

Faculté des sciences

Analysis of ancient DNA from teeth of human remains excavated from the Gallo-Roman necropolis on Rue Perdue in Tournai

Auteure : Lauranne Vermeersch

Co-promoteurs : Gontran Sonet et Bernard Knoops

Lecteurs : René Rezsöházy, Caroline Polet et Laurent Verslype

Année académique 2019-2020

Master 60 en Sciences Biologiques

Royal Belgian Institute of Natural Sciences

Louvain Institute of Biomolecular Science and Technology

Animal Molecular and Cellular Biology

Acknowledgements

I would like to first thank very much my master's thesis supervisors Gontran Sonet and Bernard Knoops, who gave me the opportunity to work on such an interesting project. I would like to thank them especially for their patience, their availability, their ever support and above all for their precious advices.

I am also very grateful to the whole team of the Royal Belgian Institute of Natural Sciences that welcomed me in their laboratory despite the challenging sanitary situation.

I would like to thank Laurent Verslype and Erika Weinkauff for giving me the opportunity to work on the archaeological remains of the Rue Perdue in Tournai and for providing me with all the necessary archaeological information for the realization of this study. This project would not have been possible without them.

I would like to thank Marianne Delcourt-Vlaeminck and Caroline Polet who gave me access to the human remains.

I would also like to thank H  l  ne Dailly and Sophie Opfergelt for their work in analyzing Strontium ratios.

I would like to thank the readers, Ren   Rezsohazy, Caroline Polet and Laurent Verslype, for the time they have devoted to read this study.

I would like to especially thank Andr   Lejeune without whom this project would never have been achieved. Since the first day, he believed in my academic project. He helped to address many obstacles and had been always very supportive and encouraging.

Finally, I must express my profound gratitude to Carinne, Benoit et Bryan for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

Lauranne Vermeersch

Abstract

This study was aimed at applying molecular biology techniques to address archaeological questions that could not be answered otherwise through classical historical and archaeological methods.

Thanks to genetic analyses performed during this work on teeth of human remains excavated between 1989 and 2010 from a Gallo-Roman necropolis discovered in Tournai (Belgium), it has been possible to acquire new information on populations living in Tournai during the Late Roman Empire (3rd-4th century AD).

The first objective of this study was to assess the feasibility of the extraction and analysis of ancient mitochondrial and nuclear DNA from teeth. After extracting and sequencing ancient DNA, we evaluated the authenticity of the results. DNA was successfully extracted. Only a few cross-contaminations were reported among some sequences amplified by PCR.

The second objective addressed the result contribution to archaeological studies. It aimed at answering several questions.

The geographical origin of individuals was first investigated by using mitochondrial haplogroups and measurements of Strontium isotopes ($^{87}\text{Sr}/^{86}\text{Sr}$). An accurate geographical origin of individuals could not be determined when using these haplogroups. However, Sr ratio analyses helped to restrict the plausible origins by excluding certain regions such as Northern Europe, Spain or the Alps. These results also demonstrated that individuals were not native of the same region. This would suggest that population movements had occurred at least at the regional scale of Tournai and its surrounding regions.

Based on mitochondrial haplogroups, the concentrations did not appear to be associated with family clusters. The presence of family vaults based on maternal lineages could be excluded.

Finally, an analysis of nuclear DNA data enabled sex identification of some individuals, which confirmed the results of previous anthropological studies.

This project outlined the relevance and the potential of the molecular biology techniques applied to the archeological field. The advanced technology could help to confirm or not historical assumptions, to complement and acquire new knowledge on past populations. This approach also contributes to value the Belgian patrimony.

More generally, new application areas could eventually emerge from such study. Based on adequate governance and methodologies, genetic results could be further structured into a large international and archeological database. New recent data visualization and interconnection techniques could be helpful to historical interpretations. They could bring innovative referential to mutually enrich local studies.

Résumé

Cette étude avait pour but d'appliquer certaines techniques de biologie moléculaire à des questions archéologiques inabordables via les méthodes historiques et archéologiques classiques.

Grâce aux analyses génétiques de dents humaines provenant d'une nécropole gallo-romaine découverte à Tournai (Belgique) entre 1989 et 2010, il a été possible d'acquérir de nouvelles informations sur les populations vivant à Tournai durant le Bas-Empire romain (3^e-4^e siècle après J.-C.).

Le premier objectif de l'étude consistait à évaluer la faisabilité de l'extraction et de l'analyse d'ADN ancien mitochondrial et nucléaire provenant de dents des squelettes de la nécropole. Une fois l'ADN extrait et séquencé, nous avons évalué l'authenticité des résultats. De l'ADN mitochondrial et nucléaire a pu être extrait avec succès. Seules quelques contaminations croisées ont été observées parmi les séquences de certaines analyses.

Le deuxième objectif concernait la contribution de ces résultats aux recherches archéologiques et devait répondre à plusieurs questions.

D'abord, l'origine géographique des individus a été étudiée à partir des haplogroupes mitochondriaux et des mesures des isotopes de Strontium ($^{87}\text{Sr}/^{86}\text{Sr}$). L'origine géographique précise n'a pas pu être déterminée à l'aide de ces haplogroupes. Les résultats des analyses du ratio de strontium ont cependant permis de restreindre les possibilités en excluant certaines régions telles que l'Europe du Nord, l'Espagne ou encore les Alpes. Ces résultats ont également montré que les individus étudiés ne sont pas originaires d'une seule et même région. Ils suggèrent ainsi que des mouvements de populations se seraient produits au minimum à l'échelle régionale de Tournai et des régions avoisinantes.

Ensuite, sur base des haplogroupes mitochondriaux il n'a pas été possible d'identifier des liens de parentés entre les individus. De plus, les concentrations ne semblent pas associées à des regroupements familiaux. La présence de caveaux familiaux associés aux lignées maternelles seraient donc à exclure.

Finalement, une analyse des données de l'ADN nucléaire a permis d'identifier le sexe de certains individus, confirmant ainsi les résultats des études anthropologiques antérieures.

Ce projet a souligné la pertinence et les perspectives de la biologie moléculaire appliquée au domaine archéologique. Ces techniques permettent de confirmer ou d'infirmer des hypothèses historiques, de compléter ou d'acquérir de nouvelles connaissances sur les populations du passé. Cette approche contribue aussi à la mise en valeur du patrimoine belge.

Plus globalement, moyennant la mise en place de gouvernance et de méthodologie appropriées, l'agrégation structurée de tels résultats pourrait aider à l'élaboration de bases de données plus larges. De nouvelles techniques de visualisation et d'interconnexion contribueraient à tester des interprétations historiques et apporteraient un nouveau référentiel aux démarches locales.

Table of Contents

1. Introduction	9
1.1. Study of ancient DNA.....	11
1.1.1. Interest of ancient DNA studies	11
1.1.2. Difficulties of ancient DNA studies	11
1.1.3. Authenticity criteria.....	14
1.1.4. Genetic markers used in ancient DNA studies	15
1.2. Strontium Analysis	16
1.3. Archaeological context	17
2. Objectives.....	21
3. Materials and Methods	25
3.1. Teeth sampling.....	27
3.1.1. Selection of individuals	27
3.1.2. Selection of teeth	27
3.1.3. Collection and storage.....	27
3.2. DNA Sequencing.....	28
3.2.1. DNA extraction	28
3.2.2. PCR strategies	30
3.2.3. Electrophoresis gel	34
3.2.4. DNA Purification.....	35
3.2.5. DNA Sequencing.....	35
3.2.6. Analysis of sequencing results	36
3.3. Strontium analysis.....	38
4. Results and Discussion.....	41
4.1. Authenticity and feasibility.....	43
4.1.1. Overall PCR results of mitochondrial and nuclear data	44
4.1.2. Results of mitochondrial data.....	45
4.1.3. Results of nuclear data.....	48
4.2. Contribution to archaeological data	50
4.3. Strontium Analysis	56
5. Conclusion and Perspectives	61
5.1. Feasibility and authentication	63
5.2. Archaeological contributions.....	63
6. Bibliography	67
7. Supplementary data	75
S1. R code for Figure 12 and 13.	77

1. Introduction

1.1. Study of ancient DNA

1.1.1. Interest of ancient DNA studies

It has always been difficult for archaeologists to apprehend some facets of past populations. Available data are sometimes not sufficient to understand some issues related to individual characteristics and their daily life.

Ancient DNA analysis can provide new and complementary information to archaeological resources that would not have been possible to acquire otherwise.

The study of DNA of individuals from extinct societies can tell us more about their identity, and a fortiori about their culture. Specific questions such as sex identification, geographic origin, kinship relationships between individuals, diet and the presence of diseases can be answered using genetic data (Adler et al. 2013, Emery et al. 2018a, Haak et al. 2008, Mann et al. 2020, Vai et al. 2020, Vanek et al. 2009, Weyrich et al. 2017).

Availability of such information can lead to new interpretations of archaeological remains and to a better understanding of ancient societies.

1.1.2. Difficulties of ancient DNA studies

The study of ancient DNA has two major challenges: DNA degradation and contamination.

1.1.2.1. Degradation

After death, enzymatic reactions and microorganism invasions lead to the decomposition of the body and to the degradation of DNA (Allentoft et al. 2012, Pääbo et al. 2004, Peyrégne and Prüfer 2020). DNA degradation is a real problem as it makes genetic analysis more challenging. DNA degradation consists in various mechanisms such as strand breaks, hydrolysis and oxidation. Strand breaks are mostly due to microorganism nucleases and depurination (Figure 1). It causes DNA fragmentation (Pääbo et al. 2004). Hydrolysis: It provokes the loss or the modification of bases via depurination (Kistler et al. 2017, Llamas et al. 2017). For example, cytosine can be deaminated and transformed into uracil which is then detected as thymine during sequencing (Figure 1) (Llamas et al. 2017, Meyer et al. 2012, Orlando et al. 2011, Peyrégne and Prüfer 2020). This is referred as miscoding or misincorporation (Mann et al. 2020). Adenine can also be transformed into hypoxanthine, guanine into xanthine and 5-methylcytosine into thymine (Pääbo et al. 2004). This results in a misreading of DNA sequences. Oxidation: It causes base or deoxyribose residue damages. This leads to base and sugar fragmentation or nucleotide modification (Lindahl 1993, Pääbo et al. 2004).

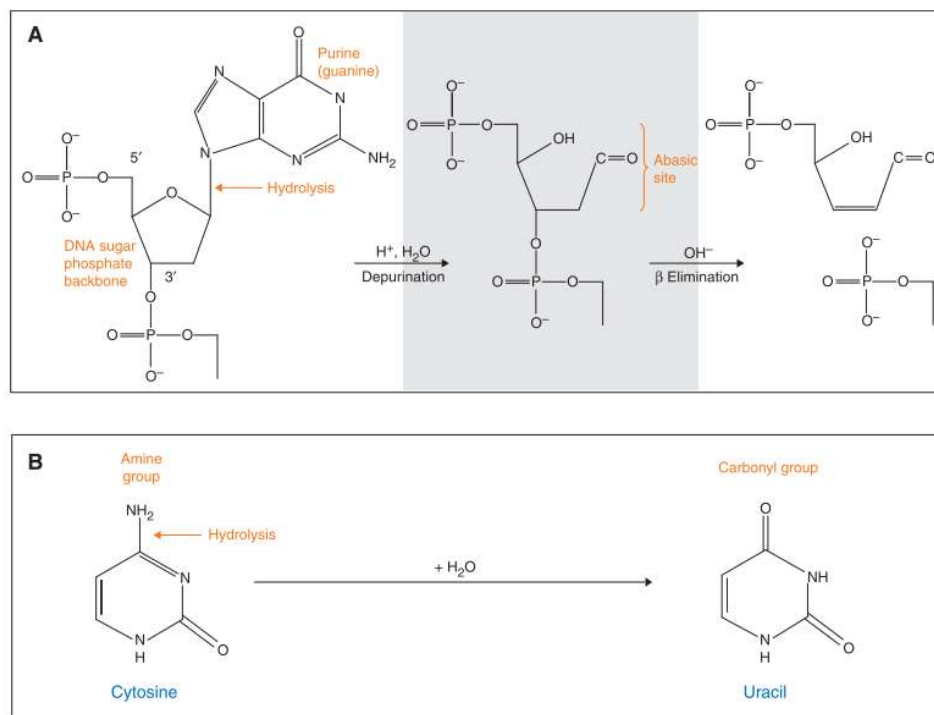


Figure 1. Fragmentation and deamination. (A) A likely cause of fragmentation in ancient DNA is depurination, in which the N-glycosyl bond between a sugar and an adenine or guanine residue is cleaved, resulting in an abasic site. The DNA strand is then fragmented through beta elimination, leaving 3'-aldehydic and 5'-phosphate ends. (B) Deamination of cytosine to uracil is the major mechanism leading to miscoding lesions in ancient DNA. 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 (Dabney et al. 2013b).

Ancient DNA that can be extracted and analyzed is therefore composed of extremely short fragments and also contains post-mortem degradation. (Dabney et al. 2013a, Mann et al. 2020)

Mitochondrial DNA is often more successfully preserved than nuclear DNA. Indeed, each cell has a large number of copies of mitochondrial DNA, whereas it has only one copy of nuclear DNA. Mitochondrial DNA is therefore much more likely to be preserved and easier to be analyzed (Latham and Miller 2019).

1.1.2.2. Contamination

Contamination is one of the most significant challenge in the analysis of ancient DNA.

The only DNA material still available in bone and teeth for genetic analysis consists of a small amount of short DNA fragments, mainly mitochondrial DNA that have been damaged (Peyrégne and Prüfer 2020). The size of ancient DNA fragments is often between 100 to 500 bp (Mohandesan et al. 2004, Pääbo et al. 2004). Even minor contamination by exogenous DNA can significantly influence the results of an analysis (Llamas et al. 2017). Contaminated DNA, even in small quantities, can become dominant in the sample compared to endogenous DNA. As recent DNA is less degraded, it is more likely to be amplified compared to endogenous DNA

(Sampietro et al. 2006). This is especially challenging in studies of human remains, which makes the detection of contamination very complicated (Peyrégne and Prüfer 2020).

“The poor DNA preservation in these samples means that the total concentration of ancient DNA is low and that the contamination can constitute a large proportion of aligning sequences.” (Peyrégne and Prüfer 2020).

Over time, several studies have already revealed the presence of contamination by modern DNA in their studies:

“Examples include human DNA PCR amplified from archeological and historical specimens of calcified tissues from pigs (Richards et al. 1995), cave bears (Hofreiter et al. 2001), foxes (Wandeler et al. 2003), dogs (Malmström et al. 2005), and Neandertals (Krings et al. 1997, Lalueza-Fox et al. 2005, Serre et al. 2006).” (Sampietro et al. 2006).

There are several possible sources of contamination, affecting the samples before their discovery, during archaeological excavations and until DNA sequencing such as environmental DNA and human DNA (Gilbert et al. 2005a, Llamas et al. 2017, Malmström et al. 2007, Pilli et al. 2013, Sampietro et al. 2006).

Indeed, microorganisms present in the direct environment of the bones, such as soil or water, can penetrate the bones and their DNA can therefore be confused with endogenous DNA. Exogenous DNA can therefore contaminate bones, be amplified during PCR and be sequenced. The step where bones are removed at the time of their discovery is therefore important, to avoid as much as possible the contact of bones with external material.

Human DNA is the most common contamination. It may occur at any stage during the treatment of bones from their discovery to the DNA sequencing. Different sources of such contamination exist:

- Modern DNA through handling, manipulation or during experiments if no particular precautions have been taken.
- Contamination through lab equipment.
Such contamination is mostly avoided by using sterilized materials and by dedicating a room exclusively to the study of ancient DNA.
- Cross-contamination between samples.
If bones of different individuals get in contact, they might contaminate each other.
- Cross Contamination between DNA extract.
DNA from one tube may contaminate another sample during laboratory analysis.
These two kinds of contaminations are the most complex to detect since the contamination is caused by ancient DNA.
- Contamination of reagents used in the different analysis.

Reagents may be contaminated, so it is important to ensure reagents are sterile. The use of new reagents that have been UV-treated can reduce the risk of contamination.

