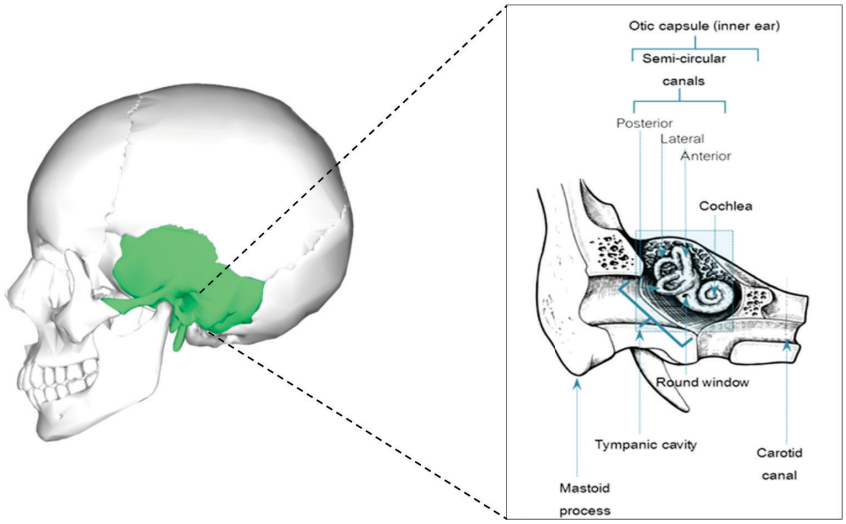


Figure 12 — Temporal bone and structure of inner ear-petrous bone. Adapted from Gonzalez et al., 2020.

In ancient research, all the studies have been focused on comparing petrous and tooth; however, it has recently been published a systematic study of human DNA preservation from different skeletal elements (molar, petrous pyramid, clavicle, rib, thoracic vertebrae, metacarpal, distal phalanx, ischial tuberosity, femur, talus) in medieval skeletons. The results confirm the best performance of the petrous pyramid and identify seven additional sampling locations across four skeletal elements (vertebra body, cementum, distal phalanx, and talus) that yield adequate aDNA for most applications in human palaeogenetics (Parker et al., 2020).

1.5.1.2 Newborn bones

The newborn bones are not completely developed, and the thickness of the cortical is deficient, are less protected than the adult, and are very limited with tiny-sized bones and lack of dental development. All these features may difficult the recovery of DNA. As

mentioned above, it has been proven that the most protected and developed cortical is more amount of endogenous DNA that may be recovered if a suitable DNA extraction protocol is used (Dabney & Meyer, 2019; Damgaard et al., 2015). However, these findings have been established for adult skeletons of various archaeological contexts from the Bronze Age until the 18th century (Damgaard et al., 2015; Edson, 2019; Parker et al., 2020). Various studies show that infants' bones have been used in ancient (Afonso et al., 2019; Faerman et al., 1998; Millett & Gowland, 2015; Teschler-Nicola et al., 2020) and forensic (Minaguchi et al., 2003; Nelson & Melton, 2007) approaches. Petrous, long bones and vertebra have been used to obtain successful results by either conventional techniques or NGS. However, there are no systematic comparative studies from newborn or infant bones to know the amount of endogenous DNA recovered and subsequently obtain alternative skeletal remains to elucidate many questions forensically and evolutive.

1.5.2 Extraction methods in ancient and forensic DNA

In the last years, ancient DNA research and forensic laboratories have developed various procedures for DNA extraction from diverse types of human tissues. For skeletal remains, the organic extraction method of Phenol-Chloroform (Ph-Cl) was the first method used in ancient DNA (Hagelberg & Clegg, 1991) and forensic genetics (Prado et al., 1997). Some forensic laboratories routinely used this organic protocol for a long time to extract DNA from degraded skeletal remains (Caputo et al., 2013a; Ferreira et al., 2013). However, this protocol has some limitations such as being high time-consuming, has additional steps for DNA purification, increasing in the risk of contamination (handling a great volume of solutions), employs of harmful reagents which may affect the health of the operator, and is less effective in the removal of a hydrophilic component as DNA and potential PCR inhibitors.

Ancient DNA community has improved their extraction methods from Ph-Cl protocol to extraction methods based on DNA adsorption to silicon dioxide particles (silica) (Paabo, 1993; Rohland & Hofreiter, 2007b; Yang et al., 1998). This is achieved by mixing a lysis buffer to release DNA from previously prepared sample powder of teeth or bones with a high-salt binding buffer. DNA binding to silica surface can be performed either by adding

a silica suspension (Rohland & Hofreiter, 2007a) or by centrifugation through silica spin columns (Figure 13) (Rohland et al., 2018). DNA adsorbed is then desalted with a wash buffer containing ethanol and eluted into a low-salt buffer.

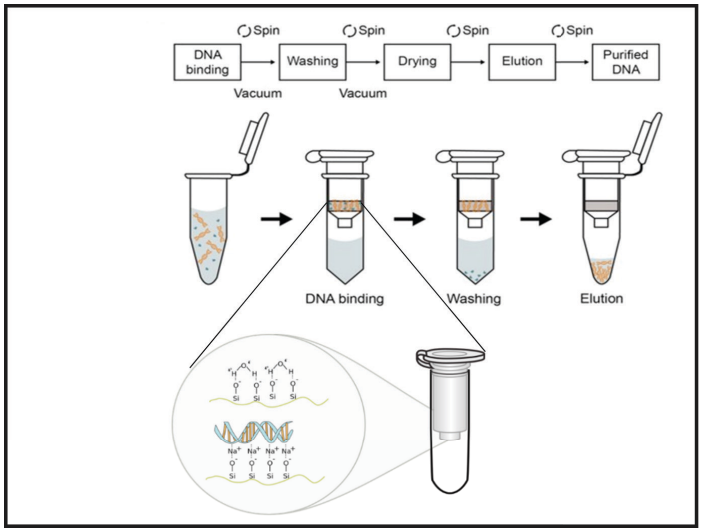


Figure 13 — The procedure of DNA binding to the silica surface. The positive ions in the buffer aid in forming the salt bridge between the DNA and silica surface. Guanidium HCl and Guanidinium thiocyanate are commonly used as chaotropic salt in this DNA extraction method. The elution is carried out using TE buffer at pH 8.4. It removes the positively charged ions and disrupts the salt bridge between the DNA and the silica surface. Adapted from Lee et al., 2018.

These protocols minimize the co-extraction of inhibitory substances when used with chaotropic binding buffers, offer ease of handling with commercially available silica spin columns, and allow efficient recovery of DNA fragments as short as 35pb (Dabney et al., 2013; Rohland et al., 2018). Therefore, these advances have made it possible to access very short DNA fragments preserved in ancient samples that cannot be targeted directly by PCR, but for NGS. Also, they proved that there is an inverse exponential correlation between fragment length and the amount of DNA in the sample (Figure. 14) (Allentoft et al., 2012; Rohland et al., 2018). In this sense, the aDNA community has accomplished to improve better its protocols by maximizing the recovery of endogenous DNA by comparing silica and Ph-Cl technique as it results that DNA binding to silica particles through using chaotropic salts is highly superior to Ph-Cl (Rohland & Hofreiter, 2007b).

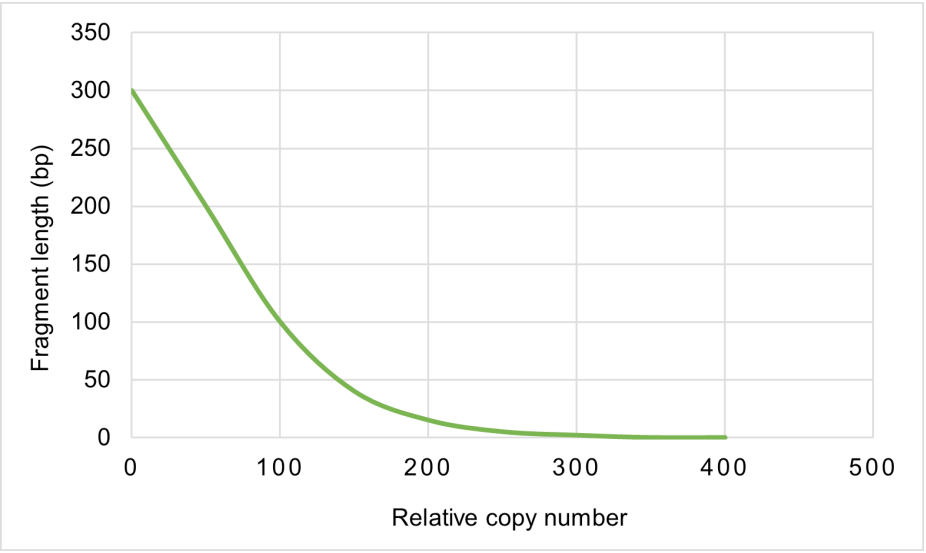


Figure 14 — DNA fragmentation theory. The template fragment length distribution follows an exponential decline determined by the proportion of copies number of DNA. Adapted from Allentoft et al., 2012.

Forensic genetics DNA isolation procedures vary depending on the type of biological evidence; for example, whole blood must be treated differently from a bloodstain or a bone fragment. For many years, organic extraction (Ph-Cl) was the most widely used for DNA extraction, involving the serial addition of several chemicals. First, sodium dodecyl sulfate (SDS) and proteinase K are added to break open the cell walls and break down the proteins (e.g., histones) that protect the DNA molecules, packed as chromosomes. Then a phenol-chloroform mixture is added to separate the proteins from the DNA. After centrifugation, the unwanted protein and cell debris are separated from the organic-aqueous mixture. The double-stranded DNA molecules can be recovered in a cleanly phase aqueous and transferred for analysis. Finally, Centricon 100 (Millipore, Billerica, MA) dialysis is performed to remove any inhibition and concentrate DNA. Another extraction method used in forensic genetics is chelex extraction. This method is an ion exchange resin that is added as a suspension to the samples. The typical sample such as bloodstain, blood, or semen is extracted by the chelex method. In the past decade, new extraction methods have evolved to various forms of solid-phase extraction where DNA is selectively bound to a substrate such as silica particles and has been demonstrated to be superior to classical organic procedures. The most widely used solid-phase extraction methods are

Qiagen columns, DNA IQ, or PrepFiler. These procedures could be applied in forensic laboratories as manual or automatized protocols.

For more than a decade, products from Qiagen like QIAamp spin columns have proven effective as a means of DNA isolation. The principle of DNA extraction is the same applied to ancient DNA silica-based protocol (adsorption to the silica in the column). Several robotic platforms have been developed to enable automated procedures of Qiagen DNA extractions, including the EZ1, M48, and QIAcube. DNA IQ ™ Promega Corporation (Madison, WI), which stands for isolation and quantification, utilizes the same silica-based DNA binding and elution chemistries as Qiagen kits but with silica-coated paramagnetic resin. With this approach, DNA isolation can be performed in a single tube by sampling and removing solutions. In 2008 Applied Biosystem released a magnetic particle-based DNA extraction, PrepFiler BTA™, similar to DNA IQ. PrepFiler enables the isolation of high-quality DNA from forensic samples in high yields at a time in less than 2.5 hours using the automated liquid-handling workstation. All these protocols have been standardized for all types of samples, including skeletal remains (Amory et al., 2012; Desmyter et al., 2017; Harrel et al., 2018; Iyavoo et al., 2013; Liu et al., 2018; Rothe & Nagy, 2016; Ye et al., 2004; Zupanič Pajnič et al., 2016).

Although the great variety of extraction protocols to extract DNA from bones or teeth has been standardized, there is still no extraction method that completely recovers the DNA from highly degraded skeletal remains. In this sense, some factors are limiting and the improvement for recovering DNA from highly degraded skeletal remains persists. Factors like the type of bone or the amount of starting sample (wide range used between 100mg-2g) and the preparation of bones and teeth samples (fine powder or thin slides of bone) also affect the DNA extraction and subsequently the typing of any genetic information such as mtDNA and STR profiling (Adler et al., 2011b; Caputo et al., 2013a).

The DNA extraction protocols based on silica that have been released from the aDNA community could be applied to forensic context, reflecting convergence in methods from ancient DNA to forensic genetics and vice versa. These methods recover increasingly small DNA

fragments (~20-80 bp) from highly degraded skeletal remains; therefore, the Sanger sequencing is not a technique adequate to use since it requires larger sites for primer binding. As forensic and aDNA genetics continues to integrate more NGS methods, those smaller fragments, which are informative, can be obtained by targeting SNPs in mtDNA to determine some genetic identification features such as biological sexing or biogeographic origin (Loreille et al., 2018).

1.6 DNA analysis in critical samples: from classical techniques to Next Generation Sequencing

1.6.1 PCR based approaches

Until now, PCR is the method of choice for in vitro amplification of DNA molecules, although the success in obtaining genetic information can be reduced if starting from DNA of low quality and quantity (Alaeddini et al., 2010a). The polymerase chain reaction has been one of the most important advances in biological sciences, including forensic genetics and aDNA. Both disciplines share a limiting and critical aspect, such as the study of critical samples due to their high level of degradation. Environmental factors break DNA molecules into smaller pieces and reduce the success of recovering targeted information. In this context, the use of older techniques such as restriction fragment length polymorphism (RFLPs) would have had little success in detecting the large variable number of tandem repeat (VNTR) alleles. However, the development of smaller short tandem repeats (STRs) has overcome this limitation (Butler, 2012).

Novel PCR approaches/methods have been optimized for the analysis of forensic degraded samples, and nowadays, multiplex PCR can amplify STRs alleles from less than one ng of DNA. This adaptation of PCR allows the simultaneous amplification of different regions by using more than one primer set to the reaction mixture (Edwards & Gibbs, 1994). Another advance and improvement in PCR technique focus on critical samples are reduced size PCR products named miniSTRs. This approach implies redesigning PCR primers that were close to the repeated sequence. (Wiegand & Kleiber, 2001) demonstrated that it could be useful for the successful typing of degraded DNA. In search of improving sensitivity DNA detection, another adaptation in PCR methodology was carried out in 1990 in the

United Kingdom's Forensic Service (Gill, 2001). The increase in PCR amplification cycles enables STR typing with samples containing less than 1000 pg of DNA.

In ancient and forensic DNA, mtDNA analysis is the solution to obtain genetic information when nuclear DNA is highly degraded. In most cases, this occurs in old specimens. PCR amplification of mtDNA is usually done with 34 to 38 cycles. Protocols for highly degraded DNA specimens even call for 42 cycles (Gabriel et al., 2001). Moreover, small amplicons to improve amplification success is employed; at the same in the STRs case, the use of smaller-sized PCR products improves the recovery of information from the original DNA template. Mini-mito primer sets have been developed to amplify smaller portions of HV1 and HV2 (Eichmann & Parson, 2008; Gabriel et al., 2001). The use of mini amplicons that overlaps one another is sometimes referred to as an "ancient DNA" approach and can recover abundant DNA in a sample that might otherwise fail to produce results with a standard protocol. This approach was used to successfully recover information from Neanderthal bones that are thousands of years old (Krings et al., 1997).

1.6.2 Sanger sequencing

In 1977 gave the place the modern history of DNA sequencing. Frederick Sanger reported his method for determining the order of nucleotides of DNA involving electrophoresis and is based on the random incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication (Sanger et al., 1977). In the same year, the first human gene was isolated and sequenced. In 1986, Hood and co-workers described an improvement in the Sanger sequencing method that included attaching fluorescent dyes to the nucleotides, which permitted them to be sequentially read by a computer. The first automated DNA sequencer was developed by Applied Biosystems in 1987. Nowadays, the performance of DNA chemical sequencing has progressed, increasing the sensitivity obtaining results of less material as little as 1ng of DNA PCR product. More recently, next-generation sequencing methods have replaced higher volume Sanger sequencing, especially for large-scale, automated genome analyses. However, the Sanger method remains widely used for smaller-scale projects and validation of deep sequencing results. It still has the advantage over

short-read sequencing technologies (such as Illumina) in that it can produce DNA sequence reads of >500 nucleotides and maintains a very low error rate with accuracies around 99.99% (Shendure & Ji, 2008). Sanger sequencing is still actively being used in efforts in medical and forensic contexts, while in ancient DNA sequencing was gradually being substituted for NGS.

1.6.3 Next Generation Sequencing-NGS

Since around 2005, massively parallel sequencing (MPS) or next-generation sequencing (NGS) technology has been commercially available and proved to be superior to Sanger sequencing. NGS enables for sequencing of millions to billions of short reads in a single run. Many different MPS platforms have been developed to address sequencing in various fields of life sciences. In forensic genetics, two benchtop sequencers are common, the Illumina (MiSeq FGx) and the ThermoFisher (Ion Torrent PGM and Ion S5), while in the aDNA community most DNA sequence data have been produced using Illumina instruments such as the HiSeq4000, HiSeqX, and NovaSeq platforms. These platforms work optimally on relatively short DNA sequences (<400 bp) and are well suited to sequence DNA in the range of 50–300 bp, which characterizes degraded DNA in ancient and forensic DNA communities (Orlando et al., 2021).

Illumina sequencers employ a sequencing-by-synthesis in which a polymerase is used along with a signal, such as a fluorophore, while the Ion Torrent instrument involves a change in pH to identify the incorporation of a nucleotide. These new sequencing methods have been improved in three aspects compared to the conventional technologies. First, NGS does not require bacterial cloning of DNA fragments, as it is normally used in typing mtDNA fragments (Hatsch et al., 2006); instead, the preparation of NGS libraries is employed in a cell-free system. Second, instead of hundreds of sequencing reactions, they can parallelize the thousands-to-many millions of sequencing reactions. Third, the sequencing output is directly detected with no need for electrophoresis. The enormous number of reads generated by NGS enables the sequencing of even entire genomes at an unprecedented speed.

Figure 15 illustrates the experimental workflow that the ancient and forensic DNA communities can follow from critical samples such as skeletal remains to obtain genetic

information by NGS. Although this illustration is taken from a review of ancient DNA analysis, the workflow in forensic genetics could have the same experimental steps (Orlando et al., 2021).

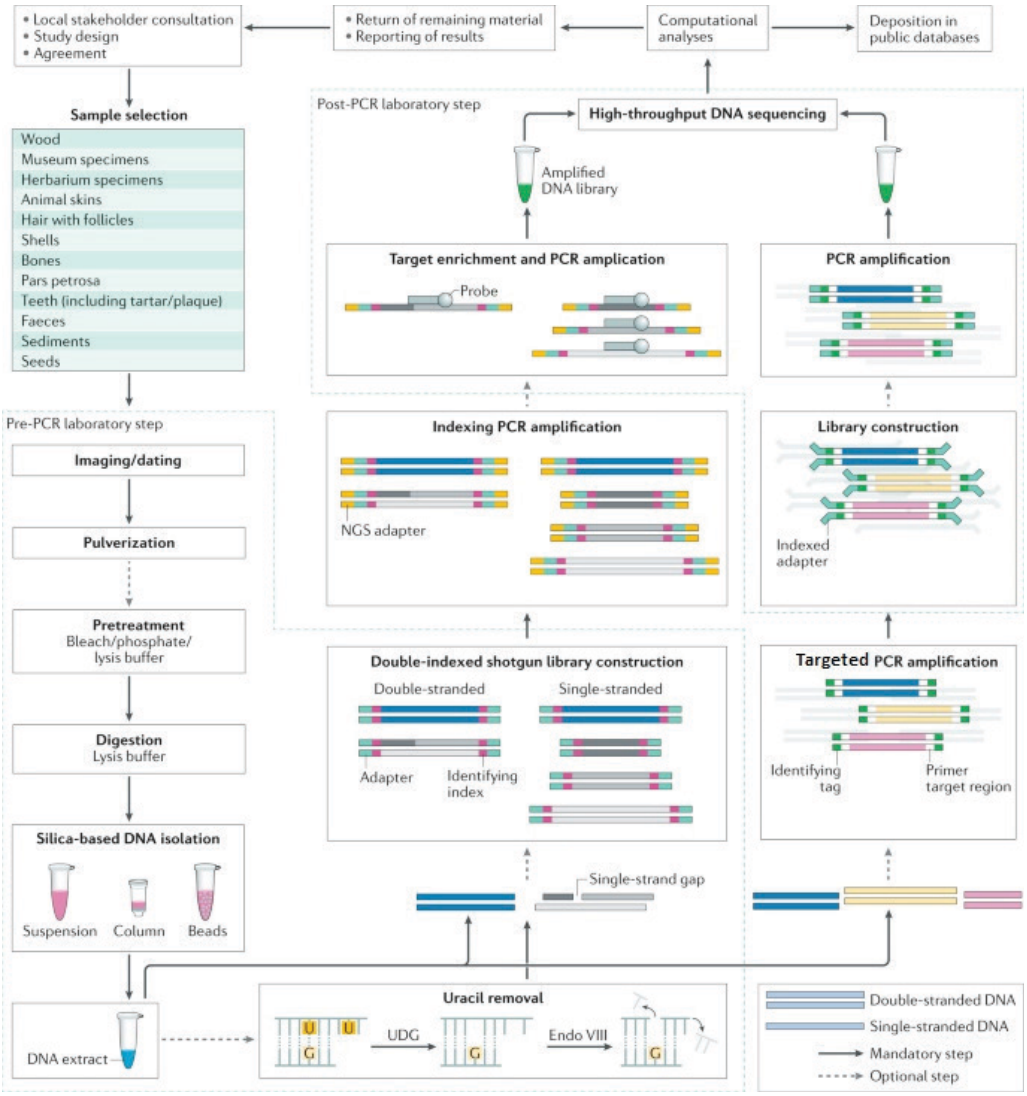


Figure 15 — Experimental workflow by NGS. An ancient and forensic community uses a wide range of samples with degraded DNA. Before sample destruction, a research plan should be performed. The different laboratory procedures must be carried out in specific separate areas. On the one hand, all pre-amplification experimental steps, including sample preparation, DNA extraction, optional USER treatment (uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII), New England Biolabs), and DNA library construction, and on the other, target enrichment and PCR amplification. Following next-generation sequencing (NGS), the sequence data are processed on computational servers and uploaded to public repositories. Adapted from Orlando et al., 2021.

NGS generates reads of samples which, through data analysis could detect the post-mortem damage showing the rate of misincorporations (G to A and C to T). These misincorporations increase towards the ends of reads when mapped against a reference sequence owing to cytosine deamination preferentially occurring in the single-stranded overhanging termini of fragmented DNA. This type of damage can induce sequence errors, but its impact decrease if DNA extracts are treated before library construction with a commercialized enzymatic mix of uracil–DNA–glycosylase (UDG) and endonuclease VIII (Endo VIII) known as the USER reagent (New England Biolabs). This reagent removes uracils and cleaves the resulting abasic sites, thereby cutting damage and shortening the DNA molecule's length.

This removing of the damage patterns could be a negative effect hindering the differentiation from contaminating DNA. In the case of mammalian nuclear DNA, this damage signal can still be observed by examining CpG dinucleotides. Another solution is to use a modified USER protocol (UDG-half) that removes most damage but retains single uracil at each end of the molecule allowing authentication of the retrieval of fragmented DNA (Briggs et al., 2010). During downstream data analysis, the single uracil can be masked or clipped; this analysis may improve calculations regarding the relative importance of sequence fidelity, sequence length, and cost for a given study design and research question.

After the treatment of molecular damage, the library construction must be performed. DNA library is a molecular construction in which DNA fragments are ligated to DNA adapters of known sequences to be amplified and optionally captured before sequencing; different sequencing platforms require different library constructs. This section will focus on the type of library constructions for Illumina platforms applied in forensic and ancient DNA research.

1.6.3.1 Library preparation and sequencing methods

Two main library preparation methods are currently available for sequencing by synthesis (double-stranded library preparation and single-stranded library preparation).

The main type of NGS used for the forensic genetics community is double-stranded library preparation by targeted PCR amplification followed by sequencing, and for the ancient DNA community is double-stranded library preparation and single-stranded library preparation followed by shotgun sequencing. In the double-stranded preparation method, DNA molecules are end-repaired and ligated to double-stranded adapters (Carøe et al., 2018; Kircher et al., 2012), whereas, in the single-stranded preparation method, the process includes heat-denatured DNA templates and adapters ligation as single-stranded molecules (Gansauge et al., 2017; Kapp et al., 2021). Each one has different characteristics in terms of costs, hands-on time, and sensitivity.

1.6.3.1.1 Targeted PCR amplification and multiplex sequencing

A convenient way to overcome the problem of retrieval of degraded DNA with low content of endogenous DNA in skeletal remains is to enrich the genomic regions of interest by PCR amplification. In forensic genetics, the regions used to identify an individual are autosomal or Y STRs, mtDNA-CR, or SNPs. The first step to enrich these genomic regions is the same used in the conventional techniques, in which pool-specific primers pairs that span these molecular markers are used. There are a variety of commercially available kits in the conventional techniques to amplify STRs markers. Some kits amplify just autosomal STRs markers, and another amplifies Y or X-STRs.

Therefore, the use of different kits to amplify various STRs markers by separate suppose high cost and time demanding with poor results compared to NGS technology. As has been mentioned in the section 1.4, recently, a new DNA profiling system has become available for the forensic community, and it is based on next-generation sequencing. This technique allows large amounts of loci to be characterized for each sample and are potentially very highly discriminating; however, their use is currently limited. NGS platforms such as the Illumina MiSeq FGx in combination with the Verogen ForenSeq DNA Signature Prep Kit, the ForenSeq mtDNA Whole Genome Kit or mtDNA Control Region Kit or MiSeq in combination with PowerSeq 46GY System are recently being tested in forensic genetics.

MPS libraries involved repair of ends and adenylate end 3' of amplicons products; subsequently, through the adenine tail in the 3'end incorporated, the paired-end adapters and indexed are ligated. Finally, the library construction is validated by qPCR. The validation quantifies molecules that do not have adapters on both ends and do not form clusters. More of these non-clusterable molecules are due to the absence of PCR enrichment; thus, these libraries should not be sequenced to avoid spending money. The validated libraries are normalized and pooled to run in the MiSeq platform.

Contrary to STRs markers, there are no commercially available kits for mtDNA typing by PCR-multiplex in the conventional technique. However, kits have already been developed for NGS technology that can amplify the control region or whole genome of mtDNA using one or two sets of primers by PCR-multiplex. Normally, the library preparation in these kits involves the same steps employed for NGS-STRs typing (amplification, library construction, library validation, and normalized/pooled). PowerSeq™CRM Nested System kit (Promega Corporation) employs fewer steps than the other kits as it links or nested the amplification of the entire mtDNA-CR and the ligation of adapters in a single multiplex-PCR. Namely, this workflow combines primers needed for amplifying the targeted sequences with primers that contain unique indexing and adapter sequences for sequencing on the MiSeq instrument (Figure 16). This process greatly simplifies library preparation by saving time, decreasing sample loss, and reducing data variability. The amplified products and library construction of mtDNA-CR typing can be quantified by qPCR or fluorometry. The latter is designed specifically for use with the Qubit Fluorometer and is highly selective to quantify double-stranded DNA (dsDNA).

1.6.3.1.2 Shotgun Sequencing

The shotgun technique is a non-targeted sequencing for determining the DNA sequence of an organism's genome. It is usually used in the ancient DNA communities to capture all the fragmented DNA converted into a library.

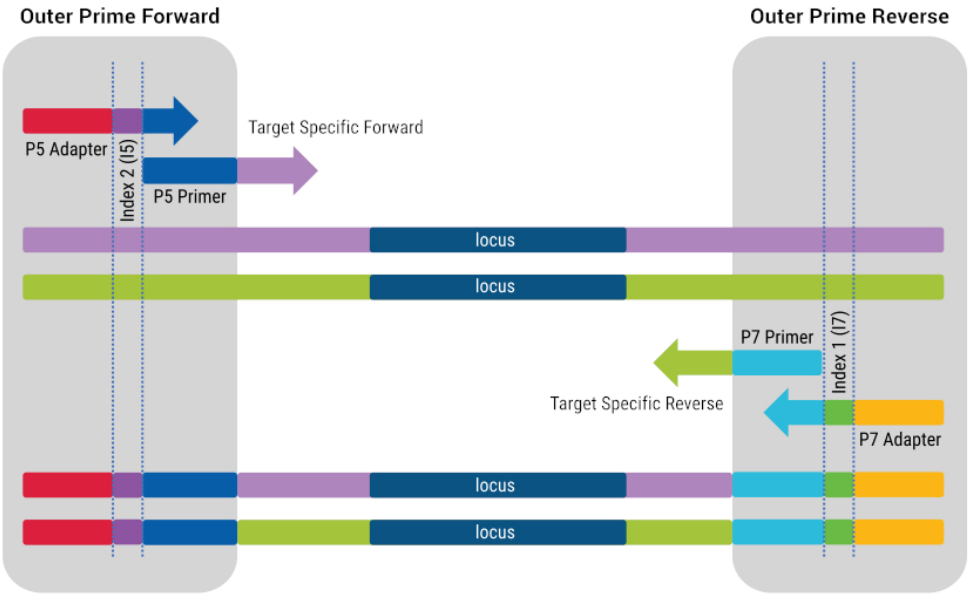


Figure 16 — Dual indexing in a 4-primer approach. The two outer and two inner primers are combined in one PCR reaction to yield a MiSeq compatible product with dual indexing. Each sample is first mixed with a unique combination of outer primer indexes. Adapted from Lange et al., 2014.

The double-stranded (dsDNA) library preparation method is optimized for degraded DNA and is the most used in aDNA. The Meyer and Kircher (MK) approach is based on double-stranded DNA library preparation (Meyer & Kircher, 2010), which includes an end-repair step that fills in bases present as single-stranded overhangs to create blunt-ended molecules onto which the sequencing adapters can be ligated. However, the blunt-end ligation of the two sequencing adapters is non-directional, which means that either of the two adapters can be added to each end of the molecule. Only molecules with one of each adapter in the correct orientation can be sequenced. Therefore, half of the molecules are lost due to incompatible adapter combinations. Additionally, this double-strand library preparation is laborious as it requires three purification steps before amplification, which can lose more DNA libraries. The BEST protocol is a more recently developed double-stranded DNA library preparation and is performed in a single tube, using heat denaturation rather than purification between reaction steps. BEST has been shown to yield higher complexity libraries than other double-stranded library protocols (Carøe et al., 2018).

Single-strand library preparation shows increased sensitivity for capturing all damaged DNA input pools in the sample, such as fragmented double-stranded DNA (dsDNA) with overhangs or blunt ends, nicked dsDNA, and ssDNA (single-strand DNA). The double-strand library preparation converts only dsDNA with blunt ends or overhangs into library-ready molecules. However, the single-stranded DNA library preparation method has been considered a revolutionary technique for aDNA research because that allows for the recovery of all the damaged DNA input pools. Another advantage of the single-stranded preparation method is that it enables the molecular selection of DNA templates carrying evidence of post-mortem DNA damage, increasing the fraction of endogenous DNA incorporated into sequencing libraries and thus reducing downstream sequencing cost (Gansauge & Meyer, 2014). The Santa Cruz Reaction or SCR based on single-stranded library preparation uses directional splinted ligation of Illumina's P5 and P7 adapters to convert natively single-stranded DNA and heat-denatured double-stranded DNA into sequencing libraries, requires no exotic reagents, is fast, simple, and has a ligation efficient of the adapter with minimal effort and increased sensitivity (Figure 17). Moreover, experimental procedures in which libraries are prepared within single tubes are also available, further reducing manipulation and hands-on time while facilitating parallelization (Kapp et al., 2021).

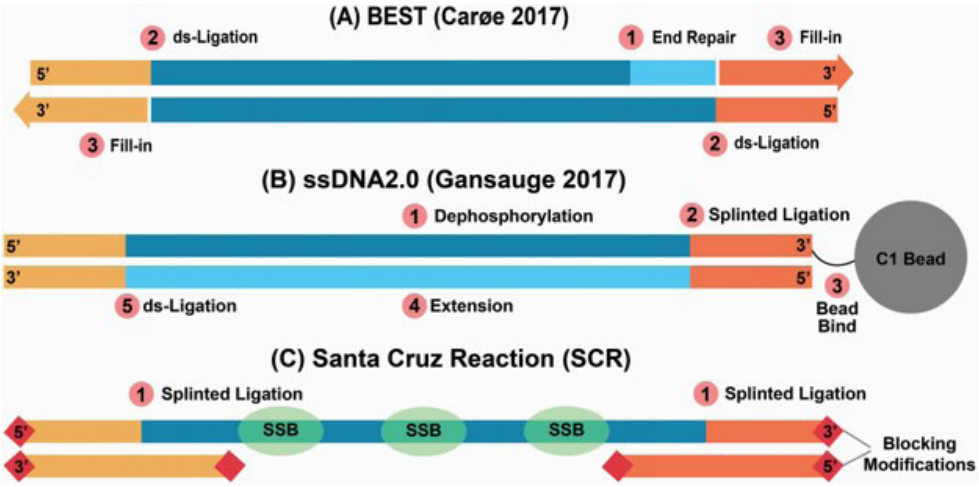


Figure 17 — Single-stranded library preparation methods as the Santa Cruz Reaction begin with a denaturation step in which all DNA molecules in the extract are converted to single-stranded form. This allows conversion of DNA preserved in a single-stranded state and separate conversion of both strands of DNA preserved in a double-stranded state. Then simultaneously is ligated Illumina's P5 and P7 adapter using splinted ligation (Kapp et al., 2021).

All NGS DNA libraries include unique identifying short sequences known as indexes integrated within their adapters to ensure traceability of the samples. The tag-indexed samples allow multiple samples to be pooled within sequencing run, greatly reducing contaminations risks within the laboratory because an adapter-index combination unique (barcode) is carried for each sample. Using indexes on both adapters is highly recommended to allow the detection of chimeric DNA templates that can form through jumping PCR and index switching during cluster generation (Kircher et al., 2012)

1.6.4 The impact of NGS on ancient DNA and forensic genetics

Next-generation sequencing has meant a tremendous technological advance in some areas of knowledge related to genetics. This technique has been more evident in scientific communities such as ancient DNA and forensic genetics, which work and study highly degraded biological samples. NGS has allowed a better understanding of the complex processes of chemical DNA degradation from fossils preserved in situ (Overballe-Petersen et al., 2012), and it has had an important impact on the understanding of phenomena such as cancer, aging, evolution, and forensic identification (Hoeijmakers, 2009; Latham & Miller, 2018).

Although ancient DNA research had been an interesting but marginal discipline for many years, the advent and application of NGS technology have made it a strategical method for studying evolutionary biology (Knapp & Hofreiter, 2010). This novel sequencing technology has impacted the fact that it allowed for contamination detection in the data presented in the past carrying out an improvement in the recovery of ancient genomes sequenced and giving ancient genomes more efficient and reliable.

Although the ancient DNA extractions have largely remained un-automated compared to forensic lab procedures (Kallupurackal et al., 2015), next-generation sequencing has motivated the improvement and development of novel protocols that allow for automated extraction (Rohland et al., 2018).

The revolutionary effect also has expanded to forensic genetic applications, despite this discipline having a shorter history than ancient DNA. NGS technology has encouraged the development of novel markers to discriminate between individuals and find out the biogeographical ancestry and deconvolute mixed samples (Bulbul et al., 2018). The adoption of NGS technology in the forensic context has been widely accepted, and consequently, the recovery of smaller DNA fragments could be possible by extraction methods applied in ancient DNA research (Dabney et al., 2013; Palomo-Díez et al., 2017; Rohland et al., 2018)

NGS technology has opened the possibility to obtain more information from unique samples in a single experiment by analyzing combinations of markers (STRs, SNPs, insertion/ deletions, mRNA), offering new possibilities for forensic genetic casework or even other applications (Børsting & Morling, 2015).

In both research areas (Forensics and Ancient DNA), the development of NGS platforms has meant an important technological and economic impact for research companies and public research institutions which have invested high amounts of resources in improving and adapting the novel and better platforms to their requirements (Aly et al., 2015; Butler, 2015).

In general, the arrival of NGS technology has had a deep impact on scientific knowledge due to the increasing speed with which whole genomes can be sequenced, and the development of instruments, kits, and software for analysis is required. At the same time, the economic impact is also noteworthy since complete genome information can be generated with high quality, precision, and in short periods, decreasing costs in forensic or ancient DNA laboratories.

1.7 Bioinformatic approaches in Forensic and aDNA

Various tools and pipelines are available for ancient and forensic DNA sequence analysis. For aDNA, the selection of appropriate procedures depends on the objectives project; for instance, if the research purpose is about modeling population history, microbial profiling,

or paleogenomic, there are numerous additional steps in the analytical workflow that should be applied to the common steps. Regarding forensic genetics, there are pipelines and software (usually paid) that have been developed and validated along with the available commercially NGS kits.

In general terms, the raw FASTQ data produced by the NGS instrument must pass by a pre-processing of reads in which adapter sequences must be trimmed from short DNA inserts. After that, the reads are mapped to either the mitochondrial or nuclear genome, depending on the recovered molecular markers. Processed reads are typically aligned using BWA (Li & Durbin, 2009) or Bowtie (Langmead & Salzberg, 2012) tools.

In aDNA analysis, the assessment of authenticity and error rates, including miscoding lesions resulting from post-mortem damage, are normally analyzed. DNA molecules can be sequenced in one or both directions, resulting in single or paired-end reads. When paired-end reads are employed, the overlapping regions may be merged while considering per-base quality scores, and PCR duplicates are also filtered. Some of the many analytical tools used in aDNA analysis and freely available are presented in Table 8.

Table 8 — Analytical tools used in aDNA analysis.

Software	Link	Description
PALEOMIX (Schubert et al., 2014)	https://paleomix.readthedocs.io/en/latest/	Read alignment and processing, phylogenomics
nf-core/EAGER1 (Peltzer et al., 2016)	https://eager.readthedocs.io/en/latest/index.html https://github.com/nf-core/eager	Read alignment and processing
mapDamage2 (Jónsson et al., 2013)	https://ginolhac.github.io/mapDamage	Post-mortem DNA damage assessment
PMDtools (Skoglund et al., 2014)	https://github.com/pontussk/PM	Selection of reads showing signatures of post-mortem DNA damage
lcMLkin (Lipatov et al., 2015)	https://github.com/COMBINE-lab/maximum-likelihood-relatedness-estimation	Kinship inference
READ (Monroy Kuhn et al., 2018)	https://bitbucket.org/tguenther/read/src/master/	Kinship inference
ROHan (Renaud et al., 2019)	http://grenaud.github.io/ROHan/	Heterozygosity estimates and runs of homozygosity

Regarding the process of forensic NGS data, STRait Razor is a software freely available based on R programming language and can analyze the NGS data for 44 STRs, including 23 autosomal and 21 Y chromosome STRs (King et al., 2017; Warshauer et al., 2013), In this software a reference allele database is available when STRs-NGS kits such as ForenSeq DNA Signature Prep Kit and PowerSeq 46GY System kit are used.

There is also a package of data analysis tools for Next-Generation Sequencing of forensic DNA markers; FDSTools include tools for characterization and filtering of PCR stutter artifacts and other systemic noise and automatic detection of the alleles in a sample. The software package is written in Python and runs on Windows, Linux, or MacOS (Hoogenboom et al., 2017, 2021).

GeneMarkerHTS software provides a validated streamlined workflow for forensic mitochondrial DNA, STR, and Y-STR casework from massively parallel sequencing platforms (MPS) - such as the Illumina and Ion Torrent. NextGENe software is also an analytical tool for analyzing desktop sequencing data produced by Illumina iSeq, Miniseq, MiSeq, NextSeq, HiSeq, and NovaSeq systems, Ion Torrent Ion GeneStudio S5, PGM, and Proton systems as well as other platforms. The latter two software are normally paid for and validated by forensic laboratories (Brandhagen et al., 2020; Holland & McElhoe, 2015; Huszar et al., 2019; McElhoe et al., 2014). ForenSeq Universal Analysis and Ion Reporter are software for analyzing mtDNA and STRs, which are included when NGS-kits are acquired to be used with MiSeqFGx and Ion Torrent sequencers.

1.8 Similarities and differences between Forensic and Ancient DNA

Retrieval DNA from human remains in ancient and forensic DNA has technically and analytically similarities and differences. However, when the identification of skeletal remains is necessary, the main aim is to join all technical efforts to obtain the most reliable genetic information which allows for their identification. Most of the techniques based on target amplification by NGS have not success and the use of shot-gun techniques and

criteria of authenticity are necessary. Table 9 shows the features of the technique that are similar or different in both research communities.

Table 9 — Similarities and differences of techniques features in forensic genetics and ancient DNA. Adapted from Capelli et al., 2003.

Techniques Features	Forensic genetics	Ancient DNA
Separated workstation and labware	Yes	Yes
Investigation of biochemical preservation (authentication of the retrieval of fragmented DNA)	–	Yes (molecular damage, deamination, depurination, etc.)
Clean extract/PCR controls	Yes	Yes
Cloning of PCR products	–	Yes, although less and less frequent
DNA quantitation	Informative	Yes
Phylogenetic test	–	Yes
Independent reproduction of the results	Yes	Recommended for human remains
Age samples	Up to 100 years and or medico-legal significance which exhibit degraded DNA characteristics	Highly always degraded over 50 years, typically much older and are not of medical-legal significance.
Elaboration of the report to the court	Yes, in the case of a missing person where the biological sample is skeletal remains.	No, just in case of having legal implications.
STRs typing	Yes	-
SNPs Typing	Yes	Yes
MtDNA typing	Yes, (Control Region-CR)	Yes, (CR and Whole-genome)
NGS based on amplicons	Yes	less and less frequent
NGS is based on the capture of all DNA	-	Yes

2

Objectives and Justification

"Everything is theoretically impossible until it is done
-Robert A. Heinlein

This dissertation aims to connect current standard practices in aDNA research and forensic genetics to increase the recovery of genetic information from highly degraded skeletal remains. Some techniques have been tested to accomplish these aims that could be alternative routes to achieve the genetic identification of victims of armed conflicts, missing persons, or disaster victims where skeletal remains are the only source of DNA.

The proposed objectives are developed in three chapters:

- To standardize DNA extraction methods from different highly degraded skeletal elements

In Chapter 1, this aim was completed by an article published in *Electrophoresis* journal titled *Human DNA Extraction from Highly Degraded Skeletal Remains: How to find a suitable method?*

Forensic genetics has grown substantially during the last decade due to improvement in the sensibility of DNA technologies. However, the generation of DNA profiles from human remains is still challenging since many factors influence the efficiency of DNA recovery. In cases of mass disasters (natural or caused), wars, or criminal cases, skeletal remains are often the only biological material remaining for the identification process. In this context, *postmortem* interval (Calacal et al. 2015; Rothe and Nagy 2016), type of skeletal remains (long bones, petrous bones, or teeth) (Hansen et al., 2017b; Mundorff & Davoren, 2014), decomposition degree (Scorrano et al., 2015), and *postmortem* conservation conditions (Desmyter et al. 2017; Jakubowska, Maciejewska, and Pawłowski 2012; Mameli, Piras, and Delogu 2014), limit DNA recovery. The degradation of biological material leads to low DNA preservation, fragmentation into small pieces of 25-30bp (Glocke and Meyer 2017),

and molecular damage, which hinders the extraction and purification of genetic material. Therefore, several strategies have been used to retrieve DNA from bones and teeth, some established in forensic laboratories (Huel et al. 2012; Marshall et al. 2014) and others coming from the ancient DNA research community (Gamba et al. 2016; Rohland et al. 2018).

In this sense, more systematic studies are needed to evaluate the efficiency of the different silica-based extraction methods considering other factors such as the type of skeletal remains. The main goal of the present study is to establish the best extraction method and the type of skeletal remains that can maximize the recovery of PCR-amplifiable DNA and the subsequent obtaining of mitochondrial and nuclear genetic information relevant for genetic identification.

- To assess the DNA preservation from newborn burials by comparing different types of skeletal remains and different age categories using the shotgun-NGS technique.

In Chapter 2, this aim was completed by an article prepared for submission titled *Newborn Skeletal Remains: an aid or a constraint to paleogenomics?*

Infant human skeletal remains from prehistory periods always raise many questions from the archaeological, biological, and cultural points of view. Some of the most prominent questions are related to ritual burials or even infanticide (Abu-Mandil Hassan et al. 2014; Afonso et al. 2019; Faerman et al. 1998; Millett and Gowland 2015). Moreover, the main burial ritual in some periods and cultures was cremation, and only a few infants' burials have been found in houses and productive areas. These represent the only biological material available to establish inferences about populations that live in such specific periods and cultures. In this context, ancient DNA (aDNA) analysis can be particularly informative in solving these questions and inferring ancient communities' population origin and structure. Also, in the forensic field, the genetic information obtained could solve forensic cases through sex determination or kinship inference (Minaguchi et al. 2003).

Authentic ancient DNA molecules, referred to as endogenous DNA content, represent less than 1% of the vast majority of DNA extracts from old specimens (Carpenter et al. 2013).

This limited fraction of endogenous DNA results from the decomposition process over time, during which the tissues are invaded by microorganisms and fungi, leading to the degradation of DNA (Allentoft et al. 2012). Several studies (Damgaard et al., 2015; Parker et al., 2020) suggest that a well-preserved endogenous DNA recovery depends on the type of skeletal remains. Some bones are likely located more protected against microbiota attacks than others. The temporal bone in the cochlear region (Nieves-Colón et al. 2018) have demonstrated to be superior in the human endogenous DNA preservation compared to the tooth (Adler et al. 2011; Hansen et al. 2017b) or postcranial bones (Hong, Kim, and Park 2017), used both in ancient DNA studies (Pinhasi et al. 2015) and forensic genetics (Gonzalez et al. 2020). These findings have been established for adult skeletons from various archeological contexts from the Bronze Age until the 18th century (Damgaard et al. 2015; Parker et al. 2020). However, there are no systematic studies on newborn or infant individuals until today. Infant skeletal remains are not entirely developed, are less protected than the adult ones, and are very limited with tiny-sized bones and lack of dental development.

In the last years, broad studies of ancient European populations have allowed thousands of low coverage ancient genomes (Gansauge et al. 2020; Olalde et al. 2019; Patterson et al. 2022). Although the great majority of samples belong to adult individuals, some studies include an important number of samples from infant individuals, particularly the recent one from Patterson et al. (2022).

In this context, our main goal is to evaluate the viability of infant remains to perform paleogenomics studies. Two specific goals were defined based on intra and inter-individual approaches: 1) to investigate if there are differences between petrous bone and vertebra shotgun-NGS libraries when both remains belong to the same infant individual; and 2) to investigate if petrous and other skeletal remains present differences and also if there are differences between performance according to the age of the individuals (perinatal, infantile, juvenile and adult).

- To standardize the recovery of the whole control region of mtDNA by Single-Multiplex Massively Parallel Sequencing from Highly Degraded Skeletal Remains.