One of the most significant contaminations comes from the handling of bones by archaeologists and anthropologists during excavations as well as during bone studies (Pilli et al. 2013). But the contamination seems to be the most important just after excavation (Gilbert et al. 2005b, Sampietro et al. 2006). Indeed, adequate measures are not yet systematized when bones are discovered. In many cases, few measures are taken during the removal of bones out of the ground. As a result, archaeological remains can easily be in contact with exogenous DNA, including environmental and human DNA from archaeologists (Gilbert et al. 2005b, Sampietro et al. 2006). Despite all precautions, a very rigorous protocol and high criteria for validating and authenticating results, it is almost impossible to completely prevent low levels of contamination (Peyrègne and Prüfer 2020, Pilli et al. 2013). Moreover, proof cannot be given that the sequences produced are authentic and free of contamination (Pilli et al. 2013).

All bones of the body are not affected equally by contamination. Some human remains are less prone to exogenous contamination, particularly teeth. In fact, teeth are very dense and are surrounded by a layer of enamel (Pilli et al. 2013, Sampietro et al. 2006). Samples for DNA extraction can be collected from an area of the tooth where the surface and the enamel are removed. This avoids external contamination as much as possible.

1.1.3. Authenticity criteria

Given the multiple possible sources of contamination, the sequences obtained should be authenticated as old DNA. Several criteria can be used to distinguish contaminated sequences from authentic sequences.

Indeed, some characteristics of ancient DNA can be used to discriminate it from modern DNA (see below). This allows the identification of contaminations and thereby checks the authenticity of results obtained by sequencing. Various methods can be used for this purpose, but as time and financing are limited, not all arguments will be explored in this work.

Negative controls should be added at all stages of the assays, i.e. DNA extraction and PCR preparation. Controls are performed following exactly the same process as the samples but without any addition of DNA. This allows identification of the step during which contamination has occurred (Peyrègne and Peter 2020).

Proceeding to several extractions and independent analyses is also interesting. The results can then be confirmed if they are identical. To avoid contamination and to confirm the results, it is even recommended to perform the analyses in different laboratories. The inclusion of a non-human control in the samples may also be of interest as a control of the presence of modern human contaminating DNA (Llamas et al. 2017, Sampietro et al. 2006).

Because of its fragmentation, ancient DNA is shorter than modern DNA. Amplifications of short sequences should therefore have a higher success rate than amplifications of longer

sequences. The shorter the sequence, the higher the success rate should be (Dabney et al. 2013a, Malmström et al. 2007, Mann et al. 2020, Renaud et al. 2019).

The presence of patterns where the cytosines are detected as thymines due to their deamination can also help to identify ancient DNA sequences. The presence of both nucleotides C and T in one position of a sequence, or a mutation presenting a Thymine instead of a cytosine can reveal this pattern (Mann et al. 2020, Peyrègne and Peter 2020).

Some programs have also been developed to estimate the rate of contamination and to identify authentic sequences. Contamix and AuthenticCT are two examples (Peyrègne and Peter 2020).

1.1.4. Genetic markers used in ancient DNA studies

The majority of human genetic variations are SNPs (Single Nucleotide Polymorphism), in other words it is the variations of a single pair of bases that can be identified in DNA and thus serve as a marker (Jobling 2009). Such variations rarely occur exactly the same way more than once. Two individuals having exactly the same SNP in their DNA have most likely inherited it from a common ancestor, even a very far distant one (Figure 2) ("Nature Education" 2014b, Jobling 2009).

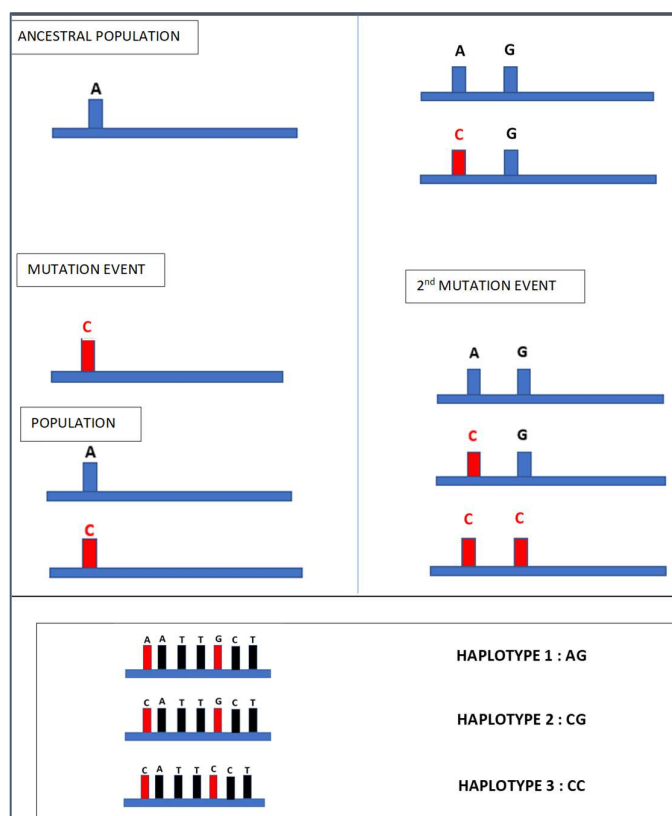


Figure 2. Schematic representation of the formation of haplotypes associated with SNPs. (inspired by Université-Pierre-et-Marie-Curie s. d.)

Genes on closely related loci are normally transmitted together during gamete formation and form what is called a haplotype (linked genes) ("Nature Education" 2014a, Gabriel et al. 2002). To be considered as a polymorphism, SNPs must be present in at least 1% of the population. SNPs account for about 10% of the available variation (Raven et al. 2007). Certain associations of SNPs are characteristic of certain populations and therefore allow the study of possible movements of these populations and the kinship relationships between different individuals.

SNPs are used as markers to identify and characterize haplotypes. Those can also be determined by some other marker, microsatellites that are short DNA sequences (typically 3 or 4 bases long) repeated once or several times (Figure 3). These are also called STR (short tandem repeat). Genetics transmission of these sequences follows Mendel's law. The variation of these markers occurs more frequently than for SNPs. Different chromosomes can carry a variable number of specific STRs. (Jobling 2009)

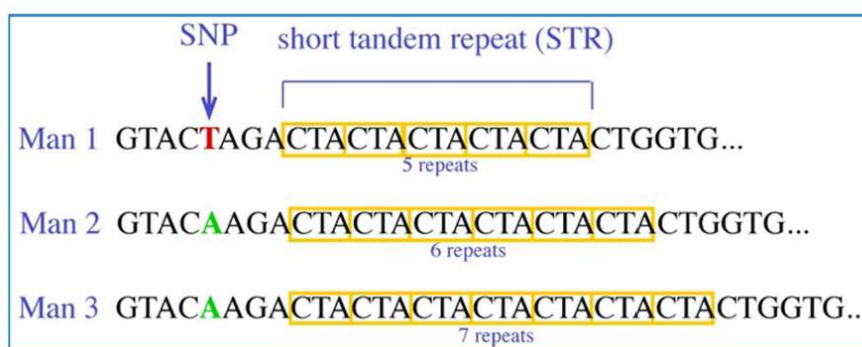


Figure 3. Example of STR (Jobling 2009).

1.2. Strontium Analysis

Individuals' origins may be hard to determine based on archaeological remains. Material goods found in graves can provide some clues about some individuals' origins. Presence of foreign artefacts may be associated with a foreign origin. The Romanization of the conquered regions resulted in the adherence to Roman traditions by local populations (Mathisen and Shanzer 2011, Pearce 2014). This is called acculturation. It is therefore possible to find typical objects of Roman culture in the graves belonging to local populations of conquered regions. Nevertheless, this cultural exchange is not a one-way street, the Roman populations appropriate sometimes some aspects of foreign traditions (Mathisen and Shanzer 2011, Pearce 2014). In this context, it is often hard to differentiate an individual of Roman or local origin.

Strontium (Sr) ratio analyses ($^{87}\text{Sr}/^{86}\text{Sr}$) are useful in studies looking at population movements or at the origin of individuals (Emery et al. 2018a, Milella et al. 2019).

Through soil erosion, strontium is integrated into plant tissues and thereby enters the food chain (Eckardt et al. 2009). Strontium is then incorporated into bones and teeth by ionic substitutions of Calcium (Emery et al. 2018b, Schweissing and Grupe 2003).

The strontium ratio varies geographically and thus enables the establishment of geographical origin of the individuals by comparison of archaeological data with regional environmental data (Barberena et al. 2017, Emery et al. 2018b, Gregoricka and Sheridan 2017, Prowse et al. 2010). The strontium ratio is related to the food ingested during tissue mineralization (Bentley 2006, Bowen 2010). All teeth are not formed at the same time in an individual's life, e.g. milk teeth are formed in utero and therefore reflect the maternal diet. It is therefore possible to obtain isotopic information from birth to approximately 14 years old. Strontium isotope analyses thus provide information on the geographical origins of individuals, i.e. where they lived between birth and childhood (Emery et al. 2018b).

The comparison of Strontium isotope ratios of different individuals may therefore reveal whether they come from similar or distinct origins. The data from the archaeological remains can then be compared with geographical maps indicating the strontium levels, which makes it possible to establish a more or less precise origin of the individuals (Bataille and Bowen 2012, Emery et al. 2018b, Hoogewerff et al. 2019, Stark et al. 2020).

1.3. Archaeological context

The Roman Empire has been studied in many ways by archaeologists and historians over time. The knowledge about Roman society is extremely rich and increasingly detailed. Nevertheless, genetic data on these populations are still scarce (Antonio et al. 2019). And yet, in many societies, economic and social relations have a biological and therefore also genetic connotation (Alzualde et al. 2007). Burials frequently provide the only evidence of the biological characteristics of individuals and of certain aspects of society (Alzualde et al. 2007, Binford 1971). Most studies investigate variability in funeral rites and in biological characteristics mainly through morphological analyses of skeletons and through a detailed study of objects discovered in graves (Alzualde et al. 2007, Binford 1971). Thanks to the development of new techniques, ancient DNA studies increase and provide the opportunity to explore new aspects of past populations, such as genetic variability (origin and kinship relationship), sex determination, diet, diseases, ... (Adler et al. 2013, Alzualde et al. 2007, Emery et al. 2018a, Haak et al. 2008, Weyrich et al. 2017).

Given the expansion of the Roman Empire and due to the roman integration policy of population, people from various origins entered into contact over time in accordance with geopolitical affiliations, notably through population movements throughout the Roman Empire including the Near East and Europe. Such contacts could be reflected in genetic flows identifiable through the genetics analysis of individuals (Antonio et al. 2019).

The site studied in this work concerns the Late Roman Empire necropolis on the Rue Perdue in Tournai, Belgium (Figure 4). The city that grew in Roman times (Brulet 1994) was strategically located on the banks of a navigable river and at the crossroads of the roads connecting Bavay, Cassel, and Arras (Brulet 2014, Faider-Feytmans 1951).

The knowledge of the Roman city of Tournai was gathered thanks to several excavation campaigns resulting from several construction projects in the urban center (Figures 4) (Brulet 1990). Hence, the site of the Rue Perdue was also subject to various excavations campaigns (Figure 4 and 5). It was determined that the necropolis was used during the 3rd and 4th centuries AD (Brulet 1994).

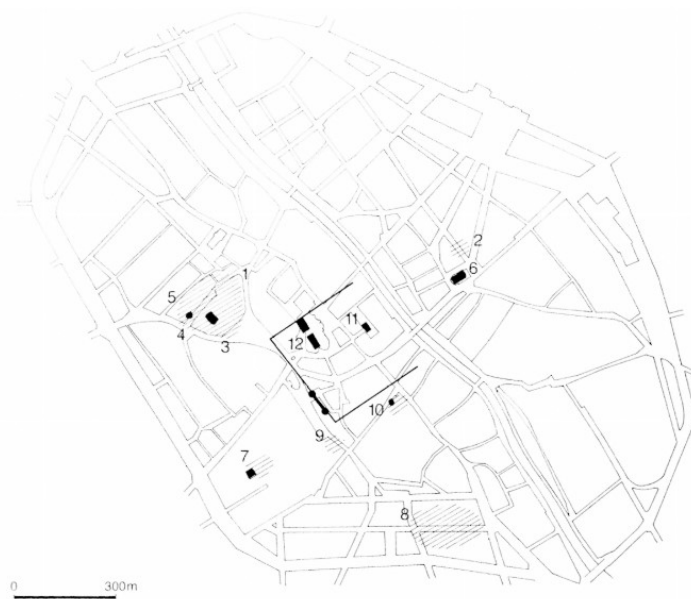


Figure 4. Schematic map of the main archaeological sites of Tournai, against the backdrop of current parcels. 1. Grand-Place; 2. Rue Monnel; 3. Saint-Quentin; 4. Ancien Théâtre; 5. Rue Perdue; 6. Saint-Brice; 7. Saint-Martin; 8. La Citadelle; 9. Rue d'Espinoy; 10. Saint-Piat; 11. Saint-Pierre; 12. Cathedral. (Brulet 1994)

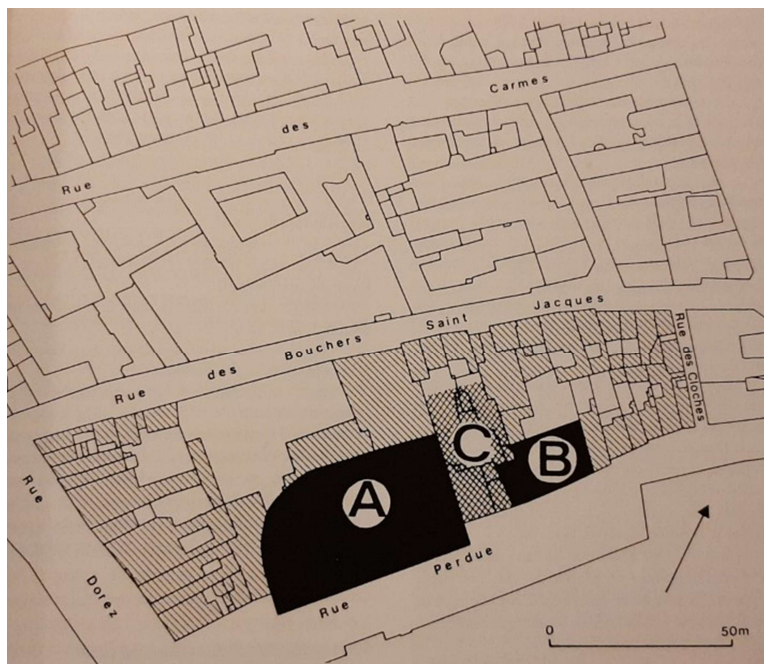


Figure 5. Location of the excavated areas on Rue Perdue: sectors A and B (excavations from 1967 to 1973), sector C (excavations from 1989) and under the street (excavations from 2010) (Brulet 1996).

Over 400 graves have been discovered during the various excavations occurred in this zone (Figure 6) (Brulet 1996). The necropolis' surface was very extensive. The burials were not arranged in any precise order nor organization. The orientation of the tombs varied, and the

whole ensemble of tombs did not seem to be part of a regular pattern that could indicate circulatory paths (Brulet 1996). Nevertheless, ten distinct zones could be distinguished in which a large number of tombs were concentrated in a more restricted space of 6 to 8m in diameter (Figure 6). Graves present in these concentrations overlapped regularly. These concentrations had therefore enabled the relative dating of a large part of tombs discovered in the necropolis (Brulet 1996).

The general organization of the necropolis with its uncommon concentrations could coincide with family concessions, where inhumations would have succeeded one after each other at the same place without taking into consideration the orientation of the tombs. Another interpretation could be the presence of plantations, which may explain the empty areas without graves. Nevertheless, this plan must be treated with caution. In fact, some of the empty burial areas could be zones that archaeologists were unable to investigate during the excavations. The reason for the necropolis' organization remains unanswered and could be resolved through genetic analysis. (Brulet 1990, Brulet 1996)

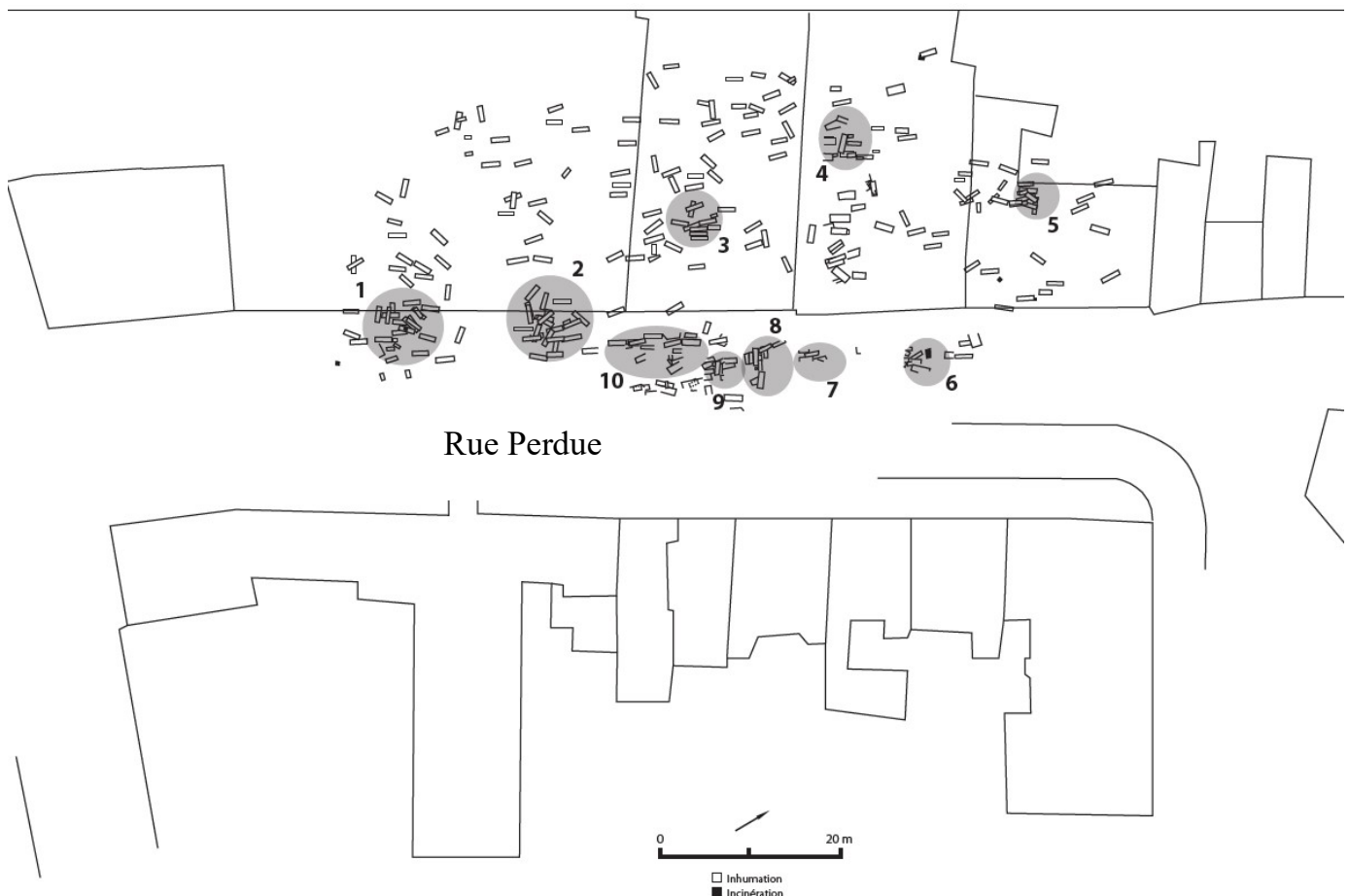


Figure 6. Overview of the concentrations. The graves are represented by rectangular boxes. The grey circles delimit the concentrations. Each one is assigned with a number. (Verslype and Weinkauff n.d.)

2. Objectives

This work is part of a current genetic analysis research on archaeological remains. It is a multi-disciplinary project linking molecular biology with archaeology. This work is conducted in collaboration with the Belgian Institute of Natural Sciences (IRSNB), where the analyses were performed, and the “Centre de Recherches d'Archéologie Nationale” (CRAN), which provided the archaeological material.

Given the complex situation with COVID-19 since March 2020, it was necessary to adapt the objectives of this work. The methods and techniques that could have been experimented and used during this project have been limited due to the reduced time, restricted access to the laboratory since mid-October 2020 and limited funding (more details below).

In this context, this work pursued two main objectives.

The first objective was to assess the feasibility of ancient mitochondrial and nuclear DNA extraction and analysis from teeth of human skeletons excavated from the Gallo-Roman necropolis on Rue Perdue in Tournai. Hence, once DNA was extracted and sequenced, we tried to evaluate the authenticity of the results.

The second objective concerned the contribution of these results to archaeological studies. It aimed at answering several questions. Firstly, the geographical origin of the individuals will be investigated by mitochondrial haplogroup. In this respect, a complementary approach was applied on the basis of Strontium ratio analysis. It could lead to the identification of population movements during the Late Roman Empire. Secondly, it will attempt to determine whether the concentrations are associated with family clusters. Finally, confirmation of the sex of individuals will be attempted on the basis of nuclear DNA data.

3. Materials and Methods

3.1. Teeth sampling

3.1.1. Selection of individuals

The selection of individuals was based on the file compiled by archaeologists Erika Weinkauf and Laurent Verslype, which includes all the data collected on the archaeological remains (Verslype and Weinkauf n.d.). They include for each individual: the burial number, relative dating, the state of conservation of the grave, the identification of the deceased (age, sex, etc.), the state of preservation of the skeleton, the state of preservation of the teeth, the pathologies detected on the skeleton, the place of conservation (Archaeological Museum of Tournai, IRSNB or lost) and the attributed number of the concentration.

A detailed study was also carried out to identify the skeletons that are kept in museum reserves, available for analysis, and the skeletons that were lost. From the available remains, the first step was to select the individuals whose teeth were preserved, as for many individuals the teeth were missing from the skeletons.

3.1.2. Selection of teeth

A visual examination of the teeth was then carried out before they were removed. The choice of the extracted teeth was based on two main criteria:

- Removability of the teeth. Teeth that were already detached from the skeleton or were easily extracted were preferred in order to manipulate the bones as little as possible.
- The state of preservation, which was determined via the general appearance of the tooth, presence of caries or cracks, and via its density, which was assessed by under-weighing the tooth in the hand. Indeed, high bone porosity is a good indicator of weak DNA preservation (Damgaard et al. 2015).

3.1.3. Collection and storage

Teeth were collected, with a mouth mask and hair protection, on a table using latex gloves and dissecting nippers. The surface and instruments were cleaned with bleach and 70% denatured ethanol before and between each individual. The gloves were also changed between each sampling. The extracted teeth were stored in individual sealed bags annotated with the burial number, the number of teeth taken and a brief description of them.

Whenever possible, the dental calculus present on the teeth was also sampled, using the dissection tools. The resulting fragments were collected in 1.5 ml Eppendorf tubes and placed in the bags containing the teeth of the same individual.

The bags were then rolled up in aluminum foil and later placed in a Styrofoam container at room temperature.

3.2. DNA Sequencing

3.2.1. DNA extraction

DNA extraction and PCR preparation were done in a room specially dedicated to ancient DNA, equipped with UV lamps, positive pressure air system and the necessary equipment to be independent and to avoid contamination as recommended (Adler et al. 2013, Weyrich et al. 2019).

Before any laboratory experiment, surfaces and necessary equipment were cleaned with bleach and exposed to UV light for at least 20 minutes. Other precautions were also taken: wearing haircap and gloves taped to the lab coat. A second pair of gloves was put on and changed before and between the manipulation of each sample.

Material

- Dremel
- Aluminium
- Crosslinker
- Lighter
- Parafilm
- Balance
- Vortex
- Eppendorf centrifuge (15.000xg)
- 50ml centrifuge (2.000xg)
- Thermoshaker
- Roche Kit
- Buffers (see buffers preparation)

Buffers preparation

Buffers are prepared before the extraction process as described in Table 1 to 5.

Table 1. Preparation of Proteinase K solution

Products	For 10 ml
Proteinase K	100 mg
H ₂ O	10 ml

The Proteinase K solution is stored at -20°C

Table 2. Preparation of Extraction buffer

Products	For 10 ml
0.45M EDTA	9.25 ml
Proteinase K solution	250 µl

EDTA has been UV-irradiated for 30 min before adding Proteinase K solution.

Table 3. Preparation of Binding buffer

Products	For 50 ml
GuHCL	23.88g
H ₂ O	Fill to 30 ml
Isopropanol	Fill to 50 ml
Tween-20	25 μ l

The buffer has been UV-irradiated for 30 minutes before adding Tween-20, which prevents the DNA from being fixed to the side of the tubes.

Table 4. Preparation of TE buffer

Products	For 12 ml
H ₂ O	9879.9 μ l
1M Tris-HCl (pH 8.0)	99.94 μ l
0.5M EDTA (pH 8.0)	19.97 μ l

The buffer has been UV-irradiated for 30 minutes.

Table 5. Preparation of TTE buffer

Products	For 12 ml
TE-buffer	11.994 ml
Tween-20	6 μ l

Method (based on Dabney et al. 2013a)

- The teeth are first placed on a sheet of aluminum foil and exposed to UV light for 10 minutes, with the surface which will be sampled facing the UV light.
- The surface of the teeth are scraped off with a drill wrapped in plastic film to avoid contamination.
- The teeth are then placed back under UV light for 10 minutes.
- The cleaned surface of the teeth is scraped off with a drill. Between 50 and 60 mg of powder is collected for each tooth. For the dental calculus, the powder collected on the bones is simply weighed to reach the same weight. Between each sample, the surfaces are cleaned with bleach and 70% ethanol, the plastic film of the drill is changed and the instruments that have been in contact with the tooth are dipped in bleach, rinsed with 70% ethanol and finally put under a flame to avoid contamination as much as possible.
- 900 μ l of bleach is added to each sample. The tubes are then centrifuged for 1 minute at 12000 rpm and the supernatant is discarded. This step is then repeated 5 times with pure water. The powder is resuspended by pipetting up and down. This cleaning step is

optional and was only performed for the first extraction. A negative control is also added at this step.

- 1 ml of extraction buffer is added to 30-50 mg of powder. The samples are vortexed, surrounded by parafilm, and incubated overnight at 37°C in the thermoshaker.
- For each 1 ml of extraction buffer, 400 µl of 3M sodium acetate pH 5.2 and 10 ml of binding buffer are added in 50 ml tubes.
- The incubated tubes are centrifuged at maximum speed for 2 minutes and the supernatant is transferred to the 50 ml tubes containing the binding buffer and sodium acetate.
- The contents are transferred to the High Pure Extender from the Roche kit and are centrifuged at 2,000 x g for 4 minutes. The columns are then rotated by 180° and centrifuged again at 2,000 x g for 2 minutes.
- The columns are detached from the extender, placed in 2ml collection tubes and dry centrifuged at 8.000 x g.
- 450µl PE buffer is added to each column. They are centrifuged at 8.000 x g for 30 seconds. The collection tube is changed and the step is repeated once. The columns are then dry centrifuged at maximum speed and placed in 1.7 ml siliconized tubes.
- 45 µl of TTE buffer is added, left for 5 minutes and centrifuged at maximum speed for 1 minute to obtain the first DNA elution. This step has been repeated to obtain a second DNA elution. Elutions are stored at -20°C.

3.2.2. PCR strategies

Material

- PCR Buffer (10x) (Life Technologies Europe/ThermoFisher Scientific. Catalog Number: 10966034)
- MgCl
- dNTP
- Polymerase (taq Platinum) (Life Technologies Europe/ThermoFisher Scientific. Catalog Number: 10966034)
- H₂O
- Primers
- DNA samples
- Thermocycler (T-Personnal Thermal Cycler Biometra)

Method

First a Master Mix is prepared as summarized in Table 6 according to the final volume of the PCR reaction (see Table 7). A working solution of primers at 20 μM are prepared from the stock (100 μM).

Table 6. Preparation of Master Mix for PCR

Products	For 1 reaction (25 μl / run)	For 1 reaction (15 μl / run)	Concentration in the working solution	Final concentration
PCR Buffer	2.5 μl	1.5 μl	10x	1x
MgCl	0.75 μl	0.45 μl	50 mM	1.5 mM
dNTP	2.5 μl	1.5 μl	2 mM each	0.2 mM
Polymerase (taq Platinum)	0.25 μl	0.45 μl	5 Units/ μl	0.05 units/ μl
H ₂ O	14.25 μl	8.8 μl	-	-
Forward primer	1.25 μl	0.45 μl	20 μM	See table 7*
Reverse primer	1.25 μl	0.3 μl	20 μM	See table 7*

* Final concentration is adapted for each PCR.

The Master Mix is then distributed in the PCR tubes (22.5 μl for the reactions of 25 μl and 10 μl for those of 15 μl volumes). Template DNA (from elution 1 or 2) is added to each tube (2.5 μl for 25 μl volumes and 5 μl for 15 μl volumes). All PCRs were performed from the first elution of the DNA extractions except for S12, S13, S14, S15, Sex100 and MAY where the second elution of the same extraction was used (because the first elution was almost completely used). After a quick centrifugation, the PCR tubes are placed in a thermocycler following the parameters of Figure 4 for PCR volumes of 25 μl and Figure 5 for PCR volumes of 15 μl MA and MAY. For every PCR a negative control PCR reaction is added where no template DNA is added in the mix.

Table 7. Primers used in the study. The primers were selected on the basis of several articles. (Haak et al. 2008, Masuyama et al. 2017, Mohandesan et al. 2004, Sampietro et al. 2005)

PCR	Primers	Sequence	Final Concentration in PCR	PCR Volume (μl)	Annealing temperature (°C)
S2	L16055	GAAGCAGATTGGGTACCAC	1 μM	25	50
	H16158	TGTGGATTGGGTTTTATGT	1 μM		
S4	L16122	CATTACTGCCAGCCACCATGAATA	1 μM	25	59
	H16233	GCTTTGGAGTTGCAGTTGATGTGT	1 μM		
S5	L16131	CACCATGAATATTGTACGGT	1 μM	25	52
	H16247	CAACTATCACACATCAACTGCAA	1 μM		
S8	L16223	AGCAAGTACAGCAATCAAC	1 μM	25	52
	H16356	GTCATCCATGGGGACGAGAA	1 μM		
S9	L16261	CAACTATCACACATCAACTGCAA	1 μM	25	55
	H16401	TGATTTACGGAGGATGGTG	1 μM		
S11	L12227	GAAAGCTCACAAGAACTGC	1 μM	25	52
	H12341	GGTTATAGTAGTGTGCATGG	1 μM		
S12	L6999	CAAACTCATCACTAGACATCG	1 μM	25	52
	H7066	GAATGAAGCCTCCTATGATGG	1 μM		
S13	L13669	CACCTTACTAACATTAACG	1 μM	25	50
	H13725	TAGTAATGAGAAATCCTGCG	1 μM		
S14	L13257	AATCGTAGCCTTCTCCACTTCA	1 μM	25	50
	H13372	TTGTAAAGTTGTGGATGAT	1 μM		
S15	L2993	TAGGGTTTACGCCTCGATGT	1 μM	25	58
	H3054	TCTGAACTCAGATCACGTAGGACT	1 μM		
H4	L16287	CACTAGGATACCAACAAACC	1 μM	25	52
	H16410	GCGGGATATTGATTTACGG	1 μM		
H5	L00034	TCTATCACCCCTATTAACCACTCAC	1 μM	25	52
	H00177	TTAGTAAGTATGTTGCGCTGTAAT	1 μM		
H6	L00144	CGCAGTATCTGTCTTTGATTCCTG	1 μM	25	57
	H00327	TTGGCAGAGATGTGTTAAGTGCT	1 μM		
H8	L00274	TGTCTGCACAGCCACTTTCCACAC	1 μM	25	57
	H00397	AGTGCATACCGCCAAAAGATAAAA	1 μM		
Sex100	British_Forward	CCCTGGGCTCTGTAAAGAATAGTG	0.4 μM*	25	60
	British_Reverse	ATCAGAGCTTAAACTGGGAAGCTG	0.4 μM*		
MS	AMELY-Exon 2-A	IIIIIGGATTTTGTTCCTGCCTTG	0.3 μM	15	66
	AMELX-Exon 2-A	GGATTTTATTTGCCTGCCTCC	0.3 μM		
	AMEL-Exon-2-A-R	IIIIIIIIACAGGCATGGCAAAAGCTG	0.4 μM		
MA	AMELY-Exon 2-B	IIIIITGGCAAAAGCTGCTCCAC	0.4 μM	15	64
	AMELX-Exon 2-B	TGGCAAAAGCTGCTCCAG	0.4 μM		
	AMEL-Exon-2-B-R	TCAAGAAATGGGGACCTGGATT	0.4 μM		
MY	SRY-F	TCGGGTAACATTGGCTACAAAGA	0.4 μM	15	66
	SRY-R	GTTATCGTAAAAAGGAGCATCTAGGTAG	0.4 μM		
MAY	AMELY-Exon-2-A	IIIIITGGCAAAAGCTGCTCCAC	0.4 μM	15	64
	AMELX-Exon-2-A	TGGCAAAAGCTGCTCCAG	0.4 μM		
	SRY-F	TCGGGTAACATTGGCTACAAAGA	0.4 μM		
	AMEL-Exon-2-A	TCAAGAAATGGGGACCTGGATT	0.4 μM		
	SRY-R	GTTATCGTAAAAAGGAGCATCTAGGTAG	0.4 μM		

* Two PCR Sex100 were done, one with 0.4 μM of final concentration (Elution 1) and another with 1.2 μM (Elution 2).