In Chapter 3, this aim was completed by an article submitted to the *Electrophoresis* journal titled *MtDNA-CR Typing from Highly Degraded Skeletal Remains by Single-Multiplex Massively Parallel Sequencing*.

The analysis of mitochondrial DNA (mtDNA) allows the retrieval of useful genetic information in forensic contexts where highly degraded human skeletal remains are the only source of genetic material. The higher copy number of mtDNA per cell (can be >1000) increases the success of DNA recovery compared to the analysis of nuclear DNA —autosomal STR (Short Tandem Repeats)— commonly used in forensic genetics (Butler and Levin 1998). MtDNA is useful at the identification level by comparing genetic polymorphisms to determinate maternal kinship and to infer the population origin of subjects.

The traditional forensic mtDNA analysis involves the amplification and sequencing of its mtDNA control region (CR). When the sample has good quality, two distinct regions of around 200-400bp are usually amplified by PCR and sequenced using Sanger technology (Wilson et al. 1995). However, this attempt to obtain >200pb amplicons in samples of poor quality may not provide any results since the DNA is highly fragmented into smaller molecules.

The recovery of small amplicons (100-130pb) has been proved successfully in ancient DNA research and forensic analysis of bones and hair shafts (Alonso et al. 2003). Since then, some authors have demonstrated that using a set of shorter amplicons is an effective alternative when highly degraded samples have been analyzed (Eichmann and Parson 2008; Gabriel et al. 2001). However, at least eight fragments must be amplified and sequenced to recover the entire Control Region representing at least 8 PCR amplification reactions and 16 Sanger sequencing reactions (Gabriel et al. 2001). Since mtDNA testing for damaged samples is generally manually performed, the procedure is very demanding and time-consuming. Over

the past five years, the analysis by massively parallel sequencing (MPS) or next-generation sequencing (NGS) has shown to be a highly robust technique for typing mtDNA sequences in forensics (Holland et al., 2017; Jäger et al. 2017; McElhoe et al. 2014; Wilson et al. 1995). NGS provides larger genetic informative data set with higher throughput. It also offers the capability to reduce workflows and have an overall lower cost per nucleotide than capillary electrophoresis-based methods (Mardis, 2008; Yang et al., 2014).

Nowadays, there are NGS commercial kits that allow mtDNA-CR analysis. The ForenSeq mtDNA Control Region Kit (Verogen) (Verogen Inc 2019) can recover 18 overlapping amplicons of 60-150pb using 18 primer pairs distributed in two multiplexes. The Precision ID mtDNA control region Panel Kit (ThermoFisher) (Xpress, Adapters, and A 2016) can recover 14 amplicons of 163pb using 14 primer pairs distributed in two multiplexes. In both kits, two PCR reactions are performed to obtain all mtDNA-CR typed, and then one additional step of PCR is required to add adapters and indexing samples.

The PowerSeq™CRM Nested System kit (Promega Corporation) employs a multiplex-PCR strategy to amplify the entire mtDNA-CR in a single multiplex that combines ten primer pairs, generating ten overlapping short amplicons (147-237pb). This kit can simplify workflow using a nested amplification protocol by incorporating indexed adapters for Illumina MiSeqTM (Illumina, Inc., San Diego, CA, USA) single-multiplex PCR reaction in only one reaction step, largely reducing amplicon manipulation and risk of contamination. FBI Laboratory has validated the technique according to validation guidelines for forensic DNA of SWGDAM (Scientific Working Group on DNA Analysis Methods). Overall, this validation established the reliability of this kit for accurate mtDNA control region typing to aid in forensic identifying procedures (Brandhagen et al., 2020). The validation has been performed for buccal swabs, hair samples, and some skeletal remains samples. However, more systematic studies are needed for critical samples.

The aims of our study were: i) to analyze the success of mtDNA-CR typing of highly degraded human skeletal remains coming from different chronological ages by single-

multiplex massively parallel sequencing using the PowerSeq™CRM Nested System kit; ii) to compare the variants or polymorphisms called using an in-house pipeline against the GeneMarker® HTS software (Holland, Pack, and McElhoe 2017) (most recently released software package used by forensic community); and iii) to evaluate the concordance of the mtDNA MPS data with Sanger technique.

3

Methodology
Results and Discussion

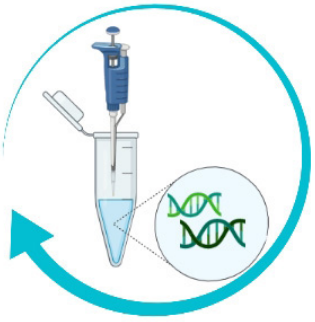
"The capacity to blunder slightly is the real marvel of DNA. Without this special attribute, we would still be anaerobic bacteria, and there would be no music."
-Lewis Thomas

3.1. Chapter 1:

Human DNA Extraction from Highly Degraded Skeletal Remains: How to find a suitable method?

Corpus of the manuscript published in Electrophoresis

Vinueza-Espinosa, D. C., Santos, C., Martínez-Labarga, C., & Malgosa, A. (2020). Human DNA extraction from highly degraded skeletal remains: How to find a suitable method? Electrophoresis, 41(24), 2149-2158. https://doi.org/10.1002/elps.202000171.



DNA neither cares nor knows. DNA just is and we dance to its music

-Richard Dawkins

3.1.1 Material and Methods

Sample Preparation

Five individuals were selected from an archaeological site in (Catalonia-Spain) dating from 5-11th centuries AD. For each individual, five samples from different skeletal regions were collected: petrous bone, pulp cavity, and cementum of tooth, rib, and at least one different limb bone such as radius, ulna, metacarpal, or phalange.

To eliminate possible exogenous contaminants, each sample was cleaned, removing a thin layer with a sterile tungsten tip placed into a micro drill up to 5000 rpm. After, careful drilling was performed in the cleaned area using a sterile tungsten tip, collecting roughly 120-800 mg of powder sample per each type of skeletal remain. Then powder was subdivided into different tubes (Table SI: 1), one for each DNA extraction method used in this study (Figure 18). All the procedures were conducted in highly sterile conditions in a specific ancient DNA laboratory.

DNA extraction

Four extraction methods were tested; three silica-based and the classical method based on Phenol-Chloroform (Figure 18). Experimental details of the extraction methods procedures and the composition of reagents are detailed in Supporting Information (Table SI: 2). A negative extraction control without powder was also included per each set of five samples, representing a total of 100 DNA extracts and 25 controls.

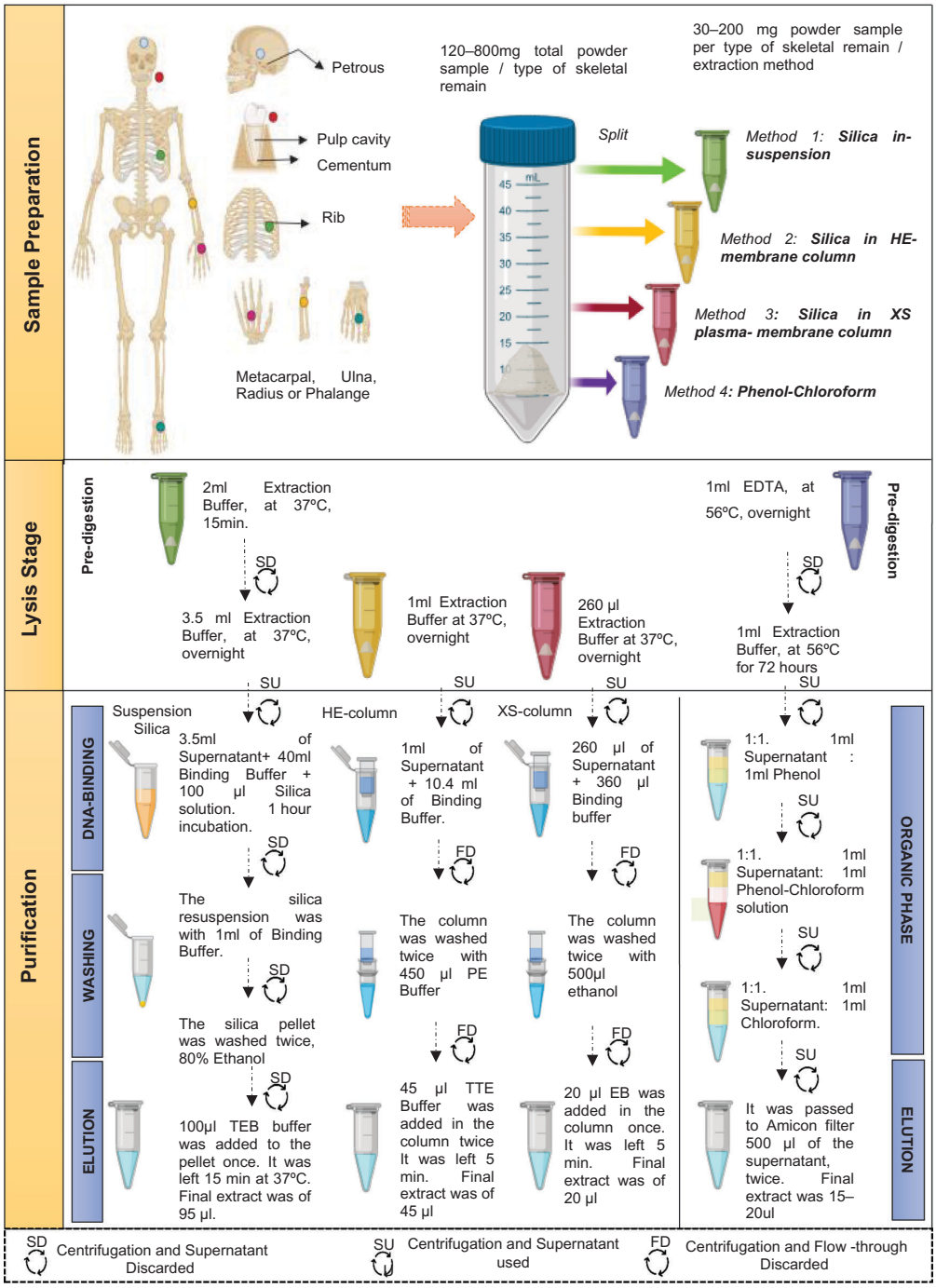


Figure 18 — Workflow of the DNA extraction from petrous, pulp cavity, cementum of the tooth, rib, and limb bones by using the four extraction methods.

Method 1: Silica in-suspension

The present DNA extraction protocol was a variant of the method described by Allentoft et al. 2015 (Allentoft et al., 2015).

Method 2: Silica in HE- membrane column

This method had been optimized in our laboratory for ancient bones and teeth using only the silica-membrane columns of the **H**igh Pure **E**xtender Assembly Kit (HE-column) (Roche, Basel Switzerland)(Rohland et al., 2010, 2018; Rohland & Hofreiter, 2007b). These spin columns efficiently purify viral nucleic acids (short molecules) of large sample volumes up to 3ml.

Method 3: Silica in XS plasma- membrane column

The NucleoSpin® Plasma XS kit is designed to isolate circulating DNA from human blood plasma efficiently. The Fragmented DNA as small as 50-1000 bp can be purified with high efficiency.

Method 4: Phenol-Chloroform

The present DNA extraction protocol was a variant of the method described by Hagelberg & Clegg (Hagelberg & Clegg, 1991).

DNA Quantification

Total DNA quantification was performed using Nanodrop ND-1000 (ThermoFisher). To know if the recovered DNA corresponds to short or long DNA fragments, DNA was sized and quantified using the Bioanalyzer High Sensitivity DNA Kit ® (Agilent Technologies).

PCR amplification and sequencing of mitochondrial DNA

For all DNA extracts were amplified and sequenced, a 203-bp region from the Hypervariable Region I (HVRI) of mitochondrial DNA (mtDNA) located between position L16030 and H16230 using primers Forward 5'-CATGGGGAAGCAGATTTGGG-3' and Reverse 5'-GATAGTTGAGGGTTGATTGCTG-3' (Simón et al., 2011) and a 238pb

region from the Hypervariable Region II (HVRII) located between position L27 and H264 using primers Forward 5'-GGTCTATCACCCTATTAACC-3' and Reverse 5'-CTTTCCACACAGACATCAT -3'(Simón et al., 2011). The PCR reactions were carried out using the Rotor-Gene®-Q in a 10µl reaction containing 2X Type-it HRM PCR Kit (Qiagen, Germany), 10 pmol/µl primers, and 1-5µl of DNA template. The conditions were: hold at 95ºC for 10 min, denaturation at 95ºC for 10 seconds, annealing at 58ºC for HVRI and HVRII for 45 seconds, and extension at 72ºC for 10 seconds with 55 cycles.

PCR products were purified using MSB® Spin PCRapace (Invitek, Germany) according to the specifications protocol and quantified using Qubit® dsDNA Assay Kit (ThermoFisher) following the manufacturer procedures. MtDNA PCR products were then sequenced using BigDye® Terminator Sequencing Kit v1.1 in 10µl reaction containing: 0.5µl Big Dye, 1µl Buffer, 1µl primer 5pmol/µl, 7µl H_2O and 0.5µl DNA. Finally, the sequence reactions were purified with BigDye ® X-terminator purification Kit. The Sequencing was accomplished in Bioinformatics and Genomic Service (SGB) at UAB using a Genetic Analyzer 3130xl (Thermo Fisher Scientific). Sequence Scanner v1.0 and BioEdit software (Hall, 2001) were used to analyze sequences.

Analysis of nuclear DNA

Nuclear DNA was only analyzed for DNA samples extracted with the methods that revealed the best performance in mtDNA analysis. First, an 85-bp region encompassing the -13910C/T polymorphism strongly associated with Lactase Persistent in Europeans, located 13910 bp upstream of the lactase (LCT) gene in chromosome 2, was amplified with primers Forward 5'-AATGTACTAGTAGGCCTCTGCG-3' and Reverse 5'-TGCAACCTAAGGAGGAGAGT-3'(Burger et al., 2007). The PCR conditions were performed as described for mtDNA using an annealing temperature of 53ºC. The PCR product quantification was performed with Qubit™dsDNA BR Assay Kit.

Second, a set of 21 autosomal STRs, 2 Y-markers, and the Amelogenin gene were analyzed using the GlobalFiler® PCR Amplification Kit (Thermo Fisher Scientific) following

the manufacturer procedure. Detection of amplified products was performed using the 3130xl Genetic Analyzer (Thermo Fisher Scientific) in the Genomics Core Facility at the Universitat Pompeu Fabra (UPF), using Gene Mapper ID-X v4.2 Software (Thermo Fisher Scientific) with a threshold established by internal laboratory validation of 20 relative fluorescent units (RFUs).

Data analysis

Evaluated parameters

The success of the DNA recovery was evaluated through total DNA concentration and the concentration of amplified mtDNA (HVRI and HVRII) and nuclear (LCT) fragments. The success of mtDNA (HVRI and HVRII) and nuclear (LCT) amplification was evaluated through the presence or absence of the amplified fragment by High Resolution Melting (HRM).

The reproducibility of the mtDNA sequence results using different DNA extracts of the same individual obtained with different extraction methods was evaluated. The quality of the STR profile was assessed through parameters such as the average of peaks height (RFUs) and peak height ratio (PHR), which indicate the height balance between heterozygotes alleles, and the number of reportable alleles, which shows the percentage of alleles amplified in the genetic profile (homozygotes and heterozygotes) (Grisedale & van Daal, 2012; Harrel et al., 2018).

Statistical Analysis

A two-way ANOVA followed by an Honestly-Significant-Difference (HSD) Turkey test for multiple comparisons was used to compare the concentration of total DNA and the concentration of mtDNA PCR fragments (HVRI and HVII) between the types of skeletal remains and the extraction methods. The concentration of nuclear DNA fragment, peak heights (RFUs), peak height ratio (%), and percentage of reportable alleles between the best types of skeletal remains and extraction methods.

A Categorical Principal Components Analysis (CPCA) was employed to represent the relationships between some of the factors analyzed in this study except the quality parameters of STR profile and concentration of nuclear fragment (LCT). All the statistical analyses were performed considering a significant level of 5% by the IBM SPSS Statistics (IBM Corp, 2013).

3.1.2 Results

The results of the total DNA quantification indicated essential differences between methods and skeletal remains (Table 10). Regarding the type of skeletal remains, the highest amount of total DNA from different individuals and methods was recovered from petrous bone ($\bar{X}$=33.20ng/µl), followed by pulp cavity ($\bar{X}$=19.87ng/µl), upper limb bones ($\bar{X}$=19.27ng/µl), cementum of tooth ($\bar{X}$=16.22ng/µl) and rib ($\bar{X}$=13.72ng/µl). The amount of total DNA recovered from all individuals and type of sample presented significant differences between skeletal remains (Two-Way ANOVA; F=4.07, df=4, p=0.005).

Table 10 — Mean ± Standard Deviation of total DNA concentration (ng/µl) obtained from five types of skeletal remains and using four extraction methods.

Extraction methods[a]	Petrous	Pulp cavity	Cementum of tooth	Limb bones	Rib	Total
1	78.08±3.14	43.64±28.64	19.34±13.18	46.12±38.11	32.68±31.54	43.97±31.6
2	39.18±18.72	18.78±5.09	31.22±28.50	19.76±9.52	17.26±25.44	25.24±19.3
3	12.58±8.27	14.26±4.92	12,20±5.30	9.38±7.56	3.72±1.25	10.42±6.65
4	2.99±3.13	2.817±2.56	2.14±1.20	1.83±1.27	1.25±1.28	2.208±1.98
Total	33.20±31.36	19.87±20.45	16.22±18.22	19.27±25.15	13.72±22.61	

a) 1: Silica in-suspension, 2: Silica in HE-membrane columns, 3: Silica in XS plasma- membrane column, 4: Phenol-Chloroform

Concerning the extraction methods, the method 1: Silica in-suspension allowed the highest DNA retrieval ($\bar{X}$=43.97ng/µl), followed by the method 2: Silica in HE-membrane columns ($\bar{X}$=25.24ng/µl) and method 3: Silica in XS plasma- membrane column ($\bar{X}$=10.42ng/µl) whereas the method 4: Phenol-Chloroform recovered the lowest amount of DNA ($\bar{X}$=2.208ng/µl). The differences between methods are statistically significant (Two-Way ANOVA; F=30.15, df= 3, p<0.001).

Considering both the extraction methods and the kind of skeletal remains, the highest amount of total DNA retrieval was $\bar{x}$=78.08ng/µl from petrous bones followed by upper limb bones ($\bar{x}$=46.12ng/µl) and pulp cavity ($\bar{x}$=43.64ng/µl) using the method 1: Silica in-suspension. The DNA amount from petrous bones displays the best results with all methods. In addition, petrous bone DNA retrieved with method 1 showed the best results and presented significant differences with all trials except the pulp cavity and limb bones extracts using the same method (Silica in-suspension) and the petrous bone extracts with the method 2 (multiple comparisons; p-values between 0.051 and 0.25).

A representative example of the DNA retrieval profile from the petrous bone using the four extraction methods is shown in figure 19. Both short (35-200bp) and long (700-10330bp) fragments were recovered using the methods 1, 2, and 3 (Figure 19A, B, and C), while the method 4 recovered a few fragments and almost all correspond to long fragments ~ 2517 up 12423pb (Figure 19D). The results also showed that methods 1 and 2 recovered the highest amount of short fragments characteristic of highly degraded DNA typical of forensic and ancient samples. However, the profile of long fragments in methods 1 and 2, may represent DNA from bacteria, fungi, or human contaminants. Method 2 Silica in the HE- membrane column (Figure 19B) presents a low number of long fragments compared to Method 1 Silica in-suspension (Figure 19A).

Concerning the success of mtDNA PCR amplification, the highest number of positive amplifications was obtained by method 1 (19/25 samples amplified for HVRI and 24/25 samples amplified in HVRII). Method 4 showed the lowest results (12/25 samples amplified for HVRI and 9/25 samples amplified in HVRII) (Table SI: 1). Concerning the concentration of mtDNA PCR products (Table 11), method 1 ($\bar{x}$ =15.72ng/µl for HVRI) ($\bar{x}$=12.45ng/µl for HVRII) and method 2 ($\bar{x}$=14.32ng/µl for HVRI) ($\bar{x}$=11.67ng/µl for HVRII) allowed the highest amplification of mtDNA. In contrast, method 4 ($\bar{x}$=1.40ng/µl for HVRI) ($\bar{x}$=1.16ng/µl for HVRII) was associated with the poorest results. Regarding the type of skeletal remains, DNA extracts from petrous bone ($\bar{x}$=18.12ng/µl for HVRI) ($\bar{x}$=14.58ng/µl for HVRII) and pulp cavity ($\bar{x}$=12.80ng/µl for HVRI) ($\bar{x}$=13.88ng/

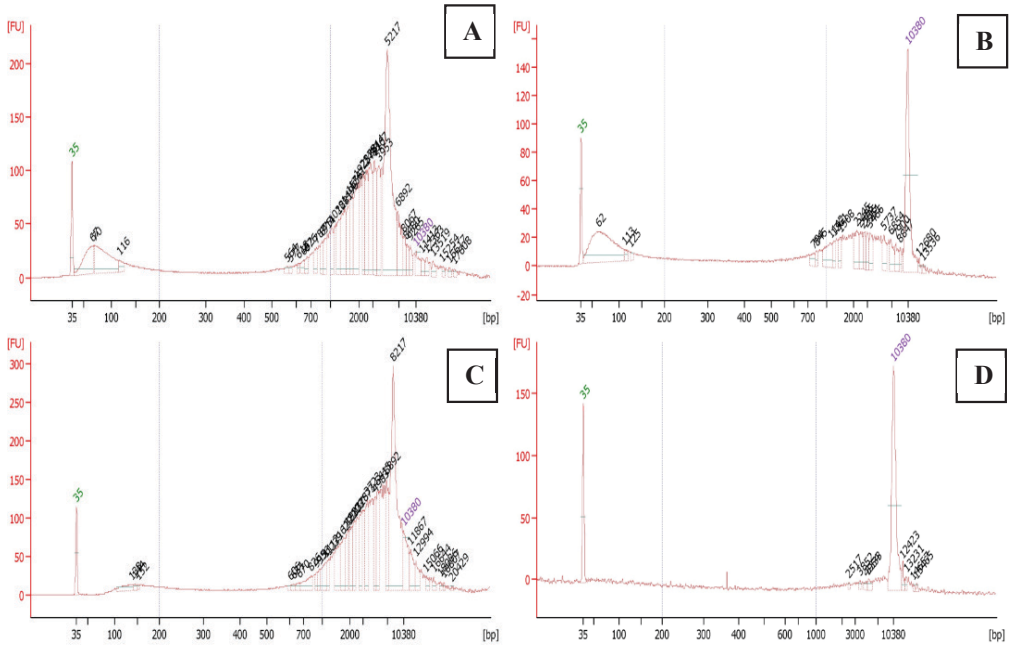


Figure 19 — Length distributions of DNA Quantification with Agilent High Sensitivity DNA kit® from the petrous DNA of the individual 2 retrieved using the four extraction methods. The size markers or ladder are sharp peaks at 35bp and 10380 bp. (A) Method 1: Silica in- suspension, (B) Method 2: Silica in HE-membrane column, (C) Method 3: Silica in XS plasma- membrane column and (D) Method 4: Phenol-Chloroform.

µl for HVRII) displayed the best results, while DNA extract from rib showed the worst ($\bar{x}$=2.29ng/µl for HVRI) ($\bar{x}$=2.27ng/µl for HVRII). Both differences between extraction methods (Two-Way ANOVA; F=14.02 for HVRI and F=18.97 for HVRII df=3 p=0.000) and types of skeletal remains (Two-Way ANOVA; F=9.17 for HVRI and F=19.29 for HVRII, df=4, p=0.000) were statistically significant.

Considering the interaction between extraction methods and kinds of skeletal remains, the highest concentration of mtDNA fragments were $\bar{x}$=33.08ng/µl for HVRI, $\bar{x}$=24.60ng/µl for HVRII and $\bar{x}$=28.88ng/µl for HVRI, $\bar{x}$=28.42ng/µl for HVRII from the petrous bone DNA using the methods 1 and 2 respectively, and $\bar{x}$=19.86ng/µl for HVRI, $\bar{x}$=23.76g/µl for HVRII and $\bar{x}$=19.19ng/µl for HVRI, $\bar{x}$=18.44g/µl for HVRII from the pulp cavity DNA with methods 1 and 2 respectively (Table 11). The rib extracts showed the lowest amount of amplified DNA in the HVRII fragment using methods 3 and 4 and did not

Table 11 — Mean ± Standard Deviation of mtDNA (HVI and HVII) PCR fragments concentration (ng/µl) obtained from five types of skeletal remains using four extraction methods.

Extraction Methods[a]	Petrous		Pulp cavity		Cementum of tooth		Limb bones		Rib		Total	Total
	HVRI[b]	HVRII[c]	HVRI	HVRII	HVRI	HVRII	HVRI	HVRII	HVRI	HVRII	HVRI	HVRII
1	33.08±6.78	24.60±8.52	19.86±11.60	23.76±14.61	14.82±16.50	3.48±4.61	4.37±4.37	5.71±5.00	6.48±9.16	4.66±5.97	15.72±14.27	12.45±12.55
2	28.88±4.40	28.42±7.54	19.19±13.50	18.44±12.26	8.48±11.62	3.87±3.96	12.36±16.23	3.40±4.53	2.70±6.03	4.25±5.01	14.32±13.83	11.67±12.30
3	8.60±12.30	3.61±5.81	10.38±11.51	10.93±12.28	5.07±8.26	2.70±3.69	7.36±8.69	0.51±1.01	NA	0.14±0.14	6.28±9.20	3.57±7.00
4	1.92±3.29	1.71±1.60	1.76 ±1.02	2.41±2.37	1.54±1.53	1.45±2.98	1.80± 1.85	0.1832±0.32	NA	0.031±0.069	1.40±1.86	1.16±1.94
Total	18.12±15.17	14.58±13.67	12.80±12.34	13.88±13.32	7.48±11.21	2.88±3.66	6.47±9.61	2.45±3.91	2.29±5.72	2.27±4.22		

a) 1: Silica in-suspension, 2: Silica in HE-membrane columns, 3: Silica in XS plasma-membrane column, 4: Phenol-Chloroform.
NA: No amplification.
b) mtDNA_HVRI PCR fragment concentration
c) mtDNA_HVRII PCR fragment concentration

show amplification in the HVRI fragment. The concentration of mtDNA amplified in HVRI coming from petrous bone, pulp cavity, and cementum extracted with method 1 did not present significant differences compared with the HVRI-mtDNA PCR concentration from petrous bones and pulp cavity extracted with method 2 (multiple comparisons; p-value between 0.17 and 1.00). While the concentration of mtDNA amplified in HVRII coming from petrous bone, pulp cavity extracted with method 1 did not present significant differences compared with the HVRII-mtDNA PCR concentration from petrous bones and pulp cavity extracted with method 2 and method 3 (multiple comparisons; p-value between 0.129 and 1.00).

Sequences of good quality were obtained from the amplification of extracts with methods 1 and 2. Also, the HVRI and HVRII mtDNA profiles obtained from different extraction methods and different bone pieces of the same individual were concordant.

The relation between extraction methods, total DNA concentration, presence/absence of mtDNA amplification, and the concentration of mtDNA PCR fragments (HVRI and HVRII), individuals, and type of skeletal remains was represented (Figure 20). It appears that the result is highly related to individuals since each individual is located in a different position in the plot. We can observe that the extraction methods 1 and 2 group together and methods 3 and 4. The presence of mtDNA amplification (HVRI and HVRII) and the highest concentration of total DNA and amplifiable mtDNA were related to the individual 2, methods 1 and 2, petrous and pulp cavity. Conversely, method 4, individual 1, and ribs showed the worst results related to the absence of mitochondrial amplification and the lowest concentration of total DNA and mtDNA amplifications.

To determine how the different factors influence the amplification of nuclear DNA, the best essays were selected to evaluate the success of amplifying the 85bp nuclear DNA fragment (LCT). In this case, the petrous bone and the pulp cavity extracted with methods 1 and 2 were chosen. All DNA extracts from petrous bone with methods 1 and 2 amplified the nuclear LCT fragment (5/5 samples), but only 3 out of 5 samples from the pulp cavity

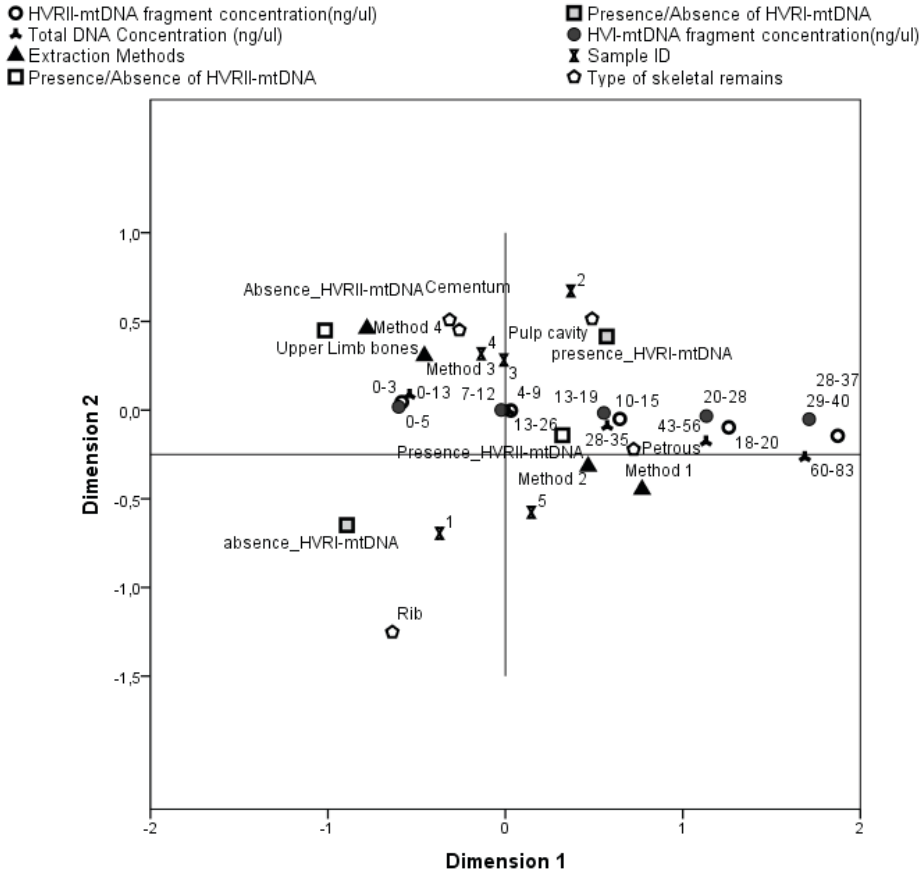


Figure 20 — Analysis of Categorical Principal Components for the total DNA and mtDNA fragment concentration (ng/µl) of HVRI and HVRII related to the presence/absence of mtDNA amplification, extraction methods, individuals (Sample ID) and type of skeletal remains.

DNA extracted with method 1 showed positive results. Moreover, DNA extracted using method 2 from the pulp cavity did not reveal amplification.

Concerning the concentration of the nuclear DNA fragment (Table 12), the highest values came from the petrous bone ($\bar{x}$=5.37ng/µl) and the pulp cavity ($\bar{x}$=4.85ng/µl) extracted with method 1. Conversely, the amplified fragment concentration was only $\bar{x}$=0.85ng/µl for the pulp cavity DNA extracted using method 2. The differences obtained were not statistically significant in any trials (paired t-test, p-value between 0.069 and 0.89), except

for the petrous bone DNA extracted with method 1 paired with the pulp cavity DNA extracted using method 2 (paired t-test, t= 4.12, p=0.015).

Regarding autosomal STRs, DNA from the petrous bone extracted with methods 1 and 2 yielded positive amplifications. The DNA from the pulp cavity only worked with method 1, whereas the extract from method 2 did not show amplification of any STR (Table 12, Figure SI: 1 and 2). Regarding peak height (measured by RFUs), the highest values were obtained from the petrous bone extracts ($\bar{x}$=498.60 RFUs and $\bar{x}$=228.50 RFUs, methods 1 and 2 respectively), and significant differences were found between them (paired t-test, t=4.02, p=0.016). The lowest height value came from the pulp cavity DNA extracted with method 1 ($\bar{x}$=57.82 RFUs). There were significant differences between the petrous bone DNA and pulp cavity DNA extracted with method 1 (paired t-test; t=4.11, p=0.015), while the pulp cavity DNA extracted with method 1 and the petrous bone DNA extracted with method 2 presented minimal significant differences (paired t-test; t=2.64, p=0.057).

Table 12 — Mean ± Standard Deviation of (LCT) nuclear DNA fragment concentration (ng/µl) and STRs parameters (peak height, percentage of peak height ratio, and reportable alleles) obtained of the DNA from petrous and pulp cavity extracted by methods 1 and 2.

Type of skeletal remain	Nuclear DNA concentration of LCT gene (ng/µl)		Peak Heights (RFUs)		% of Peak Height ratio (PHR)		% Reportable Alleles	
	M1[a]	M2	M1	M2	M1	M2	M1	M2
Petrous bone	5.37±2.37	4.10±3.27	498.60±252.74	228.50±118.88	53.08±18.94	62.89±12.01	79.09±21.22	71.81±18.01
Pulp cavity	4.85±8.80	0.85±0.47	57.82±84.22	NA	23.46±32.81	NA	24.54±37.97	NA

a) M1: Silica in-suspension extraction Method, M2: Silica in HE-membrane columns extraction Method.

Peak Height Ratio (PHR) showed the highest values from petrous bone extract using both methods 1 and 2 ($\bar{x}$=53.08% and $\bar{x}$=62.89% respectively); meanwhile, the pulp cavity DNA extracted with method 1 showed the lowest one ($\bar{x}$=23.46%). No significant differences between methods and sample type were observed (paired t-test; p-value between 0.09 and 0.47).

Finally, the DNA from petrous bones extracted with methods 1 and 2 showed the highest percentage of reportable alleles ($\bar{x}$=79.09% and $\bar{x}$=71.81%, respectively) (Table 12).

Significant differences between methods and sample type were observed (paired t-test; p-value between 0.035 and 0.049).

3.1.3 Discussion

Critical samples from forensic and ancient contexts contain small amounts of highly fragmented and damaged DNA. Therefore, the method used for DNA recovery is crucial for genetic characterization and identification. Moreover, the selection of a suitable DNA source is a crucial decision. In this study, we extracted DNA from five different types of skeletal remains using four extraction methods: three based on silica and the traditional organic method. One hundred DNA extracts were obtained, and their performance was assessed using different strategies based on PCR amplification.

First, it is essential to note that the analyses of five different regions of skeletal remains per individual were also used to determine whether the variation in the DNA recovery success was attributed to the individual (exposure conditions). Our results showed that individuals had preserved different DNA amounts. This fact is because the individuals do not come from the same sector of the necropolis; this means that although the site's environmental conditions may take as a whole, certain specific areas could be more rocky, sandy, or humid. Previous studies (Collins et al., 2002; Korlević et al., 2015b; Mundorff & Davoren, 2014) show that the tomb construction material or death cause of individuals could provide distinct environmental conditions (fungi, bacterial), so the performance of DNA recovery will be different in each individual.

Regarding DNA source, five types of skeletal remains were selected, from the most compact hydroxyapatite tissue (tooth and petrous bone), intermediate thickness (ulna and radius), to the thinnest tissue from small elements (phalanges and ribs). The DNA extract from petrous bone, followed by the pulp cavity and cementum of the tooth, showed the highest recovery values of DNA regardless of the extraction method used. These results support previous studies (Damgaard et al., 2015; Hansen et al., 2017b; Rothe & Nagy, 2016) reporting that the petrous bone has higher endogenous DNA content (5-8 fold

on average) than teeth. Bone is a connective tissue composed mainly of collagen and an inorganic mineral called hydroxyapatite (Collins et al., 2002). Studies have shown that the DNA has a strong affinity for hydroxyapatite, and its degradation is linked to the extent of crystallinity loss of this compound as well as the loss of collagen.

Therefore, the higher the bone density, the higher the amount of DNA adsorbed. The petrous bone, part of the temporal bone, is the toughest and most dense bone in the mammal's body, as well as where DNA can be more protected against moisture, sunlight, temperature, and microorganisms compared to a variety of interesting forensic materials such as other bones (ribs or limbs bones), hair, excrement or dry blood exposed to the same conditions (Adler et al., 2011a; Hansen et al., 2017b). Thus, promising results from petrous bone and not as good as other fewer dense bones have been expected. The negative results from ribs seem to be related to the thin cortical tissue, which is more taphonomically affected than other tissues.

Along with the petrous bone, the densest skeletal tissue in the body is that of teeth. Indeed, our second-best results are obtained with dental tissue. However, the amount of DNA recovered is not the same in the different regions of teeth (dentine, cementum, and pulp cavity) as other studies have also reported (Hansen et al., 2017b). Although the common forensic practice could be using the entire tooth to extract DNA, it is sometimes interesting to preserve it, whether it comes from archaeological cases or is very important or unique evidence.

Our study compared the amount of DNA recovered from the pulp cavity and cementum. The pulp cavity provided a higher DNA recovery than the cementum. There is no evidence of studies comparing DNA preservation between the pulp cavity and cementum. However, some findings show the mtDNA content of tooth cementum was five times higher than the commonly used dentine (Adler et al., 2011a; Damgaard et al., 2015). This may be due to dentine usually does not contain any nucleated cell bodies, just can contain some mitochondrial DNA-s that accumulate from the odontoblastic process (Perlich et al., 1981). In this sense, we compared cementum (high cell density) (Adler et al., 2011a) with the

pulp cavity because, as is reported (Rohland & Hofreiter, 2007a), it is a highly vascularized connective tissue containing numerous cell types and rich in DNA. The latter could explain better performance from the pulp cavity than cementum.

Degradation processes of any biological sample result in a low concentration of DNA and its fragmentation into short pieces, having a direct correlation between fragment length and abundance (Allentoft et al., 2015; Rohland et al., 2018). We evaluated the recovery profile to corroborate the rescue of highly degraded endogenous DNA from the samples. The outcomes showed that methods 1 and 2 presented the greatest recovery of short DNA. However, method 2 showed a lower concentration of long fragments than method 1. Our results agree with Allentoft et al. 2015 (Allentoft et al., 2015) who evaluated the DNA recovery profile using different binding buffer compositions with a silica in-suspension extraction method in combination with and without Min-Elute columns (Qiagen). DNA extracts that passed through a column showed less recovery of long fragments than those without passed through a column, probably because a purification step is performed by membrane columns allowing those little fragments to be retained in the filter and the longer fragments pass.

Regarding extraction methods, our results showed that methods based on silica were better than the organic extraction method. The Phenol-chloroform method produced the lowest concentration of DNA, an unquantifiable recovery of short fragments commonly related to endogenous critical DNA and the lowest concentrations of mitochondrial amplified fragments. Although there are very few studies comparing extraction methods based on silica and phenol-chloroform, our results are in agreement with these previous reports (Marshall et al., 2014b; Rohland & Hofreiter, 2007b). The success of the silica method may be related to the presence of a high concentration of chaotropic salts (binding buffer) and a pH:4 that produces a reversibly specific bind of DNA to silica, isolating it from other molecules. Moreover, using the silica in-membrane method, the ultrafiltration by columns removes contaminants or inhibitors improving the amplification of recovered DNA (Ferreira et al., 2013; Rohland & Hofreiter, 2007b). Contrariwise, the phenol-chloroform

method is so effective at removing proteins and lipids tending to be ineffective for the removal of hydrophilic compounds such as soil humic acid (PCR hydrophilic inhibitors) (Jakubowska et al., 2012).

Our results showed that the extraction with silica in-suspension determined the highest DNA yield compared with the other extraction protocols. Similar results were reported (Palomo-Díez et al., 2017), using two DNA extraction methodologies, silica in-suspension, and in columns. Their results showed a higher quantity of DNA in 68.4% of the cases using the silica in-suspension method compared to 21.05% by the silica in-column method. The reason may be related to the binding time. In the silica in-suspension method, binding of DNA is performed for 1 hour in-suspension with the silica, whereas in the silica in-columns method, the DNA-binding was performed for few seconds while it is being washed through the silica matrix. It is also known that extended binding time is critical to obtaining high DNA yields (Rohland et al., 2010). Other studies have also shown that the Silica-DNA adsorption is more efficient in the presence of buffer solutions with pH at or below the acid constant of the surface silanol groups, namely pH:4 (Rohland et al., 2010; Rothe & Nagy, 2016). Thus, the protocol of silica in-columns omits the pH adjustment after extraction buffer, silica and binding buffer are mixed. Instead of this, sodium acetate is added to the binding solution resulting in a pH between 5 and 6 that is above the ideal value. Moreover, the silica in-suspension method adds the phenol-red solution that allows establishing a redshift to yellow indicating that the solution was at pH:4 during the silica-DNA binding phase; if the solution was greenish-yellow, HCl can be added up to the suitable pH. While this exact pH adjustment could be beneficial for a better DNA yield as is in this study, an excess of HCl could cause a decrease in DNA yields and the DNA may be completely degraded in such acidic conditions (Rohland et al., 2010). In this sense, a pH regulation could be considered to improve the extraction methods based on silica in-columns. Also, the use of columns after the binding step allows reducing the workload substantially and decreases the risks of cross-contamination compared to resuspension by extensive pipetting with higher volumes, involving many time-consuming. In addition, the handling of high volumes allows a reduction in the number of samples conveniently

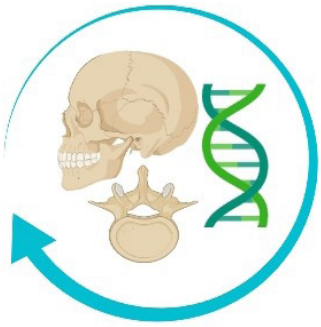
processed in parallel. As other studies have shown, the silica-column method shows a measurable amount of extracted DNA (Hasan et al., 2014; Mameli et al., 2013; Marshall et al., 2014b; Rothe & Nagy, 2016) and could be also useful for dry and highly-degraded samples. Otherwise, fresh bone tissue can block the membrane with a high amount of proteins such as collagen; these proteins simply cannot pass through the membrane being later eluted with DNA. The final result will be a low DNA yield (Rothe & Nagy, 2016).

To evaluate the quality of the DNA extracts, we amplified both mitochondrial and nuclear DNA. MtDNA consensus haplotypes were obtained, and no DNA contamination was detected; so, both methods 1 and 2 appear to be equally valid. Similar results were obtained by Palomo-Díez et al. 2017 (Palomo-Díez et al., 2017). They used the silica in-suspension and silica in-columns extraction methods and found mtDNA consensus haplotypes, but the largest amount of DNA was obtained using the silica in-suspension method. Regarding the ability of nuclear DNA recovery, the highest amount was obtained from the petrous bone and pulp cavity using method 1. The STRs profiles also showed the highest values of RFUs from petrous DNA extracted with method 1 and reportable alleles >70% with both methods. Meanwhile, the best balance of heterozygous alleles (PHR) was from petrous bone DNA extracted with method 2. Similar values were obtained by Marshall et al. 2014 (Marshall et al., 2014b) using an extraction method based on silica in-columns (Hi-Flow®). Following other studies (Caputo et al., 2013b; Marshall et al., 2014b; Rothe & Nagy, 2016; Tvedebrink et al., 2012), we can confirm by this parameter the presence of fewer inhibitory compounds in method 2 than in method 1.

3.2. Chapter 2:

Newborn Skeletal Remains: an aid or a constraint to paleogenomics?

Corpus of the manuscript prepared for submission in American Journal of Human Biology



"DNA is like a computer program but far, far more advanced than any software ever created"

-Bill Gates, The Road Ahead

3.2.1 Material and methods

Intra-individual study

Sample preparation

Nine well-preserved newborn infants were selected from the collection of Iberian Individuals (Iron Age) studied at the Universitat Autònoma de Barcelona (Spain). Eight newborns were buried under the pavement of the houses from the Camp de Les Lloses Iberian site (Tona, Barcelona), and one individual in the Can Mateu site (Cabrera de Mar, Barcelona). For each individual, two skeletal elements were selected for DNA extraction: Petrous bone (inner cochlear) and cervical or lumbar vertebral posterior arch (Table SI 3). Each skeletal element was cleaned with a wipe soaked in 0.01%v/v of bleach on the outer layer followed by exposure to UV light for 10 seconds to any residual surface contamination from modern DNA. The sampling of the petrous involved the sectioning of the petrous pyramid in having access to dense inner cochlea using a diamond-coated rotary cutting disc. The cochlea region was isolated and then crushed with a mortar, and the vertebrae were directly crushed. Each sample generated 100 to 200 mg of bone powder from both kinds of skeletal remains.

DNA Extractions, library preparation, and sequencing

All the laboratory work was performed in the dedicated clean laboratory facilities of Centre for GeoGenetics, Natural History Museum, University of Copenhagen, following strict aDNA standards (Gilbert et al., 2005; Willerslev & Cooper, 2005). Negative controls were included during the DNA extraction and library preparation steps. A 10 min pre-digestion step was applied to all specimens before beginning the DNA extraction process using the silica-in suspension method (Vinueza-Espinosa et al., 2020).

80µl of DNA was eluted. Genomic DNA was quantified using Qubit™ dsDNA BR Assay Kit (Thermo Fisher Scientific), adding 1µl of template DNA following the manufacturer's protocol. 20 µl of each DNA extract was built into a single-strand library following the protocol of Santa Cruz Reaction (SCR) (Kapp et al., 2021). 50 µl of the library was recovered and purified by SPRI beads for each sample. The optimal number of PCR cycles was determined by qPCR (MxPro 3000, Agilent Technologies). The qPCR reactions consisted of 1 µl library template with 19 µl Master Mix containing 12.3 µl H20, 2µl Taq Gold buffer, 2µl MgCl2, 0.8 µl BSA, 0.8 µl SYBR green/ROXY, 0.4 µl Forward (Meyer & Kircher, 2010) Illumina-compatible IS4 primer, 0.4 µl Reverse primer P7 index primer #50 (10µM), 0.16 µl dNTPs (25 mM), and 0.16 µl Taq Gold polymerase (Invitrogen). The indexing of libraries was performed using the KAPA library quantification kit (Roche). Finally, the amplified libraries were purified using SPRI beads and quantified on a 2200 TapeStation (Agilent Technologies) using High Sensitivity tapes. The amplified and indexed libraries were then pooled in equimolar amounts and sequenced on 1/8 of a lane of Illumina HiSeq 4000 sequencer.