In the first case (Figure 7), steps 1 and 2 correspond to the denaturation of the DNA. Step 3 stands for the hybridization of the primers. The temperature of this step varies according to the primers used (Table 7). Steps 4 and 5 are the elongation phase. Step 6 is used to stop the reaction.

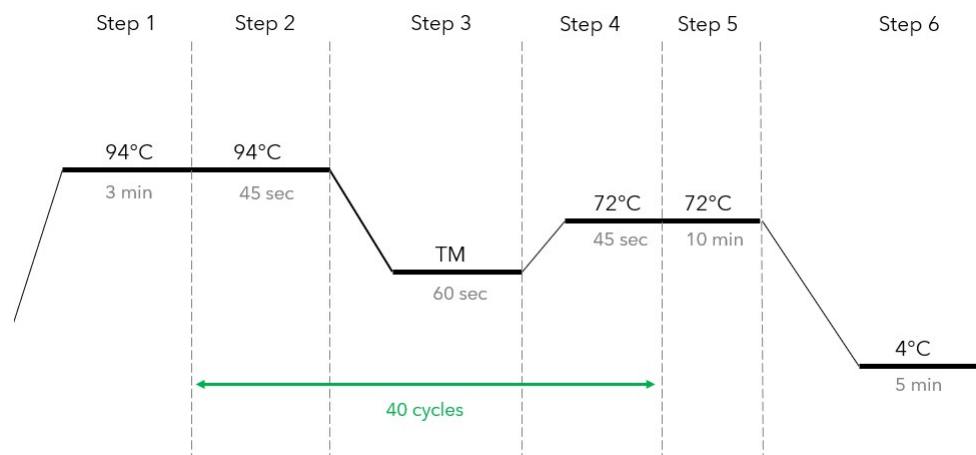


Figure 7. Parameters for PCR of 25 μ l and 15 μ l except MA and MAY.

In the second case (Figure 8), steps 1, 2 and 5 correspond to denaturation. Steps 3 and 6 are the step of primer hybridization, whose temperature also varies according to the primers. Steps 4, 7 and 8 correspond to the elongation and the last step stops the reaction. A schematic overview of all PCRs performed is summarized in Figure 9.

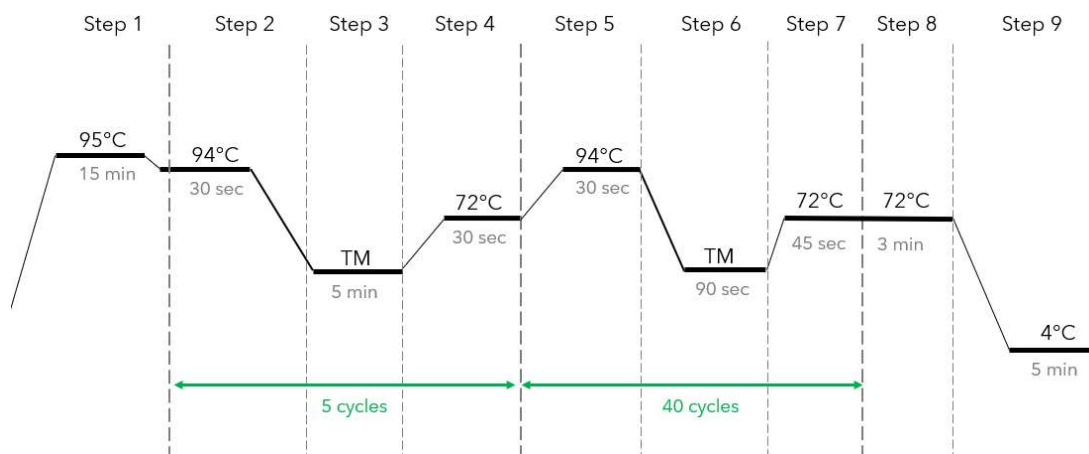


Figure 8. Parameters for PCR of 15 μ l MA and MAY.

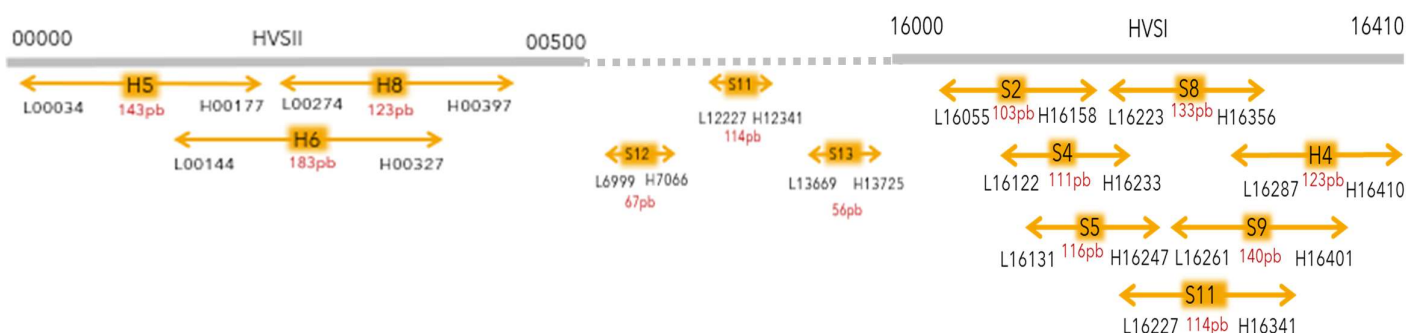


Figure 9. Schematic overview of all the PCRs. The PCR's are indicated by the orange arrows that are placed in relation to the position on the human mitochondrial genome (grey line). HVS I correspond to the mitochondrial Hypervariable Sequence I, while HVS II correspond to the mitochondrial Hypervariable Sequence II.

3.2.3. Electrophoresis gel

Material

- PCR samples
- Agarose I (Molecular Biology Grade (ThermoFisher Scientific, cat. 17852)
- TBE (Tris, Borate, EDTA buffer made in the lab of RBINS) 1x
- Midori Green (Nippongenetics, cat. MG04)
- Purple (6X) loading dye (New England Biolabs, cat. B7025S)
- Ladder (1KB Plus DNA Ladder (Invitrogen/ThermoFisher Scientific, cat. 10787018) or GeneRuler Low Range DNA Ladder (ThermoFisher Scientific, cat. SM1193))

Method

A 2% agarose gel is prepared following Table 8. The agarose is mixed with TBE 1x in an Erlenmeyer and heated in the microwave. After the mixture has cooled slightly, the midori green is added. The liquid is then poured into the tray in which the combs have been placed. After 30 minutes of polymerization, the gel is put in the electrophoresis tank. The ladder and the samples are mixed with the loading dye and loaded into the gel. The migration takes 40 minutes at 90 mV. The gel is then placed in a transilluminator to visualize the DNA. For the PCRs MS, MA and MY, a 3% agarose gel was used (1.8 g of agarose) and the migration took 55 minutes at 70V.

Table 8. Preparation of an 2% agarose gel

Products	For small gel (20 wells)	For big gel (28 wells)
agarose	1.2 g	2.4 g
TBE 1x	60 ml	120 ml
Midori Green	2.4 µl	4.8 µl

3.2.4. DNA Purification

The samples to be sequenced must be purified. The selection of samples for sequencing is based on the PCR results. The purification improves sequencing signal by removing free dNTPs, excess primers, buffer components and contaminants such as proteins or RNAs.

Material

- ExoSAP-IT PCR Product Cleanup reagents (FastAP Thermosensitive Alkaline Phosphatase (1 U/ μ L) ThermoFisher Scientific, cat. EF0651 and Exonuclease I (20 U/ μ L) ThermoFisher Scientific, cat. EN0581.
- Thermocycler (T-Personal Thermal Cycler (Biometra))

The ExoSAP mix is prepared as described in Table 9.

Table 9. Preparation of the ExoSAP mix

Products	For 600 μ l
Exonuclease I (Enzyme Exonuclease I ; EN0581 ; 4000U (20 U/ μ l))	100 μ l
FastAP (Thermosensitive alkaline phosphatase ; EF0651 ; 1000U (1 U/ μ l))	200 μ l
Buffer (exonuclease)	30 μ l
H2O	270 μ l

Method

- 15 μ l of each PCR sample is mixed with 3 μ l of ExoSAP mix.
- The mixes are incubated at 37°C for 15 minutes to degrade the remaining primers and nucleotides.
- The mixes are then incubated at 85°C for 15 minutes to inactivate the ExoSAP reaction.

3.2.5. DNA Sequencing

Sequencing was performed using the Sanger method. This technique is based on the principle of DNA replication using nucleotides (dNTPs) and dideoxynucleotide (ddNTPs). ddNTPs are labeled with four fluorescent dye, one for each type of nucleotide. When dNTPs are added the chain extension can continue, but it is stopped once a ddNTP is added. The sequencing results are based on the formation of extension products of various lengths. The extension products are separated by their sizes using capillary electrophoresis. Fluorescent ddNTPs are then detected by a laser and recorded according to size. (Figure 10)

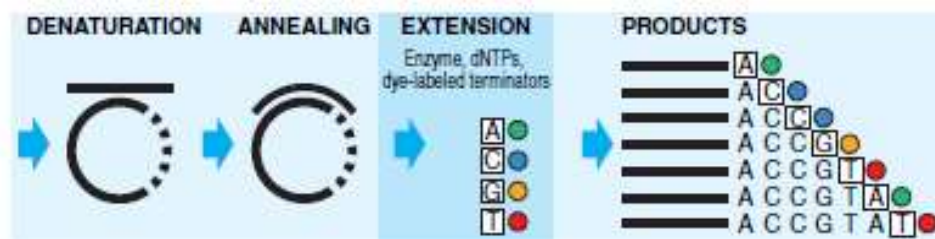


Figure 10. Diagram of dye terminator cycle sequencing (ABI 2009)

Method

Firstly, the primers are diluted to 10 μ M. On a sequencing plate, two wells are associated with a given purified PCR product one to read the sequence in one direction starting from the forward primer and the other for another read starting from the reverse primer. In each well, 5 μ l of the corresponding primer is charged. Finally, 5 μ l of the purified PCR products are transferred in the associated well.

Sequencing can be performed at the RBINS but, due to limited access to the lab, the plates have been sent to MacroGen Europe B.V. using the EZ-Seq - MacroGen Online Sequencing Order System.

3.2.6. Analysis of sequencing results

The sequencing results that were obtained by sequencing were processed on Codon Code Aligner. First of all, the sequences were aligned for each individual DNA extract and for each PCR according to the following parameters (Figure 11). These parameters have been adapted to the short length of the sequences and the degradations related to ancient DNA. However, some sequences could not be assembled. In the rest of the analysis, we used only the sequences of good quality, where peaks could be clearly distinguished. They were then cleaned and the primers were manually removed. The sequences were then checked one by one by examining the peaks.

Parameter	Value
Algorithm	Local alignments
Min. percent identity	50.0
Min. overlap length	10
Min. score	10
Max. unaligned end overlap	80.0
Bandwidth (max. gap size)	30
Word length	6
Max. successive failures	100
Match score	1
Mismatch penalty	-1
Gap penalty	-1
Additional first gap penalty	-1

Figure 11. Parameters used in Codon Code Aligner.

Throughout all sequences obtained, we have frequently observed double peaks that lead to possible misinterpretations in the sequence. When a sequence is of bad quality, multiple peaks are frequent and are due to aspecific PCR amplifications. In contrast, when the sequence is generally clear and of good quality, double peaks can occur in nuclear fragments in cases of heterozygosity, but, in mitochondrial fragments, they can be explained in several ways:

- Contamination by one or more similar sequences that only show differences at the site of double peaks (contaminant human DNA).
- Degradation (characteristic of ancient DNA) such as deamination (Peyrègne and Peter 2020).
- Mistake of the polymerase at the beginning of the PCR cycles.
- Heteroplasmy (which is relatively rare) (Tang and Huang 2010).

Ambiguities were identified on the basis of double peaks in the sequences. They were retained if the complementary sequences indicated the same ambiguity or if no complementary sequence was available to lift the ambiguity.

The sequences were then exported to Geneious. They were mapped to a reference sequence that is standardly used for human mitogenome analysis and that is called the “revised Cambridge reference sequence”, or “rCRS” (accession number in the public GenBank database: NC_012920). The substitutions between the reference sequence (rCRS) and the sequences of the individuals were identified and recorded in a table (Table 10).

Finally, these data were analyzed in Empop using SAM2 database search (Huber et al. 2018) according to the substitutions recovered in Table 10. The following data are entered in the search program: the name of the individual, the range (i.e., regions where sequenced data are available) and the profile (i.e., polymorphisms identified against rCRS). Haplogroups are then identified. When more than one haplogroup is compatible, the haplogroup with the lowest cost is selected. Empop provides the MRCA haplogroup (Most Recent Common Ancestor) if several possibilities exist (Huber et al. 2018).

Table 10. Mitochondrial haplotypes and SNPs against rCRS. For each individuals the concentration, the haplogroups estimated, the range (i.e., regions where sequenced data are available), the profile (i.e., polymorphisms identified against rCRS) and the missing and additional polymorphisms compared to the estimated haplogroup.

Individuals	Concentration	Haplogroup	Range	SNP's against rCRS	Missing Polymorphism	Additional polymorphism
T2	1	M8a	7000-7053 16222-16285 16287-16288 16290 16293 16295-16355	7026A 7028T 7041A 7053A 16223T 16298C 16319A	(none)	7026A 7041A 7053A
T4	1	H5	274-397 7000-7065 12228- 12340 13670-13724 16123-16409	16304C	315.1C	(none)
T8	1	L3	7003-7065 13670-13724 16056-16157	7028T 16126C	(none)	(none)
T10	1	JT	34-397 7000-7065 12228- 12340 13670-13724 16056-16374 16376-	73G 146C 152C 263G 310.1C 316.1C 7028T 16126C	(none)	146C 152C 317.1C
T14	1	L3	7000-7065 13695-13724		(none)	(none)
T17	1	T2b17	34-177 274-397 7000- 7065 12228-12340 13681- 13724 16056-16384 274-397 7000-7053 12228-	73G 304T 306T 310.1T 316.1C 7028T 13692T 16126C 16294T 16304C	(none)	304T 306T 311T 317.1C
T34	2	H5	274-397 7000-7053 12228- 12340 13670-13724 16056-16409	316.1C 16304C 16393T	315.1C	16393T 317.1C
T38	2	L3	7000-7065 12228-12340	7028T	(none)	(none)
T39	2	A+152+16362	7000-7065 13670-13724 16056-16157 16248- 16400	7028T 16151T 16290T 16319A 16362C	(none)	16151T
T42	2	H	303-397 7000-7065 12228- 12340 13670-13724 16056-16232 16288-	313T 316.1C	315.1C	313T 317.1C
T47D2	2	R	13670-13724		(none)	(none)
T47S	2	U5a2	286-397 7000-7036 13682- 13724 16123-16409	316.1C 7028T 16256T 16270T	315.1C	317.1C
T52D1	2	H1bc	7000-7065 13670-13724	7028T 13695T	(none)	7028T
T52D2	2	H	7020-7065 12228-12340 16222-16376	12244A 12275A 12276A 12283A	(none)	2244A 12275A 12276A 12283
T52S	2	R0	34-392 12229-12237 12248-12340 16283- 16409	263G 310.1C 316.1C 12231G	(none)	317.1C 12231G
T53	2	H	276-397 7000-7065 12228- 12340 13670-13724 16056-16355		315.1C	(none)

The identified haplogroups were then placed on a phylogenetic tree of the human mitochondria (Van Oven and Kayser 2009). These data were finally represented manually (Figure 15).

The figures (12 and 13) were realized on the R program (R_Core_Team 2020, R_Studio_Team 2020). The script is available in the appendices (S1).

3.3. Strontium analysis

The radiogenic Sr isotope measurements were performed by Hélène Dailly and Sophie Opfergelt from the MOCA platform (Mineral & Organic Chemical Analysis) of the Earth and Life Institute of UCLouvain. The protocol used for the analyses, which follows, was also written by Hélène Dailly and Sophie Opfergelt. Samples provided were prepared in a clean

laboratory (class 100, laminar flow). Sample dissolution was performed in sealed Teflon vials with HF/HNO₃/HCl (1 : 4.5 : 2) by micro-wave.

The microwave temperature program was set as followed: rising to 180°C within 5 minutes and then left at 180°C during 5 minutes. The organic matter of the samples was eliminated by the action of 2x100 µl of H₂O₂. The samples were then dried at 40°C before being placed in 1ml of 3M HNO₃. A total of 0.2 ml of this solution was then diluted 100x.

The Sr concentration in the resulting solution was measured by ICP-AES. An aliquot of 500 ng of Sr was dried down and picked up in 100 µl of 3M HNO₃. To purify Sr, the sample was loaded onto Biorad microspin column containing 500 µl of pre-cleaned Strontium specific resin (50-100 µm SR-B10S Triskem), and eluted in several stages with HNO₃ (Aciego et al. 2009). The total procedural Sr blank was less than 0.09 ppb Sr, constituting < 0.1% of the total Sr analysed for a typical Sr analysis.

Sr isotope measurements were carried out by MC-ICP-MS (Neptune Plus™ High Resolution Multicollector ICP-MS, Thermo Fisher Scientific, Earth & Life Institute, UCLouvain, Belgium) in wet plasma mode using a PFA nebulizer of 100 µl.min⁻¹ uptake rate. The analyses were performed in 2% HNO₃ matrix, with a typical sensitivity of 7V for 100 ppb Sr. Each sample was measured twice and the results are expressed as ⁸⁷Sr/⁸⁶Sr ratio. External precision on the in-house standard Sr ICP solution was 0.708611 ± 0.000010 (2SD, n = 27). The reference material SRM987 was used to assess the precision of column chemistry and the analytical procedure and yielded a value of 0.710287 ± 0.000027 (2SD, n = 5) consistent with previous studies (e.g., 0.710264 ± 0.000016) (Stevenson et al. 2016).

4. Results and Discussion

4.1. Authenticity and feasibility

The first objective of this work was to estimate the feasibility of analyzing the DNA of the tooth remains on the Rue Perdue of Tournai (Belgium). For this, we needed to obtain DNA from ancient tooth remains and evaluate its authenticity. We focused on mitochondrial analyses and sex determination on DNA extracts from 13 individuals. The extractions are summarized below in Table 11. DNA from more than 40 individuals was planned to be extracted during this project, but the ongoing pandemic restricted lab access and hence limited the number of analyses.

Table 11. Overview of the extractions. . Individuals are identified on the basis of their tomb number (e.g. T2) followed by the letters "S", "T", "D1" or "D2", which indicate which part of the sample was used for extraction. They respectively stand for non-decontaminated surface, dental calculus, decontaminated dentin of a first tooth and decontaminated dentin of a second tooth. Without any specification, it is assumed that only one decontaminated tooth has been sampled for extraction. Negative Extraction Control 1 (NExt1) corresponds to the first extraction performed (lab extraction identifier: DNA20191204), while Negative Control 2 (NExt2) corresponds to the second one (DNA20201016). The age indicated comes from anthropological studies carried out on the bones. The dates indicated are relative dates. They were estimated by archaeologists during a detailed study of the organization of the graves in relation to each other, the stratigraphic layers and the objects found in the graves.

INDIVIDUAL	DNA EXTRACTION	ID_EXTRACTION	CONCENTRATION	SEX	AGE	POSTQUEM DATATION (AD)
T2	T2	DNA20201016_1	1	Male	53	310
T4	T4	DNA20201016_2	1	Male	25	298
T8	T8	DNA20201016_3	1	Female	35	200-399
T10	T10	DNA20201016_4	1	NA	11	307
T14	T14	DNA20201016_5	1	Male	46	307
T17	T17	DNA20201016_6	1	Female	25	NA
T34	T34	DNA20201016_7	2	Female	20	200-399
T38	T38	DNA20201016_8	2	Female	40	321
T39	T39	DNA20201016_9	2	Male	50	200-399
T42	T42	DNA20201016_10	2	Male	40	307
T53	T53	DNA20201016_11	2	Female	15	NA
	NExt2	DNA20201016_12				
T47	T47S	DNA20191204_1	2	Female	50	323
	T47T	DNA20191204_2	2	Female	50	323
	T47D1	DNA20191204_3	2	Female	50	323
	T47D2	DNA20191204_4	2	Female	50	323
T52	T52S	DNA20191204_5	2	Male	50	323
	T52D1	DNA20191204_6	2	Male	50	323
	T52D2	DNA20191204_7	2	Male	50	323
	NExt1	DNA20191204_8				

4.1.1. Overall PCR results of mitochondrial and nuclear data

Several PCRs were performed during the analysis of the DNA extracts. The PCR amplification success rate differs between individuals as shown in figure 12.

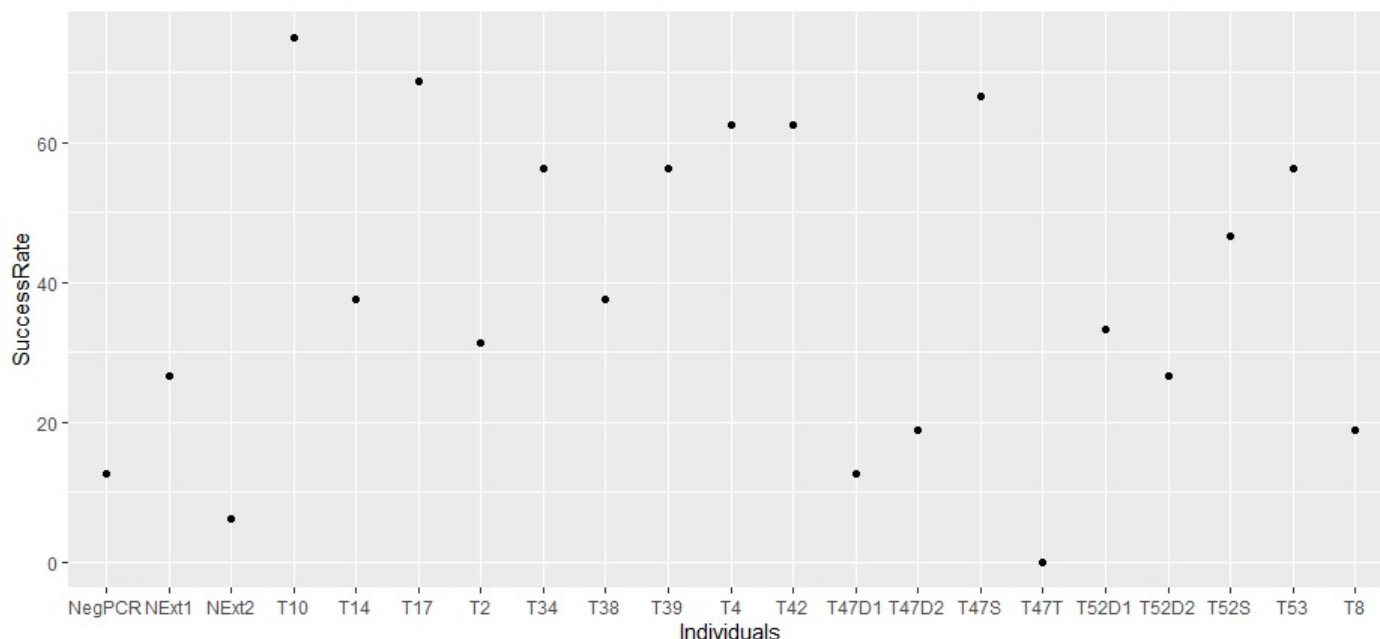


Figure 12. PCR amplification success rate for each DNA extract (sample code), negative extract (NExt1 and NExt2) and for all PCR negatives (NegPCR). All amplifications have been taken into account in the elaboration of the figure.

The negative controls for the two extractions (NExt1 and NExt2) and the PCR (NegPCR) have an amplification rate of 26%, 6% and 12% respectively. The negative control of the first extraction (NExt1) was positive for 3/12 PCR targeting mitochondrial DNA and 1/3 PCR targeting nuclear DNA. The second extraction control (NExt2) was positive for PCR MS. The negative PCR control (NegPCR) was positive for S12 and MAY. The NegPCR is an indication of the percent of false positive amplifications.

Extraction from individuals T4, T10, T17, T34, T39, T42, T47 (DNA extraction S), T53 have over 50% amplification success rate. The other individuals have a success rate between 10% and 50%, except for T47T. The latter, corresponding to the dental calculus of individual 47, failed in all amplifications. Since dental calculus theoretically doesn't contain any human DNA, it could be therefore considered as a contamination control (Sampietro et al. 2006). The absence of amplification for T47T suggests that there was no contamination by exogenous DNA. This supports the authenticity of our results obtained during sequencing, despite the fact that the negative controls show a slight rate of contamination for some PCRs (but see below).

It is highly doubtful that the rate of DNA preservation will be identical for all individuals. Therefore a heterogeneous amplification success rate would generally be observed. If homogeneous amplification success rates is observed, it is almost guaranteed that the contamination is generalized to all individuals. However, a heterogeneous success rate cannot

exclude sporadic contaminations of the samples. Nevertheless, results on Figure 7 suggest that all the sequences obtained are not solely due to contamination. The success rate does not exceed 80% for any of the samples, which would have been the case with modern contaminant DNA.

4.1.2. Results of mitochondrial data

Ancient DNA is characterized by a high degree of fragmentation (Pääbo et al. 2004). This results in a lower success rate of amplification as the size of the fragment increases. (Dabney et al. 2013a, Malmström et al. 2007, Mann et al. 2020, Renaud et al. 2019). At first glance, the data obtained in this study are consistent with this trend (Figure 13). The theoretic relationship between the success rate and the size of the fragment is a decreasing exponential (Allentoft et al. 2012). The small residuals of the exponential regression associated with our data seem to be consistent with this theoretical model. This is not in itself sufficient to prove that these sequences are indeed ancient DNA, but it reinforces the hypothesis.

The sequencing results of mitochondrial fragments are gathered in Table 12.

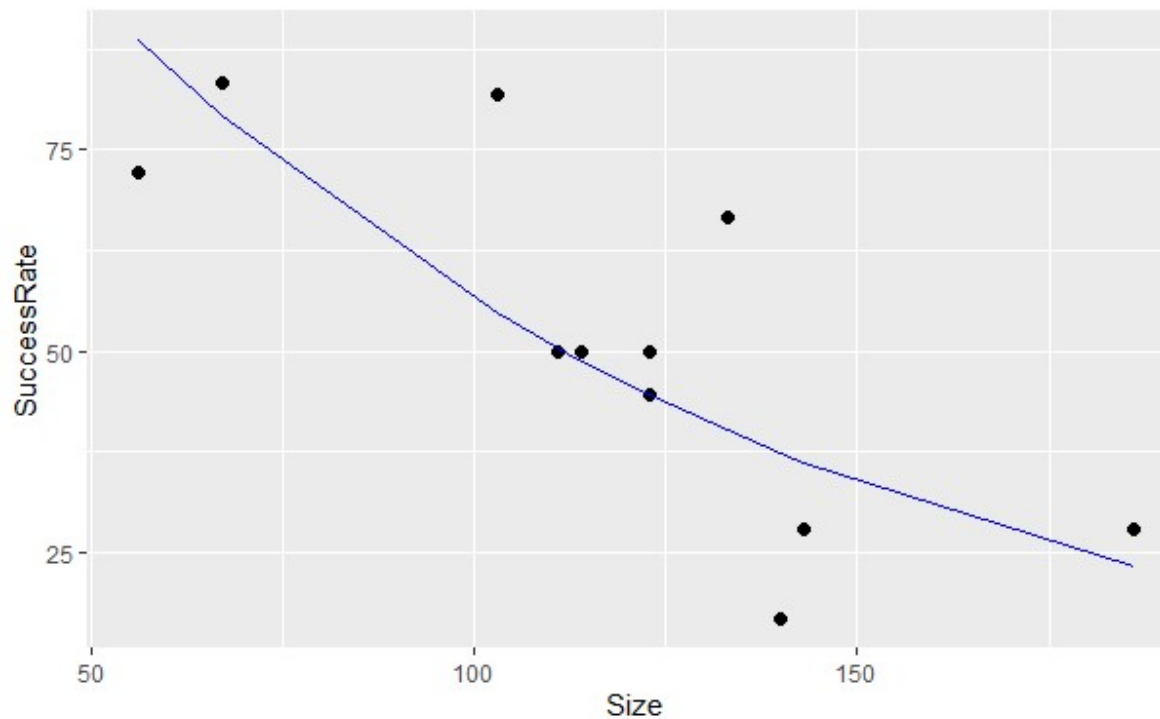


Figure 13. Success Rate of mitochondrial amplification by size (bp).

Table 12. Results of mitochondrial fragment sequencing. For each individual, an entry contains a number if a fragment was successfully sequenced and aligned to the reference sequence that is standardly used for human mitogenome analysis and that is called the “revised Cambridge reference sequence”, or “rCRS” (accession number in the public GenBank database: NC_012920), otherwise it contains nothing. The number in a given row corresponds to the number of substitutions compared to the reference sequence. It can be interpreted as the number of differences to the reference sequence. To avoid double counting of substitutions, overlaps of different fragments are placed in their own separate column.

Individuals																Total of substitutions / individual		Total of C-T	Total of G-A
T2						4						1	0	2		7	2	4	
T4					0	0	0	0		0	0	0	0	1	0	1	0	0	
T8						1		0	0	1						2	1	1	
T10	1	2	1	2	0	1	0	0	0	1	0	0	0	0	0	8	1	0	
T14						0		0								0	0	0	
T17	1	0		1	0	1	0	1	0	1	0	0	0	2	1	8	4	0	
T34				1	0	0	0	0	0	0	0	0	0	1	1	3	1	0	
T38						1	0									1	1	0	
T39						1		0	0	1		0	2	1		5	3	1	
T42				2	0	0	0	0	0	0	0	0	0	0	0	2	1	0	
T53					0	0	0	0	0	0	0	0	0	0	0	0	0	0	
T47S				1	0	1		0		0	0	0	2	0	1	5	3	0	
T47D2								0								0	0	0	
T52S	0	0	1	2	0		1						0	0	0	4	0	0	
T52D1						1		1								2	2	0	
T52D2						0	4				0	0	0			4	0	4	
Fragments	H5	H5-H6	H6	H6-H8	H8	S12	S11	S13	S2	S2-S4	S4-S8	S8-S9	S8-S9-H4	S9-H4		52	19	10	Total
Total of substitutions / fragments	2	2	2	9	0	11	5	2	0	4	1	2	8	4					

Some individuals show few sequenced fragments, notably T14 and T38. On the contrary, others have a very high number of sequencing successes, in particular T10 and T17 (Figure 12 and Table 12).

The observation of the fragments also outlines that some of them could be sequenced for a large number of individuals, such as S12 and S13, while others concern only a limited number of individuals, such as H5 and H6. These data are consistent with the results illustrated in Figure 8. Indeed, S12 and S13 target short fragments of respectively 67 bp and 56 bp. While H5 and H6 target longer fragments of 143 bp and 183 bp respectively.

It should be noted that this table has been realized only on the basis of fragments and individuals that gave results which could be aligned to the reference sequence. Individuals T47T and T47D1 gave no results for any fragment.

Some PCRs show negative controls with contamination, such as fragment S12 for which the negative control for the 1st extraction (NExt1) and for the PCR (NegPCR) were both positive. A more detailed analysis of the sequences obtained revealed that the one of NExt1 wasn't identical to any other fragment of individuals from the same PCR. Consequently, it appears that the sequence of NExt1 is different from the ones of the individuals. Concerning NegPCR, it

shows the exact same mutation (T at base 7028) as T8, T10, T17, T2, T38, T39, T47S and T52D1. It can therefore be concluded that the sequence of NegPCR and likely the other sequences obtained from the same PCR are most probably the result of a cross-contamination that occurred during the PCR set up. Nevertheless, this mutation is characteristic of ancient DNA (Peyr gne and Pr fer 2020, Renaud et al. 2019). The contamination of NegPCR could possibly be explained by cross-contamination from ancient DNA.

On the same way, NExt1 turns out to be also positive for PCR S11 and identical to all the sequences of individuals that gave a result, except for T52S and T52D2. The latter has four G-A mutations which is characteristic of ancient DNA (Dabney et al. 2013b, Renaud et al. 2019).