NGS- Data Analysis

An in-house pipeline was used to analyze all NGS data. The FastQ files generated were checked by FastQC v0.11.9 (Andrews, 2010). Adapters, bad-quality base pairs ($Q < 25$), and reads shorter than 30 base pairs from 3' end were trimmed by fastp v0.20.1 (Chen et al., 2018). A second FastQC v0.11.9 was used to check the quality of the new files. Reads were mapping by BWA v0.7.17 (Li & Durbin, 2009) with GRCh38.p13 human reference genome (hg19, GenBank accession GCA_000001405.28) for nuclear DNA (nDNA) and with revised Cambridge Reference Sequence (rCRS, GenBank accession NC_012920.1) for mitochondrial DNA (mtDNA). For nuclear and mitochondrial analysis, SAMtools v1.7 (Danecek et al., 2021) was used to generate a sorted BAM file with all the reads with a mapping quality of over 30 and indexing BAM files. MarkDuplicates removed duplicates from Picard (v2.25.0). BAM quality was checked by QualiMap v2.2.2a (Okonechnikov et al., 2016). For nDNA, post-mortem damage analysis was performed by MapDamage2 v2.2.0 (Jónsson et al., 2013).

Statistical Analysis

Comparisons between petrous bone and vertebra library NGS shotgun performance were accomplished using a paired sample t-test. The number of sequences, the percentage of endogenous DNA, the mean read length, the % of clonality reads, the mtDNA and nuclear genome coverage, the number of unique reads mapping per million of raw reads (Parker et al., 2020) (see equation a), the number of reads mapped in X and Y Chromosomes and the estimation of total genomic coverage per library (Parker et al., 2020) (see equation b), were evaluated. The endogenous DNA content (%) was calculated with and without PCR duplicates.

a) The number of unique reads mapping to the human genome per million reads sequencing effort was calculated as (Parker et al.,2020)

$$\frac{\#\ of\ reads\ mapping\ to\ reference\ after\ duplicate\ removal\ and\ quality\ filtering}{\#\ of\ reads\ generated\ prior\ to\ merging\ or\ filtering} \times 1{,}000{,}000$$

b) Total genomic coverage within a library as estimated by calculating (Parker et al.,2020):

$$\frac{\#of\ DNA\ molecules\ in\ library * protportion\ of\ human\ DNA\ recovered * Avg.length\ of\ mapping\ reads}{Length\ of\ reference\ genome}$$

Inter-individual study by bone and age category

Data from Patterson and collaborators (Patterson et al., 2022) namely, information related to each library performance (available in supplementary material, table SI 4) was used. Information about the skeletal elements used for DNA extraction and age category of each individual was also compiled from the description of each archaeological site. Recompiled information used in the present work is avaliable in Table SI 4. According to the description of the original age, the individuals were classified into 4 categories: perinatal (> 6 months), infant (<10 years); juvenile (10-17 years), and adult (> 18 years). If the authors did not specify the age, the individual was considered undetermined (n/a). The skeletal remains were also classified as *petrous bone* and *other skeletal elements* (for postcranial bones and teeth). A total of 825 individuals were considered (one individual was excluded since the library

was performed using hair): 31 perinatal, 48 infantile, 65 juveniles, 313 adults, and 368 undetermined. A total of 1019 libraries used in our comparative study

Statistical Analysis

The mtDNA coverage, 1240K coverage at targeted positions, 1240K unique SNPs hit, and 1240K sequences hitting X Chr were used as indicators of NGS sequence library quality. A Two-way ANOVA followed by post-hoc for multiple comparisons were used to compare each parameter between the type of skeletal remain and age category.

3.2.2 Results

Data from the intra-individual study

Table 13 shows the parameters calculated to evaluate the library complexity using shotgun sequencing from 18 single-stranded DNA libraries of two skeletal elements —petrous bone and vertebra— from nine ancient newborn individuals.

The results indicated that petrous bone is superior to the vertebra in the number of sequences generated and the fraction of endogenous human DNA with PCR duplicates, however, differences are not significant. When the duplicate reads were not considered, the human DNA content decreased considerably for the vertebra; meanwhile, petrous bone was comparably higher (Figure 21a), being significant the differences between petrous bone and vertebrae ($p=0.012$, $gl=8$, $t\text{-test}=3.208$). Moreover, the median fragment length was slightly higher in petrous bone. As to the percentage of clonality, the vertebra shows a higher rate of duplicate reads compared to petrous bone (Figure 21a) ($p<0.001$, $gl=8$, $t\text{-test}=-6.191$).

Figure 21b shows the number of unique human reads mapping per million for petrous bone and vertebra. The results showed that petrous bone recovered more unique human reads than vertebra ($p=0.013$, $gl=8$, $t\text{-test}=3.190$). Petrous bones provided slightly higher genomic coverage than vertebra from one individual library, but the difference is not significant (Figure 21c).

Table 13 — Averages and standard deviations (sd) of sequencing data obtained from 18 single-stranded DNA libraries of petrous bone and vertebrae from nine infant individuals.

Skeletal remains		Petrous bone	Vertebrae	Mean of paired differences
Number total of sequences		19.2million sd=7E+06	10 million sd=5E+06	10 million sd=12E+06
Fraction of human (%) endogenous DNA	With PCR Duplicates	14.09 std=11.09	10.73 std=8.86	3.36 std=9.78
	*Without PCR Duplicates	11.22 sd=9.26	0.74 sd=0.76	10.49 sd=9.81
Fragment Length (bp)		54.79 sd=8.21	53.06 sd=5.38	1.73 sd=12.27
*%Clonality		18.21 sd=10.48	78.27 sd=35.57	-60.06 sd=29.14
*Number of unique human reads mapping per million		1million sd=8.3E+05	61746 sd=7.3E+04	9.4E+05 sd=8.9E+05
Estimation of genomic coverage		25136x sd=2.5E+04	13487x sd=1.1E+04	11648.89x sd=2.3E+04
%Damage Patterns	1st base	20.26 sd=7.17	17 sd=4.15	3.26 sd=7.64
	*2nd base	14.71 sd=3.2	11.94 sd=1.9	2.77 sd=3.04
*Nuclear genome coverage		0.040x sd=0.030	0.002x sd=0.003	0.039x sd=0.032
*Mitochondrial DNA coverage		8.54x sd=6.34	0.66x sd=1.37	7.88 sd=7.28
*Chromosome X reads		88595.44 sd=59785	2574.33 sd=4065.32	86021.11 sd=63029.25
Chromosome Y reads		984.77 sd=1882.21	146 sd=404.17	21009.75 sd=838.77

**p<0.05 significant differences between petrous and vertebra paired student t-test*

The damage patterns caused by nucleotide misincorporation (C to T, complementary strand, A to G) at the ends of the reads were also demonstrated in the sequences generated. The frequency of C >T was evaluated as a damage signal. The result showed that petrous bone had a higher frequency than the vertebra when the 1st base and the 2nd base were analyzed, however, significant differences were only detected for the 2nd base ($p=0.026$, $gl=8$, $t\text{-test}=2.74$). The nuclear genome coverage was higher from petrous bone than from vertebra ($t\text{-test}=3.61$, $gl=8$, $p=0.007$); moreover, the reads mapped in chromosomes X and Y were higher from petrous bone than vertebrae ($p=0.003$, $gl=8$, $t\text{-test}=4.10$). As to the mitochondrial DNA coverage, the results had the same trend as the nuclear; the petrous bone was superior to a vertebra ($t\text{-test}=3.25$, $gl=8$, $p=0.012$).

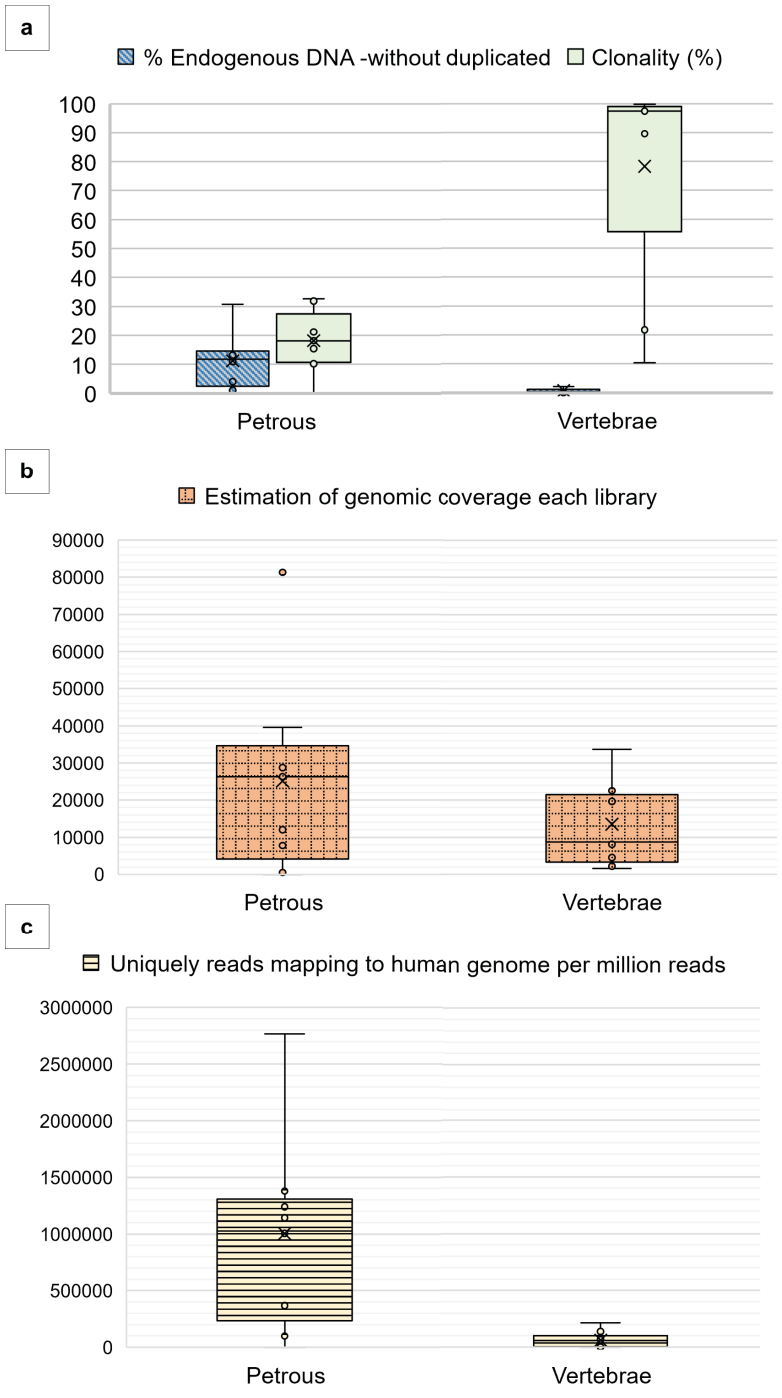


Figure 21 — Human DNA content in all 18 single-strand libraries from the petrous and vertebrae a) Fraction of Human Endogenous DNA content without duplicates and Clonality (%) b) Unique reads mapping per million raw reads c) Estimation of genomic coverage each library.

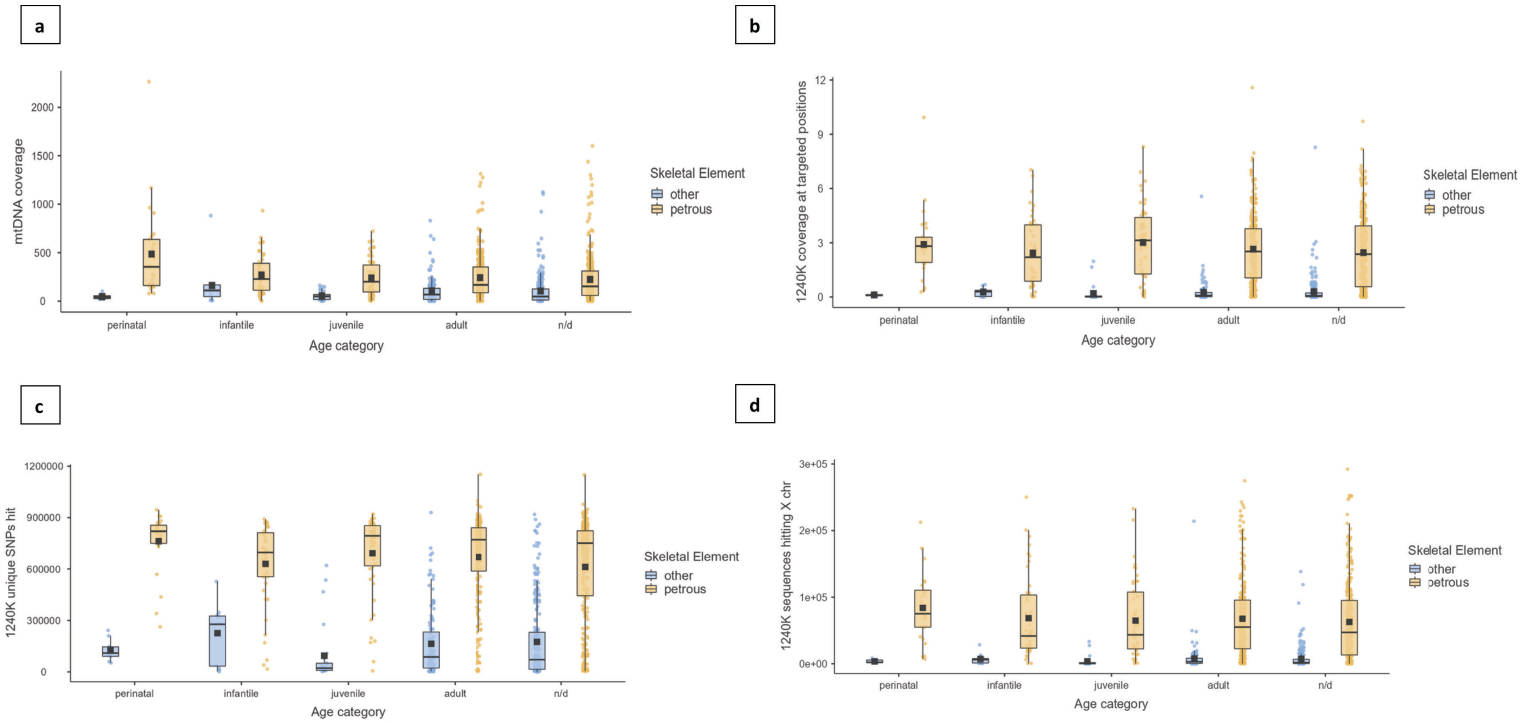


Figure 22 — Boxplot by age categories and type of skeletal remain (petrous bone and other skeletal elements) for the four indicators of library performance. a) mtDNA coverage b) 1240k coverage at targeted positions c) 1240K unique SNPs d) 1240k sequences hitting X chr.

Data from the age category by Patterson's study

Table SI 4 shows the means and standard deviations for each indicator of NGS sequence library quality evaluated in 1019 libraries from petrous and other skeletal elements of the 825 individuals previously analyzed by Patterson et al. (2022). The distribution of each parameter, considering bone type and age categories, was represented in figure 22.

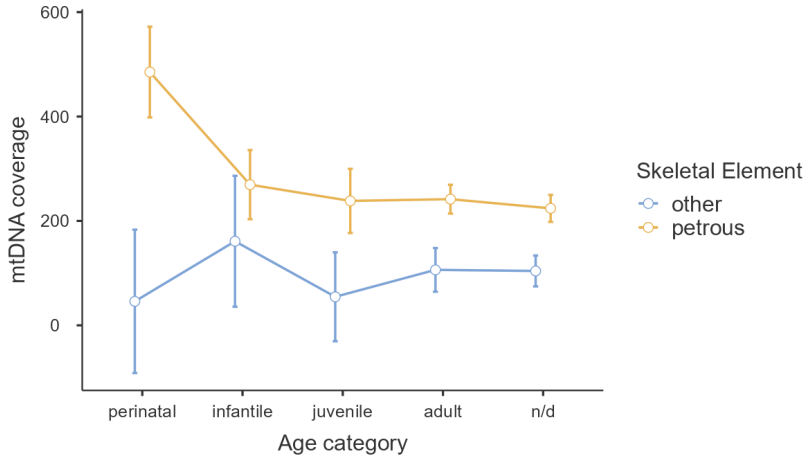


Figure 23 — Estimated marginal mean according to skeletal element and age category for mtDNA coverage.

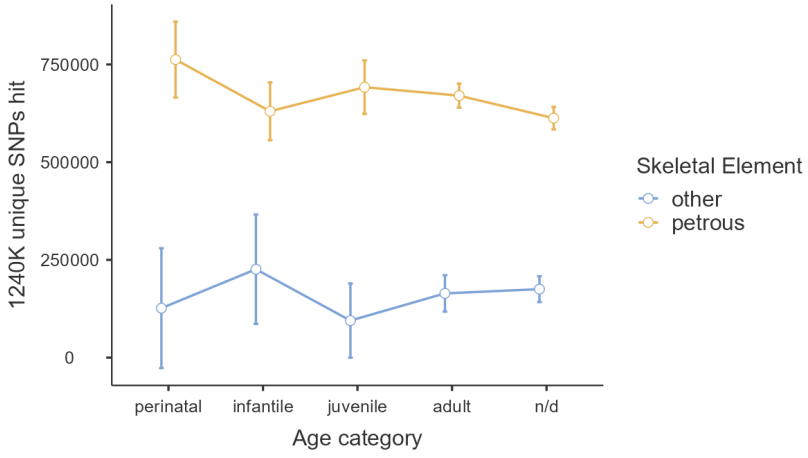


Figure 24 — Estimated marginal mean according to skeletal element and age category for 1240K unique SNPs hit

For all the parameters analyzed petrous bone showed best results than other bones (Two-way ANOVA, $p<0.001$), and all age categories appear to present similar results (Two-way ANOVA, p-value range from 0.09 to 0.843) (Figure 22). However, significant interaction is detected in mtDNA coverage (Two-way ANOVA, $p=0.005$) (Figure 23) and 1240K unique SNPs hit (Two-way ANOVA, $p=0.018$) (Figure 24). Both interactions are related to the best performance of petrous bone from perinatal individuals compared to the petrous bone from other age categories, particularly for mtDNA coverage.

3.2.3 Discussion

Evaluating DNA recovery from newborn or infant skeletal remains is necessary to know the limitations in paleogenomic and forensic analysis. Based on previous achievements in DNA recovery from adult individuals, the petrous pyramid of the temporal bone is currently the most sought-after skeletal element for aDNA analyses (Gaudio et al., 2019; Hansen et al., 2017a; Nieves-Colón et al., 2018; Pilli et al., 2018).

In the present study, we used petrous bone and vertebra from the same individuals to systematically compare skeletal elements, avoiding interindividual variation. Long bones were not selected to be analyzed to avoid the destruction of valuable samples for morphological characterization. In addition, long bones are small and have a very thin cortical, and other samples such as teeth, often used in adults, are not developed in perinatal individuals. Our results confirm the value of the petrous pyramid in recovering ancient human DNA. We obtained well-preserved endogenous DNA from this cranial vault fragment (cochlear region) in line with previous studies where petrous bones from infant individuals were also used (Patterson et al., 2022; Teschler-Nicola et al., 2020). A systematic study using eleven different skeletal remains from two juvenile and nine adult individuals showed the same tendency as the results obtained in our study where petrous bone showed to be superior to the vertebral body. Even the vertebral body was the third better skeletal element after the pulp chamber (Parker et al., 2020).

We also find that single-stranded aDNA libraries constructed from DNA extracted from the cochlear region of the petrous bone are higher in complexity in terms of the estimated

genomic coverage within each library, the number of unique human reads mapping per million reads, fragment length, and the number of reads mapping to the nuclear genome and mtDNA compared to the other sampling location (vertebra). Our findings had concordance with some comparative studies that evaluate the DNA field or genetic information obtained from petrous bone by NGS (Alberti et al., 2018; Hansen et al., 2017a; Parker et al., 2020; Pinhasi et al., 2015).

The cochlea is enveloped by the densest bone of the otic capsule, and this protection also reduces its postmortem degradation process compared to other skeletal elements (Pinhasi et al., 2015). This feature allows for better preservation and reduced exposure to contamination compared to more superficial and less dense elements. Moreover, human DNA fragments recovered from petrous bone showed a much higher frequency of cytosine deamination than vertebra, which helps to support their authenticity as ancient, while the comparatively lower deamination signal showed from vertebra may result from modern DNA contamination.

Overall, a high fragment length in conjunction with low deamination frequencies may indicate contamination with modern human DNA, which was not shown in our study as vertebra and petrous bone present small fragment length characteristic of ancient samples (Adler et al., 2011a). In terms of nuclear information recovered, petrous bone harbors far much nuclear material than vertebra and it was also showed a substantial differential in nuclear to mitochondrial mapping reads; this is expected since the number of copies of mtDNA per cell is much higher.

The comparative study performed with the data published by Patterson et al (2022) showed encouraging results when perinatal individuals are employed. The recovery of mtDNA molecules from petrous bone was much greater in perinatal than in infantile, juvenile, or adult individuals; equal results were obtained in perinatal individuals compared to the other individuals when other skeletal elements were used. A similar trend was shown in all the parameters evaluated. However, petrous bone remains as the better biological material where genetic information is better preserved. These results, demonstrate that perinatal petrous bone, can have equal or even more capability to obtain genetic information than adult skeletons.

3.2. Chapter 3:

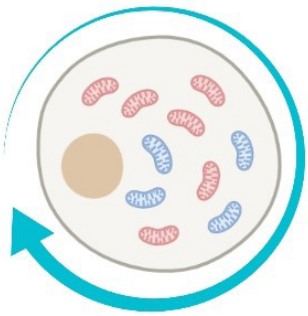
MtDNA-CR Typing from Highly Degraded Skeletal Remains by Single-Multiplex Massively Parallel Sequencing

Corpus of the manuscript submitted to Electrophoresis

Diana C. Vinueza-Espinosa[1], Daniel R. Cuesta-Aguirre[1], Assumpció Malgosa[1], Cristina Santos[1]

1Research Group in Biological Anthropology, Department of Animal Biology, Vegetal Biology and Ecology, Universitat Autònoma de Barcelona, Spain.

*Corresponding author: cristina.santos@uab.cat



"DNA is the messenger, which illuminates (our connection to the past), handed down from generation to generation, carried, literally, in the bodies of (our) ancestors. Each message a journey through time and space…"

-Byran Sykes, The Seven Daughters of Eve.

3.3.1 Material and methods

Sample preparation, DNA extraction, and quantification

Samples of bones and teeth from a total of 41 individuals were selected. Individuals were from different archaeological contexts from 5500BP-16thcentury (Ancient Samples) until the 19-20th century (Contemporary Samples) (Table SI 5 and Figure 25). The skeletal remains were cleaned, removing the external surface with a sterile tungsten tip placed into a micro drill up to 5000 rpm. Then roughly 150 mg of sample was cut by sterile tungsten disk and fragmented by forceps to get smaller pieces to make more accessible the sample digestion in the DNA extraction procedure.

DNA was extracted using a silica-based method on a HE-membrane column (Vinueza-Espinosa et al., 2020) in the ancient DNA Laboratory of the *Universitat Autònoma de Barcelona* (UAB). DNA was quantified in the Genomics Core Facility at the Universitat Pompeu Fabra using Quantifiler™ Trio kit (Thermo Fisher Scientific), adding 1µl template DNA by triplicating and following the manufacturer's protocol. The thermocycler parameters were specified by the manufacturer's recommendation in a 7500 Real-Time PCR Instrument.

Amplification and libraries preparation using PowerSeq™ CRM kit (10-plex), and sequencing

In the first experimental phase (Figure 25), 23 samples were used to test three NGS-protocols based on modifications of PCR conditions: i) The **M1** protocol is performed following the manufacturer's recommendation: 96ºC for 10 minutes, 30 cycles of 96ºC for 5 seconds, 60ºC for 35 seconds, 72ºC for 5 seconds, and 60ºC for 2 minutes; ii) The **M2** protocol is based on the same conditions as the M1 protocol, but increasing the number

of cycles to 35 times; iii) The **M3** protocol includes 35 cycles and longer times in some steps: 96ºC for **15 min,** 35 cycles of 96ºC for **15 seconds**, 60ºC for 35 seconds, 72ºC for **30 seconds**, and 60ºC for 2 min.

10 additional samples were tested using the M2 and the M3 protocols in the second phase. In total, 33 samples were tested using the M2 and M3 protocols. 8 additional samples were amplified using only the M3 protocol in the final phase. Therefore, 41 samples were tested using the M3 protocol (Figure 25).

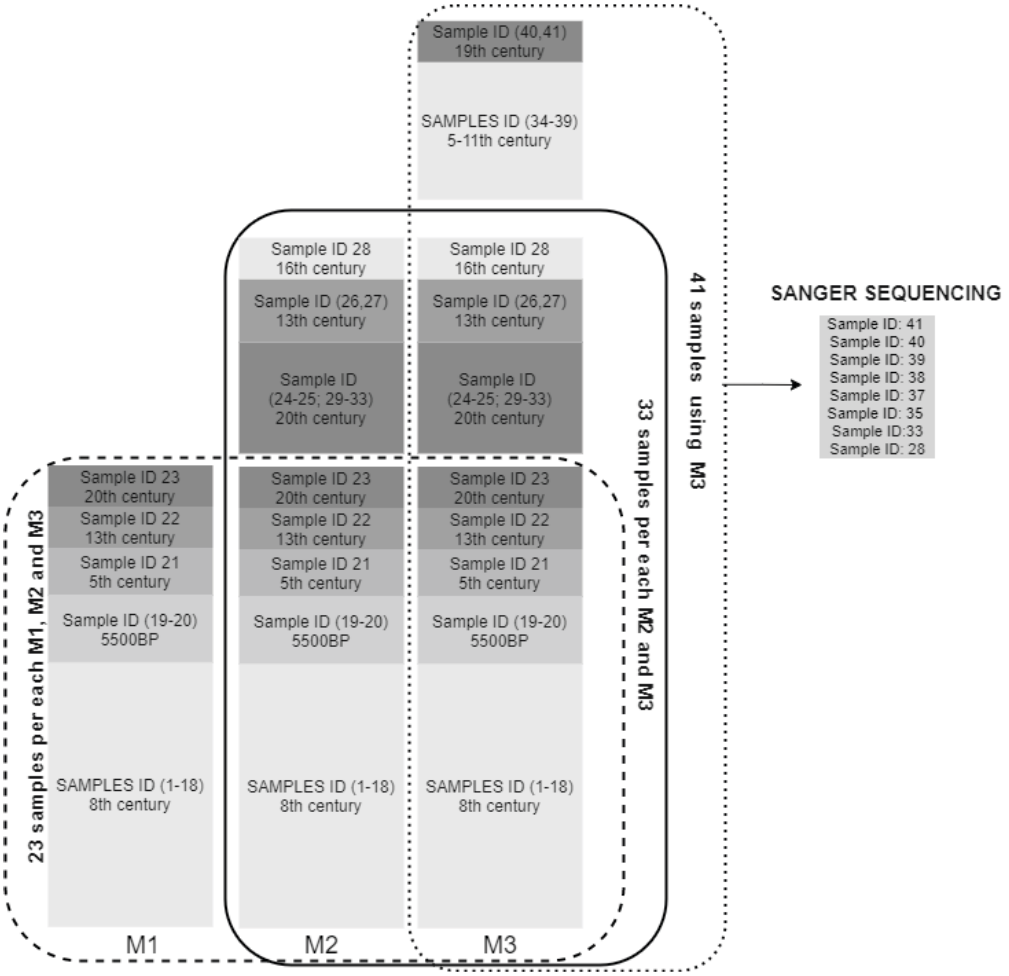


Figure 25 — Details of the samples from different chronologies used to evaluate the recovery of mtDNA-CR typing in each NGS-protocols M1, M2, and M3.

Amplification reactions-libraries were purified using the Qiagen® GeneRead™ size selection kit (column-based protocol) following the manufacturer's protocol (*GeneRead Size Selection of DNA Libraries Prepared)*. Library Quantifications were performed with Qubit™ dsDNA BR Assay Kit. To know the distributions of NGS-libraries by the length of amplicons recovered, the Bioanalyzer High Sensitivity DNA Kit ® (Agilent Technologies) was used. The library concentrations were normalized to 1nM. Three sequencing runs were performed in the Illumina MiSeq instrument using standard flow cells MiSeq® Reagent Kit v2 Nano 2×150bp.

Sanger Sequencing

To validate the variants or polymorphisms detected by the NGS technique, 8 samples sequenced using the M3 method were selected to be sequenced by the Sanger method (Figure 25). Two fragments of mtDNA were amplified: A 203bp fragment from the HVS-I region located between position 16030 and 16232, using primers Forward 5'-CATGGGGAAGCAGATTTGGG-3' and Reverse 5'-GATAGTTGAGGGTTGATTGCTG-3' (Simón et al., 2011); and a 175bp fragment from the HVS-II between position 008 and 281, using primers Forward 5'-GGTCTATCACCCTATTAACC-3' and Reverse 5'-CTTTCCACACAGACATCAT -3'(Simón et al., 2011). The PCR reactions were carried out using a Rotor-Gene®-Q in a 10µl reaction containing 2X Type-it HRM PCR Kit (Qiagen, Germany), 10pmol/µl primers, and 1-5µl of DNA template. The conditions were: hold at 95ºC for 10 min, denaturation at 95ºC for 10 seconds, annealing at 58ºC for HVSI and HVSII for 45 seconds, and extension at 72ºC for 10 seconds with 55 cycles.

The PCR products were purified using MSB® Spin PCRapace (Invitek, Germany) following the manufacturer's protocol. The sequencing reactions were prepared using BigDye® Terminator Sequencing Kit v1.1 in a 10µl reaction containing: 0.5µl Big Dye, 1µl Buffer, 1µl primer 5pmol/µl, 7µl H2O, and 0.5µl DNA of purified PCR product. Finally, the sequence reactions were purified with BigDye ® X-terminator purification Kit. The Sequencing was accomplished at the Genomic and Bioinformatics Service (SGB) of the

UAB using a Genetic Analyzer 3130xl (Thermo Fisher Scientific). The Sequence Scanner v1.0 and BioEdit software (Hall, 2001) were used to analyze sequences.

NGS-Data analysis

PowerSeq data were analyzed using the GeneMarker HTS version 2.5.0 software package and an in-house pipeline developed by our group (available online at https://github.com/DanielRCA/MTDNA-CR).

The parameters settings used in GeneMarker HTS software were a minimum base call quality score of Q<30 (bases less than 30 are trimmed from 3' end of the reads), a minimum total read depth of 10 reads, a minimum variant allele depth of 10, minimum of 30% for the minor variant frequency, and the allele score difference ≤10 (it determines if a minor allele is a true allele). Furthermore, the SNP balance ratio ≤2.5 and the indel balance ratio ≤5.0 (these values use the percentage of forward and reverse reads that include the variant) were established. Amplicon's settings were also included. The data ranges for each amplicon set were analyzed: Amplicon 1 (16013-16126), Amplicon 2 (16116-16225), Amplicon 3 (16223-16408), Amplicon 4 (16387-16486), Amplicon 5 (16474-30), Amplicon 6 (16555-152), Amplicon 7 (136-257), Amplicon 8 (246-364), Amplicon 9 (342-436) and Amplicon 10 (429-592). These ranges span sequence information (95-185pb), namely, without the primer binding sites trimming around 20bp to avoid the analyzed data are producing false heteroplasmy (Brandhagen et al., 2020).

The in-house pipeline consisted of quality-checked FastQ by FastQC v0.11.9 (Andrews, 2010). Adapters and primers were eliminated with two consecutive fastp v0.20.1 analyses. Reads were aligned to the revised Cambridge Reference Sequence (rCRS, GenBank accession NC_012920.1) using BWA v0.7.17 (Li & Durbin, 2009). The sequence was linearized from base 15901 to base 700 to avoid the problems generated by the circularity of the mtDNA and by the reads aligned in the replication origin. SAMtools v1.13 was used for generating mapped, sorted with quality checked Q>30, and indexing of BAM files. BAM-Quality was checked by QualiMap v.2.2.2a; then, the variants were called using Freebayes v0.9.21. The minimum mean depth and variant allele depth were 10 reads, and the minor

variant frequency was 30%. Indels, multi-nucleotide polymorphisms, and complex events (bases in positions 303 and 315) were ignored. PMDtools v0.60 (Carpenter et al., 2013) was used to extract all reads with molecular damage and to confirm the recovery of degraded DNA. The variants were visualized and verified using the Integrative Genomics Viewer (IGV) tool v.2.9.4 (Robinson et al., 2011). Also, Haplogrep 2 v2.4.0 was used to predict the mtDNA haplogroup (Weissensteiner et al., 2016).

Evaluation of parameters

The library quantification (nM) was evaluated for all the libraries. The libraries with a quantification higher than 0.005 ng/µl were sequenced. Those with a quantification less than 0.005 ng/µl were not sequenced and classified as no result (NR) in subsequent analysis. For each sequenced library, the total number of amplicons with '*reliable*' (>10 reads) and *'unreliable'* reads (0-9 reads) and the percentage of amplicons recovered per type of amplicon (See equation c) were evaluated. For NR libraries, it was considered that there are 0 reads per amplicon. For samples analyzed with protocols M2 and M3, the percentage of amplicons recovered was further analyzed, grouping the samples into ancient and contemporary.

There are 10 types of amplicons by sample, so, in 23 samples, 23 amplicons per type of amplicon should be expected.

Equation (c) $$\%\text{amplicons recovered per type of amplicon} = \frac{b}{23} * 100$$

Being b= number of amplicons with '*reliable reads*' in the amplicon (#1, #2, #3, #4, #5, #6, #7, #8, #9 or #10).

NGS sequence data were analyzed using two tools: GeneMarker HTS software and an in-house pipeline. The variant call was performed considering positions with 10 or more reads. Equality and discrepancies of the variants or polymorphisms detected were evaluated between both data analysis tools. The variants detected in only one software were classified as non-called (NC). Also, libraries with low coverage reads (LCR) < 10 reads in some positions were evaluated. Therefore, the percentage of libraries with NR or LCR, the total number of variants or polymorphisms generated, the equality percentage of variants

reported, the number of NC, and the percentage of reads with molecular damage, were also calculated for all the libraries sequenced using the in-house pipeline. Moreover, the base mixture was detected when the minor base exceeded a 30%

To evaluate the concordance of the NGS-technique and Sanger technology, the variant call obtained for 8 libraries sequenced using M3 protocol and Sanger technology were compared

Statistical analysis

A repeated-measures ANOVA followed by a post hoc pairwise comparison with Bonferroni correction was used to compare the libraries' quantification using the M1, M2, and M3 protocol. A Q-Cochran test followed by the McNemar test for two-two comparisons was performed to compare the total number of amplicons with reliable and unreliable reads in each protocol (M1, M2, M3). A fisher-test was performed to compare the percentage of amplicons recovered clustered by different chronologies using the M2 and M3 protocol.

3.3.2 Results

Library quantification and the total number of amplicons with reliable and unreliable reads

Table 14 shows the average and standard deviation of the library quantification (nM) and the number of amplicons with reliable and unreliable reads for the three NGS-protocols tested M1, M2, and M3 (23 samples: 69 libraries). Regarding library quantification, the lowest

Table 14 — Average ± standard deviation of library quantification (nM) and a total number of amplicons with reliable and unreliable reads from the three NGS-protocols.

NGS-Protocol	Library Quantification (nM)	Total Number of Amplicons*	
		Reliable Reads	Unreliable Reads
M1	0.36 ± 0.70	45	185
M2	3.63 ± 3.94	176	54
M3	23.89 ± 15.89*	222	8

**10 amplicons expected per library, representing a total of 230 amplicons generated from the 23 libraries of each NGS-protocol.*

average value of concentration was obtained using the M1 protocol ($\bar{x}$= 0.36 nM), followed by the M2 ($\bar{x}$= 3.63 nM), and the highest average value was achieved using the M3 ($\bar{x}$= 23.89 nM). There are significant differences in library quantification comparing M1, M2, and M3 protocols ($F=56.62$, $p<0.001$, $df=2$), and all pairs of comparisons are significant ($p<0.001$). The retrieval profile of the libraries by the length of amplicons is shown in Figure 26. Amplicons between 160-230bp were obtained using the protocol M2 and M3, while the M1 showed amplicons of 68bp that may be primer-dimers and just one peak of 220bp. Although both M2 and M3 protocols showed amplicons recovered with the required size for NGS-sequencing, the M3 produces better quality and quantity libraries for each amplicon than M2 (Figure 26.M2 and 26.M3).

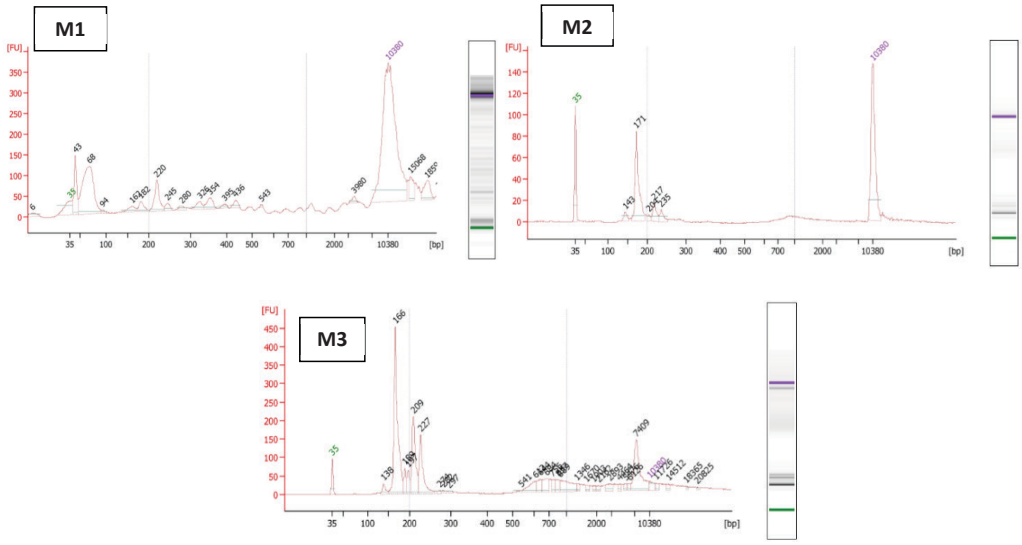


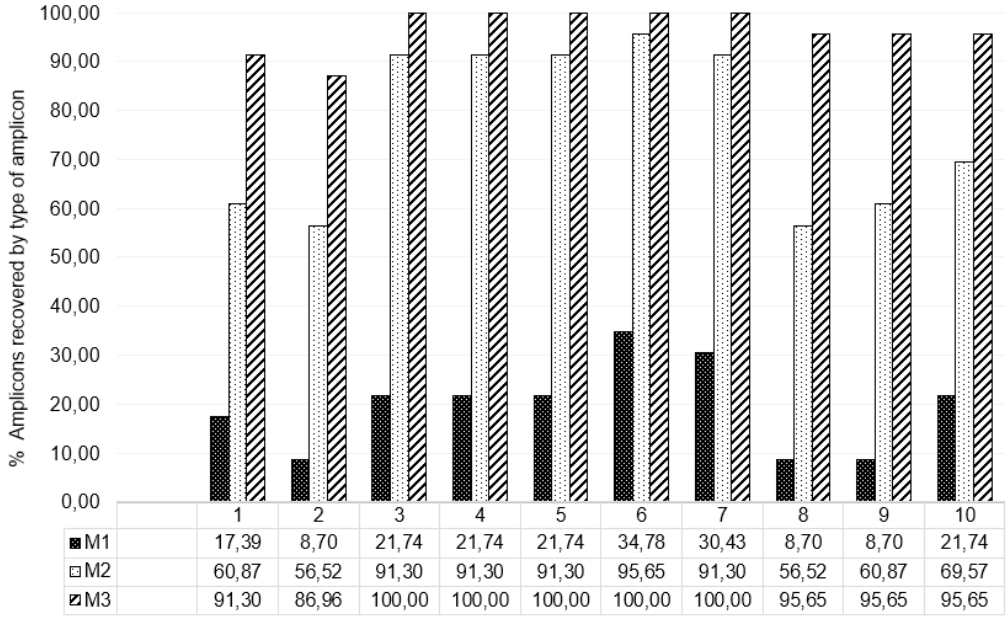
Figure 26 — Distributions of NGS-libraries by the length of amplicons recovered using the three NGS-protocols from Sample ID 4. The size markers or ladder are sharp peaks at 35bp and 10380bp. M1 Protocol: manufacturer's recommendation. M2 protocol: manufacturer's recommendation with 35 cycles. M3 protocol: 35 cycle and times longer in denaturalization and extension step.

The number of reliable amplicons obtained was 45, 176, and 222 of a total of 230 expected amplicons, using the protocol M1, M2, and M3, respectively. These results demonstrated that the protocol M3 presented a higher success in mtDNA-CR typing than the M1 and M2. Significant differences between the number of reliable and unreliable amplicons were observed (Cochran's Q test: F 282.72. df=2. $p<0.001$). Moreover, all

protocols differ from each other (pairwise comparisons McNewar test with Bonferroni correction: $p<0.001$).

Percentage of amplicons recovered

Figure 27 shows the percentage of amplicons recovered with reliable reads per type of amplicon. The results showed that the M3 protocol recovered percentages of amplicons considerably higher than the M1 protocol and slightly higher than the M2 protocol. The Amplicons #1 (16013-16126), #2 (16116-16225), #8 (246-364), #9 (342-436) and #10 (429-592), show a decline of the percentage of recovery in the three NGS-protocols used. However, the other amplicons were recovered in 91-95% and 100% using M2 and M3 protocol, respectively, and just in 21-34% with the M1 protocol. This means that M3 amplifies more than 85% of samples for all amplicons and 100% for 5 of the 10 amplicons.



	1	2	3	4	5	6	7	8	9	10
M1	17,39	8,70	21,74	21,74	21,74	34,78	30,43	8,70	8,70	21,74
M2	60,87	56,52	91,30	91,30	91,30	95,65	91,30	56,52	60,87	69,57
M3	91,30	86,96	100,00	100,00	100,00	100,00	100,00	95,65	95,65	95,65

Figure 27 — Bar Graphic of the percentage of amplicons recovered per type of amplicon (amplicon 1 to 10) using the three NGS protocols. M1 protocol: manufacturer's recommendation. M2 protocol: manufacturer's recommendation with 35 cycles. M3 protocol: 35 cycles and times longer in denaturalization and extension step.

Figure 28 shows the percentage of amplicons recovered for the 33 samples using M2 (25 ancient and 8 contemporaries) and the 41 samples (31 ancient and 10 contemporaries)

using the M3 protocol, clustered by different chronologies. The results showed that the M2 protocol recovered 82.50% of the amplicons of the contemporary samples (19th-20th century), while the M3 was 98%. The ancient samples (5500BP-16th century) showed slightly lower values than the contemporary samples in both protocols M2 (74.40%) and M3 (90.32%). Independently of the NGS protocol used, the results obtained from ancient and contemporary samples did not show significant differences between amplicons recovered (Fisher exact test: p= 0.4081).

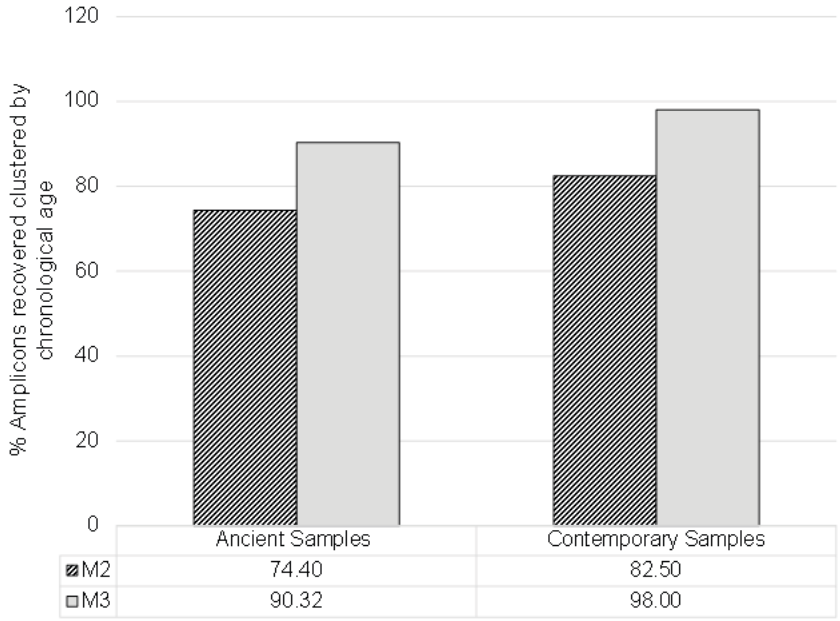


Figure 28 — Bar Graphic of the percentage of amplicons recovered clustered by different chronologies. M2 protocol: manufacturer's recommendation with 35 cycles. M3 protocol: 35 cycles and times longer in denaturalization and extension steps.

Analysis of polymorphisms

Table 15 indicates the parameters evaluated in all libraries obtained, the number of libraries generated, the percentage of libraries sequenced with LCR along with no sequenced classified as NR libraries, the total number of variants reported in the libraries sequenced, the number and percentage of the variants equality reported by both methods of analysis and discrepancy/ non-called variants (NC) between both analysis tools. Moreover, the percentage of reads with molecular damage is also presented.

Table 15 — Results from 97 libraries generated by the M1, M2, and M3 protocols, of which 78 libraries were sequenced.

NGS-protocols	Total Number of Libraries	Number of libraries with LCR or NR (%)	Total number of variants	Number equal variants in both analysis tools (%)	Number of No Called Variants (%)		% Molecular Damaged reads	
					GeneMarker-HTS	In-house pipeline	With Duplicated	Without Duplicated
M1	23	21 (91.30)	10	10 (100)	0	0	21.05	47
M2	33	13 (39.39)	104	89 (85.58)	3 (3.37)	14 (15.73)	27.17	46.29
M3	41	5 (12.20)	209	191 (91.40)	9 (4.71)	9(4.71)	25.43	48.06
All the trials	97	39 (34.05)	323	290 (89.78)	12 (3.71)	23 (7.12)	25.79	47.25

The M1 protocol shows the highest percentage of libraries with LCR or NR —21 of 23 libraries (91.30 %)—. In contrast, the M2 and the M3 show only 13 libraries (39.39%) and 5 libraries (12.20%), respectively, with poor results (Table 15). Therefore, these results indicated that the best protocol with the highest library retrieval was the M3 protocol.

A total of 78 libraries achieved quantification higher than 0.005 ng/µl (6 libraries by the M1 protocol, 31 libraries by the M2, and 41 libraries by the M3). The call variant showed a total of 323 polymorphisms, of which 89.78% (290 variants) were the same in the in-house pipeline and GeneMarker HTS software. As to the non-called variants (NC), GeneMarker HTS software did not call to 3 variants using the M2 protocol and 9 using the M3 protocol. 14 variants of the M2 protocol and 9 of the M3 were not called in the in-house pipeline. Overall, the M3 protocol showed the lowest number of NC variants which indicated the best recovery of libraries (Table SI 6).

The authenticity of the retrieved human DNA was assessed by computing molecular damage using the in-house pipeline. All 78 libraries obtained with the three NGS protocols presented reads with molecular damage. The percentage of molecular damaged reads increased almost 2-fold when PCR-duplicates were removed. In general terms, for all the reads generated, 47.25% showed patterns of molecular damage which confirms the recovery of degraded DNA and the obtaining of reliable results (Table 15 and Table SI 5: % Damage). GeneMarker HTS software did not quantify the number of damaged reads per library. However, in the interface of this software, the pile-up of the variants shows the presence of molecular damage reads (Figure SI 3).

Base mixtures were also found showing >30% of the minor variant in 6 libraries. Table 16 determines a total of 11 variants in the mixtures, which were found in the positions 150t/C, 152t/C, 16343a/G, 16390g/A, 16391g/A, and 16519t/C. The results show that 5 of 10 variants mixed were reported in both analysis tools; the base mixture remaining were not reported because they were in low coverage positions. The in-house pipeline determined more variant mixtures than GeneMarker HTS software which reported more 'NC.'

Table 16 — Bases mixture obtained from 6 libraries. NC stands for the base mix non-called in one of the analysis tools.

Sample ID	NGS-protocol	GeneMarker-HTS	In-house pipeline	Total number of variants in base mix	Number of equal base mix in both analysis tools (%)
9	M2	**150t/C**	**150t/C**	1	1(100)
10	M2	**16390g/A** NC	**16390g/A** 150t/C	2	1 (50)
	M3	**16390 g/A**	NC	1	0
14	M3	NC	**150t/C**	1	0
16	M3	**16391g/A**	**16391g/A**	1	1(100)
19	M3	NC NC **152t/C**	16343a/G 150t/C **152t/C**	3	1(33.33)
21	M3	**16519t/C**	**16519t/C**	1	1(100)
				10	5(50)

The mtDNA haplogroup of each individual was determined using a variant calling from the M3 protocol libraries (Table SI 7). The Haplogrep prediction qualities of the mtDNA haplogroups range from 50% up to 100%.

The polymorphisms obtained in the 8 samples for the mtDNA control region by Sanger were compared with those of NGS sequencing (Table 17). The results show that both techniques are concordant given that all the 33 variants detected in the regions analyzed are the same in both sequencing techniques.

Table 17 — Comparison of the control region variants obtained by Sanger sequencing and NGS.

Sample ID	Range analyzed	Polymorphisms	
		Sequencing by Sanger	**NGS**
28	16030-16230; 27-264	16111T	16111T
33	16030-16230; 16220-16420; 27-264	16169A, 16192T, 16235G, 16270T, 16304C, 73G, 150T, 228A	16169A, 16192T, 16235G, 16270T, 16304C, 73G, 150T, 228A
35	16030-16230	16129C,16145A,16183C,16189C	16129C,16145A,16183C,16189C
37	27-264	185A,188G,228A	185A,188G,228A
38	16030-16230; 27-264	185A,188G,195C, 228A	185A,188G,195C, 228A
39	16030-16230	rCRs	rCRs
40	16030-16222; 16230-16420; 27-264	16111T, 16187T, 16290T, 16325C, 16362C, 73G, 245G, 263G	16111T, 16187T, 16290T, 116325C, 16362C, 73G, 245G, 263G
41	16030-16222; 16230-16420	16111T, 16187T, 16290T, 16325C, 16362C	16111T, 16187T, 16290T, 16325C, 16362C

3.3.3 Discussion

Prior works have documented the effectiveness of using next-generation sequencing (NGS) for the genetic identification from degraded human skeletal remains (Young et al., 2019; Zupanič Pajnič & Fattorini, 2021). Most of them have reported an increase in the recovery of genetic information as the NGS technique uses more sensitive and robust commercial kits than conventional techniques. However, those studies tested several skeletal samples from only a specific chronology. They used protocols involving steps that could hinder the handling and increase the risk of contamination of critical samples.

In this study, we tested the recovery of the mtDNA control region (CR) by massively parallel sequencing from a wide range of archaeological samples (5500BP-20th century) using the PowerSeq^{TM}CRM Nested System kit. The results show that it is possible to obtain the mtDNA-CR typing from critical human skeletal remains. However, some modifications in the PCR conditions were necessary to get these successful results. We evaluated three protocols: M1 (manufacturer's recommendations), M2 (manufacturer's recommendations with 35 cycles), and M3 (also 35 cycles, and longer times in denaturalization and extension step). The M3 showed to be the best protocol with the highest percentage of recovery of the mtDNA-CR and allowing to successfully infer the mtDNA haplogroup for all the 41 individuals analyzed. The DNA from aged or degraded tissue is often highly fragmented due to autolysis, bacterial degradation, and spontaneous depurination (patterns of molecular damage). This fragmentation severely reduces the efficiency of the PCR (Golenberg et al., 1996). However, as in the present work, it has been demonstrated that the slight increase in the number of cycles (no more than 35 cycles), and longer times in the denaturalization and extension step (15-60 seconds), allow full-length polymerization and good DNA yield (Eichmann & Parson, 2008).

In the forensic context, the mtDNA-CR typing is based on the production of PCR libraries from previously amplified products of D-loop regions by next-generation sequencing. Usually, the library is prepared in two steps: the first one is the amplification of specific regions (HVSI, HVSII, or HVSIII), and the second one is the libraries preparation (adapters

addition and indexing) (Liu et al., 2012; McElhoe et al., 2014; Yang et al., 2014). In our study, we obtained good results of the control region mtDNA with the amplification and library preparation of 10 amplicons spanning all the D-loop region in just one PCR step. The results showed that the M3 protocol had more success in retrieving the 10 amplicons than the other protocols. However, regardless of the protocol used, we could observe a decrease in the number of reads at the beginning (Amplicons 1-2) and the end of the mtDNA-CR (Amplicons 8-10). Some studies (Brandhagen et al., 2020; Hofreiter, 2001) have demonstrated that most molecular damage hot spots are in amplicons 1 and 2 (HVSI). This molecular damage could inhibit the primer binding sites and produce a low coverage or no result for these regions; then, the improvement of primer design that avoids these regions could increase the success of mtDNA analysis in damaged samples.

The skeletal remains from the 20th century seem to present less degradation since they showed the highest percentage of amplicons recovered for both M2 and the M3 protocols. Many studies indicate that the degradation state or DNA preservation is low or high depending on the environmental conditions to which the biological samples are exposed (Latham & Miller, 2018). A sample buried at a low temperature (-15°C), with little humidity, and almost sterile conditions for many years could preserve its DNA much better than a sample buried only a few years ago but exposed to unfavorable conditions. However, the results did not show significant differences between the more ancient and contemporary samples.

Regarding the analysis tools used to determine the variants or polymorphisms from each library generated, we found a strong concordance of variants; therefore, both GeneMarker HTS software and the in-house pipeline are equally valuable. However, the GeneMarker HTS software does not estimate the reads with molecular damage, but the in-house pipeline did. This parameter should be evaluated for all the highly degraded samples since it indicates and confirms the recovery of degraded DNA instead of possible contamination of modern DNA (Skoglund et al., 2014). All our libraries showed molecular damage; this means that the template molecules that started the amplification were degraded DNA and not modern contaminating DNA. An advantage of both analysis tools is that both

can mitigate the sequencing errors, excluding the coordinates of the primer binding sites, avoiding overlapping regions, and providing only the informative sequence (Huszar et al., 2019). All NC variants were not called due to low coverage. This fact also occurs in some studies that used modern and degraded samples to validate this kit (Holland & McElhoe, 2015; Huszar et al., 2019). The bases mixtures were also reported, and the results showed that the in-house pipeline determined more variant mixtures than the GeneMarker HTS software, which reported more 'NC.' This fact could be for the differences in the analysis workflow of the software as mixture bases are not reported when there is a disequilibrium between the base in forward and reverse.

Regarding the coincidence of variants, they were reported with different analysis tools and sequencing techniques. The NGS technique showed comparable and reliable results to the Sanger technique. This concordance has been demonstrated in other studies that do not use the same kit but use NGS techniques to obtain mtDNA information (Parson et al., 2013).

4

General Considerations

"The nitrogen in our DNA, the calcium in our teeth, the iron in our blood, the carbon in our apple pies were made in the interiors of collapsing stars. We are made of starstuff."
- Carl Sagan, Cosmos

During the last decade, advances in molecular biology techniques have improved the genetic analyses of any DNA preserved in small amounts and with various states of degradation. In this way, genetic studies from skeletal remains are now useful for interpreting our evolutionary past through paleogenetics or to identifying a person through forensic genetics. However, despite technological advances, the recovery DNA still be challenging due to its scarcity and damage. Therefore, the first milestone to address must focus on the improvement of DNA recovery. The present doctoral thesis contributes the development and improvemente of methods in different ways: 1) comparing DNA contribution of different types of skeletal elements (explored in chapter 1), 2) searching for a useful and suitable protocol that maximizes the recovery of the major number of molecules and subsequently the genotyping of autosomal STR markers and the control region of mtDNA by conventional techniques (explored in chapter 1), 3) exploring the recovery of human whole-genome from newborn bones by NGS (explored in chapter 2), 4) probing the efficiency of a new NGS-kit to amplify control region of mtDNA by NGS (explored in chapter 3), and 5) providing a good bioinformatic tool to analyze the results of NGS-data, showing the molecular damage as a signal of the authentic recovery of endogenous DNA (explored in chapter 3). Overall, this thesis spanned comparisons of different aspects that can affect the retrieval of genetic information and that should be considered if critical samples, like bones, are used for identification purposes or population characterization.

The results showed that one of the limiting factors in the recovery of fragmented/degraded DNA is the used DNA extraction technique. Prior works have already reported differences in DNA yield when diverse extraction methods are used on critical samples (Jakubowska et al., 2012; Rohland & Hofreiter, 2007b; Rothe & Nagy, 2016). Therefore, using suitable extraction

techniques to recover as much DNA as possible is crucial for further enzymatic manipulation. The latter usually involves amplification by PCR of targeted regions for the sequencing by conventional or NGS techniques or the capture of the whole genome by shotgun-NGS.

A wide range of techniques has been published to maximize DNA yields from ancient DNA (Dabney & Meyer, 2019; Rohland et al., 2010, 2018; Rohland & Hofreiter, 2007a). Ancient DNA research shares a common problem with forensics requiring analyses of specimens highly damaged with extremely limited amounts of endogenous DNA. Thus, when compromised samples are tested, the extraction methods usually employed in ancient DNA can also be applied to the forensic community. The extraction methods based on silica-in columns and silica-in suspension have shown to be highly efficient in the recovery of PCR-amplifiable DNA from ancient bone and teeth specimens and, at the same time, in minimizing co-extraction of substances that inhibit PCR (Rohland et al., 2018; Rohland & Hofreiter, 2007a; Yang et al., 1998). Some authors have conducted studies on DNA performance comparing silica-based and Ph-Cl extraction methods. The results demonstrated that the silica-based extraction method allowed better results in STR typing from degraded bone samples than the commonly used phenol/chloroform method (Davoren et al., 2007; Marshall et al., 2014a; Ozdemir-Kaynak & Yesil-Celiktas, 2015; Rohland & Hofreiter, 2007b). The paper developed in this thesis (Vinueza-Espinosa et al., 2020) shows concordant outcomes with these findings.

The extraction protocols based on silica–DNA binding show the ability to successfully recover DNA, obtain good nuclear STR profiles, and a complete typing of the mtDNA control region by Sanger and Capillary Electrophoresis. It is important to highlight that this is the first work conducting a systematic study of recovered DNA from different types of skeletal remains taken from the same individual by methods based on silica (in-columns and in-suspension) in contrast to Ph-Cl.

The extraction protocols based on silica used in this dissertation have been standardized by the ancient DNA community, which uses a smaller amount of skeletal remains powder

(50mg -<1gr) than the forensic laboratories commonly use (>1gr) (Ferreira et al., 2013). In samples such as damaged as those employed in forensic and ancient investigations, there is often not enough biological material to extract DNA or its preservation is crucial due to its archaeological significance, so the best protocol should be one that uses less starting material like both silica-based protocols tested in this thesis.

In general, solely a few types of skeletal remains contain analyzable DNA since the state of bone preservation is highly variable and influenced by direct environmental conditions. The inner part of the petrous bone and tooth is currently recognized as the optimal substrates perserving high amount of endogenous DNA. Some studies have reported that DNA preservation in petrous bones is significantly higher than in teeth from archaeological samples (Gamba et al., 2016; Hansen et al., 2017c). All these findings evaluated the yield of DNA by the NGS technique. This dissertation compared the recovered DNA from various types of skeletal remains (including petrous and teeth) by conventional and NGS techniques in separate chapters (chapters 1 and 2, respectively).

The results showed concordance with previous ancient DNA findings that the petrous is superior compared with other types of skeletal remains, even from teeth by conventional and NGS techniques. Thus, it is important to mention that even when gold standard techniques are used to identify an individual, petrous is still the ideal substrate of DNA content, resulting in partial to complete profiles of autosomal STR and genetic information of the mtDNA control region (HVRI and HVRII).

Most the systematic studies to know the highest DNA yield by type of skeletal remains are performed on adults' skeletons. The lack of systematic technical studies on neonatal bones is evident; hence, it is important to know if the use of new born skeletal remains is a limiting factor in DNA recovery and later analysis. In this context, my research evaluated two approaches. The first is an intra-individual study comparing two skeletal remains—petrous bone and vertebra— And the second is a comparative study between age categories —infantile, juvenile, adult, and perinatal— from petrous and other skeletal

remains. Intra-individual results confirmed that petrous bones from infants showed much better results than other bones. Also, it proved to be an equally valuable sample than in adults since a large amount of endogenous DNA was recovered. These outcomes could also be highly valuable for the forensic approach because there are many kinship cases from infant identification in which the sex is difficult to obtain by conventional techniques (Čakar et al., 2020; Minaguchi et al., 2003).

Mitochondrial DNA is a profitable genetic analysis tool in forensic and paleogenetic research that allows us to establish lineage with close or distant kinship. For these reasons, mtDNA is widely used to identify missing persons (i.e., from the Spanish civil war) or population characterization.

Usually, from well-preserved samples, two distinct regions (HVI and HVII) of around 200-400bp are PCR amplified and sequenced using Sanger technology (Wilson et al., 1995). However, this attempt to obtain >200pb amplicons in samples of poor quality may not provide any results since the DNA is highly fragmented into smaller molecules. Since the advent of NGS, the whole mitogenome analysis is possible from high up to extremely low-quality samples, such as skeletal remains (Holt et al., 2021). However, this new technology is not completely implemented in forensic laboratories because comparative exercises are still needed.

The recovery of small amplicons (100-130pb) has been proved successful in ancient DNA research and forensic analysis of bones and hair shafts (Alonso et al., 2003; Gabriel et al., 2001). Since then, some authors have demonstrated that using a set of shorter amplicons is an effective alternative when highly degraded samples have been analyzed (Eichmann & Parson, 2008). However, at least eight fragments must be amplified and sequenced to recover the entire Control Region, which means at least 8 PCR reactions and 16 Sanger sequencing reactions (Gabriel et al., 2001). Since mtDNA testing for damaged samples is generally manually performed, the procedure is very demanding and time-consuming. In contrast, NGS provides larger genetic informative data set with higher throughput. It also

offers reduced workflows and lower costs per nucleotide than capillary electrophoresis-based methods (Mardis, 2008; Yang et al., 2014).

The transition from Sanger to NGS amplification can be simplified if simple kits with short protocols and no tube change are available. The PowerSeq™CRM Nested System kit (Promega Corporation) employs a multiplex-PCR strategy to amplify the entire mtDNA-CR in a single multiplex that combines ten primer pairs, generating ten overlapping short amplicons (147-237pb). The results of this thesis established the reliability of accurate mtDNA control region typing using PowerSeq™CRM Nested System kit when degraded human remains from different archaeological contexts are used to aid in forensic identifying procedures or population characterization. An accurate new protocol is proposed, modified from the proposed by the manufactures.

The outcomes confirm that this technology has been of particular benefit since a more informative sequence could be recovered than the one obtained using a traditional method based on multiple small amplicons and Sanger sequencing. Results concordance was shown between CE-Sanger and NGS, demonstrating the NGS-kit's repeatability. Notwithstanding, the use of reliable software to analyze the results is a crucial step; for this reason, a powerful forensic tool software and an in-house pipeline were compared.

The molecular damage is easily detectable and widely studied in ancient DNA because various bioinformatics pipelines have been developed. However, it is more difficult to detect it by forensic methods because most of them are based on PCR amplification, not on the capture of the whole genome.

The aDNA field has developed workflows specifically for degraded DNA. Advancements in the extraction (Dabney et al., 2013), library preparation (Gansauge et al., 2020), and enrichment of target DNA have enabled the generation of hominin nuclear and mtDNA sequence profiles from skeletal elements as well as sediment over 300,000 years old (Meyer et al., 2016; Vernot et al., 2021; Zavala et al., 2021). In aDNA, applying these methodological

advances is possible due to the degraded nature of ancient DNA samples. Forensic samples, by contrast, are much more heterogeneous in terms of DNA quality. Novel forensic methods are largely aimed to improve DNA profiling of traditional forensic targets that exceed 100 bp in size. However, these new forensic DNA methods must meet rigorous validation standards, making testing and adopting new protocols difficult. These differing forces that drive the two research fields have positioned aDNA considerably ahead of forensics in terms of method development. Yet both fields share the common problem of analyzing DNA that is (or can be) highly degraded. This commonality prompted whether methods from the aDNA field could improve success rates for generating mtDNA profiles, which has been suggested in recent studies and reviews (Hofreiter et al., 2021; Xavier et al., 2021). Zavala and collaborators tested the feasibility of integrating ancient DNA methods into a forensic casework workflow by comparing different DNA extraction methods and library preparation protocols. Their results were concordant with the reported in this dissertation in which DNA profiling is successful from highly degraded skeletal remains.

The investigations performed throughout this thesis were focused on several factors that could maximize the recovery of DNA from degraded samples. However, more factors should be considered, such as molecular taphonomy, which could predict the success of the recovery of genetic material in terms of external condition factors and soil microbiota.

Throughout this research, I also encountered challenges and additional molecular areas worthy of exploration, such as SNPs or STRs based on NGS that could not be addressed during the development of my thesis but that would serve as interesting points of investigation for future studies. Overall, this research has presented new considerations and avenues from a methodological point of view applying the knowledge about degraded DNA, showing value in enabling genetic identification or population characterization for forensic and paleogenetic purposes from degraded human skeletal remains.

5

Conclusions

"Happiness is the real sense of fulfillment that comes from hard work. Starting strong is good finishing strong is epic."
-Robin Sharma

- The comparison of DNA extraction methods analyzed in this work shows the successful use of the methods based on silica-DNA binding. The silica in-suspension and silica in the HE- membrane column methods recovered the highest amount of short fragments characteristic of highly degraded DNA, typical of forensic and ancient samples.

- The silica in-suspension method obtained the highest number of positive amplifications of the mtDNA fragments and the highest concentration of mtDNA products. In contrast, the phenol-chloroform method (Ph-Cl) was associated with the poorest results for both parameters.

- Petrous bone and pulp cavity using the methods of silica in-suspension and silica in the HE- membrane column displayed the best results in the concentration of amplified mtDNA products.

- Petrous bone extracted with the silica in-suspension method showed the highest percentage of autosomal STRs reportable alleles with the highest values of the Peak Height Ratio (PHR).

- The dense cochlear portion of the petrous pyramid is the best sampling location to recover high-quality ancient DNA from newborn or infant individuals. The single-stranded library complexity from petrous demonstrated to be superior to the vertebra.

- Perinatal petrous bones compared to petrous bones from other age categories (infantile, juvenile, adult and non-determined), appear to be equally or more valuable biological samples to obtain genetic information, disregarding the possible myth that the newborn bones could not be a good source of DNA. These results provide researchers with the first systematic study of newborn skeletal remains by NGS.

- A new modified NGS protocol, with more PCR cycles and large PCR steps, based on PowerSeqTMCRM Nested System kit, was implemented allowing the recovery of the whole mtDNA-CR from highly degraded skeletal samples of different chronologies.

- The use of PowerSeqTMCRM Nested System kit based on a single-multiplex PCR reaction appears to be an interesting tool to be applied in forensic and paleogenomics approaches because the results are highly concordant compared to Sanger Sequencing. This NGS kit reduces the workload substantially and decreases cross-contamination risks, time-consuming, and costs

- Our freely-available in-house pipeline to analyze mtDNA NGS results provides concordant variants with the forensic GeneMarker HTS software. It can confirm the presence of molecular damage of degraded DNA, contrarily to GeneMarker HTS software.

6

Supporting Information

6.1 Chapter 1:

Human DNA Extraction from Highly Degraded Skeletal Remains: How to find a suitable method?

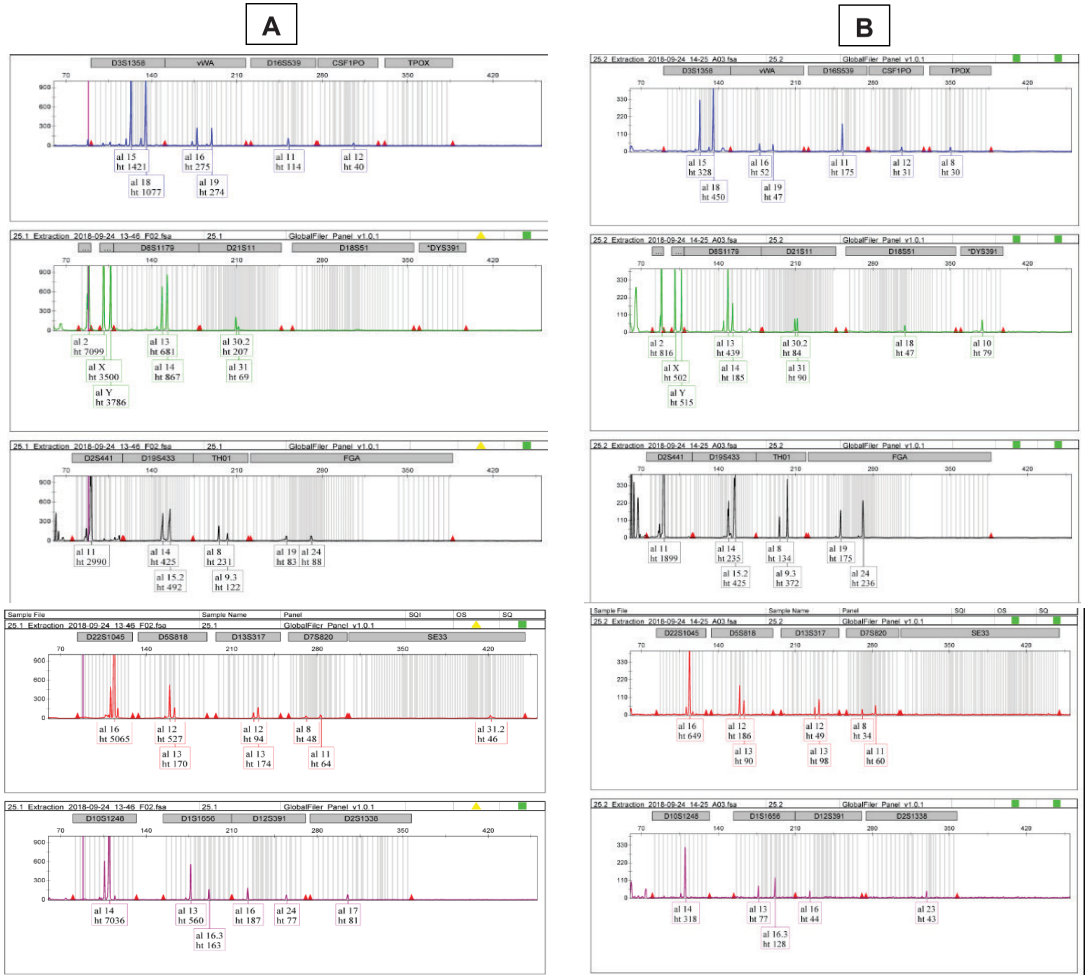


Figure SI 1 — Electropherograms obtained of the individual 5 from DNA extracted from petrous using method 1 (A) and method 2 (B).

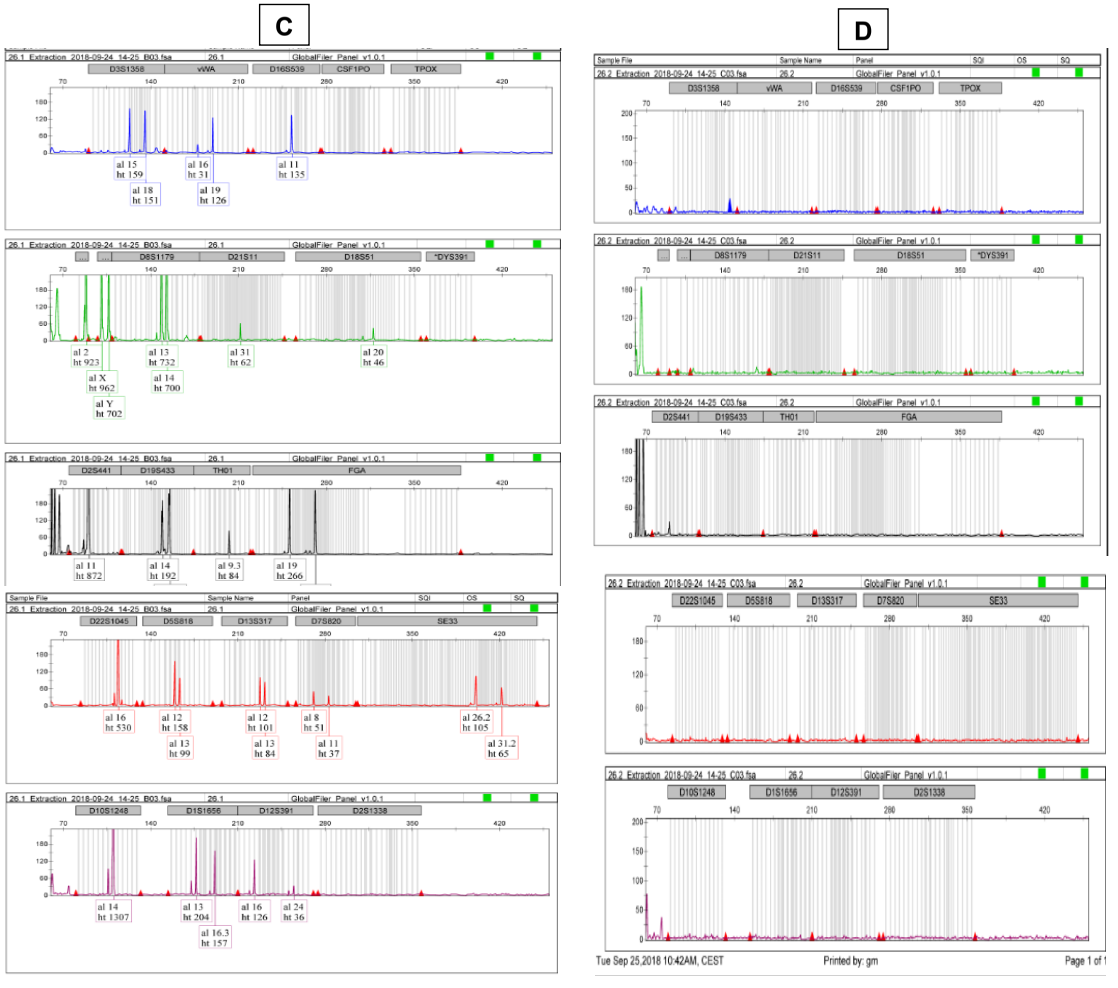


Figure SI 2 — Electropherograms obtained of the individual 5 from DNA extracted from pulp cavity using method 1 (C) and method 2 (D).

Table SI 1 — Data collected for the assessment of human DNA recovery as of five type of skeletal remains selected from each individuals. Sample ID or Laboratory internal ID for the individuals, Individual Code assigned in the archaeological site.

Sample ID	Individual Code	Type of skeletal remain	~Weight of sample powder collected per each extraction method (mg)	DNA concentration (ng/ul)				Presence (+) or Absence (-) of amplification of mtDNA fragment (HVRI)				Concentration (ng/ul) of mtDNA amplicon (HVRI)				Presence (+) or Absence (-) of amplification of mtDNA fragment (HVRII)				Concentration (ng/ul) of mtDNA amplicon (HVRII)				Presence (+) or Absence (-) of amplification of nuclear fragment		Concentration (ng/ul) of nuclear DNA amplicon		Peak Heights (RFUs)		(%) PHR		% Reportable Alleles	
				Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 3	Method 4	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2	Method 1	Method 2
1	Ind 4 camp1	Petrous	200	78	26,7	9,3	0,35	+	+	-	-	25,8	30,4	0	0	+	+	+	-	15,54	15,6	2,26	0,48	+	+	7,46	6,7	695,41	350,85	53,90722	53,219	50	50
		Pulp cavity	50	77,3	14,5	5,8	0,16	+	+	-	-	12,1	25,08	0	0	+	+	+	-	28	20,82	0,66	0	+	-	1,12	0	103,95	0	49,19728	0	36,36364	0
		Cementum	50	5,3	17,7	10,9	0,77	-	-	-	-	0	0	0	0	+	-	-	-	0,214	0	0	0										
		Radius*	150	5,7	24,2	2,7	1,23	-	+	-	-	0	28,4	0	0	+	+	-	-	0,204	0,12	0	0										
		Rib	150	25,6	62,4	5,8	0,75	+	+	-	-	1,01	13,5	0	0	+	+	+	-	0,214	0,172	0,122	0										
2	Ind 20 camp1	Petrous	200	76,3	70,2	26	5,9	+	+	+	+	40	29,3	0,31	7,6	+	+	-	+	20,4	33,2	0	4,4	+	+	2,2	1,58	667,42	309,0208	72,02175	67,7273	90,90909	81,81818
		Pulp cavity	50	10,2	19,9	17,1	3,04	+	+	+	+	10,4	28,86	12,4	2,6	+	+	+	+	5,9	30,61	18,4	5,38	-	-	0	0	0	0	0	0	0	0
		Cementum	35	32,6	60,4	17,5	3,79	+	+	+	+	39,4	24	2,18	3,14	+	+	+	-	8,04	8,76	8	6,78										
		Radius*	100	71,5	20,4	8,3	3,98	+	+	+	+	9,66	31,8	20,8	1,6	+	+	-	-	9,66	0,114	0	0										
		Rib	200	74	2,2	2,7	0,32	+	-	-	-	10,6	0	0	0	+	+	+	-	0,11	1,41	0,29	0										
3	Ind 39 camp2	Petrous	150	78	29,3	3,7	0,98	+	+	-	-	28,54	27,6	0	0	+	+	+	+	19,7	28,3	1,09	2	+	+	3,48	2,5	67,2	55,51667	25,24752	81,7137	63,63636	59,09091
		Pulp cavity	30	25,7	23,8	14,3	2,6	+	+	+	+	24,4	31,66	5,36	2,2	+	+	+	+	36,2	29,4	1,2	4,4	-	-	0	0	0	0	0	0	0	0
		Cementum	50	25	3,8	4,4	2,67	+	-	-	+	13,3	0	0	1,6	+	-	-	-	0,2	0	0	0										
		Ulna*	130	71,5	26,7	20	2	+	+	+	+	4,66	1,2	10,5	3,6	+	+	-	-	7,74	7,48	0	0										
		Rib	145	56,2	10,1	3,9	0,12	+	-	-	-	20,8	0	0	0	+	+	-	-	11	8,98	0	0										
4	Ind 131 camp2	Petrous	170	74,9	26,2	10,6	6,89	+	+	+	+	30,5	22,5	15,7	2	+	+	+	-	34,6	34,4	0,8	0,8	+	+	7,1	1,28	528,87	262,125	45,51159	58,0207	100	72,72727
		Pulp cavity	30	35,9	12,5	18	6,94	+	-	+	+	14,2	0	4,76	2,2	+	+	+	+	10,9	4,58	5,6	1,8	+	-	1,35	0	0	0	0	0	0	0
		Cementum	50	5,1	10,7	11,4	1,2	-	+	+	-	0	0,4	3,58	0	-	+	+	-	0	3,72	0,3	0										
		Phalanx*	162	78,6	24,3	13,7	1,06	+	+	+	+	7,56	0,45	5,5	3,8	+	+	+	+	10,5	9,18	2,32	0,17										
		Rib	126	3,3	8,2	3,4	1,89	-	-	-	-	0	0	0	0	+	+	-	-	11,4	10,4	0	0										
5	Ind 4 viver	Petrous	164	83,2	43,5	13,3	0,85	+	+	+	-	40,56	34,6	27	0	+	+	+	-	32,78	30,6	13,92	0,88	+	+	6,62	8,48	534,14	165	68,71422	53,8029	90,90909	95,45455
		Pulp cavity	50	69,1	23,2	16,1	1,35	+	+	+	+	38,2	10,39	29,4	1,8	+	+	+	-	37,8	6,8	28,8	0,5	+	-	20,6	0	185,18	0	68,1176	0	86,36364	0
		Cementum	30	28,7	63,5	16,8	2,3	+	+	+	+	21,4	18	19,6	3	+	+	+	+	9,04	6,8	5,2	0,5										
		Metacarpal*	120	3,3	3,2	2,2	0,89	-	-	-	-	0	0	0	0	+	+	+	+	0,488	0,12	0,228	0,74										
		Rib	150	4,3	3,4	2,8	3,2	-	-	-	-	0	0	0	0	+	+	-	+	0,576	0,288	0	0,15										
							∑	19/25	16/25	13/25	12/25					∑	24/25	23/25	15/25	9/25				∑	8/10	5/10						

** noted in this study as Upper limb bones*

Table SI 2 — Experimental details of the DNA extraction methods procedures.

		Method 1 Silica-in suspension*	Method 2 Silica in HE-membrane column	Method 3 Silica in XS plasma- membrane column		Method 4 Phenol-Chloroform
Lysis Stage		The first DNA extraction step consisted in a pre-digestion stage adding 2 ml of extraction buffer [0.46M EDTA, pH: 8, 10mM TE buffer, 0.14 mg/ml proteinase K, 0.5% SDS and 1/1000 Phenol red] to the powder sample, followed by mixing and incubation for 15 min at the 37°C. Then it was centrifuged at 13000 rpm for 2 min and the supernatant was discarded. 3.5 ml of extraction buffer were added to the pellet and incubated overnight on the rotator at 37C° until the powder was completely dissolved.	The first DNA extraction step consisted in adding 1 ml of extraction buffer [0.5M EDTA and 0.25 mg/ml of proteinase K] to each powder sample. Then, the mix was incubated at 37°C in agitation overnight.	The powder samples were digested in 260 µl of extraction buffer [0.5M EDTA and 0.25mg/µl of proteinase K]. The samples were incubated at 37°C overnight.	Lysis Stage	The powder samples were predigested in 1 ml of 0.5M EDTA, pH: 8 and incubated with agitation at 56°C overnight. Then the samples were centrifuged at 13000 rpm for 7 min. The supernatant was discarded and 1ml of extraction buffer [H2O, 5% SDS, 1M Tris-HCl, pH: 8 and 10mg/ml proteinase K] was added to the pellet. The samples were incubated for 72 hours at 56°C with agitation. After, the solution was centrifuged at 14000 rpm for 4 min and lysate was recovered.
Purification	DNA BINDING	The overnight incubated samples were centrifuged at 13000 rpm for 2 min; the supernatant was transferred to a 50 ml tube and 40 ml of binding buffer [500 ml QIAquick™ PB buffer (Qiagen), 1:1000 vol. phenol red, 24.88mM NaCl, 87.6 mM sodium acetate] and 100 µl of silica* in suspension were added, adjusting PH to 4-4.5 with 37% HCl. After, samples were incubated for 1 hour at room temperature with gently rotating.	The overnight incubated samples were centrifuged at 13.000 rpm and the supernatant transferred to a 50 ml tube followed by 10.4 ml of binding buffer [5M guanidine hydrochloride, 40% (vol/vol), 2-propanol, 0.12 M sodium acetate and 0.05% (vol/vol) Tween 20]. All the mix was transferred to the HE-column and then centrifuged at 4000 rpm for 5 min. The flow-through was discarded	The overnight incubated samples were centrifuged at 13000 rpm for 2 min and the supernatant transferred to a 2 ml tube followed by 360 µl of binding buffer [Guanidinium thiocyanate 30-60% and ethanol 35-55%]. All the mix was transferred to the XS-columns and centrifuged at 8000 rpm for 30 seconds.	ORGANIC PHASE	Then the organic phase consisted on three steps: lysate or supernatant/phenol (1:1), supernatant/phenol-chloroform solution (1:1), and finally supernatant/chloroform (1:1). The centrifugations for each organic step were at 14000 rpm for 4 min.
	WASHING	Finally, the samples were centrifuged at 8000 rpm for 5 minutes and the supernatant was discarded. The silica-pellet was resuspended in 1ml binding buffer and transferred to a 2 ml tube, then it was centrifuged, and the supernatant was totally removed. The washing was performed twice with 80% cold ethanol. The ethanol was completely discarded.	The column was twice washed adding 450 µl of PE buffer [20mM NaCl, 2Mm Tris-HCl]; a dry-spin at maximum speed was performed for each step of washed for 1 min. The flow-through was discarded.	The columns were twice washed adding 500µl ethanol (55-75%) followed by a dry spin at maximum speed for 3 minutes. The ethanol traces were evaporated of the column at 37°C. The flow-through was discarded.		
	ELUTION	Then 100µl TEB Buffer [EB buffer Qiagen® with 0.05% Tween-20] was once placed in the silica-pellet for 15 minutes at 37°C then centrifuged at 13000 rpm for 2 min. The 98 µl supernatant (DNA extract) was transferred to another tube and frozen until be used.	Finally, the DNA elution was performed adding 45µl TTE buffer [1M Tris-HCl, 0.5M EDTA pH:8 and Tween-20] in the center of the silica-membrane twice, left for 5 minutes at room temperature and centrifuged at 13000 rpm for 1 min in each step	Finally, the DNA elution was performed adding 20µl of EB elution buffer [5Mm Tris/HCl, pH: 8.5] twice in the center of the silica membrane, was left for 5 minutes at room temperature and centrifuged at 13000 rpm for 1 min in each step.	ELUTION	Finally, the DNA concentration phase was carried out using Amicon® Ultra 0.5 ml Filters (Merck, Germany). Five hundred µl of the supernatant was twice collocated into the filter following the manufacturer's recommendation. The DNA extract final volume was between 15-20 µl.

**silica: To prepare the Silica suspension, 6g of SiO2 was mixed with 50 ml of sterile H2O. The suspension was left for sedimentation 1 hour at room temperature. Then 48 ml of the supernatant were transferred to a fresh tube and left for sedimentation during 4-5 hours at room temperature. The top 43 ml (supernatant) were carefully removed and 60 µl of 37% HCl was added to resuspend and activate silica. Finally, the silica in suspension was stored at 4°C until be used.*

Supporting Information

6.2 Chapter 2:

Newborn Skeletal Remains: an aid or a constraint to paleogenomics?

Table SI 3 — Sample and sequences information in the petrous and the vertebrae bones from nine infant individuals.

Sample ID	Skeletal Remain	mg_input	PCR cycles	Molecules in library	Total Sequences	Length (bp)	% GC	% Trimmed	Reads Post Trimming	Length (bp)	% GC	Mapped Reads	Endogenous DNA Fraction with PCR-duplicates (%)	Non-clonal Sequences	Endogenous DNA Fraction Without Duplicates (%)	Clonality (%)*	Uniquely reads mapping to human genome per million reads	Estimation of genomic coverage (fold)	Mean read length (bp)	Median read length (bp)	% GC	Nuclear genome coverage*	mtDNA coverage*	5' Prime C>T 1st base	5 Prime C>T 2nd base*	Chromosome X	Chromosome Y
CL-ENT 1	Petrous-cochlear region	155	10	3,13E+12	1,96E+07	81	53%	13,50%	1,70E+07	52	53%	2,27E+06	13,36%	2,01E+06	11,8518%	11,30	1,03E+06	7,79E+03	57,73	56	37%	0,0387	10,920	22,60%	14,80%	105562,00	369
	Lumbar Vertebrae	186	19	9,45E+12	3,59E+06	81	50%	34,30%	2,36E+06	55	53%	2,28E+04	0,97%	2,37E+04	1,0062%	89,6	6,61E+04	1,57E+03	53,34	51	39%	0,0004	0,13	16,10%	11,40%	1132,00	6
CL-ENT 2	Petrous-cochlear region	162	16	1,19E+13	2,42E+07	81	52%	13,17%	2,10E+07	52	54%	4,06E+06	19,36%	2,77E+06	13,1785%	31,9	1,14E+06	3,95E+04	53,33	51	38%	0,0492	12,55	21,90%	14,10%	139754,00	508
	Lumbar Vertebrae	190	22	2,97E+12	2,87E+06	81	53%	15,22%	2,43E+06	56	54%	4,40E+05	18,09%	1,04E+04	0,4253%	97,7	3,61E+04	8,67E+03	50,06	46	37%	0,0002	0,06	15,00%	11,70%	466,00	0
CL-ENT 4	Petrous-cochlear region	195	15	1,03E+13	2,39E+07	81	56%	7,94%	2,20E+07	62	57%	3,13E+06	14,20%	2,41E+06	10,9248%	23,1	1,01E+06	2,88E+04	61,07	62	40%	0,0490	8,27	16,20%	11,70%	61399,00	5952
	Cervical Vertebrae	165	19	1,17E+13	4,79E+06	81	50%	36,23%	3,06E+06	61	54%	3,12E+05	10,21%	3,21E+04	1,0520%	89,7	6,71E+04	2,26E+04	58,46	58	40%	0,0006	0,31	11,90%	7,90%	1114,00	47
CL-ENT 6	Petrous-cochlear region	177	13	8,57E+12	2,23E+07	81	53%	15,67%	1,88E+07	53	54%	3,50E+06	18,65%	2,76E+06	14,6985%	21,2	1,24E+06	2,99E+04	57,98	57	41%	0,0534	9,65	14,40%	10,60%	136079,00	466
	Lumbar Vertebrae	158	23	7,77E+12	1,18E+07	81	50%	26,17%	8,69E+06	49	53%	1,45E+06	16,70%	1,09E+03	0,0126%	99,3	9,27E+02	2,04E+04	48,82	45	37%	0,0000	0,0051	16,10%	11,90%	516,00	2
CL-ENT 7	Petrous-cochlear region	163	14	1,19E+13	1,81E+07	81	47%	9,78%	1,63E+07	56	46%	5,92E+06	36,25%	5,01E+06	30,6818%	15,4	2,77E+06	8,13E+04	58,34	57	39%	0,0975	19,80	21,80%	15,60%	162452,00	872
	Cervical Vertebrae	198	23	6,31E+12	9,79E+06	81	50%	27,21%	7,13E+06	49	50%	1,45E+06	20,35%	3,04E+03	0,0427%	99,8	3,10E+03	1,97E+04	47,62	44	37%	0,0000	0,0211	14,40%	11,90%	172,00	1
CL-ENT 8	Petrous-cochlear region	116	15	1,07E+13	2,62E+07	81	54%	9,22%	2,38E+07	54	57%	1,43E+06	6,01%	9,65E+05	4,0524%	32,6	3,68E+05	1,20E+04	58,25	57	37%	0,0187	2,7100	28,20%	19,00%	50388,00	145
	Cervical Vertebrae	156	23	1,02E+13	1,31E+07	81	49%	30,90%	9,05E+06	47	50%	1,96E+06	21,66%	2,53E+04	0,2800%	98,7	1,93E+04	3,37E+04	47,47	44	37%	0,0004	0,1270	14,60%	11,30%	1273,00	8
CL-ENT 9	Petrous-cochlear region	185	13	4,26E+11	5,12E+03	81	49%	0,23%	5,10E+03	54	49%	1,00E+00	0,02%	1,00E+00	0,0196%	0,00	1,96E+03	1,02E+00	37,92	35	47%	0,0000	0,00	8,10%	17,50%	0,00	0
	Cervical Vertebrae	193	14	7,79E+12	2,22E+07	81	61%	2,15%	2,17E+07	67	62%	6,12E+05	2,81%	4,78E+05	2,1994%	21,8	2,15E+05	4,48E+03	63,44	65	40%	0,0101	4,2400	21,30%	13,30%	12452,00	1223
CL-19-12 UE3094	Petrous-cochlear region	171	13	7,26E+12	1,85E+07	81	55%	4,85%	1,76E+07	63	56%	3,11E+06	17,73%	2,55E+06	14,4998%	18,2	1,38E+06	2,64E+04	63,48	66	40%	0,0539	11,80	17,40%	10,70%	131287,00	498
	Vertebrae	171	22	1,17E+13	1,00E+07	81	54%	22,45%	7,76E+06	55	59%	3,15E+05	4,06%	8,10E+03	0,1044%	97,4	8,09E+03	8,07E+03	52,73	50	39%	0,0001	0,0486	18,00%	13,20%	387,00	2
CCA 2199	Petrous-cochlear region	196	9	3,11E+12	2,04E+07	81	59%	6,80%	1,90E+07	56	62%	2,27E+05	1,19%	2,04E+05	1,0717%	10,2	9,99E+04	5,39E+02	45,02	43	38%	0,0031	1,17	31,70%	18,40%	10438,00	53
	Vertebrae	196	10	7,36E+12	8,44E+06	81	59%	6,55%	7,88E+06	55	61%	1,32E+05	1,67%	1,18E+05	1,4945%	10,4	1,40E+05	2,20E+03	55,64	53	40%	0,0022	0,9774	25,60%	14,90%	5657,00	25
	Total			1,42E+14	2,60E+08				2,26E+08			3,04E+07		1,94E+07			9,59E+06	3,48E+05								8,21E+05	1,02E+04
	Average			7,92E+12	1,44E+07		53,00%	15,69%	1,25E+07		54,67%	1,69E+06	12,41%	1,08E+06	5,98%	48,24	5,33E+05	1,93E+04	53,93	52,22	39,06%	0,02	4,60	18,63%	13,33%	45584,89	565,39
	Average: Petrous			7,47E+12	1,92E+07		53,11%	9,02%	1,73E+07	55,78	54,22%	2,63E+06	14,09%	2,07E+06	11,22%	18,21	1,00E+06	2,51E+04	54,79	53,78	39,67%	0,04037	8,54111	20,26%	14,71%	88595,44	984,78
	Average: Vertebrae			8,36E+12	9,62E+06		52,89%	22,35%	7,79E+06	54,89	55,11%	7,44E+05	10,72%	7,78E+04	0,74%	78,27	6,17E+04	1,35E+04	53,06	50,67	38,44%	0,00157	0,65851	17,00%	11,94%	2574,33	146,00
																	7,51E+05	1,99E+04	6,7975476	8,335293887	0,024125218	0,028821001	6,021920829	0,059241088	0,029379476	60402,9902	1389,342127

**p<0.05 significant differences between petrous and vertebra paired student t-test*

Table SI 4 — Averages and standard deviations (sd) of sequencing data from 1019 DNA libraries of 825 individuals published by Patterson and collaborators (Patterson et al., 2022).