Regarding S8 PCR, both NegPCR and NExt1 are positive. NegPCR sequence is identical to those of T2, T17, T34, T4, T52D2 and T53 (but not Next1). This seems to suggest that these sequences are contaminated. However, none of the sequences from any of the individuals is identical to the sequence of NExt1, indicating that it is not the source of contamination within the sequences of the individuals.

Furthermore for the fragment S8-S9, T47S shows sequence that is unique in our dataset. It should be noted that for this sample data comes from the S9 PCR, whose negatives controls were negative and hence where no (cross-)contamination was suspected. In addition, the observed substitutions detected are C-T substitutions, which again can be associated with ancient DNA (Dabney et al. 2013b, Peyr gne and Peter 2020, Renaud et al. 2019). These two arguments are in favor of the authenticity of fragment S9 of the individual T47S.

Sequences of T47S come from the extraction of the non-decontaminated surface of the tooth ("S" see Table 11). It is therefore more likely for "S" individuals that the sequences obtained could be more contaminated than other samples, even if they were also treated with bleach. It can be observed that individuals T47S and T52S gave more results during sequencing than their homologous from the decontaminated tooth (T47D1, T47D2, T52D1 and T52D2), which is consistent with the fact that they contain contaminations or that UV treatment degrade the preserve DNA. If the surface is heavily contaminated, the sequencing should have worked for the vast majority of the targeted fragments, which is not the case. In addition, in the only PCR (S13) where T47S and T47D2 (two independent DNA extractions from the same sample), the sequence was identical, even if it was not much variable within our sampling. All this leads us to conclude that the S9 fragment of T47S can be a good candidate for analysis of ancient DNA.

Table 12 also shows the total number of ambiguous sites of type C-T and G-A . Some individuals are found to have neither C-T nor G-A ambiguous sites. But most individuals have at least one in one or more fragments. Globally we find more C-T than G-A ambiguities which is consistent with the data on ancient DNA in the literature (Mann et al. 2020). It is important to remind that the ambiguities reported in the columns "Total C-T" and "Total G-A" table were the ones that were clearly identified as double peaks (see material and method). Several "Y" ambiguities, corresponding to C or T in the sequence, can be identified in some fragments.

4.1.3. Results of nuclear data

We targeted 3 nuclear markers with 4 pairs of primers and 6 PCR. In total only 21 of 108 fragments of the expected size were amplified, 13 were sequenced.

It is worth noting that we have mainly done optimization and the results of the sequencing of one of the last PCR MAY are encouraging. 8 out of 18 fragments have been amplified and have the expected length. The amplification success rate was generally lower for nuclear markers than mitochondrial markers, even when targeting very short fragments. This expected because there are less copies of the nuclear genome than mitochondrial genomes in a same tissue sample. This represents one more indication that we did not amplify recent contaminant DNA, which is generally more abundant and less degraded.

The PCRs that were performed were intended to identify the individual's sex based on the number of visible bands of different length. Indeed, one of the target genes, Amelogenin, has different lengths depending on whether it is the X or Y chromosome. We focused on two approaches.

In the first one, a fragment delimited by the primers British Forward and Reverse is targeted. The fragment has a length of 106 bp for X and 112 bp for Y (Mohandesan et al. 2004). In the second approach, with the AMEL's primers we amplified a fragment of 56 bp and 60 bp in X and Y, respectively (Masuyama et al. 2017). The last gene targeted, SRY, has a length of 52 bp and is specific to the Y chromosome (Masuyama et al. 2017).

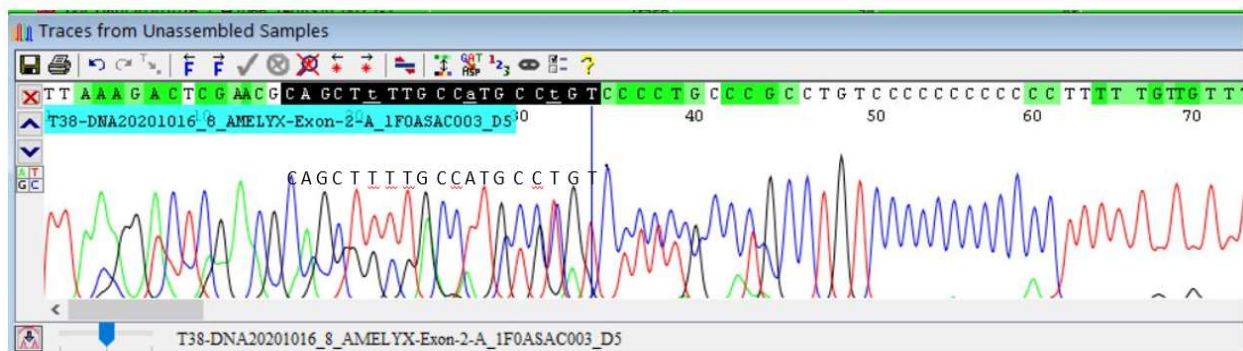
For the approach of Masuyama (Masuyama et al. 2017), we have tested several options. AMEL and SRY in the same tube (MAY) but also AMEL (MS and MA) and SRY (MY) in separate tubes. The results can be interpreted as follows:

- SRY: Male = 1 band of 52 bp; Female = 0 band
- AMEL: Male = 2 bands of 56 bp and 60 bp; Female = 1 band of 56 bp

The PCR MAY gave the most successful results with 8 out of 18 amplifications of the expected size.

Some sequences of MS, MA and MY enabled us to identify the sex of some individuals by sequencing. The sequences of MAY were planned to be analyzed with the Bioanalyzer, but due to the Covid-19 this could not be done. The DNA sequence was clear in two samples (T38? And T53?) and repeated with a four base pair interval in one sample (T39?). The first clear profile is expected to identify females whereas the second profile showing a repetition is expected to identify males (Figure 14). This enabled the identification of the sequence of T39 as male and those of T38 and T53 as females. This information is consistent with the anthropological data (Table 11).

T38 AMELYX Exon2 A



T39 AMELYX Exon2 A

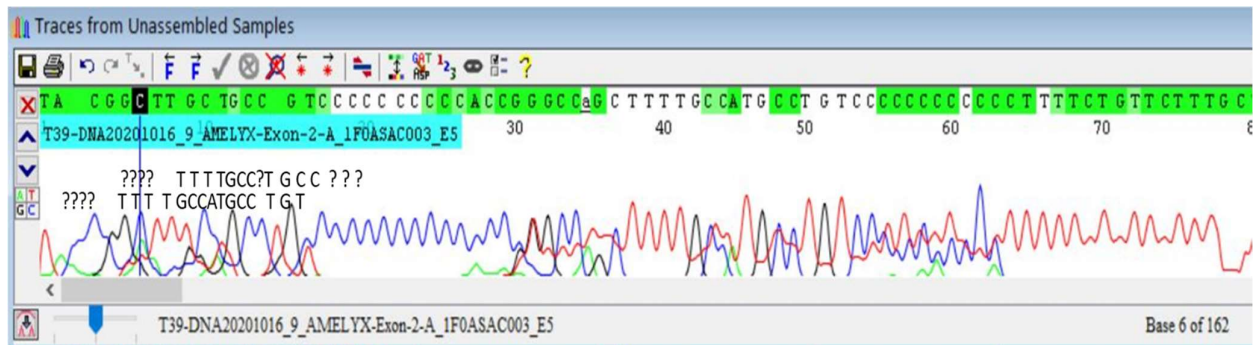


Figure 14. Example of a clear (T38) or repeated (T39) sequence for sex determination by AMELYX.

The reverse complement of T38 AMEL Exon2 A R is not repeated after ca. 4bp. (Clearly 4 T and not more). The reverse primer is also repeated after more bp but this seems to be an artefact (also observable in mitochondrial sequences). This sample is therefore expected to be a female.

The reverse complement of T39 AMEL Exon2 A R is repeated after ca. 4bp (series of 8 T). It is also repeated after more bp but this seems to be an artefact (also observable in mitochondrial sequences). This sample is expected to be a male.

4.2. Contribution to archaeological data

Haplogroup results that were obtained via Empop (Huber et al. 2018) are shown in the Table 14. The identification of the haplogroups is based on the substitutions detected in the sequences (Table 13).

Table 13. Summary table of substitutions. The numbers on the top line indicate the position for the column on the mitochondrial genome. The amplicons are located on the genome by dark gray areas and their PCR name is indicated. Substitutions are in red squares, ambiguities (double peaks) in orange squares and green squares indicate that the base could not be identified. Shaded areas show the fragments that could be sequenced and that are identical to the “rCRS”. The white areas are the parts for which we do not have information. The bases from 34 to 397 correspond to the hyper variable sequence II (HVSII). The bases from 16056 to 16409 correspond to hyper variable sequence I (HVS I). The others correspond to SNPs.

[illegible][illegible]

Table 14. Haplogroup analysis. Sequences were analyzed both with all substitutions and without the substitutions associated with post-mortem degradations. The haplogroups identified are the most likely and the lowest cost haplogroups. When more than one haplogroup were equally probable, the closest common ancestor was indicated. The substitutions were identified by the presence of double peaks in the sequences (ambiguities). The concentration is also indicated for each individual in a different color.

Individuals	Concentration	Haplogroup (considering all substitutions)	Haplogroup (considering absence of post-mortem degradation)
T2	1	M	M8a
T4	1	H5	H5
T8	1	L3	L3
T10	1	JT	JT
T14	1	L3	L3
T17	1	T2b17	T2b17
T34	2	H5	H5
T38	2	L3	L3
T39	2	A+152+16362	A+152+16362
T42	2	H	H
T47D2	2	R	R
T47S	2	U5a	U5a2
T52D1	2	H1bc	H1bc
T52D2	2	L3	H
T52S	2	R0	R0
T53	2	H	H

Both analyses present very similar results. Only some of the sub-haplogroups differ.

A first analysis with all substitutions shows that haplogroup H is present in 6 of the 16 samples analyzed. Two sub-haplogroups of H have been identified as H5 and H1bc. Two individuals had a R haplogroup. For four individuals the haplogroup result is L3. Lineage L3 includes the M and N clades which are present in the majority of individuals in Europe. Several identified haplogroups are represented by only one sample. This is the case for R, R0, M, JT, T2b17, A+152+16362, U5a.

A second analysis based on a more restricted set of substitutions that are unlikely the results of post-mortem DNA degradation shows that the haplogroups remain similar. The results are occasionally more precise reaching sub-haplogroups. This could be explained by the fact that in the first analysis, the majority of substitutions related to post-mortem degradations correspond to extra mutations which has prevented the identification of sub-haplogroups with precision. For example for T2, when the extra-mutation 16287W and 16288R are removed, it lead to the sub-haplogroup M8a. Two new haplogroups could be identified: M8a and U5a2. At least 6 samples had an H haplogroup and three had L3 haplogroup. All other haplogroups are represented by only one sample.

These data are consistent with those reported by Antonio and collaborators (Antonio et al. 2019), who analyzed 48 individuals dating from the Roman Imperial period in several archaeological sites in Italy (27 BC - 300 AD). The most frequent haplogroups were H, U and T, while J had a lower frequency. Their data in southern Italy at Vagnari showed that the most frequent haplogroups were also H and T while U was totally absent (Antonio et al. 2019). During Late Antiquity (300 AD - 700 AD) the haplogroup H remains the most frequent with T,

K, U and J coming next in order of frequency. All haplogroups mentioned are characteristic of Eurasian populations (Antonio et al. 2019). These haplogroups are already present in previous periods within the populations. The haplogroups H, T and J, K are typical of Eurasian populations (Antonio et al. 2019). Our results appear consistent with their observations.

The historical background of the necropolis on Rue Perdue is characterized by the maximal expansion of the Roman Empire in the 2nd century AD and by the transfer of the capital to Constantinople in the 4th century AD (Abulafia 2011, Antonio et al. 2019, Broodbank 2013). The Roman Empire had a population of more than 70 million people, which favours population movements throughout its territory (Abulafia 2011, Antonio et al. 2019, Broodbank 2013). According to Antonio and collaborators, this translates into important genetic flows and also into a high genetic variability among individuals (Antonio et al. 2019). Our data seem to be consistent with their conclusions. Nevertheless, a larger sample of data would be needed to confirm this trend. The lack of ancient DNA data of populations living in various European regions from the 2nd century AD to 4th century AD prevent to pinpoint a more precise origin of the individuals.

All haplogroups identified by Empop (Huber et al. 2018) (Table 14) have been retrieved on the updated comprehensive phylogenetic tree of global human mitochondrial DNA variation available in PhyloTree.org (Van Oven and Kayser 2009). A subtree of this phylogenetic tree containing the haplogroups detected in this study in Tournai has been extracted and represented in Figure 15.

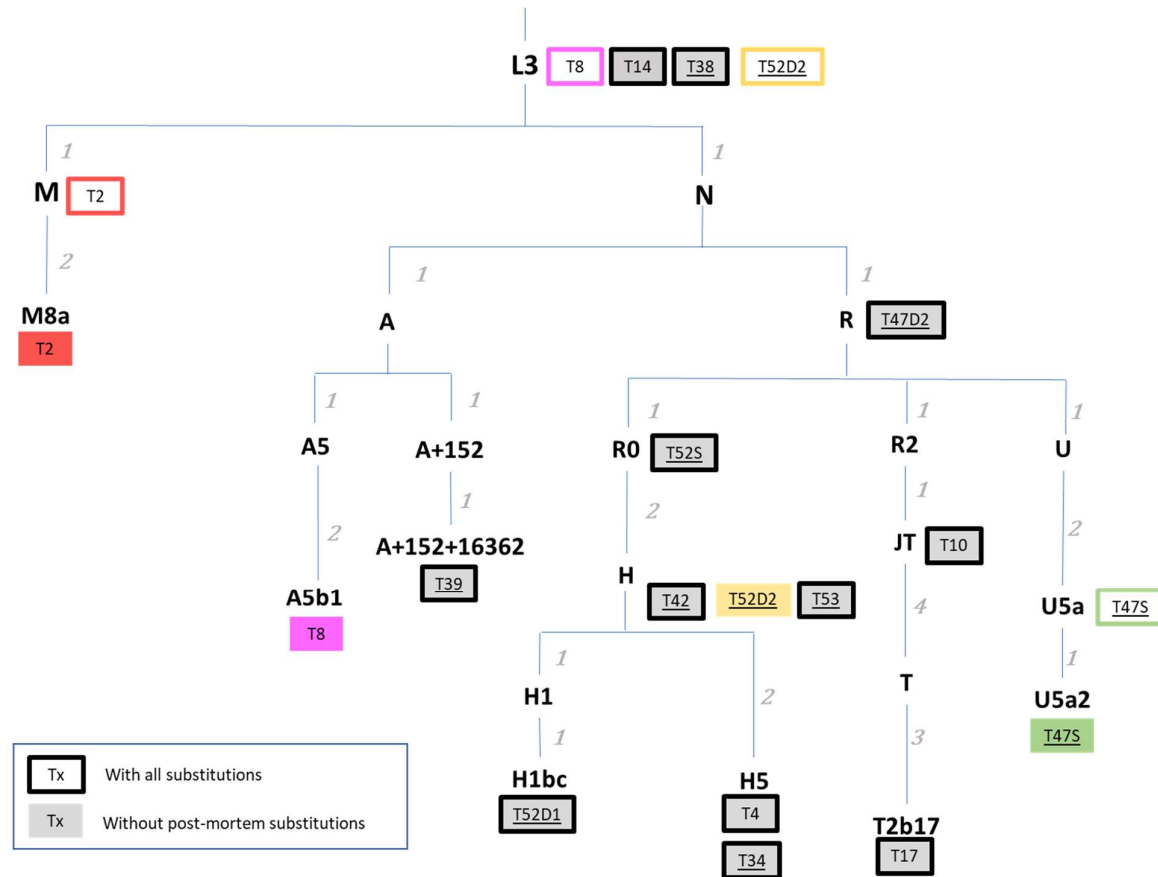


Figure 15. Phylogenetic tree of haplogroups. The number next to the branches is the number of nodes separating the haplogroups that are connected by the branches. Each individual studied here is represented by its “T” ID in a box next to the haplogroup to which it is estimated to belong (in a bold colored frame for the result of the first analysis including all substitutions and in a box filled with the same color for the result of the second analysis that do not include substitutions associated with post-mortem degradations). When the results of both analyses are identical, individuals are indicated by a grey box with a bold border. The individuals underline correspond to the second concentration (Adapted from Van Oven and Kayser 2009).

From this representation, we can see that all individuals are located in the L3 branch of the phylogenetic tree (Figure 15). Most individuals are located in the N branch of the tree. Only T2 is located in branch M. Among the 14 of branch N, 6 are located in branch H.