Variable		MtDNA coverage*			1240K coverage at targeted positions			1240K unique SNPs hit*			1240K sequences hitting X Chr		
	Skeletal remains	Petrous bone	Other	Total mean	Petrous bone	Other	Total mean	Petrous bone	Other	Total mean	Petrous bone	Other	Total mean
Age category	Perinatal	485 sd=476	45.9 sd=24.4	360 sd=448	2.91 sd=1.94	0.114 sd=0.0599	2.11 sd=2.08	762387 sd=174367	126482 sd=61406	580700 sd=327738	762387 sd=174367	126482 sd=61406	580700 sd=327738
	Infantile	269 sd=210	161 sd=236	246 sd=219	2.43 sd=1.90	0.275 sd=0.232	1.96 sd=1.90	630173 sd=239876	226069 sd=169248	542005 sd=280998	630173 sd=239876	226069 sd=169248	542005 sd=280998
	Juvenile	238 sd=187	54.7 sd=59.5	176 sd=177	3.02 sd=2.02	0.191 sd=0.494	2.05 sd=2.14	691794 sd=238548	94746 sd=174620	487541 sd=358666	691794 sd=238548	94746 sd=174620	487541 sd=358666
	Adult	242 sd=233	106 sd=143	201 sd=218	2.65 sd=1.91	0.256 sd=0.611	1.92 sd=1.96	670220 sd=251885	164291 sd=199612	516736 sd=332266	670220 sd=251885	164291 sd=199612	516736 sd=332266
	n/d	224 sd=258	104 sd=161	172 sd=229	2.45 sd=2.02	0.291 sd=0.730	1.51 sd=1.92	612627 sd=290220	175163 sd=226119	422502 sd=341806	612627 sd=290220	175163 sd=226119	422502 sd=341806
Total mean		245 sd=257	102 sd=152		2.59 sd=1.97	0.269 sd=0.661		647774 sd=267138	166720 sd=211380		647774 sd=267138	166720 sd=211380	

**p<0.05 significant differences between petrous bone from perinatal individuals compared to petrous bone from other age categories*

Supporting Information

6.3 Chapter 3:

MtDNA-CR Typing from Highly Degraded Skeletal Remains
by Single-Multiplex Massively Parallel Sequencing

Supplementary Tables

Table SI 5 — Sample information obtained regarding the concentration of DNA genomic and DNA library, the number of reads by type of amplicon using GeneMarker HTS software and the number of total and damaged reads using in-house pipeline.
Number of reads per type of amplicon with >10 reads are given as 'reliable' highlighted in green and 0-9 reads as 'unreliable' highlighted red by GeneMarker-HTS software. NR highlighted in orange stands for 'NO RESULTS' and it was given for samples with <0.005 library quantification (these samples were discarded for sequencing showing in the subsequent analysis of polymorfisms and the number of total and damaged reads as NR). The number of damaged reads and total reads are given in the right on the table by the in-house pipeline to calculate the percentage of damage.

Table SI 6 — Comparison of the control region variants called by GeneMarker-HTS software and in-house pipeline.
Homoplasmic variants highlighted in bold are obtained in both analysis tools from three NGS-protocols. NC in green font stands for 'variant not called because of low coverage'. The variants not called are highlighted in gray, variants with base mixture are highlighted in blue with >30% in minority base, LCR stands for 'low coverage reads', NR stands for 'no result'. The total number of variants detected, the number of equal variants and the number of discrepancies or variants no called are given in the right on the table.

Table SI 7 — Prediction of mtDNA haplogroup by Haplogrep 2.

Table SI 5 — Sample information obtained regarding the concentration of DNA genomic and DNA library, the number of reads by type of amplicon using GeneMarker HTS software and the number of total and damaged reads using in-house pipeline.

Sample ID	Chronological Age	NGS-Protocols	Type of bone	Cronological Age	Run	DNA genomic Concentration pg/ul	Library Concentration (nM)
1	8th century	M1	Premolar tooth (Lower right)	8th century	1	0,005	1,27
		M2			2		1,64
		M3			2		1,89
2	8th century	M1	Incisor tooth (upper left)	8th century	NR	1,83	<0,005
		M2			1		2,37
		M3			3		12,7
3	8th century	M1	Canine tooth (upper left)	8th century	NR	0,83	<0,005
		M2			1		1,79
		M3			3		28,85
4	8th century	M1	Canine tooth (upper right)	8th century	1	1	2,66
		M2			2		14,45
		M3			2		35,6
5	8th century	M1	Molar tooth upper right	8th century	NR	2,17	<0,005
		M2			1		5,75
		M3			3		19,44
6	8th century	M1	Fibula fragment (left)	8th century	NR	0,005	<0,005
		M2			NR		<0,005
		M3			3		18,18
7	8th century	M1	Incisor tooth (upper left)	8th century	1	12,7	0,84
		M2			2		2,26
		M3			2		12,26
8	8th century	M1	Metatarsal bone (I Right)	8th century	NR	2,16	<0,005
		M2			3		4,14
		M3			3		57,47
9	8th century	M1	Canine tooth (upper right)	8th century	1	16,83	<0,005
		M2			2		3,8
		M3			3		16,72
10	8th century	M1	Canine tooth (lower left)	8th century	NR	0,16	<0,005
		M2			1		1,12
		M3			2		25,84
11	8th century	M1	Metatarsal bone (V left)	8th century	NR	0,16	<0,005
		M2			1		2,27
		M3			2		14,45
12	8th century	M1	Proximal phalanx bone (right hand)	8th century	NR	0,005	<0,005
		M2			1		1,83
		M3			2		10,53
13	8th century	M1	Incisor tooth (lower right)	8th century	NR	0,16	<0,005
		M2			1		2,07
		M3			3		3,74

Table SI 5 — Continuation.

Sample ID	Chronological Age	NGS-Protocols	Type of bone	Cronological Age	Run	DNA genomic Concentration pg/ul	Library Concentration (nM)
14	8th century	M1	Incisor tooth (upper left)	8th century	NR	4,16	<0,005
		M2			1		1,41
		M3			2		20,15
15	8th century	M1	Canine tooth (upper left)	8th century	1	0,005	0,95
		M2			1		12,3
		M3			2		23,01
16	8th century	M1	Fragment of femoral shaft	8th century	3	0,005	1
		M2			1		2,15
		M3			2		11,53
17	8th century	M1	Molar tooth	8th century	2	27,2	1,53
		M2			1		12,6
		M3			2		48,76
18	8th century	M1	Incisive tooth (upper right)	8th century	NR	32,33	<0,005
		M2			1		2,17
		M3			2		49,68
19	5500 BP	M1	Metatarsal bone (left)	5500BP	NR	0,17	<0,005
		M2			1		0,99
		M3			3		6
20	5500 BP	M1	Metatarsal bone (left)	5500BP	NR	1,2	<0,005
		M2			1		1,32
		M3			2		47,22
21	5th century	M1	Petrous	5th century	NR	0,005	<0,005
		M2			1		2,48
		M3			2		13,69
22	13th century	M1	Canine tooth	13th century	NR	0,005	<0,005
		M2			1		2,61
		M3			2		31,84
23	20th century	M1	Petrous	20thcentury	NR	34,6	<0,005
		M2			3		1,87
		M3			1		40
24	20th century	M2	Pulpa	20thcentury	3	1,5	3,36
		M3			3		10,65
25	20th century	M2	Incisive tooth	20thcentury	1	4,5	6,03
		M3			3		10,45
26	13th century	M2	Femur	13th century	2	0,005	1,92
		M3					6,84
27	13th century	M2	SemiluNRr bone	13th century	2	0,005	3,46
		M3			3		4,08
28	16th century	M2	Fragment of phalanx	16th century	2	0,005	14,07
		M3			3		40

Table SI 5 — Continuation.

Sample ID	Chronological Age	NGS-Protocols	Type of bone	Cronological Age	Run	DNA genomic Concentration pg/ul	Library Concentration (nM)
29	20th century	M2	Pulpa	20thcentury	3	4	0,63
		M3			3		6,53
30	20th century	M2	Pulpa	20thcentury	3	16,2	2,45
		M3			3		43,46
31	20th century	M2	Pulpa	20thcentury	3	12,3	<0,005
		M3			3		12,5
32	20th century	M2	Pulpa	20thcentury	3	32,5	2,97
		M3			3		29,36
33	20th century	M2	Pulpa	20thcentury	3	23,5	1,43
		M3			3		1,46
34	5-11th centuries	M3	Petrous	5-11th centuries	3	0,005	1,23
35	5-11th centuries	M3	Petrous	5-11th centuries	3	0,005	0,78
36	5-11th centuries	M3	Petrous	5-11th centuries	3	1,3	1,88
37	5-11th centuries	M3	Pulpa	5-11th centuries	3	1,6	0,5
38	5-11th centuries	M3	Petrous	5-11th centuries	3	0,005	3,58
39	5-11th centuries	M3	Pulpa	5-11th centuries	3	0,005	0.50
40	19th century	M3	Premolar tooth	19thcentury	2	1,4	20,99
41	19th century	M3	Radius bone	19thcentury	2	0,16	5,07

Table SI 5 — Continuation.

ID	Amplicon 1 F16013-R16126				Amplicon 2 F16116-R16225				Amplicon 3 F16223-R16408				Amplicon 4 F16387-R16486				Amplicon 5 F16474-R30				Amplicon 6 F16555-R152				Amplicon 7 F136-R257				Amplicon 8 F246-R364				Amplicon 9 F342-R436				Amplicon 10 F429-R592			
		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R
1	Reliable	1340	674	666	Reliable	546	270	276	Reliable	3158	130	3028	Reliable	3982	3088	894	Reliable	3528	1078	2450	Reliable	4982	2394	2588	Reliable	4453	3432	1021	Reliable	238	110	128	Reliable	1584	797	787	Reliable	1685	819	866
	Reliable	164	82	82	Reliable	90	44	46	Reliable	11550	44	11506	Reliable	15301	11717	3584	Reliable	8974	3672	5302	Reliable	12359	5219	7140	Reliable	7320	7288	32	Reliable	28	2	26	Reliable	166	94	72	Reliable	478	226	252
	Reliable	268	134	134	Reliable	121	58	63	Reliable	11091	84	11007	Reliable	13833	11233	2600	Reliable	11004	2724	8280	Reliable	15519	8215	7304	Reliable	7643	7528	115	Reliable	52	6	46	Reliable	250	144	106	Reliable	831	402	429
2	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	280	142	138	Reliable	116	56	60	Reliable	3561	4	3557	Reliable	4913	3719	1194	Reliable	4112	1234	2878	Reliable	6269	2823	3446	Reliable	3666	3592	74	Reliable	108	48	60	Reliable	254	134	120	Reliable	257	126	131
	Reliable	6779	3406	3373	Reliable	1920	960	960	Reliable	6076	360	5716	Reliable	8650	6542	2108	Reliable	11888	3810	8078	Reliable	13448	7598	5850	Reliable	14074	8888	5186	Reliable	6198	3051	3147	Reliable	9237	4652	4585	Reliable	17608	8699	8909
3	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Reliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	196	100	96	Reliable	68	30	38	Reliable	4618	46	4572	Reliable	5832	4708	1124	Reliable	4606	1164	3442	Reliable	6957	3404	3553	Reliable	3760	3698	62	Reliable	71	23	48	Reliable	316	168	148	Reliable	354	176	178
	Reliable	1660	834	826	Reliable	492	246	246	Reliable	4204	924	3280	Reliable	3705	2792	913	Reliable	4220	1338	2882	Reliable	5624	3008	2616	Reliable	4772	3298	1474	Reliable	1383	677	706	Reliable	2695	1360	1335	Reliable	3947	1959	1988
4	Reliable	16	8	8	Unreliable	0	0	0	Reliable	1208	6	1202	Reliable	1594	1224	370	Reliable	1218	364	854	Reliable	1712	860	852	Reliable	872	856	16	Reliable	22	8	14	Reliable	10	6	4	Reliable	36	18	18
	Reliable	192	96	96	Reliable	122	60	62	Reliable	8298	10	8288	Reliable	11058	8546	2512	Reliable	7181	2595	4586	Reliable	10376	4532	5844	Reliable	6122	6058	64	Reliable	36	10	26	Reliable	252	136	116	Reliable	227	113	114
	Reliable	100	50	50	Reliable	66	32	34	Reliable	10028	8	10020	Reliable	12441	10263	2178	Reliable	8244	2228	6016	Reliable	11756	5962	5794	Reliable	5910	5898	12	Reliable	18	2	16	Reliable	118	64	54	Reliable	62	32	30
5	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	382	192	190	Reliable	140	68	72	Reliable	646	53	593	Reliable	788	616	172	Reliable	962	266	696	Reliable	1264	668	596	Reliable	1138	818	320	Reliable	300	149	151	Reliable	443	224	219	Reliable	458	228	230
	Reliable	159	80	79	Reliable	22	8	14	Reliable	4383	18	4365	Reliable	5572	4502	1070	Reliable	3938	1118	2820	Reliable	6338	2784	3554	Reliable	3724	3688	36	Reliable	44	12	32	Reliable	108	64	44	Reliable	120	60	60
6	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Reliable	NR	NR	NR
	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Reliable	NR	NR	NR
	Reliable	224	114	110	Reliable	112	54	58	Reliable	2376	102	2274	Reliable	2692	2286	406	Reliable	2344	482	1862	Reliable	3115	1823	1292	Reliable	1398	1316	82	Reliable	190	94	96	Reliable	26	14	12	Reliable	336	166	170
7	Unreliable	0	0	0	Unreliable	0	0	0	Reliable	42	0	42	Reliable	74	42	32	Reliable	38	30	8	Reliable	52	8	44	Reliable	48	46	2	Unreliable	0	0	0	Unreliable	0	0	0	Unreliable	6	2	4
	Reliable	44	22	22	Reliable	52	24	28	Reliable	6513	44	6469	Reliable	8977	6600	2377	Reliable	5266	2418	2848	Reliable	6736	2800	3936	Reliable	4061	4038	23	Reliable	19	7	12	Reliable	158	82	76	Reliable	252	122	130
	Reliable	508	254	254	Reliable	157	76	81	Reliable	14815	120	14695	Reliable	19324	15042	4282	Reliable	15393	4435	10958	Reliable	20123	10797	9326	Reliable	9835	9647	188	Reliable	136	51	85	Reliable	903	470	433	Reliable	1455	722	733
8	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	202	104	98	Reliable	946	472	474	Reliable	4522	1198	3324	Reliable	3394	2669	725	Reliable	5356	1654	3702	Reliable	6272	3638	2634	Reliable	6276	3816	2460	Reliable	1642	810	832	Reliable	1700	858	842	Reliable	7012	3478	3534
	Reliable	20	10	10	Reliable	322	156	166	Reliable	5979	104	5875	Reliable	7831	6005	1826	Reliable	7446	1906	5540	Reliable	10220	5450	4770	Reliable	5158	5002	156	Reliable	265	111	154	Reliable	630	332	298	Reliable	668	332	336
9	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	80	40	40	Reliable	46	24	22	Reliable	3375	18	3357	Reliable	4414	3450	964	Reliable	3129	1015	2114	Reliable	2146	2094	52	Reliable	90	62	28	Reliable	32	0	32	Reliable	284	160	124	Reliable	86	44	42
	Reliable	4988	2516	2472	Reliable	1562	782	780	Reliable	10130	813	9317	Reliable	13280	9322	3958	Reliable	15362	4980	10382	Reliable	11371	10125	1246	Reliable	8268	4186	4082	Reliable	38	18	20	Reliable	1749	878	871	Reliable	8197	4053	4144
10	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Unreliable	0	0	0	Unreliable	0	0	0	Reliable	256	2	254	Reliable	342	256	86	Reliable	106	82	24	Reliable	24	24	0	Unreliable	8	4	4	Unreliable	0	0	0	Unreliable	0	0	0	Unreliable	4	2	2
	Reliable	86	44	42	Unreliable	8	4	4	Reliable	15284	10	15274	Reliable	18067	15423	2644	Reliable	11196	2632	8564	Reliable	8628	8522	106	Reliable	152	126	26	Reliable	22	6	16	Reliable	134	72	62	Reliable	176	86	90
11	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Unreliable	4	2	2	Reliable	10	4	6	Reliable	717	0	717	Reliable	928	734	194	Reliable	360	198	162	Reliable	402	162	240	Reliable	244	244	0	Unreliable	0	0	0	Unreliable	0	0	0	Unreliable	2	2	0
	Reliable	28	14	14	Reliable	12	6	6	Reliable	11557	4	11553	Reliable	13656	11788	1868	Reliable	6005	1879	4126	Reliable	7492	4114	3378	Reliable	3470	3464	6	Reliable	14	2	12	Reliable	62	36	26	Reliable	16	8	8
12	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Reliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Unreliable	0	0	0	Unreliable	6	4	2	Reliable	92	0	92	Reliable	106	92	14	Reliable	60	14	46	Reliable	56	48	8	Reliable	16	12	4	Unreliable	0	0	0	Unreliable	2	0	2	Unreliable	4	2	2
	Reliable	76	38	38	Reliable	32	16	16	Reliable	3293	34	3259	Reliable	4170	3390	780	Reliable	2652	736	1916	Reliable	3070	1912	1158	Reliable	1194	1176	18	Reliable	25	9	16	Reliable	58	32	26	Reliable	120	60	60
13	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Unreliable	3	2	1	Unreliable	0	0	0	Reliable	1293	4	1289	Reliable	1594	1296	298	Reliable	418	306	112	Reliable	198	110	88	Reliable	94	94	0	Unreliable	0	0	0	Unreliable	0	0	0	Unreliable	0	0	0
	Reliable	182	92	90	Reliable	104	52	52	Reliable	1695	18	1677	Reliable	2140	1726	414	Reliable	1598	442	1156	Reliable	2218	1152	1066	Reliable	1342	1180	162	Reliable	106	52	54	Reliable	114	58	56	Reliable	492	246	246

Table SI 5 — Continuation.

ID	Amplicon 1 F16013-R16126				Amplicon 2 F16116-R16225				Amplicon 3 F16223-R16408				Amplicon 4 F16387-R16486				Amplicon 5 F16474-R30				Amplicon 6 F16555-R152				Amplicon 7 F136-R257				Amplicon 8 F246-R364				Amplicon 9 F342-R436				Amplicon 10 F429-R592			
		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R
14	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Unreliable	0	0	0	Unreliable	0	0	0	Reliable	534	0	534	Reliable	714	550	164	Reliable	362	172	190	Reliable	190	184	6	Reliable	14	10	4	Unreliable	2	0	2	Unreliable	6	4	2	Unreliable	8	4	4
	Reliable	100	50	50	Unreliable	0	0	0	Reliable	12508	8	12500	Reliable	15543	12772	2771	Reliable	9341	2819	6522	Reliable	6568	6474	94	Reliable	136	106	30	Reliable	12	2	10	Reliable	96	52	44	Reliable	144	72	72
15	Unreliable	0	0	0	Unreliable	0	0	0	Reliable	398	0	398	Reliable	530	400	130	Reliable	200	128	72	Reliable	240	72	168	Reliable	172	172	0	Unreliable	0	0	0	Unreliable	6	2	4	Reliable	16	8	8
	Reliable	120	60	60	Reliable	40	22	18	Reliable	15418	10	15408	Reliable	19770	15782	3988	Reliable	11309	4097	7212	Reliable	13824	7060	6764	Reliable	6844	6832	12	Reliable	34	12	22	Reliable	160	86	74	Reliable	768	382	386
	Reliable	644	324	320	Reliable	190	108	82	Reliable	16114	12	16102	Reliable	21962	16597	5365	Reliable	22139	5565	16574	Reliable	29637	16297	13340	Reliable	14063	13793	270	Reliable	246	96	150	Reliable	1204	631	573	Reliable	1484	736	748
16	Unreliable	0	0	0	Unreliable	0	0	0	Unreliable	6	0	6	Unreliable	8	6	2	Unreliable	8	2	6	Reliable	16	8	8	Reliable	10	8	2	Unreliable	0	0	0	Unreliable	0	0	0	Unreliable	0	0	0
	Unreliable	0	0	0	Unreliable	0	0	0	Unreliable	4	0	4	Unreliable	6	4	2	Unreliable	8	2	6	Reliable	10	8	2	Reliable	12	8	4	Unreliable	4	2	2	Unreliable	0	0	0	Reliable	22	10	12
	Unreliable	0	0	0	Unreliable	4	2	2	Reliable	228	0	228	Reliable	288	244	44	Reliable	194	40	154	Reliable	200	154	46	Reliable	44	44	0	Unreliable	0	0	0	Unreliable	8	4	4	Unreliable	0	0	0
17	Reliable	60	30	30	Reliable	24	12	12	Reliable	3841	20	3821	Reliable	4730	3890	840	Reliable	2616	842	1774	Reliable	3702	1766	1936	Reliable	2124	2052	72	Unreliable	2	0	2	Unreliable	4	2	2	Reliable	192	96	96
	Reliable	118	58	60	Reliable	98	44	54	Reliable	14442	24	14418	Reliable	18116	14830	3286	Reliable	13561	3366	10195	Reliable	20510	10192	10318	Reliable	10642	10576	66	Reliable	58	14	44	Reliable	416	224	192	Reliable	172	84	88
	Reliable	170	86	84	Reliable	128	64	64	Reliable	18248	48	18200	Reliable	22287	18800	3487	Reliable	18097	3585	14512	Reliable	25454	14462	10992	Reliable	11424	11310	114	Reliable	76	32	44	Reliable	252	135	117	Reliable	332	164	168
18	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Reliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Unreliable	0	0	0	Unreliable	4	2	2	Reliable	3370	10	3360	Reliable	4452	3447	1005	Reliable	2048	1020	1028	Reliable	1060	1018	42	Reliable	68	54	14	Unreliable	8	4	4	Unreliable	0	0	0	Reliable	42	22	20
	Unreliable	4	2	2	Reliable	26	4	22	Reliable	22562	22	22540	Reliable	31801	23119	8682	Reliable	13257	8855	4402	Reliable	4650	4404	246	Reliable	252	250	2	Reliable	14	0	14	Reliable	14	14	0	Reliable	22	10	12
19	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Unreliable	0	0	0	Unreliable	4	2	2	Reliable	90	0	90	Reliable	100	92	8	Reliable	52	8	44	Reliable	58	46	12	Reliable	10	10	0	Unreliable	4	2	2	Unreliable	4	2	2	Unreliable	4	2	2
	Reliable	228	114	114	Reliable	76	38	38	Reliable	662	206	456	Reliable	392	306	86	Reliable	416	124	292	Reliable	692	472	220	Reliable	592	282	310	Reliable	116	56	60	Reliable	92	46	46	Reliable	664	330	334
20	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	2524	1264	1260	Reliable	1028	514	514	Reliable	5103	812	4291	Reliable	5368	4306	1062	Reliable	4397	1434	2963	Reliable	3974	2990	984	Reliable	4208	2060	2148	Reliable	1381	686	695	Reliable	1418	710	708	Reliable	5063	2527	2536
	Reliable	300	152	148	Reliable	202	100	102	Reliable	23242	108	23134	Reliable	27032	23446	3586	Reliable	18073	3576	14497	Reliable	14456	14300	156	Reliable	362	218	144	Reliable	100	44	56	Reliable	230	114	116	Reliable	1009	507	502
21	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	12	6	6	Unreliable	4	2	2	Reliable	1625	0	1625	Reliable	2178	1642	536	Reliable	984	538	446	Reliable	1646	448	1198	Reliable	1218	1214	4	Unreliable	0	0	0	Reliable	14	8	6	Unreliable	8	4	4
	Reliable	48	24	24	Reliable	24	12	12	Reliable	10485	26	10459	Reliable	12675	10667	2008	Reliable	8864	2054	6810	Reliable	11908	6804	5104	Reliable	5214	5212	2	Reliable	12	2	10	Reliable	50	26	24	Reliable	78	38	40
22	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	20	10	10	Unreliable	8	4	4	Reliable	1762	14	1748	Reliable	2158	1792	366	Reliable	1040	370	670	Reliable	1450	674	776	Reliable	816	800	16	Reliable	20	8	12	Reliable	140	74	66	Reliable	48	26	22
	Reliable	68	34	34	Reliable	44	22	22	Reliable	6951	2	6949	Reliable	8351	7137	1214	Reliable	5136	1212	3924	Reliable	7492	3896	3596	Reliable	3744	3692	52	Reliable	98	36	62	Reliable	540	284	256	Reliable	118	58	60
23	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	106	52	54	Reliable	60	26	34	Reliable	7857	14	7843	Reliable	10173	8033	2140	Reliable	8696	2200	6496	Reliable	13472	6470	7002	Reliable	7214	7178	36	Reliable	85	28	57	Reliable	208	116	92	Reliable	58	30	28
	Reliable	3894	1956	1938	Reliable	1954	972	982	Reliable	13593	484	13109	Reliable	17022	13624	3398	Reliable	15500	4274	11226	Reliable	21001	10177	10824	Reliable	18370	14472	3898	Reliable	3158	1540	1618	Reliable	4358	2200	2158	Reliable	3475	1731	1744
24	Reliable	14	6	8	Unreliable	6	4	2	Reliable	1212	0	1212	Reliable	1504	1266	238	Reliable	888	262	626	Reliable	1464	622	842	Reliable	880	872	8	Unreliable	8	2	6	Reliable	54	32	22	Unreliable	8	2	6
	Reliable	60	32	28	Reliable	26	14	12	Reliable	1963	2	1961	Reliable	2642	2092	550	Reliable	1714	590	1124	Reliable	2812	1124	1688	Reliable	1836	1800	36	Reliable	46	20	26	Reliable	120	64	56	Unreliable	8	4	4
25	Reliable	150	76	74	Reliable	62	30	32	Reliable	796	58	738	Reliable	870	728	142	Reliable	936	160	776	Reliable	1082	756	326	Reliable	408	356	52	Reliable	82	34	48	Reliable	490	252	238	Reliable	312	152	160
	Reliable	6090	3078	3012	Reliable	1954	972	982	Reliable	10864	5558	5306	Reliable	2586	1662	924	Reliable	8866	4288	4578	Reliable	4580	4580	0	Reliable	12972	6106	6866	Reliable	4759	2316	2443	Reliable	6178	3144	3034	Reliable	19424	9576	9848
26	Unreliable	0	0	0	Unreliable	0	0	0	Reliable	44	0	44	Reliable	52	46	6	Reliable	24	6	18	Reliable	26	16	10	Reliable	12	10	2	Unreliable	4	2	2	Unreliable	6	4	2	Unreliable	0	0	0
	Reliable	24	12	12	Unreliable	4	2	2	Reliable	110	0	110	Reliable	122	110	12	Reliable	202	18	184	Reliable	234	180	54	Reliable	56	54	2	Unreliable	0	0	0	Unreliable	4	2	2	Unreliable	8	4	4
27	Unreliable	0	0	0	Unreliable	0	0	0	Reliable	34	0	34	Reliable	36	34	2	Reliable	144	18	126	Reliable	136	110	26	Reliable	24	24	0	Unreliable	0	0	0	Unreliable	8	4	4	Unreliable	2	2	0
	Unreliable	0	0	0	Unreliable	4	2	2	Reliable	12	0	12	Reliable	10	10	0	Reliable	14	2	12	Reliable	22	10	12	Reliable	16	14	2	Unreliable	4	2	2	Unreliable	0	0	0	Unreliable	6	4	2
28	Reliable	554	278	276	Reliable	26	10	16	Reliable	7064	38	7026	Reliable	8916	7186	1730	Reliable	10138	1804	8334	Reliable	14052	8258	5794	Reliable	6192	6014	178	Reliable	118	50	68	Reliable	160	90	70	Reliable	444	220	224
	Reliable	644	326	318	Reliable	40	20	20	Reliable	18014	40	17974	Reliable	22215	18479	3736	Reliable	24158	3806	20352	Reliable	34198	20238	13960	Reliable	14483	14352	131	Reliable	34	12	22	Reliable	38	24	14	Reliable	320	160	160

Table SI 5 — Continuation.

ID	Amplicon 1 F16013-R16126				Amplicon 2 F16116-R16225				Amplicon 3 F16223-R16408				Amplicon 4 F16387-R16486				Amplicon 5 F16474-R30				Amplicon 6 F16555-R152				Amplicon 7 F136-R257				Amplicon 8 F246-R364				Amplicon 9 F342-R436				Amplicon 10 F429-R592			
		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R		Total	F	R
29	Reliable	115	58	57	Reliable	72	36	36	Reliable	1124	92	1032	Reliable	1180	1016	164	Reliable	1910	252	1658	Reliable	2102	1608	494	Reliable	740	576	164	Reliable	44	22	22	Unreliable	0	0	0	Reliable	870	432	438
	Reliable	450	228	222	Reliable	159	80	79	Reliable	15210	164	15046	Reliable	17887	15409	2478	Reliable	15430	2654	12776	Reliable	18035	12535	5500	Reliable	6466	5984	482	Reliable	150	74	76	Reliable	100	52	48	Reliable	1797	886	911
30	Reliable	868	434	434	Reliable	180	90	90	Reliable	7706	312	7394	Reliable	8930	7399	1531	Reliable	11458	1748	9710	Reliable	14200	9736	4464	Reliable	5588	4864	724	Reliable	134	66	68	Reliable	28	14	14	Reliable	3256	1616	1640
	Reliable	4481	2252	2229	Reliable	898	446	452	Reliable	11220	1576	9644	Reliable	12392	9039	3353	Reliable	14191	4060	10131	Reliable	20221	11230	8991	Reliable	12844	9810	3034	Reliable	2584	1266	1318	Reliable	422	228	194	Reliable	10514	5216	5298
31	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR	Unreliable	NR	NR	NR
	Reliable	4888	2460	2428	Reliable	2210	1102	1108	Reliable	40400	1865	38535	Reliable	46242	38630	7612	Reliable	54069	8483	45586	Reliable	67045	45491	21554	Reliable	26969	23480	3489	Reliable	3641	1788	1853	Reliable	1230	628	602	Reliable	16366	8114	8252
32	Reliable	686	344	342	Reliable	214	106	108	Reliable	4843	312	4531	Reliable	5536	4485	1051	Reliable	5350	1320	4030	Reliable	6457	3918	2539	Reliable	3318	2780	538	Reliable	311	151	160	Reliable	58	30	28	Reliable	3258	1610	1648
	Reliable	2324	1176	1148	Reliable	698	344	354	Reliable	17226	618	16608	Reliable	20304	16732	3572	Reliable	24724	3972	20752	Reliable	30845	20655	10190	Reliable	12548	11130	1418	Reliable	1478	720	758	Reliable	346	180	166	Reliable	7783	3847	3936
33	Reliable	210	110	100	Reliable	60	30	30	Reliable	2028	162	1866	Reliable	2070	1810	260	Reliable	2424	388	2036	Reliable	2074	1960	114	Reliable	446	224	222	Reliable	65	31	34	Reliable	262	132	130	Reliable	400	198	202
	Reliable	46	24	22	Reliable	12	6	6	Reliable	194	4	190	Reliable	242	200	42	Reliable	314	54	260	Reliable	276	258	18	Reliable	68	38	30	Unreliable	8	4	4	Reliable	44	22	22	Reliable	50	24	26
34	Unreliable	0	0	0	Unreliable	0	0	0	Reliable	485	0	485	Reliable	626	498	128	Reliable	440	140	300	Reliable	322	294	28	Reliable	52	42	10	Reliable	16	6	10	Reliable	12	8	4	Unreliable	0	0	0
35	Unreliable	0	0	0	Reliable	94	42	52	Reliable	6165	52	6113	Reliable	7706	6332	1374	Reliable	5098	1516	3582	Reliable	6784	3462	3322	Reliable	3924	3624	300	Reliable	226	111	115	Reliable	536	272	264	Reliable	160	80	80
36	Unreliable	8	4	4	Unreliable	0	0	0	Reliable	35	2	33	Reliable	40	32	8	Reliable	34	8	26	Reliable	74	34	40	Reliable	42	38	4	Reliable	12	6	6	Unreliable	2	0	2	Unreliable	4	2	2
37	Reliable	14	8	6	Unreliable	4	2	2	Reliable	556	4	552	Reliable	704	560	144	Reliable	752	164	588	Reliable	1546	574	972	Reliable	1016	1000	16	Unreliable	8	4	4	Reliable	28	14	14	Reliable	22	12	10
38	Reliable	224	112	112	Reliable	114	56	58	Reliable	6144	52	6092	Reliable	7476	6320	1156	Reliable	5038	1174	3864	Reliable	6902	3824	3078	Reliable	3602	3328	274	Reliable	282	138	144	Reliable	310	156	154	Reliable	266	134	132
39	Reliable	24	12	12	Unreliable	8	4	4	Reliable	318	0	318	Reliable	399	327	72	Reliable	148	72	76	Reliable	294	76	218	Reliable	222	220	2	Unreliable	8	4	4	Reliable	16	8	8	Unreliable	8	2	6
40	Reliable	194	96	98	Reliable	32	16	16	Reliable	35255	34	35221	Reliable	41466	35914	5552	Reliable	27081	5533	21548	Reliable	35230	21448	13782	Reliable	14144	14080	64	Reliable	52	22	30	Reliable	66	36	30	Reliable	296	144	152
41	Reliable	226	114	112	Reliable	64	32	32	Reliable	5360	6	5354	Reliable	7145	5521	1624	Reliable	8470	1654	6816	Reliable	12464	6784	5680	Reliable	5880	5836	44	Reliable	30	14	16	Reliable	14	8	6	Reliable	186	92	94

Table SI 5 — Continuation.

ID	Total reads GeneMarker HTS	Inhouse-pipeline with PCR-Duplicates		
		Total reads inhouse-pipeline	Damage reads in house pipeline	% Damage
1	12768	12301	4846	39,39517112
	28230	26861	5996	22,32232605
	30329	30593	6827	22,31556238
2	NR	NR	NR	NR
	11776	12508	2987	23,88071634
	48019	42796	17572	41,05991214
3	NR	NR	NR	NR
	13393	13838	2658	19,20797803
	16362	15287	4725	30,90861516
4	3344	3335	588	17,63118441
	21951	24586	4685	19,05556007
	24383	21403	4106	19,18422651
5	NR	NR	NR	NR
	3265	3149	1019	32,3594792
	12210	12911	2557	19,8048176
6	NR	NR	NR	NR
	NR	NR	NR	NR
	6410	6875	1480	21,52727273
7	130	95	18	18,94736842
	16049	14826	3371	22,7370835
	41351	41252	9485	22,99282459
8	NR	NR	NR	NR
	18668	17293	5187	29,99479558
	19278	19739	3315	16,79416384
9	NR	NR	NR	NR
	6850	6834	2027	29,66052092
	37493	34753	14490	41,69424222
10	NR	NR	NR	NR
	370	283	192	67,84452297
	26900	25558	12963	50,71993114
11	NR	NR	NR	NR
	1334	1171	417	35,61058924
	21179	20337	6396	31,45006638
12	NR	NR	NR	NR
	171	150	34	22,66666667
	7357	7276	1415	19,44749863
13	NR	NR	NR	NR
	1804	1545	407	26,34304207
	4997	4894	1488	30,40457703

Table SI 5 — Continuation.

ID	Total reads GeneMarker HTS	Inhouse-pipeline with PCR-Duplicates		
		Total reads inhouse-pipeline	Damage reads in house pipeline	% Damage
	NR	NR	NR	NR
14	915	815	181	22,20858896
	22233	21486	5123	23,84343293
	781	677	149	22,00886263
15	34169	33135	8075	24,37000151
	53862	53833	10662	19,80569539
	24	34	5	14,70588235
16	33	34	15	44,11764706
	483	494	160	32,38866397
	8649	8343	1134	13,59223301
17	39074	40020	7895	19,72763618
	48270	49755	8498	17,07969048
	NR	NR	NR	NR
18	5531	6149	1335	21,71084729
	36368	40642	10015	24,64199596
	NR	NR	NR	NR
19	163	208	120	57,69230769
	1965	2012	737	36,63021869
	NR	NR	NR	NR
20	17236	17296	10032	58,00185014
	42533	48856	33070	67,68871786
	NR	NR	NR	NR
21	3845	3501	807	23,05055698
	24692	25188	5472	21,72463078
	NR	NR	NR	NR
22	3732	3731	995	26,6684535
	16279	17052	4272	25,05277973
	NR	NR	NR	NR
23	23978	25520	5473	21,44592476
	51178	50719	14916	29,40909718
24	3019	3544	765	21,58577878
	5619	7063	1124	15,9139176
25	2595	2852	756	26,50771388
	41327	27492	14916	54,2557835
26	84	90	30	33,33333333
	382	430	27	6,279069767
27	44	40	3	7,5
	193	210	23	10,95238095
28	23833	24554	2689	10,95137249
	57089	57464	5319	9,256229987

Table SI 5 — Continuation.

ID	Total reads GeneMarker HTS	Inhouse-pipeline with PCR-Duplicates		
		Total reads inhouse-pipeline	Damage reads in house pipeline	% Damage
29	4081	4294	594	13,83325571
	37890	41155	9248	22,47114567
30	26181	25748	1819	7,064626379
	44896	41377	2198	5,312129927
31	NR	NR	NR	NR
	131564	133903	29248	21,84267716
32	15024	14666	3053	20,81685531
	59154	59865	12658	21,14424121
33	5020	990	299	30,2020202
	56021	6144	1904	30,98958333
34	977	1531	778	50,81645983
35	15367	16480	2753	16,70509709
36	126	132	15	11,36363636
37	2326	2578	576	22,34290147
38	15184	15954	2823	17,69462204
39	723	702	182	25,92592593
40	76938	79869	20907	26,17661421
41	19925	19543	1325	6,779921199

Table SI 6 — Comparison of the control region variants called by GeneMarker-HTS software and in-house pipeline

Sample ID	Chronological Age	NGS-Protocols	Polymorphisms — GeneMarker-HTS software	Polymorphisms — in-house pipeline	Total Number of variants called	Number of concordant variants	%Variants concordant	Number of discrepencies o variants no called NC	
								GeneMarker HTS	In house pipeline
1	8th century	M1	16069T 16126C 73G 185A 228A 263G 295T 462T 489C	16069T 16126C 73G 185A 228A 263G 295T 462T 489C	9	9	100,00%	0	0
		M2	16069T 16126C 73G 185A 228A NC 462T 489C	16069T 16126C 73G 185A 228A 263G 462T 489C	8	7	87,50%	1	0
		M3	16069T 16126C 73G 185A 228A 263G 462T 489C	16069T 16126C 73G 185A 228A 263G 462T 489C	8	8	100,00%	0	0
2	8th century	M1	NR	NR			NR		
		M2	16172C 16219G 73G 263G	16172C 16219G 73G 263G	4	4	100,00%	0	0
		M3	16172C 16219G 16278T 73G 263G	16172C 16219G 16278T 73G 263G	5	5	100,00%	0	0
3	8th century	M1	NR	NR			NR		
		M2	16051G 16183C 16189C 16356C 16519C 263G	16051G NC NC 16356C 16519C 263G	6	4	66,67%	0	2
		M3	16051G 16183C 16189C 16356C 16519C 263G	16051G 16183C 16189C 16356C 16519C 263G	6	6	100,00%	0	0
4	8th century	M1	LCR	LCR			LCR		
		M2	16172 C 16186T	16172 C 16186T	2	2	100,00%	0	0
		M3	16172 C 16186T NC	16172C 16186T 263G	3	2	66,67%	1	0
5	8th century	M1	NR	NR			NR		
		M2	16093C 16168T 16519C 263G	16093C 16168T 16519C 263G	4	4	100,00%	0	0
		M3	16093C 16519C 263G	16093C 16519C 263G	3	3	100,00%	0	0
6	8th century	M1	NR	NR			NR		
		M2	NR	NR			NR		
		M3	16274A 16304C 16311C 263G 456T	16274A 16304C 16311C 263G 456T	5	5	100,00%	0	0
7	8th century	M1	LCR	LCR			LCR		
		M2	16051G 16183C 16189C 16356C 16519C NC	16051G NC NC 16356C 16519C 263G	6	3	50,00%	1	2
		M3	16051G 16183C 16189C 16356C 16519C 263G	16051G 16183C 16189C 16356C 16519C 263G	6	6	100,00%	0	0
8	8th century	M1	NR	NR			NR		
		M2	16126C 16147T 16294T 16296T 16297C 16304C 16519C 73G 263G	16126C 16147T 16294T 16296T 16297C 16304C 16519C 73G 263G	9	9	100,00%	0	0
		M3	16126C 16147T 16294T 16296T 16297C 16304C 16519C 73G 263G	16126C 16147T 16294T 16296T 16297C 16304C 16519C 73G 263G	9	9	100,00%	0	0
9	8th century	M1	LCR	LCR			LCR		
		M2	16223T 16519C 73G 150t/C 152C 235G 494T	16223T 16519C 73G 150t/C 152C NC 494T	7	6	85,71%	0	1
		M3	16223T 16355T 16519C 73G 150T 152C 235G 263G 494T	16223T 16355T 16519C 73G 150T 152C 235G 263G 494T	9	9	100,00%	0	0
10	8th century	M1	NR	NR			NR		
		M2	16390 a/G	16390 a/G	1	1	100,00%	0	0
		M3	16069T 16126C 16390 a/G 16519C NC 152C 189G 195C NC 489C	16069T 16126C 16390 a/G 16519C 150t/C 152C 189G 195C 263G 489C	10	8	80,00%	2	0
11	8th century	M1	NR	NR			NR		
		M2	LCR	LCR			LCR		
		M3	16126C 16519C	NC 16519C	2	1	50,00%		1
12	8th century	M1	NR	NR			NR		
		M2	LCR	LCR			LCR		
		M3	NC 16519C 152C	16362C 16519C 152C	3	2	66,67%	1	0
13	8th century	M1	NR	NR			NR		
		M2	LCR	LCR			LCR		
		M3	16519C 263G	16519C 263G	2	2	100,00%	0	0

Table SI 6 — Continuation.

Sample ID	Chronological Age	NGS-Protocols	Polymorphisms — GeneMarker-HTS software										Polymorphisms — in-house pipeline										Total Number of variants called	Number of concordant variants	%Variants concordant	Number of discrepencies o variants no called NC	
																										GeneMarker HTS	In house pipeline
14	8th century	M1	NR										NR												NR		
		M2	LCR										LCR												LCR		
		M3	16069T	16126C	16519C	73G	NC	152C	189G	195C	NC	489C	16069T	16126C	16519C	73G	150t/C	152C	189G	195C	263G	489C	10	8	80,00%	2	0
15	8th century	M1	LCR										LCR												LCR		
		M2	16172C	16183C		16219G	73G	263G					NC	NC		NC	73G	NC					5	1	20,00%	0	4
		M3	16172C	16183C	16189C	16219G	73G	263G					16172C	16183C	16189C	NC	73G	263G					6	5	83,33%	0	1
16	8th century	M1	LCR										LCR												LCR		
		M2	LCR										LCR												LCR		
		M3	16391g/A	152C									16391g/A	NC									2	1	50,00%	0	1
17	8th century	M1			263G										263G								1	1	100,00%	0	0
		M2		72C	263G									72C	263G								2	2	100,00%	0	0
		M3	16298C	72C	263G								16298C	72C	263G								3	3	100,00%	0	0
18	8th century	M1	NR										NR												NR		
		M2	NC										146C										1	0	LCR	1	
		M3	NC										146C										1	0	LCR	1	
19	5500 BP	M1	NR										NR												NR		
		M2	LCR										LCR												LCR	0	0
		M3	16176T	NC	16390A	16519C	NC	152t/C	263G				16176T	16343a/G	16390A	16519C		150t/C	152t/C	263G			7	5	71,43%	2	0
20	5500 BP	M1	NR										NR												NR		
		M2	NC	16390A	16519C	73G	150T	263G						16343G	16390A	16519C	73G	150T	263G				6	6	100,00%	0	0
		M3	16343G	16390A	16519C	73G	150T	263G						16343G	16390A	16519C	73G	150T	263G				6	6	100,00%	0	0
21	5th century	M1	NR										NR												NR		
		M2	LCR										LCR												LCR		
		M3	16093C	16223T	16298C	16325C	16311C	16519t/C					16093C	16223T	16298C	NC	16311C	16519t/C					6	5	83,33%	0	1
22	13th century	M1	NR										NR												NR		
		M2		16311C		195C		497T						NC		195C		NC						3	1	33,33%	0	
		M3	16224C		16519C	195C	263G	497T					NC		16519C	195C	263G	497T						5	4	80,00%	0	
23	20th century	M1	NR										NR												NR		
		M2	16093C	16311C	16519C	263G							16093C	16311C	16519C	263G							4	4	100,00%	0	0
		M3	16093C	16311C	16519C	263G							16093C	16311C	16519C	263G							3	3	100,00%	0	0
24	20th century	M2	LCR										LCR												LCR		
		M3	16519C	263G									16519C	263G									2	2	100,00%	0	0
25	20th century	M2	16126C	16163G	16186T	16189C	16294T	16519C	73G	152C	195C	198T	263G	16126C	16163G	16186T	16189C	16294T	16519C	73G	152C	195C	198T	263G	11	11	100,00%	0	0
		M3	16126C	16163G	16186T	16189C	16294T	16519C	73G	152C	195C	198T	263G	16126C	16163G	16186T	16189C	16294T	16519C	73G	152C	195C	198T	263G	11	11	100,00%	0	0
26	13th century	M2	LCR										LCR												LCR		
		M3	LCR										LCR												LCR		
27	13th century	M2	LCR										LCR												LCR		
		M3	16519C	152C									16519C	152C									2	2	100,00%	0	0

Table SI 6 — Continuation.