More specifically, the results with or without post-mortem substitutions are identical for most individuals (12/16). For others, the second analysis (without substitutions) made it possible to specify the sub-haplogroups. For example, the first analysis (with substitutions) placed T8 in the L3 branch while the second analysis placed it in A5b1. In the same way T52D2 placed in L3 by the first analysis but appears in branch H in the second analysis.

Individuals for which multiple extractions were performed (T47 and T52), show results that vary amongst samples. The results for T52 remain consistent as they are all located on the same R0 branch at different degrees of precision. In the same way T47 shows consistent results indicating either the U5a2 or R haplogroup depending on the sample. Indeed, individuals

belonging to the L3 lineage may still be assigned to more precise haplogroup M, N, A, R, H, J, T, U if more data are collected. Idem for R haplogroups that can be assigned to H, J, T or U.

Regardless of both analyses, it would not be possible to establish direct kinship relationships between individuals. Nevertheless, a closer relationship among the maternal lineages of some individuals could be excluded. The results also highlight a greater degree of proximity between the maternal origins of some individuals compared to others. For example, two individuals in the H branch have a more recent maternal ancestor (on the maternal genealogy) than two individuals belonging to H and T haplogroups.

The most frequent haplogroup (H) is present in both concentrations. Even if only one individual in the first concentration has this haplogroup (Figure 16).

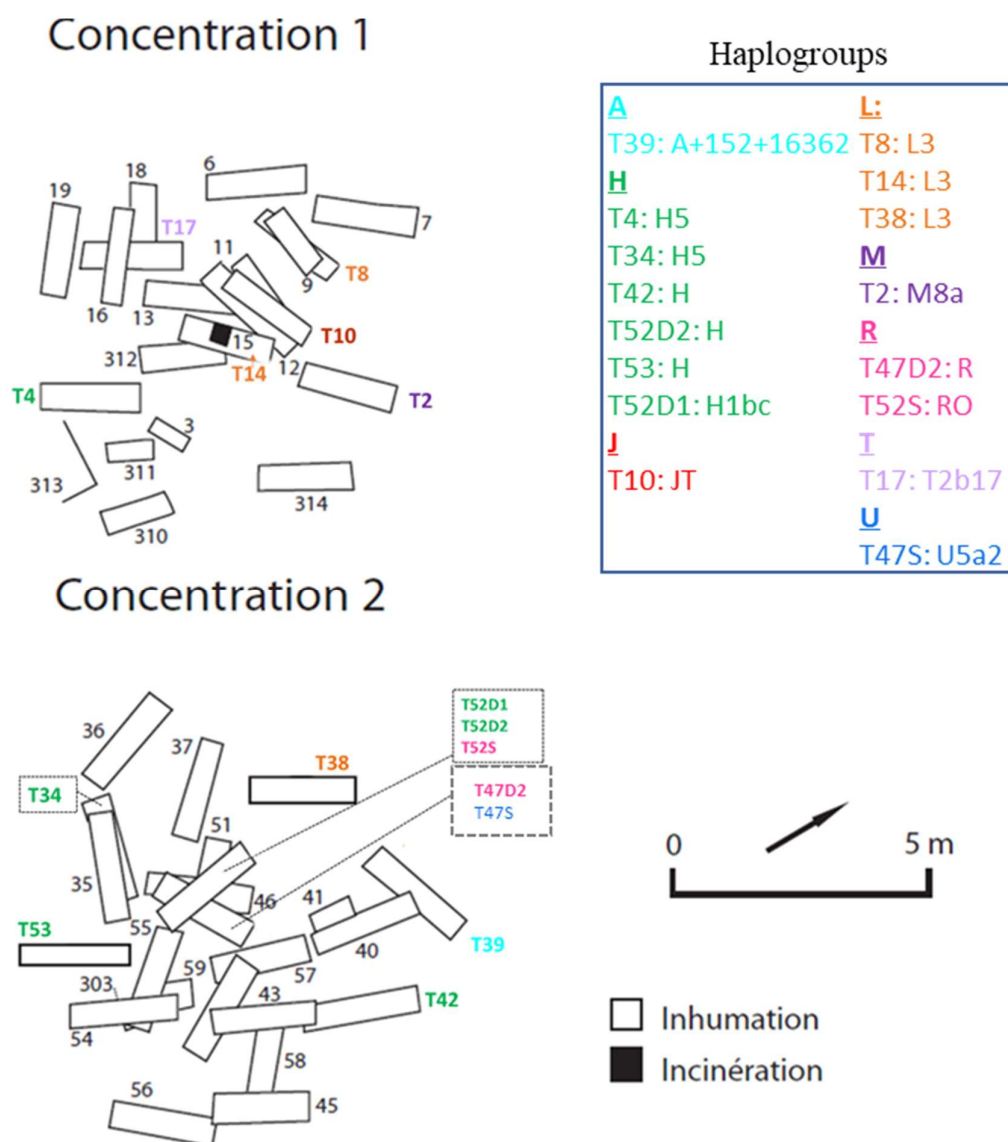


Figure 16. Map of the concentration 1 and 2 of the cemetery of Rue Perdue (Modified from Verslype and Weinkauff n.d.). mtDNA haplogroups are superimposed and colour coded over associated grave structures represented by a rectangular box for burials and a black square for incineration. Only

burials were considered in this study.. Each individual studied here is represented by its “T” ID next to his graves and in the colour of the haplogroup to which it is estimated to belong. The other numbers show individuals who have not been studied. We have to note that a different haplogroup code does not necessarily represent a different haplogroup. Indeed, individuals identified as belonging to L3 may still be assigned to haplogroups M, A, H, T, U, etc. if more data is collected. Idem for R that can be assigned to H, J, T or U. This last observation is important because it means that the different results obtained for individual T47 (from DNA extractions S and D2) are not contradictory. For interpretation, please refer to the genealogy represented in fig. 15.

Individuals in the first concentration have haplogroups in different branches. Each individual studied belongs to a different branch (M8a, A5b1, H5, JT, T2B17 and L3). Some have closer common ancestors (e.g. JT and T2b17), others are very distant from each other, e.g. T2 has a haplogroup in branch M while the others have mostly haplogroups in branch N. Individuals in the 2nd concentration, if we consider samples from the same individual (T52 and T47) together, have more frequently a haplogroup H (4/6 individuals). Only T39 and T47 are in different branches.

No direct kinship relationship can be established between individuals (identical haplogroups). Some individuals appear to be closer to other individuals (individuals of branch H). On the other hand, we can exclude kinship links on the basis of maternal genealogy between some individuals. Individuals who have a distant common ancestor such as T2 with individuals of branch N are less related to each other concerning their maternal genealogy compared to two individuals showing the same haplogroup.

Concentrations could not be associated with family clusters based on maternal genealogy. Since Roman society is generally considered as patriarchal, it is possible that if family clusters exist, they may be based on paternal lineages that could not be investigated here.

Based on our limited sampling, individuals of concentration 2 are closer to each other than are those within concentration 1. Since haplogroup H seems to be the most frequent haplogroup in the Roman populations of the imperial and Late Roman Empire (Antonio et al. 2019) and considering the limited number of individuals studied, it is possible that a large part of the individuals in the necropolis have a haplogroup H, independently of the concentration, as suggested by T4 of concentration 1.

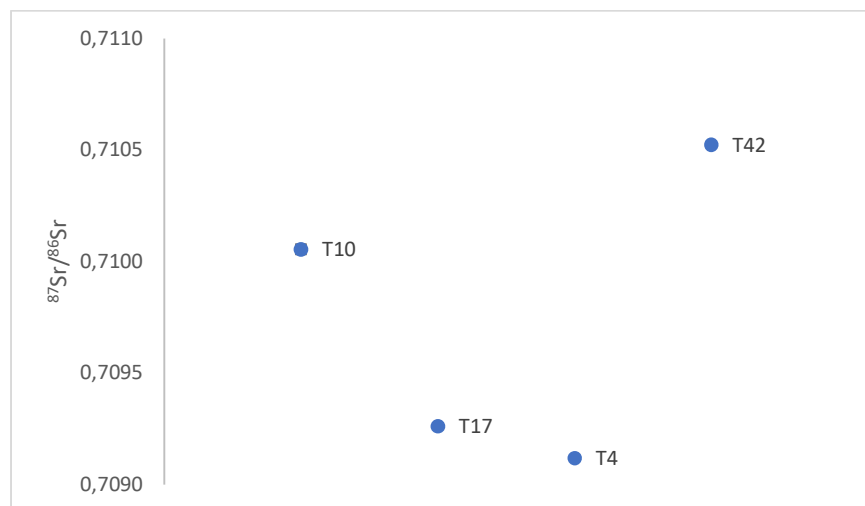
Some haplogroups are rare and are specific to one of the concentrations. The haplogroups T, J and M are only present in concentration 1 while A is only found in concentration 2. As mentioned above haplogroups H and T seems to be the most frequent haplogroups in roman population while J is present at a lower frequency (Antonio et al. 2019). In ancient Roman DNA studies, there is very little information available for haplogroups A and M. This might suggest that they were probably rare in the roman populations since they were rarely detected in the studies.

The identified mitochondrial haplogroups, did not enable an association of concentrations with familial clusters due to the limited number of both individuals (samples extracted) and available data (markers sequenced).

Since the individuals are located on the phylogenetic tree with a varying degree of precision, the final interpretation could still be improved and would deserve further investigation with additional data.

4.3. Strontium Analysis

Table 15. Results of radiogenic Sr isotopes measurements. The ratio, the standard error and the number of runs are indicated for each sample. This Table was provided by H  l  ne Dailly and Sophie Opfergelt .



These results can be compared with geographical data from different regions of Europe (Emery et al. 2018b, Hoogewerff et al. 2019, Willmes et al. 2014).

No map of Sr values for Belgium exists. Nevertheless, a study by Demaiffe and Dejonghe gives isotopic values for the city of Tournai according to different geological formations (Demaiffe and Dejonghe 1990). Tournai's values are 0.70814 ± 0.00004 for limestone formations and between 0.70793 ± 0.00006 and 0.70804 ± 0.00003 for anhydrite formations (Demaiffe and Dejonghe 1990).

Maps of neighboring regions (Willmes et al. 2014), Europe at a larger scale (Hoogewerff et al. 2019) or other regions (Emery et al. 2018b) are also available. The values for regions near Tournai in France range from 0.708792 to 0.7113211 (Figure 18). The more general map of Europe shows approximate values for Belgium between 0.705 and 0.712 (Figure 19). The isotopic map of Italy, on the other hand, is quite detailed. For all of Italy the values are between 0.7037 and 0.7400 (Figure 20).

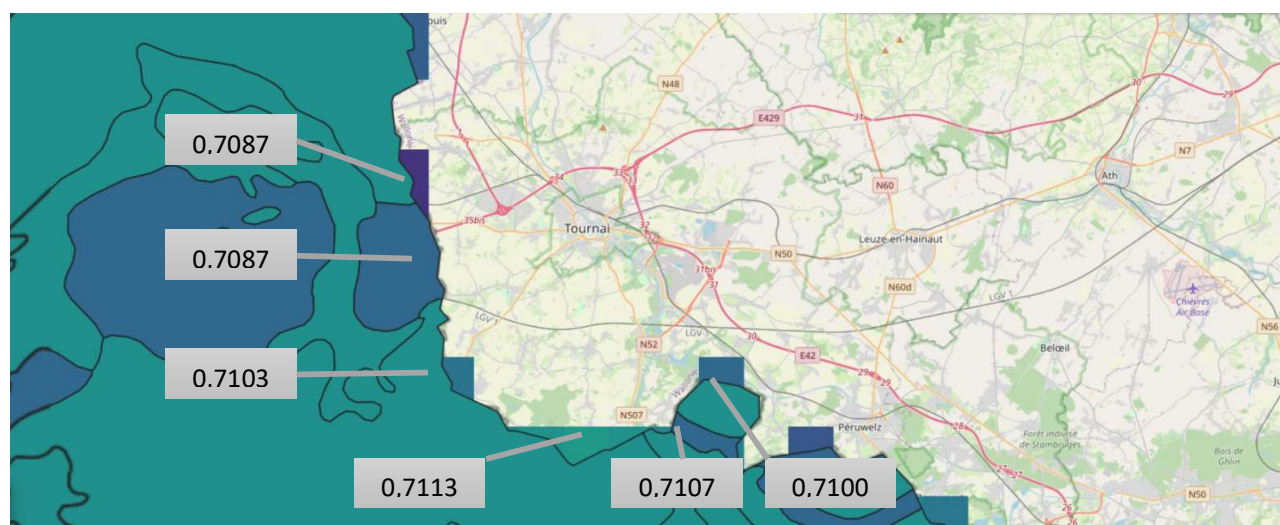


Figure 18. $^{87}\text{Sr}/^{86}\text{Sr}$ variation map of France in the vicinity of the region of Tournai. Sr ratio measurements for some regions have been shown in boxes on the map (modified from Willmes et al. 2014).

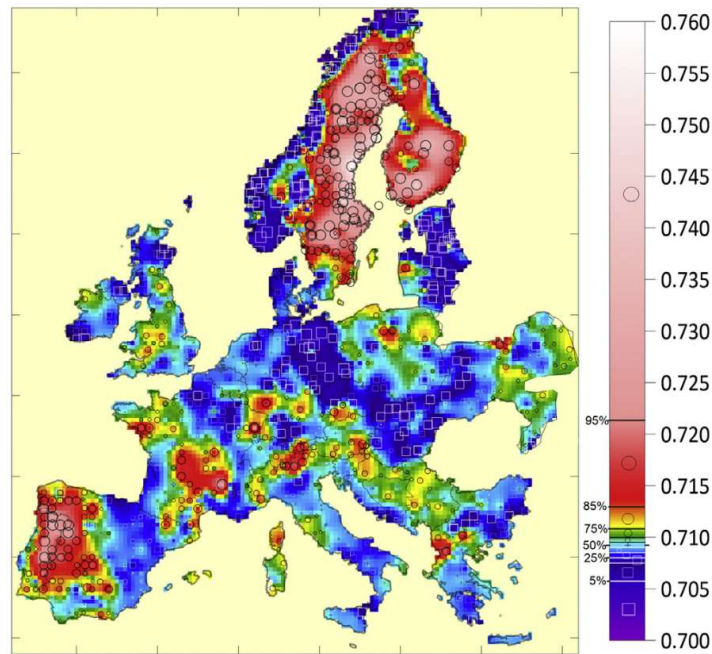


Figure 19 (adapted from Hoogewerff et al. 2019). Locations of the 1187 analysed sites with scaled symbols according to $^{87}\text{Sr}/^{86}\text{Sr}$ values percentiles and ordinary kriging interpolation in Surfer ("Golden Software" 2017) on a 25×25 km grid, with 400 km search radius. The circles (values above median 0.7091) and squares (values below median) show the locations of the measured $50 \text{ km} \times 50 \text{ km}$ sites and $^{87}\text{Sr}/^{86}\text{Sr}$ value for each grid cell is indicated on the colour legend (adapted from Hoogewerff et al. 2019).

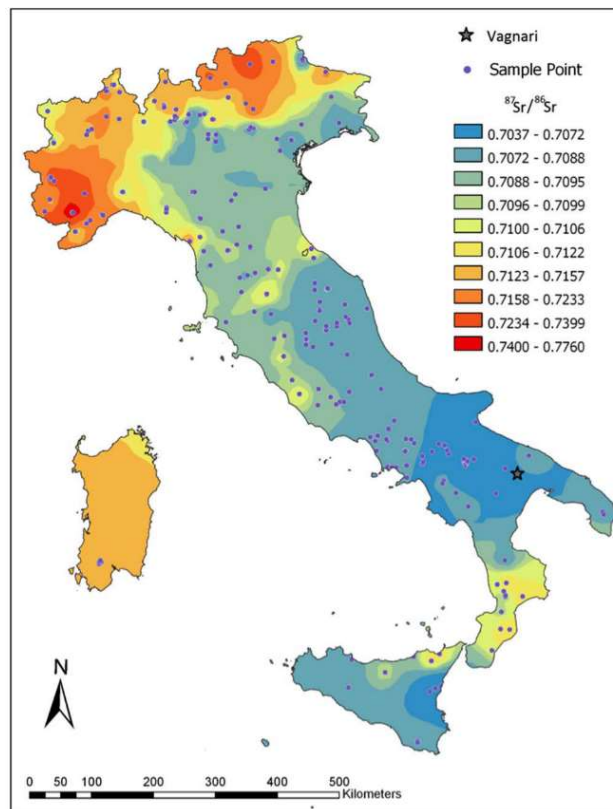


Figure 20. $^{87}\text{Sr}/^{86}\text{Sr}$ variation map of Italy. This map was produced using disparate sources of $^{87}\text{Sr}/^{86}\text{Sr}$ sources and interpolated using a bounded inverse distance weighting (IDW) algorithm (Emery et al. 2018b).