Sample ID	Chronological Age	NGS-Protocols	Polymorphisms (GeneMarker-HTS software)	Polymorphisms (in-house pipeline)	Total Number of variants called	Number of concordant variants	%Variants concordant	Number of discrepencies o variants no called NC	
								GeneMarker HTS	In house pipeline
28	16th century	M2	16111T 16357C 16519C 263G	16111T 16357C 16519C 263G	4	4	100,00%		2
		M3	16111T 16357C 16519C 263G	16111T 16357C 16519C 263G	4	4	100,00%	0	0
29	20th century	M2	16051G 16129C 16183C 16189C 16362C 16519C 73G 152C 217C 263G 340T 508G	16051G 16129C 16183C 16189C 16362C 16519C 73G 152C 217C 263G NC 508G	12	11	91,67%	0	1
		M3	16051G 16129C 16183C 16189C 16362C 16519C 73G 152C 217C 263G 340T 508G	16051G 16129C 16183C 16189C 16362C 16519C 73G 152C 217C 263G 340T 508G	12	12	100,00%	0	0
30	20th century	M2	16093C 16519C 263G	16093C 16519C 263G	3	3	100,00%	0	0
		M3	16093C 16519C 263G	16093C 16519C 263G	3	3	100,00%	0	0
31	20th century	M2	NR	NR			NR		
		M3	16298C 72C 195C 263G	16298C 72C 195C 263G	4	4	100,00%	0	0
32	20th century	M2	16189C 16519C 263G	16189C 16519C 263G	3	3	100,00%	0	0
		M3	16189C 16519C 263G	16189C 16519C 263G	3	3	100,00%	0	0
33	20th century	M2	150T 228A 263G	150T 228A 263G	3	3	100,00%	0	0
		M3	16169A 16192T 16235G 16270T 16304C 73G 150T 228A 263G	16169A 16192T 16235G 16270T 16304C 73G 150T 228A 263G	9	9	100,00%	0	0
34	5-11th centuries	M3	LCR	LCR			LCR		
35	5-11th centuries	M3	16129C 16145A 16183C 16189C 16362C 16519C 73G 152C 217C 263G 340T 508G	16129C 16145A NC 16189C 16362C 16519C 73G 152C 217C 263G NC 508G	12	10	83,33%		2
36	5-11th centuries	M3	LCR	LCR			LCR		
37	5-11th centuries	M3	16519C 185A 188G 195C 228A 263G	16519C 185A 188G NC 228A NC	6	4	66,67%	0	2
38	5-11th centuries	M3	16293G 16311C 195C 263G	16293G 16311C 195C 263G	4	4	100,00%	0	0
39	5-11th centuries	M3	LCR	LCR			LCR		
40	19th century	M3	16111T 16187T 16223T 16290T 16325C 16362C 73G 245G 263G 489C	16111T 16187T 16223T 16290T 16325C 16362C 73G 245C 263G 489C	10	10	100,00%	0	0
41	19th century	M3	16111T 16187T 16223T 16290T 245G 263G 489C	16111T 16187T 16223T 16290T 245C 263G 489C	7	7	100,00%	0	0
					323	290	89,78%	12	23

Table SI 7 — Prediction of mtDNA haplogroup by Haplogrep 2.

Sample ID	Chronological Age	NGS-Protocol	Polymorphisms									Haplogroup	Prediction Quality (%)
1	8th century	M3	16069T	16126C	73G	185A	228A	263G	462T	489C		100	J1c
2	8th century	M3	16172C	16219G	16278T	73G	263G					100	U6a
3	8th century	M3	16051G	16183C	16189C	16356C	16519C	263G				89,25	H1b
4	8th century	M3	16172C	16186T								63,39	H2a2a1d
5	8th century	M3	16093C	16519C	263G							100	H1+16189*
6	8th century	M3	16274A	16304C	16311C	263G	456T					92,44	H5+16311
7	8th century	M3	16051G	16183C	16189C	16356C	16519C	263G				89,25	H1b
8	8th century	M3	16126C	16147T	16294T	16296T	16297C	16304C	16519C	73G	263G	100	T2b23
9	8th century	M3	16223T	16355T	16519C	73G	150T	152C	235G	263G	494T	100	L3k1
10	8th century	M3	16069T	16126C	16390R	16519C	152C	189G	195C	489C		84,7	J2a2b1a*
11	8th century	M3	16519C									0,5	H2a2a1
12	8th century	M3	16519C	152C								100	H2a2a2
13	8th century	M3	16519C	263G								100	H2a2a
14	8th century	M3	16069T	16126C	16519C	73G	152C	189G	195C	489C		86,64	J2a2b1a*
15	8th century	M3	16172C	16183C	16189C	73G	263G					100	U6a1a
16	8th century	M3	16391R										H5s
17	8th century	M3	16298C	72C	263G							100	HV0
19	5500 BP	M3	16176T	16390A	16519C	152Y	263G					89,23	H7a2
20	5500 BP	M3	16343G	16390A	16519C	73G	150T	263G				100	U3a
21	5th century	M3	16093C	16223T	16298C	16311C	16519Y					94,88	C1a*
22	13th century	M3	16519C	195C	263G	497T						94,26	K1a4a*2
23	20th century	M3	16093C	16311C	16519C	263G						100	H3h2
24	20th century	M3	16519C	263G								100	H2a2a
25	20th century	M3	16126C	16163G	16186T	16189C	16294T	16519C	73G	152C	195C	198T	95,38	T1a1'3
27	13th century	M3	16519C	152C								100	H2a2a2
28	16th century	M3	16111T	16357C	16519C	263G						87,57	H1af
29	20th century	M3	16051G	16129C	16183C	16189C	16362C	16519C	73G	152C	217C	263G	100	U2e1'2'3

Table SI 7 — Continuation.

Sample ID	Chronological Age	NGS-Protocol	Polymorphisms										Haplogroup	Prediction Quality (%)
30	20th century	M3	16093C	16519C	263G								100	H10+(16093)
31	20th century	M3	16298C	72C	195C	263G							100	HV0+195
32	20th century	M3	16189C	16519C	263G								100	H1+16189
33	20th century	M3	16169A	16192T	16235G	16270T	16304C	73G	150T	228A	263G		100	U5b3a1
35	5-11th centuries	M3	16129C	16145A	16189C	16362C	16519C	73G	152C	217C	263G	508G	100	U2e1e
37	5-11th centuries	M3	16519C	185A	188G	228A							100	J1c2
38	5-11th centuries	M3	16293G	16311C	195C	263G							100	H11a
40	19th century	M3	16111T	16187T	16223T	16290T	16325C	16362C	73G	245G	263G	489C	90,39	D1g
41	19th century	M3	16111T	16187T	16223T	16290T	245G	263G	489C				60	D1g4

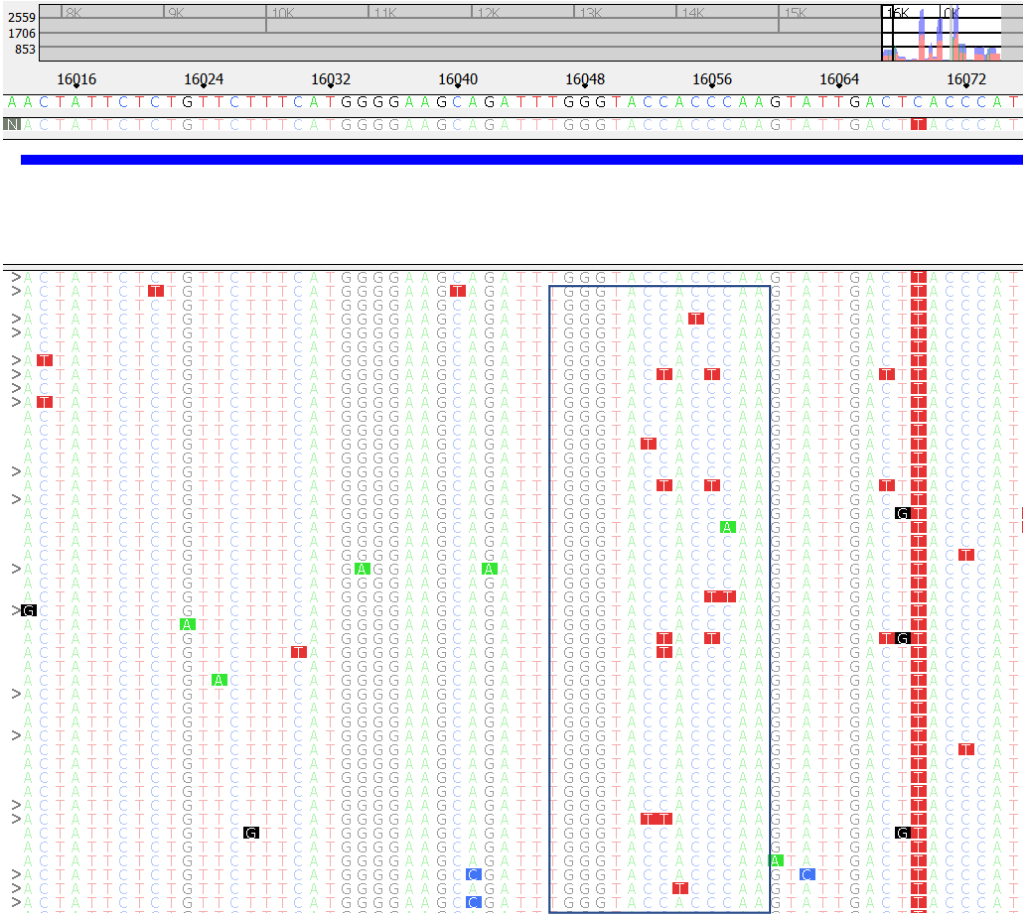


Figure SI 3 — Pile-up of GeneMarker HTS software: Molecular damage is remarked in the box where some miscoding lesions (nucleotide substitutions from C to T) are shown in sequences from degraded DNA from Sample ID 5 in the 16048-16056 region.

7

References

Abu-Mandil Hassan, N., Brown, K. A., Eyers, J., Brown, T. A., & Mays, S. (2014). Ancient DNA study of the remains of putative infanticide victims from the Yewden Roman villa site at Hambleden, England. Journal of Archaeological Science, 43, 192–197. https://doi.org/https://doi.org/10.1016/j.jas.2013.12.017.

Adler, C. J., Haak, W., Donlon, D., & Cooper, A. (2011a). Survival and recovery of DNA from ancient teeth and bones. Journal of Archaeological Science, 38(5), 956–964. https://doi.org/10.1016/j.jas.2010.11.010.

Adler, C. J., Haak, W., Donlon, D., & Cooper, A. (2011b). Survival and recovery of DNA from ancient teeth and bones. Journal of Archaeological Science, 38(5), 956–964. https://doi.org/10.1016/j.jas.2010.11.010.

Adler, C. J., Haak, W., Donlon, D., & Cooper, A. (2011c). Survival and recovery of DNA from ancient teeth and bones. Journal of Archaeological Science. https://doi.org/10.1016/j.jas.2010.11.010.

Afonso, C., Nociarova, D., Santos, C., Martinez-Labarga, C., Mestres, I., Duran, M., & Malgosa, A. (2019). Sex selection in late Iberian infant burials: Integrating evidence from morphological and genetic data. American Journal of Human Biology, 31(1), e23204. https://doi.org/https://doi.org/10.1002/ajhb.23204.

Alaeddini, R., Walsh, S. J., & Abbas, A. (2010a). Forensic implications of genetic analyses from degraded DNA--a review. Forensic Science International. Genetics, 4(3), 148–157. https://doi.org/10.1016/j.fsigen.2009.09.007.

Alaeddini, R., Walsh, S. J., & Abbas, A. (2010b). Forensic implications of genetic analyses from degraded DNA-A review. Forensic Science International: Genetics, 4(3), 148–157. https://doi.org/10.1016/j.fsigen.2009.09.007.

Alberti, F., Gonzalez, J., Paijmans, J. L. A., Basler, N., Preick, M., Henneberger, K., Trinks, A., Rabeder, G., Conard, N. J., Münzel, S. C., Joger, U., Fritsch, G., Hildebrandt, T., Hofreiter, M., & Barlow, A. (2018). Optimized DNA sampling of ancient bones using Computed Tomography scans. Molecular Ecology Resources, 18(6), 1196–1208. https://doi.org/https://doi.org/10.1111/1755-0998.12911.

Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M., & Gerstein, M. B. (2010). Annotating non-coding regions of the genome. Nature Reviews Genetics, 11(8), 559–571. https://doi.org/10.1038/nrg2814.

Alford, A. I., Kozloff, K. M., & Hankenson, K. D. (2015). Extracellular matrix networks in bone remodeling. The International Journal of Biochemistry & Cell Biology, 65, 20–31. https://doi.org/https://doi.org/10.1016/j.biocel.2015.05.008.

Allard, M. W., Polanskey, D., Miller, K., Wilson, M. R., Monson, K. L., & Budowle, B. (2005). Characterization of human control region sequences of the African American SWGDAM forensic mtDNA data set. Forensic Science International, 148(2), 169–179. https://doi.org/https://doi.org/10.1016/j.forsciint.2004.06.001.

Allard, M. W., Wilson, M. R., Monson, K. L., & Budowle, B. (2004). Control region sequences for East Asian individuals in the Scientific Working Group on DNA Analysis Methods forensic mtDNA data set. Legal Medicine, 6(1), 11–24. https://doi.org/https://doi.org/10.1016/j.legalmed.2003.08.004.

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., Samaniego, J. A., Gilbert, T. P. M., Willerslev, E., Zhang, G., Scofield, R. P., Holdaway, R. N., & Bunce, M. (2012). The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B: Biological Sciences, 279(1748), 4724–4733. https://doi.org/10.1098/rspb.2012.1745.

Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. Della, Dąbrowski, P., … Willerslev, E. (2015). Population genomics of Bronze Age Eurasia. Nature, 522, 167. https://doi.org/10.1038/nature14507.

Alonso, A., Albarrán, C., Martín, P., García, P., García, O., de la Rúa, C., Alzualde, A., Fernández de Simón, L., Sancho, M., & Fernández Piqueras, J. (2003). Multiplex–PCR of short amplicons for mtDNA sequencing from ancient DNA. International Congress Series, 1239, 585–588. https://doi.org/https://doi.org/10.1016/S0531-5131(02)00401-6.

Aly, S., Mostafa, E., & Bolbol, D. (2015). Overviews of "next-generation sequencing." Research and Reports in Forensic Medical Science, 1. https://doi.org/10.2147/RRFMS.S57998.

Amory, S., Huel, R., Bilić, A., Loreille, O., & Parsons, T. J. (2012). Automatable full demineralization DNA extraction procedure from degraded skeletal remains. Forensic Science International: Genetics, 6(3), 398–406. https://doi.org/10.1016/j.fsigen.2011.08.004.

Anders, E., & Andrea, M. (2012). Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. Proceedings of the National Academy of Sciences, 109(35), 13956–13960. https://doi.org/10.1073/pnas.1200567109.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R., & Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. Nature, 290(5806), 457–465. https://doi.org/10.1038/290457a0.

Andrews, R., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics, 23(2), 147. https://doi.org/10.1038/13779.

Andrews, S. (2010). FASTQC. A quality control tool for high throughput sequence data.

Andronowski, J. M., Mundorff, A. Z., Pratt, I. V, Davoren, J. M., & Cooper, D. M. L. (2017). Evaluating differential nuclear DNA yield rates and osteocyte numbers among human bone tissue types: A synchrotron radiation micro-CT approach. Forensic Science International. Genetics, 28, 211–218. https://doi.org/10.1016/j.fsigen.2017.03.002

Archibald, J. M. (2015). Endosymbiosis and Eukaryotic Cell Evolution. Current Biology, 25(19), R911–R921. https://doi.org/https://doi.org/10.1016/j.cub.2015.07.055.

Bandelt, H. J., Alves-Silva, J., Guimaraes, P. E. M., Santos, M. S., Brehm, A., PEREIRA, L., Coppa, A., Larruga, J. M., Rengo, C., Scozzari, R., Torroni, A., Prata, M. J., Amorim, A., Prado, V. F., & Pena, S. D. J. (2001). Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. Annals of Human Genetics, 65(6), 549–563. https://doi.org/https://doi.org/10.1046/j.1469-1809.2001.6560549.x.

Barta, J. L., Monroe, C., Crockford, S. J., & Kemp, B. M. (2014). Mitochondrial DNA preservation across 3000-year-old northern fur seal ribs is not related to bone density: Implications for forensic investigations. Forensic Science International, 239, 11–18. https://doi.org/10.1016/j.forsciint.2014.02.029.

Behar, D. M., van Oven, M., Rosset, S., Metspalu, M., Loogväli, E.-L., Silva, N. M., Kivisild, T., Torroni, A., & Villems, R. (2012). A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. The American Journal of Human Genetics, 90(4), 675–684. https://doi.org/https://doi.org/10.1016/j.ajhg.2012.03.002.

Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., Bertranpetit, J., Quintana-Murci, L., Tyler-Smith, C., Wells, R. S., & Rosset, S. (2008). The Dawn of Human Matrilineal Diversity. The American Journal of Human Genetics, 82(5), 1130–1140. https://doi.org/10.1016/j.ajhg.2008.04.002.

Bendall, K. E., Macaulay, V. A., Baker, J. R., & Sykes, B. C. (1996). Heteroplasmic point mutations in the human mtDNA control region. American Journal of Human Genetics, 59(6), 1276–1287. https://pubmed.ncbi.nlm.nih.gov/8940273.

Bendall, K. E., & Sykes, B. C. (1995). Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. American Journal of Human Genetics, 57(2), 248–256.

Bertranpetit, J., Sala, J., Calafell, F., Underhill, P. A., Moral, P., & Comas, D. (1995). Human mitochondrial DNA variation and the origin of Basques. Annals of Human Genetics, 59(1), 63–81. https://doi.org/10.1111/j.1469-1809.1995.tb01606.x.

Birky, C. W. J. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proceedings of the National Academy of Sciences of the United States of America, 92(25), 11331–11338. https://doi.org/10.1073/pnas.92.25.11331.

Bodner, M., Bastisch, I., Butler, J. M., Fimmers, R., Gill, P., Gusmão, L., Morling, N., Phillips, C., Prinz, M., Schneider, P. M., & Parson, W. (2016). Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). Forensic Science International: Genetics, 24, 97–102. https://doi.org/10.1016/J.FSIGEN.2016.06.008.

Børsting, C., & Morling, N. (2015). Next generation sequencing and its applications in forensic genetics. Forensic Science International: Genetics, 18, 78–89. https://doi.org/10.1016/j.fsigen.2015.02.002.

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics, 32(3), 314–331. https://pubmed.ncbi.nlm.nih.gov/6247908.

Bower, M. A., & Jones, M. K. (2010). DNA: Ancient. In Encyclopedia of Archaeology (pp. 1094–1101). https://doi.org/10.1016/B978-012373962-9.00196-5.

Brandhagen, M. D., Just, R. S., & Irwin, J. A. (2020). Validation of NGS for mitochondrial DNA casework at the FBI Laboratory. Forensic Science International: Genetics, 44, 102151. https://doi.org/https://doi.org/10.1016/j.fsigen.2019.102151.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Research, 38(6), e87. https://doi.org/10.1093/nar/gkp1163.

Brookes, A. J. (1999). The essence of SNPs. Gene, 234(2), 177–186. https://doi.org/https://doi.org/10.1016/S0378-1119(99)00219-X.

Brown, T. (2002). The Human Genome. In Genomes. (2nd editio). https://www.ncbi.nlm.nih.gov/books/NBK21134/#.

Bruder, C. E. G., Piotrowski, A., Gijsbers, A. A. C. J., Andersson, R., Erickson, S., Diaz de Ståhl, T., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A., Crowley, M., Crasto, C., Partridge, E. C., Tiwari, H., Allison, D. B., Komorowski, J., van Ommen, G.-J. B., Boomsma, D. I., Pedersen, N. L., … Dumanski, J. P. (2008). Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles. The American Journal of Human Genetics, 82(3), 763–771. https://doi.org/10.1016/j.ajhg.2007.12.011.

Budowle, B, Moretti, T. R., Baumstark, A. L., Defenbaugh, D. A., & Keys, K. M. (1999). Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. Journal of Forensic Sciences, 44(6), 1277–1286.

Budowle, Bruce, Allard, M. W., & Wilson, M. R. (2002). Critique of interpretation of high levels of heteroplasmy in the human mitochondrial DNA hypervariable region I from hair. Forensic Science International, 126(1), 30–33. https://doi.org/10.1016/s0379-0738(02)00019-1.

Budowle, Bruce, Allard, M. W., Wilson, M. R., & Chakraborty, R. (2003). Forensics and Mitochondrial DNA: Applications, Debates, and Foundations. Annual Review of Genomics and Human Genetics, 4(1), 119–141. https://doi.org/10.1146/annurev.genom.4.070802.110352.

Bulbul, O., Pakstis, A. J., Soundararajan, U., Gurkan, C., Brissenden, J. E., Roscoe, J. M., Evsanaa, B., Togtokh, A., Paschou, P., Grigorenko, E. L., Gurwitz, D., Wootton, S., Lagace, R., Chang, J., Speed, W. C., & Kidd, K. K. (2018). Ancestry inference of 96 population samples using microhaplotypes. International Journal of Legal Medicine, 132(3), 703–711. https://doi.org/10.1007/s00414-017-1748-6.

Burger, J., Kirchner, M., Bramanti, B., Haak, W., & Thomas, M. G. (2007). Absence of the lactase-persistence-associated allele in early Neolithic Europeans. Proceedings of the National Academy of Sciences, 104(10), 3736–3741. https://doi.org/10.1073/pnas.0607187104.

Butler, J. M. (2012). Advanced Topics in Forensic DNA Typing: Methodology (1st edition).

Butler, J. M. (2015). The future of forensic DNA analysis. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1674), 20140252. https://doi.org/10.1098/rstb.2014.0252.

Butler, J. M., & Levin, B. C. (1998). Forensic applications of mitochondrial DNA. Trends in Biotechnology, 16(4), 158–162. https://doi.org/https://doi.org/10.1016/S0167-7799(98)01173-1.

Čakar, J., Džehverović, M., & Pilav, A. (2020). DNA analysis of thirty-eight years old stillborn's skeletal remains in case of disputed maternity. Forensic Science International: Genetics, 47. https://doi.org/10.1016/j.fsigen.2020.102294.

Calacal, G. C., Apaga, D. L. T., Salvador, J. M., Jimenez, J. A. D., Lagat, L. J., Villacorta, R. P. F., Lim, M. C. F., Fortun, R. d. R., Datar, F. A., & De Ungria, M. C. A. (2015). Comparing different post-mortem human samples as DNA sources for downstream genotyping and identification. Forensic Science International: Genetics, 19, 212–220. https://www.sciencedirect.com/science/article/pii/S187249731530051X?via%3Dihub.

Capelli, C., Tschentscher, F., & Pascali, V. L. (2003). "Ancient" protocols for the crime scene?: Similarities and differences between forensic genetics and ancient DNA analysis. Forensic Science International, 131(1), 59–64. https://doi.org/10.1016/S0379-0738(02)00396-1.

Caputo, M., Irisarri, M., Alechine, E., & Corach, D. (2013a). A DNA extraction method of small quantities of bone for high-quality genotyping. Forensic Science International: Genetics, 7(5), 488–493. https://doi.org/10.1016/j.fsigen.2013.05.002.

Caputo, M., Irisarri, M., Alechine, E., & Corach, D. (2013b). A DNA extraction method of small quantities of bone for high-quality genotyping. Forensic Science International: Genetics, 7(5), 488–493. https://doi.org/10.1016/J.FSIGEN.2013.05.002.

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., Sicheritz-Pontén, T., & Gilbert, M. T. P. (2018). Single-tube library preparation for degraded DNA. Methods in Ecology and Evolution, 9(2), 410–419. https://doi.org/https://doi.org/10.1111/2041-210X.12871.

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova,

D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., … Bustamante, C. D. (2013). Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. American Journal of Human Genetics, 93(5), 852–864. https://doi.org/10.1016/j.ajhg.2013.10.002.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics (Oxford, England), 34(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Chen, Y.-S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K., & Wallace, D. C. (2000). mtDNA Variation in the South African Kung and Khwe—and Their Genetic Relationships to Other African Populations. The American Journal of Human Genetics, 66(4), 1362–1383. https://doi.org/10.1086/302848.

Chiego, J. (1994). Histology of Pulp. Oral Development and Histology, 262–281. http://ci.nii.ac.jp/naid/10008396357/en/.

Coble, M. D., Loreille, O. M., Wadhams, M. J., Edson, S. M., Maynard, K., Meyer, C. E., Niederstätter, H., Berger, C., Berger, B., Falsetti, A. B., Gill, P., Parson, W., & Finelli, L. N. (2009). Mystery Solved: The Identification of the Two Missing Romanov Children Using DNA Analysis. PLOS ONE, 4(3), e4838. https://doi.org/10.1371/journal.pone.0004838.

Collins, M. J., Nielsen–Marsh, C. M., Hiller, J., Smith, C. I., Roberts, J. P., Prigodich, R. V, Wess, T. J., Csapò, J., Millard, A. R., & Turner–Walker, G. (2002). The survival of organic matter in bone: a review. Archaeometry, 44(3), 383–394. https://doi.org/10.1111/1475-4754.t01-1-00071.

Condemi, S., Mounier, A., Giunti, P., Lari, M., Caramelli, D., & Longo, L. (2014). Correction: Possible Interbreeding in Late Italian Neanderthals? New Data from the Mezzena Jaw (Monti Lessini, Verona, Italy). PLOS ONE, 9(1), 10.1371/annotation/b8ec29b1-e71f-4148-baad-1441fca. https://doi.org/10.1371/annotation/b8ec29b1-e71f-4148-baad-1441fca03ed9.

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: do it right or not at all. In Science (New York, N.Y.) (Vol. 289, Issue 5482, p. 1139). https://doi.org/10.1126/science.289.5482.1139b.

Court, D. S. (2021). Mitochondrial dna in forensic use. Emerging Topics in Life Sciences, 5(3), 415–426. https://doi.org/10.1042/ETLS20210204.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M., Weihmann, A., & Nickel, B. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. 1–6. https://doi.org/10.1073/pnas.1314445110.

Dabney, J., & Meyer, M. (2019). Extraction of Highly Degraded DNA from Ancient Bones and Teeth. Methods in Molecular Biology (Clifton, N.J.), 1963, 25–29. https://doi.org/10.1007/978-1-4939-9176-1_4.

Dabney, J., Meyer, M., & Paabo, S. (2013). Ancient DNA Damage. Cold Spring Harbor Perspectives in Biology, 5(7), a012567–a012567. https://doi.org/10.1101/cshperspect.a012567.

Damgaard, P. B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., & Allentoft, M. E. (2015). Improving access to endogenous DNA in ancient bones and teeth. Scientific Reports, 5, 1–12. https://doi.org/10.1038/srep11184.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience, 10(2). https://doi.org/10.1093/gigascience/giab008.

Davoren, J., Vanek, D., Konjhodzic, R., Crews, J., Huffine, E., & Parsons, T. J. (2007). Highly effective DNA extraction method for nuclear short tandem repeat testing of skeletal remains from mass graves. Croat Med J, 48(4), 478–485. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17696302.

Desmyter, S., De Cock, G., Moulin, S., & Noël, F. (2017). Organic extraction of bone lysates improves DNA purification with silica beads. Forensic Science International, 273, 96–101. https://doi.org/10.1016/J.FORSCIINT.2017.02.003.

Diroma, M. A., Modi, A., Lari, M., Sineo, L., Caramelli, D., & Vai, S. (2021). New Insights Into Mitochondrial DNA Reconstruction and Variant Detection in Ancient Samples. Frontiers in Genetics, 12. https://doi.org/10.3389/fgene.2021.619950.

Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O., & Raoult, D. (1998). Detection of 400-year-old Yersinia pestis DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. Proceedings of the National Academy of Sciences, 95(21), 12637 LP – 12640. https://doi.org/10.1073/pnas.95.21.12637.

Edson, S. M. (2019). Extraction of DNA from Skeletonized Postcranial Remains: A Discussion of Protocols and Testing Modalities. Journal of Forensic Sciences, 64(5), 1312–1323. https://doi.org/10.1111/1556-4029.14050.

Edwards, M. C., & Gibbs, R. A. (1994). Multiplex PCR: advantages, development, and applications. Genome Research, 3(4), S65–S75. http://genome.cshlp.org/content/3/4/S65.short.

Eichmann, C., & Parson, W. (2008). 'Mitominis': multiplex PCR analysis of reduced size amplicons for compound sequence analysis of the entire mtDNA control region in highly degraded samples. International Journal of Legal Medicine, 122(5), 385–388. https://doi.org/10.1007/s00414-008-0227-5.

ENFSI. (2012). DNA Database Management. Review and Reccommendation. April, 1–85.

Ermini, L., Der Sarkissian, C., Willerslev, E., & Orlando, L. (2015). Major transitions in human evolution revisited: a tribute to ancient DNA. Journal of Human Evolution, 79, 4–20. https://doi.org/10.1016/j.jhevol.2014.06.015.

Faerman, M., Bar-Gal, G. K., Filon, D., Greenblatt, C. L., Stager, L., Oppenheim, A., & Smith, P. (1998). Determining the Sex of Infanticide Victims from the Late Roman Era through Ancient DNA Analysis. Journal of Archaeological Science, 25(9), 861–865. https://doi.org/https://doi.org/10.1006/jasc.1997.0231.

Ferreira, S. T. G., Paula, K. A., Maia, F. A., & Moraes, A. V. (2013). A comparative study between two protocols for DNA extraction from bones. Forensic Science International: Genetics Supplement Series, 4(1), 374–375. https://doi.org/10.1016/j.fsigss.2013.10.190.

Frantz, L. A. F., Bradley, D. G., Larson, G., & Orlando, L. (2020). Animal domestication in the era of ancient genomics. Nature Reviews. Genetics, 21(8), 449–460. https://doi.org/10.1038/s41576-020-0225-0.

Fu, Q., Rudan, P., Pääbo, S., & Krause, J. (2012). Complete Mitochondrial Genomes Reveal Neolithic Expansion into Europe. PLOS ONE, 7(3), e32473. https://doi.org/10.1371/journal.pone.0032473.

Fulton, T. L., & Shapiro, B. (2019). Setting Up an Ancient DNA Laboratory BT - Ancient DNA: Methods and Protocols (B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans, & A. E. R. Soares (eds.); pp. 1–13). Springer New York. https://doi.org/10.1007/978-1-4939-9176-1_1.

Gabriel, M., Huffine, E., Ryan, J., Holland, M., & Parsons, T. (2001). Improved MtDNA Sequence Analysis of Forensic Remains Using a “Mini-Primer Set” Amplification Strategy. Journal of Forensic Sciences, 46(2), 247–253. https://doi.org/10.1520/JFS14957J.

Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-Rasheid, K. A. S., Bradley, D. G., & Orlando, L. (2016). Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. Molecular Ecology Resources, 16(2), 459–469. https://doi.org/10.1111/1755-0998.12470.

Gamba, C., Jones, E. R., Teasdale, M. D., McLaughlin, R. L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kővári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T. F. G., Hofreiter, M., Bradley, D. G., & Pinhasi, R. (2014). Genome flux and stasis in a five millennium transect of European prehistory. Nature Communications, 5(1), 5257. https://doi.org/10.1038/ncomms6257.

Gansauge, M.-T., Aximu-Petri, A., Nagel, S., & Meyer, M. (2020). Manual and automated preparation of single-stranded DNA libraries for the sequencing of DNA from ancient biological remains and other sources of highly degraded DNA. Nature Protocols, 15(8), 2279–2300. https://doi.org/10.1038/s41596-020-0338-0.

Gansauge, M.-T., & Meyer, M. (2014). Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Research, 24(9), 1543–1549. https://doi.org/10.1101/gr.174201.114.

Gansauge, M. T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., Riehl, L. M., Schmidt, A., & Meyer, M. (2017). Single-stranded DNA library preparation from

highly degraded DNA using T4 DNA ligase. Nucleic Acids Research, 45(10). https://doi.org/10.1093/nar/gkx033.

Gaudio, D., Fernandes, D. M., Schmidt, R., Cheronet, O., Mazzarelli, D., Mattia, M., O'Keeffe, T., Feeney, R. N. M., Cattaneo, C., & Pinhasi, R. (2019). Genome-Wide DNA from Degraded Petrous Bones and the Assessment of Sex and Probable Geographic Origins of Forensic Cases. Scientific Reports, 9(1), 8226. https://doi.org/10.1038/s41598-019-44638-w.

Gettings, K. B., Kiesler, K. M., & Vallone, P. M. (2015). Performance of a next generation sequencing SNP assay on degraded DNA. Forensic Science International: Genetics, 19. https://doi.org/10.1016/j.fsigen.2015.04.010.

Gilbert, M. T. P., Bandelt, H.-J., Hofreiter, M., & Barnes, I. (2005). Assessing ancient DNA studies. Trends in Ecology & Evolution, 20(10), 541–544. https://doi.org/https://doi.org/10.1016/j.tree.2005.07.005.

Gilbert, M. T. P., Willerslev, E., Hansen, A. J., Barnes, I., Rudbeck, L., Lynnerup, N., & Cooper, A. (2003). Distribution patterns of postmortem damage in human mitochondrial DNA. American Journal of Human Genetics, 72(1), 32–47. https://doi.org/10.1086/345378.

Gilbert, M. T. P., Wilson, A. S., Bunce, M., Hansen, A. J., Willerslev, E., Shapiro, B., Higham, T. F. G., Richards, M. P., O'Connell, T. C., Tobin, D. J., Janaway, R. C., & Cooper, A. (2004). Ancient mitochondrial DNA from hair. Current Biology, 14(12), R463–R464. https://doi.org/https://doi.org/10.1016/j.cub.2004.06.008.

Gilder, J. R., Doom, T. E., Inman, K., & Krane, D. E. (2007). Run-specific limits of detection and quantitation for STR-based DNA testing. Journal of Forensic Sciences, 52(1), 97–101. https://doi.org/10.1111/j.1556-4029.2006.00318.x.

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. Proceedings of the National Academy of Sciences of the United States of America, 77(11), 6715–6719. https://doi.org/10.1073/pnas.77.11.6715.

Gill, P. (2001). Application of low copy number DNA profiling. Croatian Medical Journal, 42(3), 229–232.

Gill, P., Ivanov, P. L., Kimpton, C., Piercy, R., Benson, N., Tully, G., Evett, I., Hagelberg, E., & Sullivan, K. (1994). Identification of the remains of the Romanov family by DNA analysis. Nature Genetics, 6(2), 130–135. https://doi.org/10.1038/ng0294-130.

Gill, P., Jeffreys, A. J., & Werrett, D. J. (1985). Forensic application of DNA 'fingerprints.' Nature, 318(6046), 577–579. https://doi.org/10.1038/318577a0.

Ginther, C., Issel-Tarver, L., & King, M.-C. (1992). Identifying individuals by sequencing mitochondrial DNA from teeth. Nature Genetics, 2(2), 135–138. https://doi.org/10.1038/ng1092-135.

Glocke, I., & Meyer, M. (2017). Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. 1230–1237. https://doi.org/10.1101/gr.219675.116.1230.

Golenberg, E. M., Bickel, A., & Weihs, P. (1996). Effect of Highly Fragmented DNA on PCR. Nucleic Acids Research, 24(24), 5026–5033. https://doi.org/10.1093/nar/24.24.5026.

Gonder, M. K., Mortensen, H. M., Reed, F. A., de Sousa, A., & Tishkoff, S. A. (2007). Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages. Molecular Biology and Evolution, 24(3), 757–768. https://doi.org/10.1093/molbev/msl209.

Gonzalez, A., Cannet, C., Zvénigorosky, V., Geraut, A., Koch, G., Delabarde, T., Ludes, B., Raul, J.-S., & Keyser, C. (2020). The petrous bone: Ideal substrate in legal medicine? Forensic Science International: Genetics, 47. https://doi.org/10.1016/j.fsigen.2020.102305.

González, M. del M., Ramos, A., Aluja, M. P., & Santos, C. (2020). Sensitivity of mitochondrial DNA heteroplasmy detection using Next Generation Sequencing. Mitochondrion, 50, 88–93. https://doi.org/10.1016/j.mito.2019.10.006.

Götherström, A., Collins, M. J., Angerbjörn, A., & Lidén, K. (2002). Bone preservation and DNA amplification. Archaeometry, 44(3), 395–404. https://doi.org/10.1111/1475-4754.00072.

Greenberg, B. D., Newbold, J. E., & Sugino, A. (1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene, 21(1), 33–49. https://doi.org/https://doi.org/10.1016/0378-1119(83)90145-2.

Grisedale, K. S., & van Daal, A. (2012). Comparison of STR profiling from low template DNA extracts with and without the consensus profiling method. Investigative Genetics, 3(1), 1. https://doi.org/10.1186/2041-2223-3-14.

Guillén, M., Lareu, M. V., Pestoni, C., Salas, A., & Carracedo, A. (2000). Ethical-legal problems of DNA databases in criminal investigation. Journal of Medical Ethics, 26(4), 266 LP – 271. https://doi.org/10.1136/jme.26.4.266.

Gutiérrez, G., & Marín, A. (1998). The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. Molecular Biology and Evolution, 15(7), 926–929. https://doi.org/10.1093/oxfordjournals.molbev.a025998.

Habelberg, E., Sykes, B., & Hedges, R. (1989). Ancient bone DNA amplified. Nature, 342(6249), 485. https://doi.org/10.1038/342485a0.

Hagelberg, E., & Clegg, J. B. (1991). Isolation and characterization of DNA from archaeological bone. Proceedings of the Royal Society B: Biological Sciences, 244(1309), 45–50. https://doi.org/10.1098/rspb.1991.0049.

Hall, T. (2001). BioEdit 5.0.6. http://www.mbio.ncsu.edu/BioEdit/BioDoc.pdf

Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., & Allentoft, M. E. (2017a). Comparing ancient DNA preservation in petrous bone and tooth cementum. PLoS ONE, 12(1), 1–18. https://doi.org/10.1371/journal.pone.0170940.

Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., & Allentoft, M. E. (2017b). Comparing ancient DNA preservation in petrous bone and tooth cementum. PLoS ONE, 12(1), 1–18. https://doi.org/10.1371/journal.pone.0170940.

Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., & Allentoft, M. E. (2017c). Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. 1–18. https://doi.org/10.1371/journal.pone.0170940.

Harrel, M., Mayes, C., Gangitano, D., & Hughes-Stamm, S. (2018). Evaluation Of A Powder-Free DNA Extraction Method For Skeletal Remains. Journal of Forensic Sciences, 63(6), 1819–1823. https://doi.org/10.1111/1556-4029.13749.

Hasan, M. M., Hossain, T., Majumder, A., Momtaz, P., Sharmin, T., Sufian, A., & Akhteruzzaman, S. (2014). An efficient DNA extraction method from bone and tooth samples by complete demineralization followed by the use of silica-based columns. Dhaka University Journal of Biological Sciences, 23(2 SE-Articles). https://doi.org/10.3329/dujbs.v23i2.20089.

Hatsch, D., Amory, S., Keyser-Tracqui, C., Hienne, R., Crubézy, E., & Ludes, B. (2006). High-throughput mitochondrial DNA cloning in forensic and anthropological studies. International Congress Series, 1288, 118–120. https://doi.org/https://doi.org/10.1016/j.ics.2005.09.096.

Hecht, N. B., Liem, H., Kleene, K. C., Distel, R. J., & Ho, S. (1984). Maternal inheritance of the mouse mitochondrial genome is not mediated by a loss or gross alteration of the paternal mitochondrial DNA or by methylation of the oocyte mitochondrial DNA. Developmental Biology, 102(2), 452–461. https://doi.org/https://doi.org/10.1016/0012-1606(84)90210-0.

Higgins, D., & Austin, J. J. (2013). Teeth as a source of DNA for forensic identification of human remains: A Review. Science & Justice, 53(4), 433–441. https://doi.org/10.1016/j.scijus.2013.06.001.

Higgins, D., Kaidonis, J., Townsend, G., Hughes, T., & Austin, J. J. (2013). Targeted sampling of cementum for recovery of nuclear DNA from human teeth and the impact of common decontamination measures. Investigative Genetics, 4(1), 18. https://doi.org/10.1186/2041-2223-4-18.

Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. A., & Wilson, A. C. (1984). DNA sequences from the quagga, an extinct member of the horse family. Nature, 312(5991), 282–284. https://doi.org/10.1038/312282a0.

Hoeijmakers, J. H. J. (2009). DNA Damage, Aging, and Cancer. New England Journal of Medicine, 361(15), 1475–1485. https://doi.org/10.1056/NEJMra0804615.

Hofreiter, M. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Research, 29(23), 4793–4799. https://doi.org/10.1093/nar/29.23.4793.

Hofreiter, Michael. (2008). Long DNA sequences and large data sets: investigating the Quaternary via ancient DNA. Quaternary Science Reviews, 27(27), 2586–2592. https://doi.org/https://doi.org/10.1016/j.quascirev.2008.09.012.

Hofreiter, Michael, Paijmans, J. L. A., Goodchild, H., Speller, C. F., Barlow, A., Fortes, G. G., Thomas, J. A., Ludwig, A., & Collins, M. J. (2015). The future of ancient DNA: Technical advances and conceptual shifts. BioEssays, 37(3), 284–293. https://doi.org/10.1002/bies.201400160.

Hofreiter, Michael, Sneberger, J., Pospisek, M., & Vanek, D. (2021). Progress in forensic bone DNA analysis: Lessons learned from ancient DNA. Forensic Science International. Genetics, 54, 102538. https://doi.org/10.1016/j.fsigen.2021.102538.

Holland, C. A., McElhoe, J. A., Gaston-Sanchez, S., & Holland, M. M. (2021). Damage patterns observed in mtDNA control region MPS data for a range of template concentrations and when using different amplification approaches. International Journal of Legal Medicine, 135(1), 91–106. https://doi.org/10.1007/s00414-020-02410-0.

Holland, M. M., Pack, E. D., & Mcelhoe, J. A. (2017). Forensic Science International : Genetics Evaluation of GeneMarker 1 HTS for improved alignment of mtDNA MPS data , haplotype determination , and heteroplasmy assessment. Forensic Science International: Genetics, 28, 90–98. https://doi.org/10.1016/j.fsigen.2017.01.016.

Holland, M. M., Pack, E. D., & McElhoe, J. A. (2017). Evaluation of GeneMarker® HTS for improved alignment of mtDNA MPS data, haplotype determination, and heteroplasmy assessment. Forensic Science International: Genetics, 28, 90–98. https://doi.org/10.1016/j.fsigen.2017.01.016.

Holland, M., & McElhoe, J. (2015). A custom software solution for forensic mtDNA analysis of MiSeq data. Forensic Science International: Genetics Supplement Series, 5, e614–e616. https://doi.org/10.1016/j.fsigss.2015.09.242.

Holt, C. L., Stephens, K. M., Walichiewicz, P., Fleming, K. D., Forouzmand, E., & Wu, S.-F. (2021). Human Mitochondrial Control Region and mtGenome: Design and Forensic Validation of NGS Multiplexes, Sequencing and Analytical Software. In Genes (Vol. 12, Issue 4). https://doi.org/10.3390/genes12040599.

Hong, S. B., Kim, Y., & Park, J. H. (2017). High-efficiency automated DNA extraction method for degraded old skeletal samples. Forensic Science International: Genetics Supplement Series, 6(September), e365–e367. https://doi.org/10.1016/j.fsigss.2017.09.108.

Hoogenboom, J., Sijen, T., & van der Gaag, K. J. (2021). STRNaming: Generating simple, informative names for sequenced STR alleles in a standardised and automated manner. Forensic Science International: Genetics, 52. https://doi.org/10.1016/j.fsigen.2021.102473.

Hoogenboom, J., van der Gaag, K. J., de Leeuw, R. H., Sijen, T., de Knijff, P., & Laros, J. F. J. (2017). FDSTools: A software package for analysis of massively parallel sequencing data with the ability to recognise and correct STR stutter and other PCR or sequencing noise. Forensic Science International: Genetics, 27, 27–40. https://doi.org/10.1016/j.fsigen.2016.11.007.

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., … Flicek, P. (2021). Ensembl 2021. Nucleic Acids Research, 49(D1), D884–D891. https://doi.org/10.1093/nar/gkaa942.

Huel, R., Amory, S., Bili, A., Vidovi, S., Jasaragi, E., & Parsons, T. J. (2012). DNA Extraction from Aged Skeletal Samples for STR Typing by Capillary Electrophoresis. DNA Electrophoresis Protocols for Forensic Genetics, 830, 185–198. https://doi.org/10.1007/978-1-61779-461-2.

Huszar, T. I., Wetton, J. H., & Jobling, M. A. (2019). Mitigating the effects of reference sequence bias in single-multiplex massively parallel sequencing of the mitochondrial DNA control region. Forensic Science International. Genetics, 40, 9–17. https://doi.org/10.1016/j.fsigen.2019.01.008.

IBM Corp. (2013). IBM SPSS Statistics for Windows (Version 22.0.).

Irwin, J., Just, R., Scheible, M., & Loreille, O. (2011). Assessing the potential of next generation sequencing technologies for missing persons identification efforts.

Forensic Science International: Genetics Supplement Series, 3(1), e447–e448. https://doi.org/https://doi.org/10.1016/j.fsigss.2011.09.085.

Iyavoo, S., Hadi, S., & Goodwin, W. (2013). Evaluation of five DNA extraction systems for recovery of DNA from bone. Forensic Science International: Genetics Supplement Series, 4(1), e174–e175. https://doi.org/10.1016/j.fsigss.2013.10.090.

Jäger, A. C., Alvarez, M. L., Davis, C. P., Guzmán, E., Han, Y., Way, L., Walichiewicz, P., Silva, D., Pham, N., Caves, G., Bruand, J., Schlesinger, F., Pond, S. J. K., Varlaro, J., Stephens, K. M., & Holt, C. L. (2017). Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. Forensic Science International: Genetics, 28, 52–70. https://doi.org/10.1016/j.fsigen.2017.01.011.

Jagoda, E., Xue, J. R., Reilly, S. K., Dannemann, M., Racimo, F., Huerta-Sanchez, E., Sankararaman, S., Kelso, J., Pagani, L., Sabeti, P. C., & Capellini, T. D. (2022). Detection of Neanderthal Adaptively Introgressed Genetic Variants That Modulate Reporter Gene Expression in Human Immune Cells. Molecular Biology and Evolution, 39(1), msab304. https://doi.org/10.1093/molbev/msab304.

Jakubowska, J., Maciejewska, A., & Pawłowski, R. (2012). Comparison of three methods of DNA extraction from human bones with different degrees of degradation. International Journal of Legal Medicine, 126(1), 173–178. https://doi.org/10.1007/s00414-011-0590-5.

Jeffreys, A. J., Brookfield, J., & Semeonoff, R. (1985). Positive identification of an immigration test-case using human DNA fingerprints. Nature, 317(6040), 818–819. https://doi.org/10.1038/317818a0.

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. Nature, 314(6006), 67–73. https://doi.org/10.1038/314067a0.

Jehaes, E., Decorte, R., Peneau, A., Petrie, J. H., Boiry, P. A., Gilissen, A., Moisan, J. P., Van den Berghe, H., Pascal, O., & Cassiman, J.-J. (1998). Mitochondrial DNA analysis on remains of a putative son of Louis XVI, King of France and Marie-Antoinette. European Journal of Human Genetics, 6(4), 383–395. https://doi.org/10.1038/sj.ejhg.5200227.

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 (Oxford, England), 29(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193.

Kallupurackal, V., Kausel, T., Sulzer, A., Voegeli, P., & Kratzer, A. (2015). Automating PrepFiler® forensic DNA extraction kit: Optimization and validation on Freedom EVO® 150. Forensic Science International: Genetics Supplement Series, 5, e101–e103. https://doi.org/https://doi.org/10.1016/j.fsigss.2015.09.041.

Kapp, J. D., Green, R. E., & Shapiro, B. (2021). A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. Journal of Heredity, 112(3), 241–249. https://doi.org/10.1093/jhered/esab012.

Katsanis, S. H., & Wagner, J. K. (2013). Characterization of the standard and recommended CODIS markers. Journal of Forensic Sciences, 58 Suppl 1(Suppl 1), S169-72. https://doi.org/10.1111/j.1556-4029.2012.02253.x.

King, J. L., Wendt, F. R., Sun, J., & Budowle, B. (2017). STRait Razor v2s: Advancing sequence-based STR allele reporting and beyond to other marker systems. Forensic Science International: Genetics. https://doi.org/10.1016/j.fsigen.2017.03.013.

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research, 40(1), e3. https://doi.org/10.1093/nar/gkr771.

Kivisild, T. (2015). Maternal ancestry and population history from whole mitochondrial genomes. Investigative Genetics, 6(1), 3. https://doi.org/10.1186/s13323-015-0022-2.

Klug, A. (1968). Rosalind Franklin and the Discovery of the Structure of DNA. Nature, 219(5156), 808–810. https://doi.org/10.1038/219808a0.

Knapp, M., & Hofreiter, M. (2010). Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. Genes, 1(2), 227–243. https://doi.org/10.3390/genes1020227.

Korlević, P., Gerber, T., Gansauge, M. T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., & Meyer, M. (2015). Reducing microbial and human contamination in dna extractions

from ancient bones and teeth. In BioTechniques (Vol. 59, Issue 2, pp. 87–93). https://doi.org/10.2144/000114320.

Kowalczyk, M., Zawadzka, E., Szewczuk, D., Gryzińska, M., & Jakubczak, A. (2018). Molecular markers used in forensic genetics. Medicine, Science and the Law, 58(4), 201–209. https://doi.org/10.1177/0025802418803852.

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R. E., Burbano, H. A., Hublin, J.-J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., & Pääbo, S. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. Current Biology : CB, 17(21), 1908–1912. https://doi.org/10.1016/j.cub.2007.10.008.

Krings, M., Stone, A., Schmitz, R. W., Krainitzki, H., Stoneking, M., & Pääbo, S. (1997). Neandertal DNA Sequences and the Origin of Modern Humans. Cell, 90(1), 19–30. https://doi.org/10.1016/S0092-8674(00)80310-4.

Ladoukakis, E. D., & Zouros, E. (2001). Direct evidence for homologous recombination in mussel (Mytilus galloprovincialis) mitochondrial DNA. Molecular Biology and Evolution, 18(7), 1168–1175. https://doi.org/10.1093/oxfordjournals.molbev.a003904.

Lalueza-Fox, C., Römpler, H., Caramelli, D., Stäubert, C., Catalano, G., Hughes, D., Rohland, N., Pilli, E., Longo, L., Condemi, S., de la Rasilla, M., Fortea, J., Rosas, A., Stoneking, M., Schöneberg, T., Bertranpetit, J., & Hofreiter, M. (2007). A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. Science (New York, N.Y.), 318(5855), 1453–1455. https://doi.org/10.1126/science.1147417.

Lalueza-Fox, C., Sampietro, M. L., Caramelli, D., Puder, Y., Lari, M., Calafell, F., Martinez-Maza, C., Bastir, M., Fortea, J., & Rasilla, M. de la. (2005). Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian Peninsula. Molecular Biology and Evolution, 22(4), 1077–1081.

Lange, V., Böhme, I., Hofmann, J., Lang, K., Sauter, J., Schöne, B., Paul, P., Albrecht, V., Andreas, J. M., Baier, D. M., Nething, J., Ehninger, U., Schwarzelt, C., Pingel, J., Ehninger, G., & Schmidt, A. H. (2014). Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics, 15(1), 63. https://doi.org/10.1186/1471-2164-15-63.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359. https://doi.org/10.1038/nmeth.1923

Latham, K. E., & Miller, J. J. (2018). DNA recovery and analysis from skeletal material in modern forensic contexts. Forensic Sciences Research, 4(1), 1–9. https://doi.org/10.1080/20961790.2018.1515594.

Lee, H., Na, W., Park, C., Park, K. H., & Shin, S. (2018). Centrifugation-free extraction of circulating nucleic acids using immiscible liquid under vacuum pressure. Scientific Reports, 8(1), 5467. https://doi.org/10.1038/s41598-018-23766-9.

Leonard, J. A., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, R. K., & Fleischer, R. C. (2007). Animal DNA in PCR reagents plagues ancient DNA research. Journal of Archaeological Science, 34(9), 1361–1366. https://doi.org/https://doi.org/10.1016/j.jas.2006.10.023.

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.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature, 362(6422), 709–715. https://doi.org/10.1038/362709a0.

Lipatov, M., Sanjeev, K., Patro, R., & Veeramah, K. R. (2015). Maximum Likelihood Estimation of Biological Relatedness from Low Coverage Sequencing Data. BioRxiv, 23374. https://doi.org/10.1101/023374.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. Journal of Biomedicine and Biotechnology, 2012. https://doi.org/10.1155/2012/251364.

Liu, Q., Liu, L., Zhang, M., Zhang, Q., Wang, Q., Ding, X., Shao, L., Zhou, Z., & Wang, S. (2018). A Simple and Efficient Method of Extracting DNA from Aged Bones and Teeth. Journal of Forensic Sciences, 63(3), 824–828. https://doi.org/https://doi.org/10.1111/1556-4029.13603.

Loreille, O., Ratnayake, S., Bazinet, A. L., Stockwell, T. B., Sommer, D. D., Rohland, N., Mallick, S., Johnson, P. L. F., Skoglund, P., Onorato, A. J., Bergman, N. H., Reich, D., & Irwin, J. A. (2018). Biological Sexing of a 4000-Year-Old Egyptian Mummy Head

to Assess the Potential of Nuclear DNA Recovery from the Most Damaged and Limited Forensic Specimens. In Genes (Vol. 9, Issue 3). https://doi.org/10.3390/genes9030135.

Lott, M. (2019). Human MitoSeq MITOMAP (revision 5).

Luo, S., Valencia, C. A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J., Chen, S. M., Chien, Y.-H., Hwu, W.-L., Fan, P.-C., Wong, L.-J., Atwal, P. S., & Huang, T. (2018). Biparental Inheritance of Mitochondrial DNA in Humans. Proceedings of the National Academy of Sciences, 115(51), 13039 LP – 13044. https://doi.org/10.1073/pnas.1810946115.

Lutz, S., Wittig, H., Weisser, H.-J., Heizmann, J., Junge, A., Dimo-Simonin, N., Parson, W., Edelmann, J., Anslinger, K., Jung, S., & Augustin, C. (2000). Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? Forensic Science International, 113(1), 97–101. https://doi.org/https://doi.org/10.1016/S0379-0738(00)00222-X.

Mafessoni, F., Grote, S., de Filippo, C., Slon, V., Kolobova, K. A., Viola, B., Markin, S. V, Chintalapati, M., Peyrégne, S., Skov, L., Skoglund, P., Krivoshapkin, A. I., Derevianko, A. P., Meyer, M., Kelso, J., Peter, B., Prüfer, K., & Pääbo, S. (2020). A high-coverage Neandertal genome from Chagyrskaya Cave. Proceedings of the National Academy of Sciences, 117(26), 15132 LP – 15136. https://doi.org/10.1073/pnas.2004944117.

Malgosa, A., Vinueza, D., Rasal, R., Santos, C., Armentano, N., Galtés, I., & Casals, F. (2021). Els inicis de les identificacions genètiques de les víctimes de la Guerra Civil de 1936-1939 a Catalunya The beginnings of genetic identification of the victims of the Spanish Civil War in Catalonia. 31, 259–272. https://doi.org/10.21001/rap.2021.31.14.

Mameli, A., Piras, G., & Delogu, G. (2013). The Successful Recovery of Low Copy Number and Degraded DNA from Bones Exposed to Seawater Suitable for Generating a DNA STR Profile. Journal of Forensic Sciences, 59. https://doi.org/10.1111/1556-4029.12323.

Mameli, A., Piras, G., & Delogu, G. (2014). The successful recovery of low copy number and degraded DNA from bones exposed to seawater suitable for generating a DNA STR profile. Journal of Forensic Sciences, 59(2), 470–473. https://doi.org/10.1111/1556-4029.12323.

Marciniak, S., Klunk, J., Devault, A., Enk, J., & Poinar, H. N. (2015). Ancient human genomics: the methodology behind reconstructing evolutionary pathways. Journal of Human Evolution, 79, 21–34. https://doi.org/https://doi.org/10.1016/j.jhevol.2014.11.003.

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics, 9(1), 387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359.

Marshall, P. L., Stoljarova, M., Schmedes, S. E., King, J. L., & Budowle, B. (2014a). A high volume extraction and purification method for recovering DNA from human bone. Forensic Science International. Genetics, 12, 155–160. https://doi.org/10.1016/j.fsigen.2014.06.011.

Marshall, P. L., Stoljarova, M., Schmedes, S. E., King, J. L., & Budowle, B. (2014b). A high volume extraction and purification method for recovering DNA from human bone. Forensic Science International: Genetics, 12, 155–160. https://doi.org/10.1016/J.FSIGEN.2014.06.011.

McElhoe, J. A., Holland, M. M., Makova, K. D., Su, M. S. W., Paul, I. M., Baker, C. H., Faith, S. A., & Young, B. (2014). Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. Forensic Science International: Genetics, 13, 20–29. https://doi.org/10.1016/j.fsigen.2014.05.007.

Melton, T. (2004). Mitochondrial DNA Heteroplasmy. Forensic Science Review, 16(1), 1–20.

Meyer, M., Arsuaga, J., Filippo, C. De, Nagel, S., Aximu-petri, A., Nickel, B., Martínez, I., Gracia, A., María, J., Castro, B. De, Carbonell, E., Viola, B., Kelso, J., Prüfer, K., & Pääbo, S. (2016). Nuclear DNA sequences from the Middle Pleistocene Sima

de los Huesos hominins. Nature, 531(7595), 504–507. https://doi.org/10.1038/nature17405.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols, 2010(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448.

Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., … Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science (New York, N.Y.), 338(6104), 222–226. https://doi.org/10.1126/science.1224344.

Millett, M., & Gowland, R. (2015). Infant and Child Burial Rites in Roman Britain: a Study from East Yorkshire. Britannia, 46, 171–189. https://doi.org/DOI: 10.1017/S0068113X15000100.

Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. (2007). Which transposable elements are active in the human genome? Trends in Genetics, 23(4), 183–191. https://doi.org/10.1016/j.tig.2007.02.006.

Minaguchi, K., Hanaoka, Y., Maruyama, S., Nonaka, I., Kajiwara, M., Takagi, T., & Sato, Y. (2003). DNA analysis of neonatal human remains wrapped and kept in a vinyl bag for 15 years. Legal Medicine, 5, S183–S186. https://doi.org/https://doi.org/10.1016/S1344-6223(02)00106-2.

Mitchell, S. L., Goodloe, R., Brown-Gentry, K., Pendergrass, S. A., Murdock, D. G., & Crawford, D. C. (2014). Characterization of mitochondrial haplogroups in a large population-based sample from the United States. Human Genetics, 133(7), 861–868. https://doi.org/10.1007/s00439-014-1421-9.

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

Montiel, R., Francalacci, P., & Malgosa, A. (2007). Ancient DNA and Biological Anthropology: believers vs. skeptics.

Mullis, K. B. (1990). The Unusual Origin of the Polymerase Chain Reaction. Scientific American, 262(4), 56–65. http://www.jstor.org/stable/24996713.

Mundorff, A., & Davoren, J. M. (2014). Examination of DNA yield rates for different skeletal elements at increasing post mortem intervals. Forensic Science International: Genetics, 8(1), 55–63. https://doi.org/10.1016/J.FSIGEN.2013.08.001.

Mundorff, A. Z., Bartelink, E. J., & Mar-Cash, E. (2009). DNA Preservation in Skeletal Elements from the World Trade Center Disaster: Recommendations for Mass Fatality Management. Journal of Forensic Sciences, 54(4), 739–745. https://doi.org/10.1111/j.1556-4029.2009.01045.x.

Nanci, A. (2008). Enamel : composition, formation, and structure. Ten Cate's Oral Histology Development, Structure, and Function, 141–190. http://ci.nii.ac.jp/naid/10029473220/en/.

Nelson, K., & Melton, T. (2007). Forensic Mitochondrial DNA Analysis of 116 Casework Skeletal Samples*. Journal of Forensic Sciences, 52(3), 557–561. https://doi.org/https://doi.org/10.1111/j.1556-4029.2007.00407.x.

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. Nature, 541(7637), 302–310. https://doi.org/10.1038/nature21347.

Nieves-Colón, M. A., Ozga, A. T., Pestle, W. J., Cucina, A., Tiesler, V., Stanton, T. W., & Stone, A. C. (2018). Comparison of two ancient DNA extraction protocols for skeletal remains from tropical environments. American Journal of Physical Anthropology, 166(4), 824–836. https://doi.org/10.1002/ajpa.23472.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Andrey, B., Mikheenko, A., Mitchell, V., Nicolas, A., Lev, U., Ariel, G., Sergey, A., J., H. S., Mark, D., A., L. G., Michael, A., E., A. S., Matthew, B., G., B. G., Y., B. S., … M., P. A. (2022). The complete sequence of a human genome. Science, 376(6588), 44–53. https://doi.org/10.1126/science.abj6987.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics (Oxford, England), 32(2), 292–294. https://doi.org/10.1093/bioinformatics/btv566.

Olalde, I., Mallick, S., Patterson, N., Rohland, N., Villalba-Mouco, V., Silva, M., Dulias, K., Edwards, C. J., Gandini, F., Pala, M., Soares, P., Ferrando-Bernal, M., Adamski, N., Broomandkhoshbacht, N., Cheronet, O., Culleton, B. J., Fernandes, D., Lawson, A. M., Mah, M., … Reich, D. (2019). The genomic history of the Iberian Peninsula over the past 8000 years. Science (New York, N.Y.), 363(6432), 1230–1234. https://doi.org/10.1126/science.aav4040.

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., Krause, J., Willerslev, E., Stone, A. C., & Warinner, C. (2021). Ancient DNA analysis. Nature Reviews Methods Primers, 1(1), 14. https://doi.org/10.1038/s43586-020-00011-0.

Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. Nature Reviews Genetics, 16(7), 395–408. https://doi.org/10.1038/nrg3935.

Overballe-Petersen, S., Orlando, L., & Willerslev, E. (2012). Next-generation sequencing offers new insights into DNA degradation. Trends in Biotechnology, 30(7), 364–368. https://doi.org/https://doi.org/10.1016/j.tibtech.2012.03.007.

Ozdemir-Kaynak, E., & Yesil-Celiktas, O. (2015). Microwave-assisted digestion combined with silica-based spin column for DNA isolation from human bones. Analytical Biochemistry, 486, 44–50. https://doi.org/10.1016/j.ab.2015.06.034.

Paabo, S. (1993). DNA extraction from Pleistocene bones by purification method. 21(16), 3913–3914.

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences, 86(6), 1939 LP – 1943. https://doi.org/10.1073/pnas.86.6.1939.

Pääbo, Svante. (2014). The Human Condition-A Molecular Approach. Cell, 157(1), 216–226. https://doi.org/10.1016/j.cell.2013.12.036.

Pääbo, Svante, Poinar, H., Serre, D., Jaenicke-Despres, V., Hebler, J., Rohland, N., Kuch, M., Krause, J., Vigilant, L., & Hofreiter, M. (2004). Genetic analyses from ancient DNA. Annual Review of Genetics, 38, 645–679. https://doi.org/10.1146/annurev.genet.37.110801.143214.

Pakendorf, B., & Stoneking, M. (2005). MITOCHONDRIAL DNA AND HUMAN EVOLUTION. Annual Review of Genomics and Human Genetics, 6(1), 165–183. https://doi.org/10.1146/annurev.genom.6.080604.162249.

Palkopoulou, E., Mallick, S., Skoglund, P., Enk, J., Rohland, N., Li, H., Omrak, A., Vartanyan, S., Poinar, H., Götherström, A., Reich, D., & Dalén, L. (2015). Complete Genomes Reveal Signatures of Demographic and Genetic Declines in the Woolly Mammoth. Current Biology, 25(10), 1395–1400. https://doi.org/10.1016/j.cub.2015.04.007.

Palomo-Díez, S., Gomes, C., López-Parra, A. M., Baeza-Richer, C., Cuscó, I., Raffone, C., García-Arumí, E., Vinueza-Espinosa, D. C., Santos, C., Montes, N., Rasal, R., Escala, O., Cuellar, J., Subirá, E., Casals, F., Malgosa, A., Tizzano, E., Tartera, E., Domenech, G., & Arroyo-Pardo, E. (2019). Genetic identification of Spanish civil war victims. The state of the art in Catalonia (Northeastern Spain). Forensic Science International: Genetics Supplement Series, 7(1). https://doi.org/10.1016/j.fsigss.2019.10.035.

Palomo-Díez, Sara, Martínez-Labarga, C., Gomes, C., Esparza-Arroyo, Á., Rickards, O., & Arroyo-Pardo, E. (2017). Comparison of two different DNA extraction methodologies for critical bone or teeth samples. Forensic Science International: Genetics Supplement Series, 6, e359–e361. https://doi.org/10.1016/J.FSIGSS.2017.09.110.

Parakatselaki, M. E., & Ladoukakis, E. D. (2021). mtDNA heteroplasmy: Origin, detection, significance, and evolutionary consequences. Life, 11(7). https://doi.org/10.3390/life11070633.

Parker, C., Rohrlach, A. B., Friederich, S., Nagel, S., Meyer, M., Krause, J., Bos, K. I., & Haak, W. (2020). A systematic investigation of human DNA preservation in medieval skeletons. Scientific Reports, 10(1), 1–16. https://doi.org/10.1038/s41598-020-75163-w.

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

mitochondrial DNA typing. Forensic Science International: Genetics, 13, 134–142. https://doi.org/10.1016/j.fsigen.2014.07.010.

Parson, Walther, & Dür, A. (2007). EMPOP—A forensic mtDNA database. Forensic Science International: Genetics, 1(2), 88–92. https://doi.org/10.1016/j.fsigen.2007.01.018.

Parson, Walther, Strobl, C., Huber, G., Zimmermann, B., Gomes, S. M., Souto, L., Fendt, L., Delport, R., Langit, R., Wootton, S., Lagacé, R., & Irwin, J. (2013). Reprint of: Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). Forensic Science International: Genetics, 7(6), 632–639. https://doi.org/10.1016/j.fsigen.2013.09.007.

Patterson, N., Isakov, M., Booth, T., Büster, L., Fischer, C.-E., Olalde, I., Ringbauer, H., Akbari, A., Cheronet, O., Bleasdale, M., Adamski, N., Altena, E., Bernardos, R., Brace, S., Broomandkhoshbacht, N., Callan, K., Candilio, F., Culleton, B., Curtis, E., … Reich, D. (2022). Large-scale migration into Britain during the Middle to Late Bronze Age. Nature, 601(7894), 588–594. https://doi.org/10.1038/s41586-021-04287-4.

Pedersen, J. S., Valen, E., Velazquez, A. M. V., Parker, B. J., Rasmussen, M., Lindgreen, S., Lilje, B., Tobin, D. J., Kelly, T. K., Vang, S., Andersson, R., Jones, P. A., Hoover, C. A., Tikhonov, A., Prokhortchouk, E., Rubin, E. M., Sandelin, A., Gilbert, M. T. P., Krogh, A., … Orlando, L. (2014). Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. Genome Research, 24(3), 454–466. https://doi.org/10.1101/gr.163592.113.

Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X. J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., Reymond, A., Hubbard, T. J., Harrow, J., & Gerstein, M. B. (2012). The GENCODE pseudogene resource. Genome Biology, 13(9), R51–R51. https://doi.org/10.1186/gb-2012-13-9-r51.

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biology, 17(1), 60. https://doi.org/10.1186/s13059-016-0918-z.

Perlich, M. A., Reader, A., & Foreman, D. W. (1981). A scanning electron microscopic investigation of accessory foramens on the pulpal floor of human molars. Journal of Endodontics, 7(9), 402–406. https://doi.org/https://doi.org/10.1016/S0099-2399(81)80038-6.

Pervaiz, N., Kang, H., Bao, Y., & Abbasi, A. A. (2021). Molecular evolutionary analysis of human primary microcephaly genes. BMC Ecology and Evolution, 21(1), 76. https://doi.org/10.1186/s12862-021-01801-0.

Picard, M., Wallace, D. C., & Burelle, Y. (2016). The rise of mitochondria in medicine. Mitochondrion, 30, 105–116. https://doi.org/https://doi.org/10.1016/j.mito.2016.07.003.

Pierce, B. A. (2003). Genetics: A conceptual Approaches. In Archives of Biochemistry and Biophysics (4th ed., Vol. 308, Issue 1). https://doi.org/10.1006/abbi.1994.1027.

Pilli, E., Vai, S., Caruso, M. G., D'Errico, G., Berti, A., & Caramelli, D. (2018). Neither femur nor tooth: Petrous bone for identifying archaeological bone samples via forensic approach. Forensic Science International, 283, 144–149. https://doi.org/10.1016/j.forsciint.2017.12.023.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietrusewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H., & Hofreiter, M. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. PLoS ONE, 10(6), 1–13. https://doi.org/10.1371/journal.pone.0129102.

Poinar, H N, Höss, M., Bada, J. L., & Pääbo, S. (1996). Amino acid racemization and the preservation of ancient DNA. Science (New York, N.Y.), 272(5263), 864–866. https://doi.org/10.1126/science.272.5263.864.

Poinar, Hendrik N, Schwarz, C., Qi, J., Shapiro, B., Macphee, R. D. E., Buigues, B., Tikhonov, A., Huson, D. H., Tomsho, L. P., Auch, A., Rampp, M., Miller, W., & Schuster, S. C. (2006). Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science (New York, N.Y.), 311(5759), 392–394. https://doi.org/10.1126/science.1123360.

Prado, V. F., Castro, A. K. F., Oliveira, C. L., Souza, K. T., & Pena, S. D. J. (1997). Extraction of DNA from human skeletal remains : practical applications in forensic sciences. 14, 41–44.

Ramos, A., Santos, C., & Aluja, M. P. (2012). Freqüència i patró de l'heteroplàsmia mitocondrial humana. Universitat Autònoma de Barcelona. https://books.google.es/books?id=I4RRygEACAAJ.

Renaud, G., Hanghøj, K., Korneliussen, T. S., Willerslev, E., & Orlando, L. (2019). Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. Genetics, 212(3), 587–614. https://doi.org/10.1534/genetics.119.302057.

Rius, R., Cowley, M. J., Riley, L., Puttick, C., Thorburn, D. R., & Christodoulou, J. (2019). Biparental inheritance of mitochondrial DNA in humans is not a common phenomenon. Genetics in Medicine, 21(12), 2823–2826. https://doi.org/10.1038/s41436-019-0568-0

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. Nature Biotechnology, 29(1), 24–26. https://doi.org/10.1038/nbt.1754.

Röck, A. W., Dür, A., van Oven, M., & Parson, W. (2013). Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). Forensic Science International: Genetics, 7(6), 601–609. https://doi.org/https://doi.org/10.1016/j.fsigen.2013.07.005.

Rohland, N., Glocke, I., Aximu-petri, A., & Meyer, M. (2018). Extraction of highly degraded DNA from ancient bones , teeth and sediments for high-throughput sequencing. 13(November), 2447–2461.

Rohland, N., & Hofreiter, M. (2007a). Ancient dna extraction from bones and teeth. Nature Protocols, 2(7), 1756–1762. https://doi.org/10.1038/nprot.2007.247.

Rohland, N., & Hofreiter, M. (2007b). Comparison and optimization of ancient DNA extraction. BioTechniques, 42(3), 343–352. https://doi.org/10.2144/000112383.

Rohland, N., Siedel, H., & Hofreiter, M. (2010). A rapid column-based ancient DNA extraction method for increased sample throughput. Molecular Ecology Resources, 10(4), 677–683. https://doi.org/10.1111/j.1755-0998.2009.02824.x.

Rosa, A., Brehm, A., Kivisild, T., Metspalu, E., & Villems, R. (2004). MtDNA Profile of West Africa Guineans: Towards a Better Understanding of the Senegambia Region. Annals of Human Genetics, 68(4), 340–352. https://doi.org/https://doi.org/10.1046/j.1529-8817.2004.00100.x.

Rothe, J., & Nagy, M. (2016). Comparison of two silica-based extraction methods for DNA isolation from bones. Legal Medicine, 22, 36–41. https://doi.org/10.1016/J.LEGALMED.2016.07.008.

Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J. C., Brandon, M. C., Mishmar, D., Yi, C., Kreuziger, J., Baldi, P., & Wallace, D. C. (2007). An enhanced MITOMAP with a global mtDNA mutational phylogeny. Nucleic Acids Research, 35(suppl_1), D823–D828. https://doi.org/10.1093/nar/gkl927.

Salas, A., Lareu, V., Calafell, F., Bertranpetit, J., & Carracedo, A. (2000). mtDNA hypervariable region II (HVII) sequences in human evolution studies. European Journal of Human Genetics : EJHG, 8(12), 964–974. https://doi.org/10.1038/sj.ejhg.5200563.

Sampietro, M. L., Lao, O., Caramelli, D., Lari, M., Pou, R., Martí, M., Bertranpetit, J., & Lalueza-Fox, C. (2007). Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. Proceedings of the Royal Society B: Biological Sciences, 274(1622), 2161–2167. https://doi.org/10.1098/rspb.2007.0465.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, 74(12), 5463 LP – 5467. https://doi.org/10.1073/pnas.74.12.5463.

Santos, C., Montiel, R., Sierra, B., Bettencourt, C., Fernandez, E., Alvarez, L., Lima, M., Abade, A., & Aluja, M. P. (2005). Understanding differences between phylogenetic and pedigree-derived mtDNA mutation rate: a model using families from the Azores Islands (Portugal). Molecular Biology and Evolution, 22(6), 1490–1505. https://doi.org/10.1093/molbev/msi141.

Santos, C., Sierra, B., Alvarez, L., Ramos, A., Fernández, E., Nogués, R., & Aluja, M. P. (2008). Frequency and pattern of heteroplasmy in the control region of human

mitochondrial DNA. Journal of Molecular Evolution, 67(2), 191–200. https://doi.org/10.1007/s00239-008-9138-9.

Sawyer, S., Renaud, G., Viola, B., Hublin, J.-J., Gansauge, M.-T., Shunkov, M. V, Derevianko, A. P., Prüfer, K., Kelso, J., & Pääbo, S. (2015). Nuclear and mitochondrial DNA sequences from two Denisovan individuals. Proceedings of the National Academy of Sciences, 112(51), 15696 LP – 15700. https://doi.org/10.1073/pnas.1519905112.

Scally, A. (2016). The mutation rate in human evolution and demographic inference. Current Opinion in Genetics & Development, 41, 36–43. https://doi.org/https://doi.org/10.1016/j.gde.2016.07.008.

Schubert, M., Ermini, L., Sarkissian, C. Der, Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M. D., Fernández, R., Kircher, M., McCue, M., Willerslev, E., & Orlando, L. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nature Protocols, 9(5), 1056–1082. https://doi.org/10.1038/nprot.2014.063.

Scorrano, G., Valentini, F., Martínez-labarga, C., Rolfo, F., Fiammenghi, A., Vetro, D. Lo, Martini, F., Casoli, A., Ferraris, G., Palleschi, G., Palleschi, A., Scorrano, G., Valentini, F., Martínez-labarga, C., Rolfo, F., Fiammenghi, A., Vetro, D. Lo, Martini, F., Casoli, A., … Rickards, O. (2015). Annals of Human Biology Methodological strategies to assess the degree of bone preservation for ancient DNA studies Methodological strategies to assess the degree of bone preservation for ancient DNA studies. 4460. https://doi.org/10.3109/03014460.2014.954614.

Scorrano, G., Valentini, F., Martínez-Labarga, C., Rolfo, M. F., Fiammenghi, A., Lo Vetro, D., Martini, F., Casoli, A., Ferraris, G., Palleschi, G., Palleschi, A., & Rickards, O. (2015). Methodological strategies to assess the degree of bone preservation for ancient DNA studies. Annals of Human Biology, 42(1), 10–19. https://doi.org/10.3109/03014460.2014.954614.

Selberg, A. G. A., Gaucher, E. A., & Liberles, D. A. (2021). Ancestral Sequence Reconstruction: From Chemical Paleogenetics to Maximum Likelihood Algorithms and Beyond. Journal of Molecular Evolution, 89(3), 157–164. https://doi.org/10.1007/s00239-021-09993-1.

Shanmugam, A., Nagarajan, A., & Pramanayagam, S. (2017). Non-coding DNA – a brief review. Journal of Applied Biology & Biotechnology, 5(05), 42–47. https://doi.org/10.7324/jabb.2017.50507.

Shapiro, B., & Hofreiter, M. (2014). A Paleogenomic Perspective on Evolution and Gene Function: New Insights from Ancient DNA. Science, 343(6169), 1236573. https://doi.org/10.1126/science.1236573.

Sharma, S., Saha, A., Rai, E., Bhat, A., & Bamezai, R. (2005). Human mtDNA hypervariable regions, HVR I and II, hint at deep common maternal founder and subsequent maternal gene flow in Indian population groups. Journal of Human Genetics, 50(10), 497–506. https://doi.org/10.1007/s10038-005-0284-2.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. Nature Biotechnology, 26(10), 1135–1145. https://doi.org/10.1038/nbt1486.

Simón, M., Jordana, X., Armentano, N., Santos, C., Díaz, N., Solórzano, E., López, J. B., González-Ruiz, M., & Malgosa, A. (2011). The presence of nuclear families in prehistoric collective burials revisited: The bronze age burial of montanissell cave (Spain) in the light of aDNA. American Journal of Physical Anthropology, 146(3), 406–413. https://doi.org/10.1002/ajpa.21590.

Sitnik, R., Torres, M., Bacal, N., & Pinho, J. (2006). Using PCR for molecular monitoring of post-transplantation chimerism. Einstein, 4.

Skoglund, P., Northoff, B. H., Shunkov, M. V, Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences, 201318934. https://doi.org/10.1073/pnas.1318934111.

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., De La Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., De Lumley, H., Perrenoud, C., Gušić, I., … Meyer, M. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. Science, 356(6338), 605–608. https://doi.org/10.1126/science.aam9695.

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.

Stringer, C. B., & Andrews, P. (1988). Genetic and Fossil Evidence for the Origin of Modern Humans. Science, 239(4845), 1263–1268. http://www.jstor.org/stable/1700885.

SWGDAM. (2010). SWGDAM Interpretation Guidelines. Interpretation Guidelines. http://swgdam.org/Interpretation_Guidelines_January_2010.pdf

Taanman, J.-W. (1999). The mitochondrial genome: structure, transcription, translation and replication. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1410(2), 103–123. https://doi.org/https://doi.org/10.1016/S0005-2728(98)00161-3.

Templeton, N. S. (1992). The polymerase chain reaction. History, methods, and applications. Diagnostic Molecular Pathology : The American Journal of Surgical Pathology, Part B, 1(1), 58–72. https://doi.org/10.1097/00019606-199203000-00008.

Teschler-Nicola, M., Fernandes, D., Händel, M., Einwögerer, T., Simon, U., Neugebauer-Maresch, C., Tangl, S., Heimel, P., Dobsak, T., Retzmann, A., Prohaska, T., Irrgeher, J., Kennett, D. J., Olalde, I., Reich, D., & Pinhasi, R. (2020). Ancient DNA reveals monozygotic newborn twins from the Upper Palaeolithic. Communications Biology, 3(1), 650. https://doi.org/10.1038/s42003-020-01372-8.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., Powell, K., Mortensen, H. M., Hirbo, J. B., Osman, M., Ibrahim, M., Omar, S. A., Lema, G., Nyambo, T. B., Ghori, J., Bumpstead, S., Pritchard, J. K., Wray, G. A., & Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. Nature Genetics, 39(1), 31–40. https://doi.org/10.1038/ng1946.

Trent, R. (2012). Chapter 1 - Genes to Personalized Medicine (R. J. B. T.-M. M. (Fourth E. Trent (ed.); pp. 1–37). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-381451-7.00001-3.

Tully, L. A., Parsons, T. J., Steighner, R. J., Holland, M. M., Marino, M. A., & Prenger, V. L. (2000). A sensitive denaturing gradient-gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA

control region. American Journal of Human Genetics, 67(2), 432–443. https://doi.org/10.1086/302996.

Tvedebrink, T., Eriksen, P. S., Mogensen, H. S., & Morling, N. (2012). Statistical model for degraded DNA samples and adjusted probabilities for allelic drop-out. Forensic Science International: Genetics, 6(1), 97–101. https://doi.org/10.1016/j.fsigen.2011.03.001.

Vaissier Welborn, V. (2020). Enamel synthesis explained. Proceedings of the National Academy of Sciences, 117(36), 21847 LP – 21848. https://doi.org/10.1073/pnas.2014394117.

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation, 30(2), E386–E394. https://doi.org/https://doi.org/10.1002/humu.20921.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. Science (New York, N.Y.), 291(5507), 1304–1351. https://doi.org/10.1126/science.1058040.

Vernot, B., Zavala, E. I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., Mafessoni, F., Romagné, F., Pearson, A., Petr, M., Sala, N., Pablos, A., Aranburu, A., de Castro, J. M. B., Carbonell, E., Li, B., Krajcarz, M. T., Krivoshapkin, A. I., Kolobova, K. A., Kozlikin, M. B., … Meyer, M. (2021). Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. Science (New York, N.Y.), 372(6542). https://doi.org/10.1126/science.abf1667.

Verogen Inc. (2019). ForenSeq mtDNA Control Region Kit Reference Guide (Document#VD2019001Rev.A). June, 38.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K., & Wilson Allan, C. (1991). African Populations and the Evolution of Human Mitochondrial DNA. Science, 253(5027), 1503–1507. https://doi.org/10.1126/science.1840702.

Vinueza-Espinosa, D. C., Santos, C., Martínez-Labarga, C., & Malgosa, A. (2020). Human DNA extraction from highly degraded skeletal remains: How to find a suitable method? Electrophoresis. https://doi.org/10.1002/elps.202000171.

Wallace, D., & Chalkia, D. (2013). Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. Cold Spring Harbor Perspectives in Biology, 5(11), a021220. https://doi.org/10.1101/cshperspect.a021220.

Wallace, H. M., Jackson, A. R., Gruber, J., & Thibedeau, A. D. (2014). Forensic DNA databases–Ethical and legal standards: A global review. Egyptian Journal of Forensic Sciences, 4(3), 57–63. https://doi.org/https://doi.org/10.1016/j.ejfs.2014.04.002.

Warshauer, D. H., Lin, D., Hari, K., Jain, R., Davis, C., LaRue, B., King, J. L., & Budowle, B. (2013). STRait Razor: A length-based forensic STR allele-calling tool for use with second generation sequencing data. Forensic Science International: Genetics, 7(4), 409–417. https://doi.org/10.1016/j.fsigen.2013.04.005.

Weber-Lehmann, J., Schilling, E., Gradl, G., Richter, D. C., Wiehler, J., & Rolf, B. (2014). Finding the needle in the haystack: Differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. Forensic Science International: Genetics. https://doi.org/10.1016/j.fsigen.2013.10.015.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., & Schönherr, S. (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.

Whitaker, J. P., Clayton, T. M., Urquhart, A. J., Millican, E. S., Downes, T. J., Kimpton, C. P., & Gill, P. (1995). Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. BioTechniques, 18(4), 670–677.

Wiegand, P., & Kleiber, M. (2001). Less is more – length reduction of STR amplicons using redesigned primers. International Journal of Legal Medicine, 114(4), 285–287. https://doi.org/10.1007/s004140000162.

Willerslev, E., & Cooper, A. (2005). Review Paper. Ancient DNA. Proceedings of the Royal Society B: Biological Sciences, 272(1558), 3–16. https://doi.org/10.1098/rspb.2004.2813.

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., Bunce, M., Wiuf, C., Gilichinsky, D. A., & Cooper, A. (2003). Diverse Plant and Animal

Genetic Records from Holocene and Pleistocene Sediments. Science, 300(5620), 791 LP – 795. https://doi.org/10.1126/science.1084114.

William Gahl, M. (2021). National Human Genome-Mitochondrial DNA. Mitochondrial DNA. https://www.genome.gov/genetics-glossary/Mitochondrial-DNA.

Wilson, M. R., DiZinno, J. A., Polanskey, D., Replogle, J., & Budowle, B. (1995). Validation of mitochondrial DNA sequencing for forensic casework analysis. International Journal of Legal Medicine, 108(2), 68–74. https://doi.org/10.1007/BF01369907.

Xavier, C., Eduardoff, M., Bertoglio, B., Amory, C., Berger, C., Casas-Vargas, A., Pallua, J., & Parson, W. (2021). Evaluation of DNA Extraction Methods Developed for Forensic and Ancient DNA Applications Using Bone Samples of Different Age. Genes, 12(2), 146. https://doi.org/10.3390/genes12020146.

Xpress, I., Adapters, B., & A, C. N. (2016). Precision ID mtDNA Panel Analysis user guide. Thermofisher Scientific.

Yair, S., Lee, K. M., & Coop, G. (2021). The timing of human adaptation from Neanderthal introgression. Genetics, 218(1), iyab052. https://doi.org/10.1093/genetics/iyab052.

Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C., & Saunders, S. R. (1998). Technical Note : Improved DNA Extraction From Ancient Bones Using Silica-Based Spin Columns. 543(December 1997), 539–543.

Yang, Y., Xie, B., & Yan, J. (2014). Application of next-generation sequencing technology in forensic science. Genomics, Proteomics and Bioinformatics, 12(5), 190–197. https://doi.org/10.1016/j.gpb.2014.09.001.

Yao, Y., & Zhang, Y. (2003). Pitfalls in the analysis of ancient human mtDNA. Chinese Science Bulletin, 48(8), 826–830.

Ye, J., Ji, A., Parra, E. J., Zheng, X., Jiang, C., Zhao, X., Hu, L., & Tu, Z. (2004). A simple and efficient method for extracting DNA from old and burned bone. Journal of Forensic Sciences, 49(4), 754–759.

Young, J. M., Higgins, D., & Austin, J. J. (2019). Hybridization Enrichment to Improve Forensic Mitochondrial DNA Analysis of Highly Degraded Human Remains . In Frontiers in Ecology and Evolution (Vol. 7, p. 450). https://www.frontiersin.org/article/10.3389/fevo.2019.00450.

Yu, C., & Abbott, P. V. (2007). An overview of the dental pulp: its functions and responses to injury. Australian Dental Journal, 52(s1), S4–S6. https://doi.org/https://doi.org/10.1111/j.1834-7819.2007.tb00525.x.

Zavala, E. I., Jacobs, Z., Vernot, B., Shunkov, M. V, Kozlikin, M. B., Derevianko, A. P., Essel, E., de Fillipo, C., Nagel, S., Richter, J., Romagné, F., Schmidt, A., Li, B., O'Gorman, K., Slon, V., Kelso, J., Pääbo, S., Roberts, R. G., & Meyer, M. (2021). Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave. Nature, 595(7867), 399–403. https://doi.org/10.1038/s41586-021-03675-0.

Zavala, E. I., Thomas, J. T., Sturk-Andreaggi, K., Daniels-Higginbotham, J., Meyers, K. K., Barrit-Ross, S., Aximu-Petri, A., Richter, J., Nickel, B., Berg, G. E., McMahon, T. P., Meyer, M., & Marshall, C. (2022). Ancient DNA Methods Improve Forensic DNA Profiling of Korean War and World War II Unknowns. Genes, 13(1). https://doi.org/10.3390/genes13010129.

Zischler, H., Höss, M., Handt, O., von Haeseler, A., van der Kuyl, A. C., & Goudsmit, J. (1995). Detecting dinosaur DNA. In Science (New York, N.Y.) (Vol. 268, Issue 5214, pp. 1192–1193; author reply 1194). https://doi.org/10.1126/science.7605504.

Zupanič Pajnič, I., Debska, M., Gornjak Pogorelc, B., Vodopivec Mohorčič, K., Balažic, J., Zupanc, T., Štefanič, B., & Geršak, K. (2016). Highly efficient automated extraction of DNA from old and contemporary skeletal remains. Journal of Forensic and Legal Medicine, 37, 78–86. https://doi.org/10.1016/j.jflm.2015.11.001.

Zupanič Pajnič, I., & Fattorini, P. (2021). Strategy for STR typing of bones from the Second World War combining CE and NGS technology: A pilot study. Forensic Science International: Genetics, 50(September 2020). https://doi.org/10.1016/j.fsigen.2020.102401.

8

Appendix

Electrophoresis 2020, *0*, 1–10

Diana C. Vinueza-Espinosa[1]
Cristina Santos[1]
Cristina Martínez-Labarga[2]
Assumpció Malgosa[1]

[1]Biology Anthropology Research Group, Department of Animal Biology, Vegetal Biology and Ecology, Universidad Autónoma de Barcelona, Barcelona, Spain
[2]Department of Biology, Centre of Molecular Anthropology for Ancient DNA Studies, University of Rome Tor Vergata, Rome, Italy

Received June 18, 2020
Revised August 27, 2020
Accepted September 24, 2020

Research Article

Human DNA extraction from highly degraded skeletal remains: How to find a suitable method?

Retrieving DNA from highly degraded human skeletal remains is still a challenge due to low concentration and fragmentation, which makes it difficult to extract and purify. Recent works showed that silica-based methods allow better DNA recovery and this fact may be attributed to the type of bones and the quality of the preserved tissue. However, more systematic studies are needed to evaluate the efficiency of the different silica-based extraction methods considering the type of bones. The main goal of the present study is to establish the best extraction method and the type of bone that can maximize the recovery of PCR-amplifiable DNA and the subsequent retrieval of mitochondrial and nuclear genetic information. Five individuals were selected from an archaeological site located in Catalonia–Spain dating from 5th to 11th centuries AD. For each individual, five samples from different skeletal regions were collected: petrous bone, pulp cavity and cementum of tooth, and rib and limb bones. Four extraction methods were tested, three silica-based (silica in-suspension, HE column and XS plasma column) and the classical method based on phenol–chloroform. Silica in-suspension method from petrous bone and pulp cavity showed the best results. However, the remains preservation will ultimately be the key to the molecular result success.

Keywords:
DNA extraction / Highly degraded human skeletal remains / mtDNA / Silica-based extraction methods / STRs
DOI 10.1002/elps.202000171

Additional supporting information may be found online in the Supporting Information section at the end of the article.

1 Introduction

The field of forensic genetics has grown substantially during the last decade due to improving sensibility of DNA technologies. However, the generation of DNA profiles from human remains is still challenging since many factors influence the efficiency of DNA recovery. In cases of mass disasters (natural or caused), wars or criminal cases, skeletal remains are often the only biological material remaining for the identification process. In this context, *postmortem* interval [1,2], type of skeletal remains (long bones, petrous bones, or teeth) [3,4], decomposition degree [5], and *postmortem* conservation conditions [6–8] limit DNA recovery.

Correspondence:
*Dr. Assumpció Malgosa,
Biology Anthropology Research Group, Department of Animal Biology, Vegetal Biology and Ecology, Universidad Autónoma de Barcelona, Barcelona, Spain
E-mail: assumpcio.malgosa@uab.cat

Abbreviations: **HVR**, hypervariable region; **mtDNA**, mitochondrial DNA; **PHR**, peak height ratio; **RFU**, relative fluorescent unit

The degradation of biological material leads to low DNA preservation, its fragmentation in small pieces of 2–30 bp [9], and its molecular damage, which hinders extraction and purification of genetic material. Therefore, several strategies have been used to retrieve DNA from bones and teeth, some established in forensic laboratories [10,11], and others coming from the ancient DNA research community [12,13].

In forensic genetic laboratories, the phenol/chloroform method has been extensively used to extract DNA [14–16]. Nonetheless, recent works showed that silica-based methods allow better DNA recovery, more efficient removal of inhibitors, and a decrease in the number of handling steps, diminishing the risk of cross-contamination [7,10,17]. In this context, the silica-based extraction leads to more STR reportable alleles (above the stochastic threshold of detection) than DNA extracts obtained by the organic methods.

Although silica methods are based on the same chemistry (combination with a chaotropic salt-containing binding buffer), they can have different effects on DNA yields.

Color online: See article online to view Figs. 1 and 2 in color.

The specific literature shows a high variation in efficiency of DNA recovery when using extraction methods based on silica beads, silica in-suspension, or silica in-membrane columns (included in various extraction or purification kits). Recent studies also have demonstrated that recovery success by silica-based extraction methods may be attributed to the type of bones and the quality of the preserved tissue. The silica beads method has shown the highest DNA recovery for more recent samples, whereas for samples with low amounts of DNA the silica-membrane method has been the most efficient [2]. Other studies, however, mentioned that the silica in-suspension method shows the higher DNA yields and better mitochondrial DNA (mtDNA) typing than the silica in-membrane [18].

In this sense, more systematic studies are needed to evaluate the efficiency of the different silica-based extraction methods considering other factors such as the type of skeletal remains. The main goal of the present study is to establish the best extraction method and the type of skeletal remains that can maximize the recovery of PCR-amplifiable DNA and the subsequent obtaining of mitochondrial and nuclear genetic information relevant for genetic identification.

2 Materials and methods

2.1 Sample preparation

Five individuals were selected from an archaeological site in Catalonia-Spain dating from 5th to 11th centuries AD. For each individual, five samples from different skeletal regions were collected: petrous bone, pulp cavity and cementum of tooth, rib, and at least one different limb bones such as radius, ulna, metacarpal, or phalange.

To eliminate possible exogenous contaminants, each sample was cleaned removing a thin layer with a sterile tungsten tip placed into a micro drill up to 5,000 rpm. After careful drilling was performed in the cleaned area using a sterile tungsten tip, collecting roughly 120–800 mg of powder sample per each type of skeletal remain. Then powder was subdivided into different tubes (Supporting Information Table 1), one for each DNA extraction method used in this study (Fig. 1). All the procedures were conducted in highly sterile conditions in a specific ancient DNA laboratory.

2.2 DNA extraction

Four extraction methods were tested; three silica-based and the classical method based on phenol–chloroform (Fig. 1). Experimental details of the extraction methods procedures and the composition of reagents are detailed in Supporting Information Table 2. A negative extraction control without powder was also included per each set of five samples, representing a total of 100 DNA extracts and 25 controls.

2.2.1 Method 1: Silica in-suspension

The present DNA extraction protocol was a variant of the method described by Allentoft et al. [19].

2.2.2 Method 2: Silica in HE-membrane column

This method had been optimized in our laboratory for ancient bones and teeth using only the silica-membrane columns of the High Pure Extender Assembly Kit (HE-column) (Roche, Basel Switzerland) [12,20,21]. These spin columns efficiently purify viral nucleic acids (short molecules) of large sample volumes up to 3 mL.

2.2.3 Method 3: Silica in XS plasma-membrane column

The NucleoSpin® Plasma XS kit is designed for the efficient isolation of circulating DNA from human blood plasma. The Fragmented DNA as small as 50–1000 bp can be purified with high efficiency.

2.2.4 Method 4: Phenol–chloroform

The present DNA extraction protocol was a variant of the method described by Hagelberg and Clegg [22].

2.3 DNA Quantification

Total DNA quantification was performed using Nanodrop ND-1000 (ThermoFisher). To know if the recovered DNA corresponds to short or long DNA fragments, DNA was sized and quantified using the Bioanalyzer High Sensitivity DNA Kit® (Agilent Technologies).

2.4 PCR amplification and sequencing of mtDNA

All DNA extracts were amplified and sequenced: a 203-bp region from the hypervariable region I (HVRI) of mtDNA located between position L16030 and H16230 using primers forward 5′-CATGGGGAAGCAGATTTGGG-3′ and reverse 5′-GATAGTTGAGGGTTGATTGCTG-3′ [23] and a 238-bp region from the HVRII located between position L27 and H264 using primers forward 5′-GGTCTATCACCCTATTAACC-3′ and reverse 5′-CTTTCCACACAGACATCAT-3′ [23]. The PCR reactions were carried out using the Rotor-Gene®-Q in 10 µL reaction containing 2X Type-it HRM PCR Kit (Qiagen, Germany), 10 pmol/µL primers, and 1–5 µL of DNA template. The conditions were as follows: hold at 95°C for 10 min, denaturation at 95°C for 10 s, annealing at 58°C for HVRI and HVRII for 45 s and extension at 72°C for 10 s with 55 cycles.

PCR products were purified using MSB® Spin PCRapace (Invitek, Germany) according to the specifications protocol and quantified using Qubit® dsDNA Assay Kit

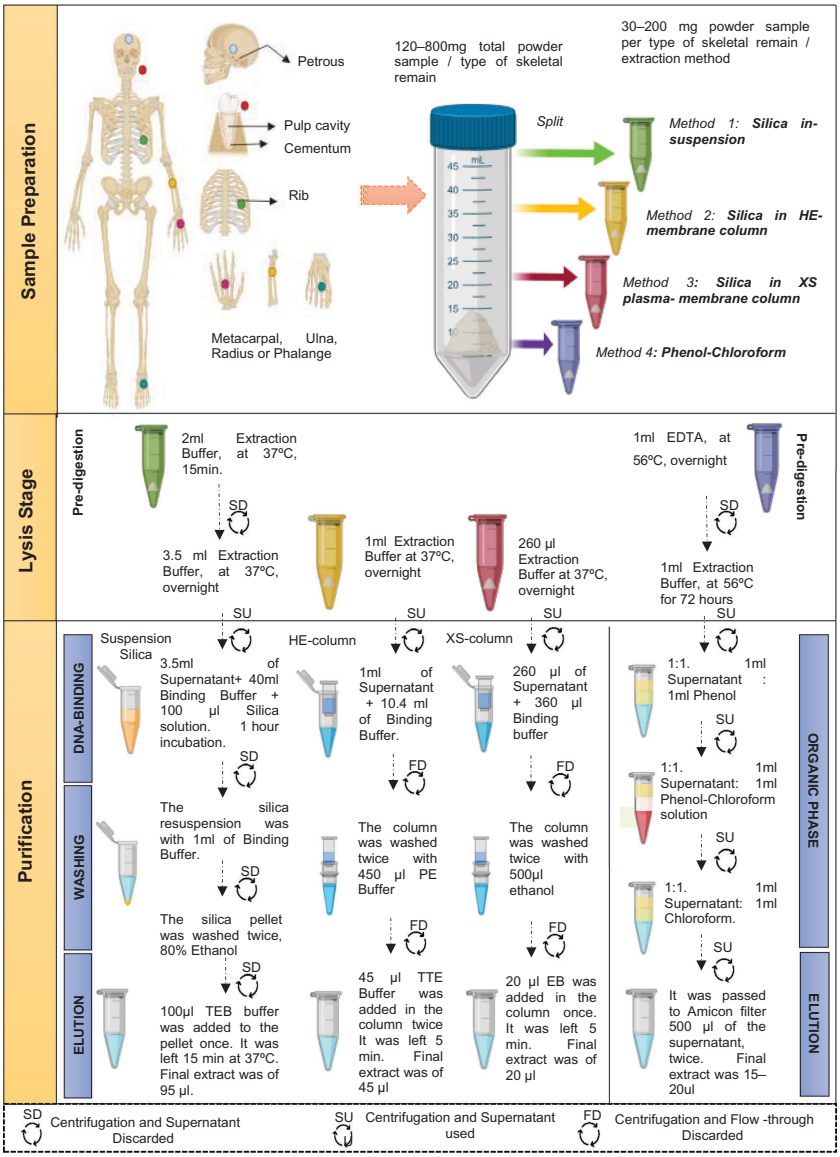


Figure 1. Workflow of the DNA extraction from petrous, pulp cavity, cementum of tooth, rib, and limb bones by using the four extraction methods.

(ThermoFisher) following the manufacturer procedures. MtDNA PCR products were then sequenced using BigDye® Terminator Sequencing Kit v1.1 in 10 µL reaction containing 0.5 µL Big Dye, 1 µL buffer, 1 µL primer 5 pmol/µL, 7 µL H_2O, and 0.5 µL DNA. Finally, the sequence reactions were purified with BigDye® X-terminator purification Kit. The sequencing was accomplished in Bioinformatics and Genomic Service (SGB) at UAB using a Genetic Analyzer 3130xl (Thermo Fisher Scientific). Sequence Scanner v1.0 and BioEdit software [24] were used for the analysis of sequences.

2.5 Analysis of nuclear DNA

Nuclear DNA was only analyzed for DNA samples extracted with the methods that revealed the best performance in mtDNA analysis. First, a 85-bp region that encompasses the −13 910C/T polymorphism strongly associate with Lactase Persistent in Europeans, located 13 910 bp upstream of the lactase (LCT) gene in chromosome 2, was amplified with primers: forward 5′-AATGTACTAGTAGGCCTCTGCG-3′ and reverse 5′-TGCAACCTAAGGAGGAGAGT-3′ [25]. The PCR conditions were performed as described for mtDNA using an annealing temperature of 53°C. The PCR product quantification was performed with Qubit™dsDNA BR Assay Kit.

Second, a set of 21 autosomal STRs, two Y-markers, and Amelogenin gene were analyzed using the GlobalFiler® PCR amplification Kit (Thermo Fisher Scientific) following the manufacturer procedure. Detection of amplified products was performed using the 3130xl Genetic Analyzer (Thermo Fisher Scientific) in the Genomics Core Facility at the Universitat Pompeu Fabra (UPF), using Gene Mapper ID-X v4.2 Software (Thermo Fisher Scientific) with a threshold established by internal laboratory validation of 20 relative fluorescent units (RFUs).

2.6 Data analysis

2.6.1 Evaluated parameters

The success of the DNA recovery was evaluated through total DNA concentration and the concentration of amplified mtDNA (HVRI and HVRII) and nuclear (LCT) fragments. The success of mtDNA (HVRI and HVRII) and nuclear (LCT) amplification was evaluated through the presence or absence of the amplified fragment by high-resolution melting.

The reproducibility of the mtDNA sequence results using different DNA extracts of the same individual obtained with different extraction methods was evaluated. The quality of STR profile was assessed through parameters as the average of peaks height (RFUs) and peak height ratio (PHR), which indicate the height balance between heterozygotes alleles, and the number of reportable alleles, which shows the percentage of alleles amplified in the genetic profile (homozygotes and heterozygotes) [26,27].

2.6.2 Statistical analysis

A two-way ANOVA followed by honestly-significant-difference Turkey's test for multiple comparisons was used to compare the concentration of total DNA and the concentration of mtDNA PCR fragments (HVRI and HVII) between type of skeletal remains and the extraction methods. The nuclear DNA concentration , peak heights (RFUs), peak height ratio (%), and percentage of reportable alleles between the best types of skeletal remains and extraction methods.

A categorical principal components analysis was employed to represent the relationships between some of the factors analyzed in this study except to the quality parameters of STR profile and concentration of nuclear fragment (LCT). All the statistical analyses were performed considering a significant level of 5% by the IBM SPSS Statistics [28].

3 Results

The results of the total DNA quantification indicated important differences between methods and skeletal remains (Supporting Information Table 3). Regarding the type of skeletal remains, the highest amount of total DNA from different individuals and methods was recovered from petrous bone ($\bar{x} =$ 33.20 ng/µL), followed by pulp cavity ($\bar{x} = 19.87$ ng/µL), upper limb bones ($\bar{x} = 19.27$ ng/µL), cementum of tooth ($\bar{x} =$ 16.22 ng/µL), and rib ($\bar{x} = 13.72$ ng/µL). The amount of total DNA recovered from all individuals and type of sample presented significant differences between skeletal remains (two-way ANOVA; $F = 4.07$, df $= 4$, $P = 0.005$).

Concerning the extraction methods, the method (1) silica in-suspension allowed the highest DNA retrieval ($\bar{x} = 43.97$ ng/µL), followed by the method (2) silica in HE-membrane columns ($\bar{x} = 25.24$ ng/µL) and method (3) silica in XS plasma-membrane column ($\bar{x} = 10.42$ ng/µL), whereas the method (4) phenol–chloroform recovered the lowest amount of DNA ($\bar{x} = 2.208$ ng/µL). The differences between methods are statistically significant (two-way ANOVA; $F = 30.15$, df $= 3$, $P < 0.001$).

Considering both, the extraction methods and the kind of skeletal remains, the highest amount of total DNA retrieval was of $\bar{x} = 78.08$ ng/µL from petrous bones followed by upper limb bones ($\bar{x} = 46.12$ ng/µL) and pulp cavity ($\bar{x} =$ 43.64 ng/µL) using the method (1) silica in-suspension. The DNA amount from petrous bones displays the best results with all methods. In addition, petrous bone DNA retrieved with method 1 showed the best results and presented significant differences with all trials except the pulp cavity and limb bones extracts using the same method (silica in-suspension), and the petrous bone extracts with the method 2 (multiple comparisons; *P*-values between 0.051 and 0.25).

A representative example of the DNA retrieval profile from petrous bone by using the four extraction methods is shown in Fig. 2. Both short (35–200 bp) and long (700–10 330 bp) fragments were recovered using the methods 1–3 (Fig. 2A–C), while the method 4 recovered few fragments and

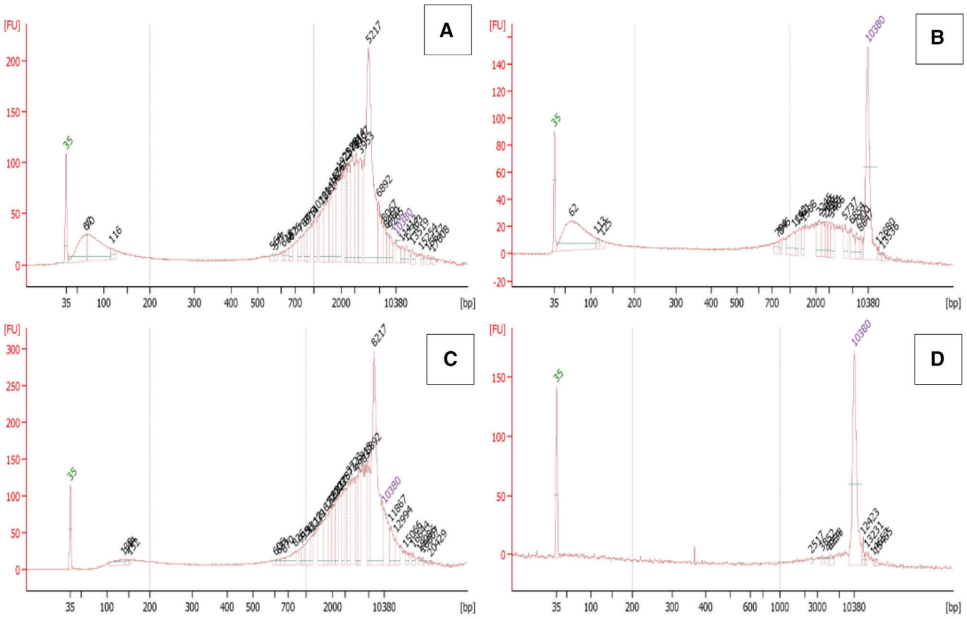


Figure 2. Length distributions of DNA quantification with Agilent High Sensitivity DNA kit® from the petrous DNA of the individual 2 retrieved using the four extraction methods. The size markers or ladder are sharp peaks at 35 bp and 10 380 bp. (A) Method 1: Silica in-suspension, (B) method 2: silica in HE-membrane column, (C) method 3: silica in XS plasma-membrane column, and (D) method 4: phenol–chloroform.

almost all correspond to long fragments ~2517 up 12 423 bp (Fig. 2D). The results also showed that methods 1 and 2 recovered the highest amount of short fragments characteristic of highly degraded DNA typical of forensic and ancient samples. However, the profile of long fragments in methods 1 and 2, which may represent modern DNA from bacteria, fungi or from human contaminants, was different. Method 2, silica in HE-membrane column (Fig. 2B), presents a low amount of high fragments compared to method 1, silica in-suspension (Fig. 2A).

In relation to the success of mtDNA PCR amplification, the highest number of positive amplifications was obtained by method 1 (19/25 samples amplified for HVRI and 24/25 samples amplified in HVRII). Method 4 showed the lowest results (12/25 samples amplified for HVRI and 9/25 samples amplified in HVRII) (Supporting Information Table 1). Concerning the concentration of mtDNA PCR products (Table 1), method 1 ($\bar{x}$ = 15.72 ng/μL for HVRI) ($\bar{x}$ = 12.45 ng/μL for HVRII) and method 2 ($\bar{x}$ = 14.32 ng/μL for HVRI; $\bar{x}$ = 11.67 ng/μL for HVRII) allowed the highest amplification of mtDNA, while method 4 ($\bar{x}$ = 1.40 ng/μL for HVRI; $\bar{x}$ = 1.16 ng/μL for HVRII) was associated to the poorest results. Regarding the type of skeletal remains, DNA extracts from petrous bone ($\bar{x}$ = 18.12 ng/μL for HVRI; $\bar{x}$ = 14.58 ng/μL for HVRII) and pulp cavity ($\bar{x}$ = 12.80 ng/μL for HVRI; $\bar{x}$ = 13.88 ng/μL for HVRII) displayed the best results, while DNA extract from rib showed the worst ($\bar{x}$ = 2.29 ng/μL for HVRI; $\bar{x}$ = 2.27 ng/μL for HVRII). Both differences between extraction methods (two-way ANOVA; F = 14.02 for HVRI and F = 18.97 for HVRII df = 3, P = 0.000) and types of skeletal remains (two-way ANOVA; F = 9.17 for HVRI and F = 19.29 for HVRII, df = 4, P = 0.000) were statistically significant.

Considering the interaction between extraction methods and kinds of skeletal remains, the highest concentration of mtDNA fragments were of $\bar{x}$ = 33.08 ng/μL for HVRI, $\bar{x}$ = 24.60 ng/μL for HVRII and $\bar{x}$ = 28.88 ng/μL for HVRI, $\bar{x}$ = 28.42 ng/μL for HVRII from the petrous bone DNA using the methods 1 and 2, respectively, and $\bar{x}$ = 19.86 ng/μL for HVRI, $\bar{x}$ = 23.76 g/μL for HVRII and $\bar{x}$ = 19.19 ng/μL for HVRI, $\bar{x}$ = 18.44 g/μL for HVRII from the pulp cavity DNA with methods 1 and 2, respectively (Table 1). The rib extracts showed the lowest amount of amplified DNA in HVRII fragment using methods 3 and 4 and did not show amplification in HVRI fragment. The concentration of mtDNA amplified in HVRI coming from petrous bone, pulp cavity, and cementum extracted with method 1 did not present significant differences compared with the HVRI-mtDNA PCR concentration from petrous

Table 1. Mean ± standard deviation of mtDNA (HVI and HVII) PCR fragments concentration (ng/μL) obtained from five types of skeletal remains using four extraction methods

Extraction methods[a)]	Petrous		Pulp cavity		Cementum of tooth		Limb bones		Rib		Total	Total
	HVRI[b)]	HVRII[c)]	HVRI	HVRII	HVRI	HVRII	HVRI	HVRII	HVRI	HVRII	HVRI	HVRII
1	33.08 ± 6.78	24.60 ± 8.52	19.86 ± 11.60	23.76 ± 14.61	14.82 ± 16.50	3.48 ± 4.61	4.37 ± 4.37	5.71 ± 5.00	6.48 ± 9.16	4.66 ± 5.97	15.72 ± 14.27	12.45 ± 12.55
2	28.88 ± 4.40	28.42 ± 7.54	19.19 ± 13.50	18.44 ± 12.26	8.48 ± 11.62	3.87 ± 3.96	12.36 ± 16.23	3.40 ± 4.53	2.70 ± 6.03	4.25 ± 5.01	14.32 ± 13.83	11.67 ± 12.30
3	8.60 ± 12.30	3.61 ± 5.81	10.38 ± 11.51	10.93 ± 12.28	5.07 ± 8.26	2.70 ± 3.69	7.36 ± 8.69	0.51 ± 1.01	NA	0.14 ± 0.14	6.28 ± 9.20	3.57 ± 7.00
4	1.92 ± 3.29	1.71 ± 1.60	1.76 ± 1.02	2.41 ± 2.37	1.54 ± 1.53	1.45 ± 2.98	1.80 ± 1.85	0.1832 ± 0.32	NA	0.031 ± 0.069	1.40 ± 1.86	1.16 ± 1.94
Total	18.12 ± 15.17	14.58 ± 13.67	12.80 ± 12.34	13.88 ± 13.32	7.48 ± 11.21	2.88 ± 3.66	6.47 ± 9.61	2.45 ± 3.91	2.29 ± 5.72	2.27 ± 4.22		

[a)] 1: Silica in-suspension, 2: silica in HE-membrane columns, 3: silica in XS plasma-membrane column, 4: phenol–chloroform. NA: No amplification.
[b)] mtDNA_HVRI PCR fragment concentration.
[c)] mtDNA_HVRII PCR fragment concentration.

bones and pulp cavity extracted with method 2 (multiple comparison; *P*-value between 0.17 and 1.00). Further, the concentration of mtDNA amplified in HVRII coming from petrous bone and pulp cavity extracted with method 1 did not present significant differences compared with the HVRII-mtDNA PCR concentration from petrous bones and pulp cavity extracted with methods 2 and 3 (multiple comparison; *P*-value between 0.129 and 1.00).

Sequences of good quality were obtained from amplification of extracts with methods 1 and 2. Also, the HVRI and HVRII mtDNA profiles obtained from different extraction methods and different bone pieces of the same individual were concordant.

The relation between extraction methods, total DNA concentration, presence/absence of mtDNA amplification and the concentration of mtDNA PCR fragments (HVRI and HVRII), individuals, and type of skeletal remains was represented (Fig. 3). It appears that the result is highly related to individual since each individual is located in a different position in the plot. We can observe that the extraction methods 1 and 2 group together and also the methods 3 and 4. The presence of mtDNA amplification (HVRI and HVRII) and the highest concentration of total DNA and amplifiable mtDNA were related to the individual 2, methods 1 and 2, petrous and pulp cavity. Conversely, method 4, individual 1, and ribs showed the worst results, which were related to the absence of mitochondrial amplification and the lowest concentration of total DNA and mtDNA amplifications.

In order to find out how the different factors influence the amplification of nuclear DNA, the best essays were selected to evaluate the success of the amplification of the 85 bp nuclear DNA fragment (LCT). In this case, the petrous bone and the pulp cavity extracted with methods 1 and 2 were chosen. All DNA extracts from petrous bone with methods 1 and 2 amplified the nuclear LCT fragment (5/5 samples), but only three of five samples from the pulp cavity DNA extracted with method 1 showed positive results. Moreover, DNA extracted using method 2 from the pulp cavity did not reveal amplification.

Concerning the concentration of the nuclear DNA fragment (Table 2), the highest values came from the petrous bone ($\bar{x} = 5.37$ ng/μL) and the pulp cavity ($\bar{x} = 4.85$ ng/μL) extracted with method 1. Conversely, the amplified fragment concentration was only $\bar{x} = 0.85$ ng/μL for the pulp cavity DNA extracted using the method 2. The differences obtained were not statistically significant in any trials (paired *t*-test, *P*-value between 0.069 and 0.89), except for the petrous bone DNA extracted with the method 1 paired with the pulp cavity DNA extracted using method 2 (paired *t*-test, $t = 4.12$, $P = 0.015$).

In relation to autosomal STRs, DNA from the petrous bone extracted with methods 1 and 2 yielded positive amplifications. The DNA from pulp cavity only worked with method 1, whereas the extract from method 2 did not show amplification of any STR (Table 2, Supporting Information Figs. SI 1 and 2). Regarding peak height (measured by RFUs), the highest values were obtained from the petrous bone extracts

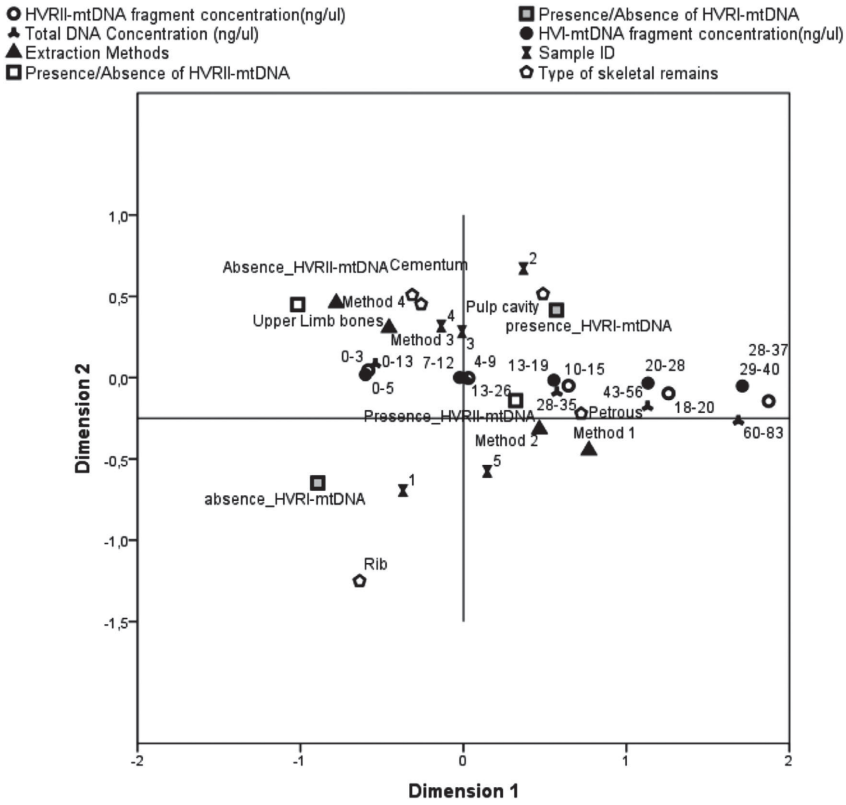


Figure 3. Analysis of categorical principal components for the total DNA and mtDNA fragment concentration (ng/μL) of HVRI and HVRII related to the presence/absence of mtDNA amplification, extraction methods, individuals (sample ID), and type of skeletal remains.

($\bar{x} = 498.60$ RFUs and $\bar{x} = 228.50$ RFUs, methods 1 and 2, respectively) and significant differences were found between them (paired *t*-test, $t = 4.02$, $P = 0.016$). The lowest height value came from the pulp cavity DNA extracted with method 1 ($\bar{x} = 57.82$ RFUs). There were significant differences between the petrous bone DNA and pulp cavity DNA extracted with method 1 (paired *t*-test; $t = 4.11$, $P = 0.015$), while the pulp cavity DNA extracted with method 1 and the petrous bone DNA extracted with method 2 presented minimal significant differences (paired *t*-test; $t = 2.64$, $P = 0.057$).

PHR showed the highest values from petrous bones extract using both methods 1 and 2 ($\bar{x} = 53.08\%$ and $\bar{x} = 62.89\%$, respectively), meanwhile the pulp cavity DNA extracted with method 1 showed the lowest one ($\bar{x} = 23.46\%$). No significant differences between methods and sample type were observed (paired *t*-test; *P*-value between 0.09 and 0.47).

Finally, the DNA from petrous bones extracted with both methods 1 and 2 showed the highest percentages of reportable alleles ($\bar{x} = 79.09\%$ and $\bar{x} = 71.81\%$, respectively) (Table 2). Significant differences between methods and sample type were observed (paired *t*-test; *P*-value between 0.035 and 0.049).

4 Discussion

Critical samples from forensic and ancient contexts contain small amounts of DNA that are highly fragmented and damaged. Therefore, the method used for DNA recovery is critical for genetic characterization and identification. Moreover, the selection of a suitable DNA source is a crucial decision. In this study, we extracted DNA from five different types of skeletal remains using four extraction methods: three based on silica and the traditional organic method. One hundred DNA extracts were obtained, and their performance assessed using different strategies based on PCR amplification.

First, it is important to note that the analyses of five different regions of skeletal remains per individual were also used to determinate whether the variation in the DNA

Table 2. Mean ± standard deviation of (LCT) nuclear DNA fragment concentration (ng/μL) and STRs parameters (peak height, percentage of peak height ratio, and reportable alleles) obtained of the DNA from petrous and pulp cavity extracted by methods 1 and 2

Type of skeletal remain	Nuclear DNA concentration of LCT gene (ng/μL)		Peak heights (RFUs)		Peak height ratio (PHR, %)		Reportable alleles (%)	
	M1[a]	M2	M1	M2	M1	M2	M1	M 2
Petrous bone	5.37 ± 2.37	4.10 ± 3.27	498.60 ± 252.74	228.50 ± 118.88	53.08 ± 18.94	62.89 ± 12.01	79.09 ± 21.22	71.81 ± 18.01
Pulp cavity	4.85 ± 8.80	0.85 ± 0.47	57.82 ± 84.22	NA	23.46 ± 32.81	NA	24.54 ± 37.97	NA

[a] M1: Silica in-suspension extraction method, M2: silica in HE-membrane columns extraction method.

recovery success was attributed to the individual (exposure conditions). Our results showed that individuals have preserved different DNA amounts. This fact is due to the individuals do not come from the same sector of the necropolis; this means that although the environmental conditions of the site may take as a whole, certain specific areas could be more rocky, sandy, or humid. Previous studies [4,29,30] show that the tomb construction material or death cause of individuals could provide distinct environmental conditions (fungi, bacterial), so the performance of DNA recovery will be different in each individual.

Regarding DNA source, five types of skeletal remains were selected, from the most compact hydroxyapatite tissue (tooth and petrous bone), intermediate thickness (ulna and radius), to the thinnest tissue from small elements (phalanges and ribs). The DNA extract from petrous bone, followed by the pulp cavity and cementum of the tooth, showed the highest recovery values of DNA regardless of the extraction method used. These results support previous studies [2,3,31] reporting that the petrous bone has higher endogenous DNA content (five- to eightfold on average) than teeth. Bone is a connective tissue largely composed of collagen and an inorganic mineral called hydroxyapatite [29]. Studies have shown that the DNA has a strong affinity for hydroxyapatite and its degradation is linked to the extent of crystallinity loss of this compound as well as the loss of collagen.

Therefore, the higher the bone density, the higher the amount of DNA adsorbed. The petrous bone, part of the temporal bone, is the toughest and most dense bone in the mammal's body, as well as where DNA can be more protected against moisture, sunlight, temperature, and microorganisms compared to a variety of interesting forensic material such as other bones (ribs or limbs bones), hair, excrement, or dry blood exposed to the same conditions [3,32]. Thus, good results from petrous bone and not as good at other fewer dense bones have been expected. The negative results from ribs seem to be related to the thin cortical tissue, which is more taphonomically affected than other tissues.

Along with the petrous bone, the densest skeletal tissue in the body is that of teeth. Indeed, our second-best results are obtained with dental tissue. However, the amount of DNA recovered is not the same in the different regions of teeth (dentine, cementum, and pulp cavity) as other studies have also reported [3]. Although the common forensic practice could be the use of the entire tooth to extract DNA, it is sometimes interesting to preserve it, whether it comes from archaeological cases or it is very important or unique evidence.

In our study, we compared the amount of DNA recovered from pulp cavity and cementum. The pulp cavity provided a higher DNA recovery than cementum. There is no evidence of studies that compare the preservation of DNA between the pulp cavity and cementum. However, there are some findings [31,32] that show the mtDNA content of tooth cementum was five times higher than the commonly used dentine. This may be due to dentine usually does not contain any nucleated cell bodies; it contains some mtDNAs that accumulate from the odontoblastic process [33]. In this sense,

we compared cementum (high cell density) [32] with the pulp cavity because, as is reported [34], it is a highly vascularized connective tissue containing numerous cell types and rich in DNA. The latter could explain better performance from the pulp cavity than cementum.

Degradation processes of any biological sample result in a low concentration of DNA and its fragmentation in short pieces, having a direct correlation between fragment length and abundance [12,19]. We evaluated the recovery profile to corroborate the rescue of highly degraded endogenous DNA from the samples. The outcomes showed that methods 1 and 2 presented the greatest recovery of short DNA. However, method 2 showed a lower concentration of long fragments than method 1. Our results agree with Allentoft et al. [19], which evaluated the DNA recovery profile using different binding buffer compositions with a silica in-suspension extraction method in combination with and without MinElute columns (Qiagen). DNA extracts that passed through a column showed less recovery of long fragments than those without pass through a column, probably because a purification step is performed by membrane columns allowing that little fragments are retained in the filter and the longer fragments pass.

Regarding extraction methods, our results showed that methods based on silica were better than the organic extraction method. The phenol–chloroform method produced the lowest concentration of DNA, an unquantifiable recovery of short fragments commonly related to endogenous critical DNA and the lowest concentrations of mitochondrial amplified fragments. Although there are very few studies comparing extraction methods based on silica and phenol–chloroform, our results are in agreement with these previous reports [10,20]. The success with silica method may be related to the presence of a high concentration of chaotropic salts (binding buffer) and a pH 4 that produces a reversibly specific bind of DNA to silica, isolating it from other molecules. Moreover, using the silica in-membrane method, the ultrafiltration by columns removes contaminants or inhibitors improving the amplification of recovered DNA [16,20]. Contrariwise, the phenol–chloroform method is so effective at removing proteins and lipids tending to be ineffective for the removal of hydrophilic compounds such as soil humic acid (PCR hydrophilic inhibitors) [6].

Our results showed that the extraction with silica in-suspension determined the highest DNA yield compared with the other extraction protocols. Similar results were reported [18], using two DNA extraction methodologies, silica in-suspension and in columns. Their results showed a higher quantity of DNA in 68.4% of the cases using the silica in-suspension method compared to 21.05% by the silica in-column method. The reason may be related to the binding time. In the silica in-suspension method, binding of DNA is performed for 1 h in-suspension with the silica, whereas the silica in-columns method, the DNA binding was performed during few seconds while it is being washed through the silica matrix. It is also known that extended binding time is critical to obtain high DNA yields [21]. Other studies have also shown that the silica-DNA adsorption is more efficient in the presence of buffer solutions with pH at or below the acid constant of the surface silanol groups, namely pH 4 [2,21]. Thus, the protocol of silica in-columns omits the pH adjustment after extraction buffer, silica and binding buffer are mixed. Instead of this, sodium acetate is added to the binding solution resulting in a pH between 5 and 6 that is above the ideal value. Moreover, the silica in-suspension method adds the phenol-red solution that allows establishing a redshift to yellow, indicating that the solution was at pH 4 during the silica–DNA binding phase; if the solution was greenish-yellow, HCl can be added up to the suitable pH. While this exact pH adjustment could be beneficial for a better DNA yield as is in this study, an excess of HCl could cause a decrease in DNA yields and the DNA may be completely degraded in such acidic conditions [21]. In this sense, a pH regulation could be considered to improve the extraction methods based on silica in-columns. Also, the use of columns after the binding step allows reducing the workload substantially and decreasing the risks of cross-contamination compared to resuspension by extensive pipetting with higher volumes, involving many time-consuming steps. In addition, the handling of high volumes allows a reduction of the number of samples conveniently processed in parallel. As other studies have shown, the silica-column method shows a measurable amount of extracted DNA [2,10,35,36] and could be also useful for dry and highly degraded samples. Otherwise, fresh bone tissue can block the membrane with a high amount of proteins such as collagen; these proteins simply cannot pass through the membrane being later eluted with DNA. The final result will be a low DNA yield [2].

To evaluate the quality of the DNA extracts, we amplified both mitochondrial and nuclear DNA. MtDNA consensus haplotypes were obtained and no DNA contamination was detected; so, both methods 1 and 2 appear to be equally valid. Similar results were obtained by Palomo-Díez et al. [18]. They used the silica in-suspension and silica in-columns extraction methods and found mtDNA consensus haplotypes, but the largest amount of DNA was obtained using silica in-suspension method. Regarding the ability of nuclear DNA recovery, the highest amount was obtained from the petrous bone and pulp cavity using method 1. The STRs profiles also showed the highest values of RFUs from petrous DNA extracted with method 1 and reportable alleles >70% with both methods. Meanwhile, the best balance of heterozygous alleles (PHR) was from petrous bone DNA extracted with method 2. Similar values were obtained by Marshall et al. [10] using an extraction method based on silica in-columns (Hi-Flow®). Following other studies [2,10,15,37], we can confirm by this parameter the presence of less inhibitory compounds in method 2 than in method 1.

In summary, the comparison of DNA extraction methods analyzed in this work shows the successful use of the methods based on silica–DNA binding, specifically with the method (1): silica in-suspension from petrous bone and pulp cavity. An improvement of the method (2): silica in HE-membrane column may be considered since it reduces

the workload substantially, decreases the risks of cross-contamination, and retrieves small fragments, facilitating the recovery of endogenous DNA. However, it is important to bear in mind that the preservation of the remains will ultimately be the key to the success of molecular results.

The work was supported by Generalitat of Cataluña (AGAUR. Ref 2017SGR1630)

The authors have declared no conflict of interest.

5 References

[1] Calacal, G. C., Apaga, D. L. T., Salvador, J. M., Jimenez, J. A. D., Lagat, L. J., Villacorta, R. P. F., Lim, M. C. F., Fortun, R. d. R., Datar, F. A., De Ungria, M. C. A., *Forensic Sci. Int. Genet.* 2015, *19*, 212–220.

[2] Rothe, J., Nagy, M., *Legal Med.* 2016, *22*, 36–41.

[3] Hansen, H. B., Damgaard, P. B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., Allentoft, M. E., *PLoS One* 2017, *12*, 1–18.

[4] Mundorff, A., Davoren, J. M., *Forensic Sci. Int. Genet.* 2014, *8*, 55–63.

[5] Scorrano, G., Valentini, F., Martínez-Labarga, C., Rolfo, M. F., Fiammenghi, A., Lo Vetro, D., Martini, F., Casoli, A., Ferraris, G., Palleschi, G., Palleschi, A., Rickards, O., *Ann. Hum. Biol.* 2015, *42*, 10–19.

[6] Jakubowska, J., Maciejewska, A., Pawłowski, R., *Int. J. Legal Med.* 2012, *126*, 173–178.

[7] Desmyter, S., De Cock, G., Moulin, S., Noël, F., *Forensic Sci. Int.* 2017, *273*, 96–101.

[8] Mameli, A., Piras, G., Delogu, G., *J. Forensic Sci.* 2014, *59*, 470–473.

[9] Glocke, I., Meyer, M., *Genome Res.* 2017, 1230–1237.

[10] Marshall, P. L., Stoljarova, M., Schmedes, S. E., King, J. L., Budowle, B., *Forensic Sci. Int. Genet.* 2014, *12*, 155–160.

[11] Huel, R., Amory, S., Bili, A., Vidovi, S., Jasaragi, E., Parsons, T. J., *DNA Electrophoresis Protoc. Forensic Genet.* 2012, *830*, 185–198.

[12] Rohland, N., Glocke, I., Aximu-petri, A., Meyer, M., *Nat. Protoc.* 2018, *13*, 2447–2461.

[13] Gamba, C., Hanghøj, K., Gaunitz, C., Alfarhan, A. H., Alquraishi, S. A., Al-rasheid, K. A. S., Bradley, D. G., Orlando, L., *Mol. Ecol. Res.* 2016, 459–469.

[14] Prado, V. F., Castro, A. K., Oliveira, C. L., Souza, K. T., Pena, S. D., *Genet. Anal. Biomol. Eng.* 1997, *14*, 41–44.

[15] Caputo, M., Irisarri, M., Alechine, E., Corach, D., *Forensic Sci. Int. Genet.* 2013, *7*, 488–493.

[16] Ferreira, S. T. G., Paula, K. A., Maia, F. A., Moraes, A. V, *Forensic Sci. Int. Genet. Suppl. Ser.* 2013, *4*, e374–e375.

[17] Davoren, J., Vanek, D., Konjhodzic, R., Crews, J., Huffine, E., Parsons, T., *Croat. Med. J.* 2007, *48*, 478–485.

[18] Palomo-Díez, S., Martínez-Labarga, C., Gomes, C., Esparza-Arroyo, Á., Rickards, O., Arroyo-Pardo, E., *Forensic Sci. Int. Genet. Suppl. Ser.* 2017, *6*, e359–e361.

[19] Allentoft Morten E., Sikora M., Sjögren K.-G., Rasmussen S., Rasmussen M., Stenderup J., Damgaard P. B., Schroeder H., Ahlström T., Vinner L., Malaspinas A.-S., Margaryan A., Higham T., Chivall D., Lynnerup N., Harvig L., Baron J., Casa P. D., Dąbrowski P., Duffy P. R., Ebel A. V., Epimakhov A., Frei K., Furmanek M., Gralak T., Gromov A., Gronkiewicz S., Grupe G., Hajdu T., Jarysz R., Khartanovich V., Khokhlov A., Kiss V., Kolář J., Kriiska A., Lasak I., Longhi C., McGlynn G., Merkevicius A., Merkyte I., Metspalu M., Mkrtchyan R., Moiseyev V., Paja L., Pálfi G., Pokutta D., Pospieszny Ł., Price T. D., Saag L., Sablin M., Shishlina N., Smrčka V., Soenov V. I., Szeverényi V., Tóth G., Trifanova S. V., Varul L., Vicze M., Yepiskoposyan L., Zhitenev V., Orlando L., Sicheritz-Pontén T., Brunak S., Nielsen R., Kristiansen K., Willerslev E., Population genomics of Bronze Age Eurasia *Nature* 2015, *522*, (7555), 167.172http://doi.org/10.1038/nature14507.

[20] Rohland, N., Hofreiter, M., *BioTechniques* 2007, *42*, 343–352.

[21] Rohland, N., Siedel, H., Hofreiter, M., *Mol. Ecol. Res.* 2010, *10*, 677–683.

[22] Hagelberg, E., Clegg, J. B., *Proc. R. Soc. B* 1991, *244*, 45–50.

[23] Simón, M., Jordana, X., Armentano, N., Santos, C., Díaz, N., Solórzano, E., López, J. B., González-Ruiz, M., Malgosa, A., *Am. J. Phys. Anthropol.* 2011, *146*, 406–413.

[24] Hall, T.A., *Nucleic Acids Symp. Ser.* 1999, *41*, 95–98.

[25] Burger, J., Kirchner, M., Bramanti, B., Haak, W., Thomas, M. G., *Proc. Natl. Acad. Sci. USA* 2007, *104*, 3736–3741.

[26] Grisedale, K. S., van Daal, A., *Investig. Genet.* 2012, *3*, 14.

[27] Harrel, M., Mayes, C., Gangitano, D., Hughes-Stamm, S., *J. Forensic Sci.* 2018, *63*, 1819–1823.

[28] IBM Corperation. *IBM SPSS Statistics for Windows, Version 22.0*, IBM Corporation, Armonk, New York, 2013.

[29] Collins, M. J., Nielsen–Marsh, C. M., Hiller, J., Smith, C. I., Roberts, J. P., Prigodich, R. V, Wess, T. J., Csapò, J., Millard, A. R., Turner–Walker, G., *Archaeometry* 2002, *44*, 383–394.

[30] Korlević, P., Gerber, T., Gansauge, M. T., Hajdinjak, M., Nagel, S., Aximu-Petri, A., Meyer, M., *BioTechniques* 2015, *59*, 87–93.

[31] Damgaard, P. B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., Allentoft, M. E., *Sci. Rep.* 2015, *5*, 11184.

[32] Adler, C. J., Haak, W., Donlon, D., Cooper, A., *J. Archaeol. Sci.* 2011, *38*, 956–964.

[33] Perlich, M. A., Reader, A., Foreman, D. W., *J. Endodontics* 1981, *7*, 402–406.

[34] Rohland, N., Hofreiter, M., *Nat. Protoc.* 2007, *2*, 1756–1762.

[35] Hasan, M. M., Hossain, T., Majumder, A., Momtaz, P., Sharmin, T., Sufian, A., Akhteruzzaman, S., *Dhaka Univ. J. Biol. Sci.* 2014, *23*, 101–107.

[36] Mameli, A., Piras, G., Delogu, G., *J. Forensic Sci.* 2013, *59*, 470–473.

[37] Tvedebrink, T., Eriksen, P. S., Mogensen, H. S., Morling, N., *Forensic Sci. Int. Genet.* 2012, *6*, 97–101.

ELECTROPHORESIS

MtDNA-CR Typing from Highly Degraded Skeletal Remains by Single-Multiplex Massively Parallel Sequencing.

Journal:	*ELECTROPHORESIS*
Manuscript ID	elps.202200052
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	28-Feb-2022
Complete List of Authors:	Vinueza-Espinosa, Diana; Universidad Autónoma de Barcelona, Dep. Biologia Animal, Biologia Vegetal i Ecologia Unitat Antropologia Biològica R. Cuesta-Aguirre, Daniel ; Universidad Autónoma de Barcelona, Dep. Biologia Animal, Biologia Vegetal i Ecologia Unitat Antropologia Biològica Malgosa, Assumpcio Santos, Cristina; Universitat Autònoma de Barcelona, Departament Biologia Animal, Biologia Vegetal i Ecologia
Keywords:	NGS, human skeletal remain, critical samples, PowerSeqCRM Nested System Kit, protocol optimization

MtDNA-CR Typing from Highly Degraded Skeletal Remains by Single-Multiplex Massively Parallel Sequencing.

Diana C. Vinueza-Espinosa[1], Daniel R. Cuesta-Aguirre[1], Assumpció Malgosa[1], Cristina Santos[1]

[1]Research Group in Biological Anthropology, Biological Anthropology Unit, Department of Animal Biology, Vegetal Biology and Ecology, Universitat Autònoma de Barcelona, Spain.

Corresponding Author

Cristina Santos, PhD

Research Group in Biological Anthropology

Unitat Antropologia Biològica

Dep. Biologia Animal, Biologia Vegetal i Ecologia

Facultat Biociències Edifici C

Universitat Autònoma de Barcelona

08193 Cerdanyola del Vallès

Barcelona (SPAIN)

Tel.: +34 93 581 13 17

Fax: +34 93 581 13 21

cristina.santos@uab.cat

Abstract

Poor nuclear DNA preservation from highly degraded skeletal remains is the most limiting factor for the genetic identification of individuals. The mitochondrial DNA (mtDNA) typing, and especially of the control region (CR), using next-generation sequencing, allows retrieving valuable genetic information in forensic contexts where highly degraded human skeletal remains are the only source of genetic material. Nowadays, NGS commercial kits allow all mtDNA-CR typed through fewer steps than the Sanger conventional technique. The PowerSeq™CRM Nested System kit (Promega Corporation) employs a multiplex-PCR strategy to amplify the entire mtDNA-CR in a single reaction. This kit has the advantage of simplifying workflow using a nested amplification protocol by incorporating indexed adapters. Our study aimed to analyze the success of mtDNA-CR typing from highly degraded human skeletal using the PowerSeq™CRM Nested System kit. We used samples from 41 individuals of different chronologies to test three NGS protocols (M1, M2, and M3) based on modifications

ABSTRACT

In the last 40 years, ancient and forensic DNA fields have developed extraordinarily. The gain was fast and large from simple PCR to NGS, cost and time invested, and information obtained. The goal is different between the two disciplines, aDNA and Forensics, but the methodologies are similar. Forensic genetics focuses on identification for legal purposes, while aDNA has broader goals, such as population reconstruction. In this way, genetic studies from skeletal remains could be useful for interpreting our evolutionary past through paleogenetics or identifying a person through forensic genetics. Despite the technical advances, the recovery of degraded DNA is still a difficult task due to its scarcity and molecular damage. Therefore, the first mission to address must focus on this problem: ameliorate its recovery.

The present doctoral thesis contributes to this task in different methodological ways: 1) comparing DNA contribution of different types of skeletal elements (**explored in chapter 1**), 2) searching for a good and easy protocol that maximizes the recovery of the major number of molecules through strategies of the genotyping of autosomal STR markers and the typing of mtDNA control region by conventional techniques (**explored in chapter 1**), 3) exploring the recovery of human whole-genome from newborn bones comparing the recovery efficiency by age categories and type of skeletal remains based on shot-gun NGS (**explored in chapter 2**), 4) probing the efficiency of a new NGS-kit to amplify the control region of mtDNA, typically used in forensic and paleogenetic researches (**explored in chapter 3**), 5) and providing a useful bioinformatic tool to analyze the NGS data, showing molecular damage as a signal of the authentic recovery of the true endogenous DNA (**explored in chapter 3**). Overall, this thesis spanned comparisons of different aspects that can affect the retrieval of genetic information, and that should be considered if these types of critical samples are used for identification purposes or population characterization.