Given all these values in comparison with the values of the archaeological remains, it is difficult to determine precisely the geographical origin of the individuals analyzed in our study. Indeed, they could be from the local region of Tournai as well as from the region of Lazio, Tuscany, Veneto, Emilia-Romagna or other regions of Europe. On the other hand, some regions such as the Alps region, West Spain, North Europe or some others regions of Europe can be excluded.

It can however be noted that all individuals do not have exactly the same geographical origin in which case they would have almost identical strontium ratio values. This suggest that population movements occurred at least at regional/local scale.

5. Conclusion and Perspectives

The objectives of this Master thesis were to demonstrate the feasibility of analyzing ancient mitochondrial and nuclear DNA extracted from ancient human remains, to authenticate the results and to estimate their contribution to archaeological data.

5.1. Feasibility and authentication

We successfully extracted mitochondrial and nuclear DNA from skeletal teeth dating back to the 3rd and 4th century AD. Not all sequences obtained are authentic. Some sequences were identified as contaminations.

The analysis established that contaminations were not generalized and could be generally identified. Cross-contamination was noted during the experimental analyses as a few negative controls (extraction and PCR) showed amplification for some PCRs. Contamination at both DNA extraction and PCR amplification could be partly tracked thanks to the negative dental calculus control (T47T) being negative in all PCRs performed. A second independent extraction from the same individuals could also have been performed. The analyses repeated independently would have increased the confidence level.

5.2. Archaeological contributions

The identified mitochondrial haplogroups, did not enable an association of concentrations with familial clusters due to the limited number of both individuals (samples extracted) and available data (markers sequenced). Inferring family groupings would require more data coming from more individuals, different concentrations and different types of DNA fragments including for example, nuclear DNA.

Only maternal lineages could be studied from the identification of mitochondrial haplogroups. The study of the Y chromosome would allow the analysis of paternal lineages and would bring complementary information to the current results.

The mitochondrial data did not allow the identification of kinship relationships between individuals. This results to the fact that the analysis did not reach a uniform degree of precision amongst the samples and therefore prevented a more precise identification of kinship relationships. Nevertheless, it was possible to establish that some individuals were genetically closer than others and that some individuals from the same concentration were less related, based on their maternal genealogy, than other individuals found in different concentrations.

The precise geographical origin of the individuals could not be determined using mitochondrial haplogroups, the possibilities are still numerous. The results of Sr ratio analyses have limited these possibilities by excluding some geographical regions such as Northern Europe, Spain, the Alps, ... At this stage, an hypothesis on an exact geographical origin could neither be refuted nor confirmed. The results of Sr ratio showed that individuals do not come from a single region. The results suggest that population movements took place at least at the regional scale of Tournai and its vicinity, but could also take place at a larger scale. The individuals studied in this work could either illustrate the movements of these populations or could result from the

acculturation of indigenous populations. The results obtained are consistent with both hypotheses and also with the results of the haplogroups identified.

The analysis of Sr ratios ($^{87}\text{Sr}/^{86}\text{Sr}$) could be coupled with measurements of the isotope ($^{18}\text{O}/^{16}\text{O}$) of Oxygen (Emery et al. 2018a, Laffoon et al. 2017). These additional data would further precise the geographical origin determined by Sr. The principle of both isotope analysis is similar. Crossing the results from these two measurements could discriminate more precisely the geographical origin of individuals in cases in which it remains undetermined following the Sr study only.

Sequenced nuclear data allowed the identification of the sex of some individuals, which confirmed the results of anthropological studies. The limited results obtained for nuclear DNA suggest that an optimization of the protocols and the techniques used would be desirable. The latest PCR performed shows encouraging results.

The Bioanalyzer approach could contribute to a better interpretation of the PCR results of the sex determination data. Indeed, the Bioanalyzer precisely gives the length of the DNA fragments present in the post-PCR samples. Given the very small discrepancy in length between the bands expected to distinguish sex (less than 10 base pairs) and the available 25 bp ladder it was impossible to interpret the results obtained by visual inspection of stained agarose gels.

The constraints related to the Covid-19 pandemic prevented the pursuit of additional specific objectives related to the diet and the health of individuals. This is a very little known aspect of ancient populations whose current knowledge is mainly based on archaeological remains, for example animal bones, and historical sources. The DNA analysis could have improved and confirmed the understanding of the ancient population in these areas. Dental calculus taken from skeletons could have been studied and analyzed via Metabarcoding. It could lead to the identification of some species of bacteria and viruses that are characteristic of certain foods or diseases (Adler et al. 2013, Power et al. 2018, Weyrich et al. 2017).

Initially, the teeth of more than 40 individuals were supposed to be sampled for analysis. All were from the necropolis of Rue Perdue in Tournai and were discovered during various rescue excavations between 1967 and 2010 (Brulet 2019). In addition to the Sanger sequencing of short DNA fragments amplified from the extracted DNA, a shotgun sequencing (using next-generation sequencing (NGS) technologies) of some DNA extracts was also planned. Unfortunately, due to the Covid-19 situation, only 13 individuals were studied from which 16 DNA samples could be extracted and NGS could not be applied.

More recent techniques than the Sanger sequencing used in this study, would be interesting to explore. The Next Generation Sequencing (NGS), as the Illumina and the method introduced by Khun in 2018 via shotgun, could provide more complete data (Monroy Kuhn et al. 2018). For example, they could enable a more accurate analysis of the nuclear genome. These techniques are in fact often used in similar studies.

The limited number of data on past population genetics, even though they have been significantly increasing over the past several years, is a real constraint to interpret the geographical origin of individuals. The establishment of an ancient DNA database of all known archaeological sites would be a real asset in this type of study. It could lead to much more precise and complementary interpretations of existing archaeological material. Such a database should be open, public and progressively enriched with authenticated data and metadata (place, period, date, type of data, haplogroups, degree of certainty, ...). For example, the "Ancient mtDNA database" (<https://amtdb.org/>) pursues such goals. New visualization techniques would be very meaningful to the interpretations and the interactive interconnections of the collected data. Important developments are currently in progress in data management at a European level (EU Dataviz) and could be applied to the archaeological and genetic fields. These approaches applied to the archaeological field would constitute a new development layer in the enrichment of historical and genetic classical methodologies and would open up new analysis perspectives. They would bring mutual benefit between local and general studies.

In conclusion, this study outlined both the feasibility and the potentialities of a genetic approach complementing the classical historical and archaeological methods. The results complemented the current archaeological findings and helped to (in)validate some previous historical assumptions. This project opened the door to the acquisition of new knowledge of past populations and contributed to highlighting some Belgian patrimony. Following the Covid-19 pandemic, several aspects of this study could not be completed. However, beyond the direct results gained from this study, the approach underlined some promising developments that deserve to be considered in the fields of the ancient DNA analysis, the sequencing techniques and in the exploitation of large genetic and archaeological data, opening the way to new perspectives of historical interpretations.

6. Bibliography

- ABULAFIA, D. (2011). The great sea: A human history of the Mediterranean, Oxford University Press.
- ACIEGO, S. M., B. BOURDON, M. LUPKER AND J. RICKLI (2009). "A new procedure for separating and measuring radiogenic isotopes (U, Th, Pa, Ra, Sr, Nd, Hf) in ice cores." Chemical Geology **266**(3-4): 194-204.
- ADLER, C. J., K. DOBNEY, L. S. WEYRICH, J. KAIDONIS, A. W. WALKER, W. HAAK, C. J. BRADSHAW, G. TOWNSEND, A. SOŁTYSIAK AND K. W. ALT (2013). "Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions." Nature genetics **45**(4): 450.
- ALLENTOFT, M. E., M. COLLINS, D. HARKER, J. HAILE, C. L. OSKAM, M. L. HALE, P. F. CAMPOS, J. A. SAMANIEGO, M. T. P. GILBERT AND E. WILLERSLEV (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.
- ALZUALDE, A., N. IZAGIRRE, S. ALONSO, N. RIVERA, A. ALONSO, A. AZKARATE AND C. DE LA RÚA (2007). "Influences of the European kingdoms of Late Antiquity on the Basque Country: an ancient-DNA study." Current Anthropology **48**(1): 155-162.
- ANTONIO, M. L., Z. GAO, H. M. MOOTS, M. LUCCI, F. CANDILIO, S. SAWYER, V. OBERREITER, D. CALDERON, K. DEVITOFranceschi AND R. C. AIKENS (2019). "Ancient Rome: A genetic crossroads of Europe and the Mediterranean." Science **366**(6466): 708-714.
- BARBERENA, R., V. A. DURAN, P. NOVELLINO, D. WINOCUR, A. BENÍTEZ, A. TESSONE, M. N. QUIROGA, E. J. MARSH, A. GASCO AND V. CORTEGOSO (2017). "Scale of human mobility in the southern Andes (Argentina and Chile): A new framework based on strontium isotopes." American Journal of Physical Anthropology **164**(2): 305-320.
- BATAILLE, C. P. AND G. J. BOWEN (2012). "Mapping $^{87}\text{Sr}/^{86}\text{Sr}$ variations in bedrock and water for large scale provenance studies." Chemical Geology **304**: 39-52.
- BENTLEY, R. A. (2006). "Strontium isotopes from the earth to the archaeological skeleton: a review." Journal of archaeological method and theory **13**(3): 135-187.
- BINFORD, L. R. (1971). "Mortuary practices: their study and their potential." Memoirs of the Society for American Archaeology: 6-29.
- BOWEN, G. J. (2010). "Isoscapes: spatial pattern in isotopic biogeochemistry." Annual Review of Earth and Planetary Sciences **38**: 161-187.
- BROODBANK, C. (2013). The Making of the Middle Sea: A History of the Mediterranean from the Beginning to the Emergence of the Classical World, Thames & Hudson.
- BRULET, R. (1990). Le sarcophage gallo-romain de Tournai. Louvain-La-Neuve, UCL. Département d'archéologie et d'histoire de l'art (Louvain-La-Neuve).
- BRULET, R. (1994). "Le cimetière dans la ville de Tournai." Actes des congrès de la Société d'Archéologie Médiévale **4**(1): 133-142.
- BRULET, R. (1996). La nécropole du Bas-Empire romain de la rue Perdue à Tournai (Fouilles 1989).

- BRULET, R. (2014). La cathédrale Notre-Dame de Tournai: le programme des fouilles archéologiques. L'empreinte chrétienne en Gaule, du IV^e au IX^e siècle: 215-232.
- BRULET, R. (2019). "Evolution du paysage urbain et fonctions de la ville de Tournai au Bas-Empire."
- DABNEY, J., M. KNAPP, I. GLOCKE, M.-T. GANSAUGE, A. WEIHMANN, B. NICKEL, C. VALDIOSERA, N. GARCÍA, S. PÄÄBO AND J.-L. ARSUAGA (2013a). "Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments." Proceedings of the National Academy of Sciences **110**(39): 15758-15763.
- DABNEY, J., M. MEYER AND S. PÄÄBO (2013b). "Ancient DNA damage." Cold Spring Harbor perspectives in biology **5**(7): a012567.
- DAMGAARD, P. B., A. MARGARYAN, H. SCHROEDER, L. ORLANDO, E. WILLERSLEV AND M. E. ALLENTOFT (2015). "Improving access to endogenous DNA in ancient bones and teeth." Scientific reports **5**: 11184.
- DEMAIFFE, D. AND L. DEJONGHE (1990). "Géochimie isotopique du strontium des barites, anhydrites, calcites et fluorites de Belgique." Annales de la Société géologique de Belgique.
- DNA Sequencing by Capillary Electrophoresis Chemistry Guide (2009). Applied Biosystems.
- ECKARDT, H., C. CHENERY, P. BOOTH, J. EVANS, A. LAMB AND G. MÜLDNER (2009). "Oxygen and strontium isotope evidence for mobility in Roman Winchester." Journal of Archaeological Science **36**(12): 2816-2825.
- EMERY, M. V., A. T. DUGGAN, T. J. MURCHIE, R. J. STARK, J. KLUNK, J. HIDER, K. EATON, E. KARPINSKI, H. P. SCHWARCZ AND H. N. POINAR (2018a). "Ancient Roman mitochondrial genomes and isotopes reveal relationships and geographic origins at the local and pan-Mediterranean scales." Journal of Archaeological Science: Reports **20**: 200-209.
- EMERY, M. V., R. J. STARK, T. J. MURCHIE, S. ELFORD, H. P. SCHWARCZ AND T. L. PROWSE (2018b). "Mapping the origins of Imperial Roman workers (1st–4th century CE) at Vagnari, Southern Italy, using ⁸⁷Sr/⁸⁶Sr and $\delta^{18}\text{O}$ variability." American Journal of Physical Anthropology **166**(4): 837-850.
- FAIDER-FEYTMANS, G. (1951). "Sépultures du IV^e siècle à Tournai." Latomus **10**(Fasc. 1): 29-52.
- GABRIEL, S. B., S. F. SCHAFFNER, H. NGUYEN, J. M. MOORE, J. ROY, B. BLUMENSTIEL, J. HIGGINS, M. DEFELICE, A. LOCHNER AND M. FAGGART (2002). "The structure of haplotype blocks in the human genome." Science **296**(5576): 2225-2229.
- GILBERT, M. T. P., H.-J. BANDELT, M. HOFREITER AND I. BARNES (2005a). "Assessing ancient DNA studies." Trends in ecology & evolution **20**(10): 541-544.
- GILBERT, M. T. P., L. RUDBECK, E. WILLERSLEV, A. J. HANSEN, C. SMITH, K. E. PENKMAN, K. PRANGENBERG, C. M. NIELSEN-MARSH, M. E. JANS AND P. ARTHUR (2005b). "Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy." Journal of Archaeological Science **32**(5): 785-793.
- "Golden Software." (2017). from www.goldensoftware.com/products/surfer.

GREGORICKA, L. A. AND S. G. SHERIDAN (2017). "Continuity or conquest? A multi-isotope approach to investigating identity in the Early Iron Age of the Southern Levant." American Journal of Physical Anthropology **162**(1): 73-89.

HAAK, W., G. BRANDT, H. N. DE JONG, C. MEYER, R. GANSLMEIER, V. HEYD, C. HAWKESWORTH, A. W. PIKE, H. MELLER AND K. W. ALT (2008). "Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age." Proceedings of the National Academy of Sciences **105**(47): 18226-18231.

"haplotype / haplotypes " (2014a). Retrieved 14-12-2020, from <https://www.nature.com/scitable/definition/haplotype-haplotypes-142/>.

HOFREITER, M., D. SERRE, H. N. POINAR, M. KUCH AND S. PÄÄBO (2001). "ancient DNA." Nature Reviews Genetics **2**(5): 353-359.

HOOGWERFF, J. A., C. REIMANN, H. UECKERMAN, R. FREI, K. M. FREI, T. VAN ASWEGEN, C. STIRLING, M. REID, A. CLAYTON AND A. LADENBERGER (2019). "Bioavailable $^{87}\text{Sr}/^{86}\text{Sr}$ in European soils: A baseline for provenancing studies." Science of The Total Environment **672**: 1033-1044.

HUBER, N., W. PARSON AND A. DÜR (2018). "Next generation database search algorithm for forensic mitogenome analyses." Forensic Science International: Genetics **37**: 204-214.

JOBLING, M. (2009). "Surnames, Genes and the History of Britain." Retrieved 14-12-2020, from <https://www.le.ac.uk/ge/maj4/NewWebSurnames041008.html>.

KISTLER, L., R. WARE, O. SMITH, M. COLLINS AND R. G. ALLABY (2017). "A new model for ancient DNA decay based on paleogenomic meta-analysis." Nucleic acids research **45**(11): 6310-6320.

KRINGS, M., A. STONE, R. W. SCHMITZ, H. KRAINITZKI, M. STONEKING AND S. PÄÄBO (1997). "Neandertal DNA sequences and the origin of modern humans." cell **90**(1): 19-30.

LAFFOON, J. E., T. F. SONNEMANN, T. SHAFIE, C. L. HOFMAN, U. BRANDES AND G. R. DAVIES (2017). "Investigating human geographic origins using dual-isotope ($^{87}\text{Sr}/^{86}\text{Sr}$, $\delta^{18}\text{O}$) assignment approaches." PLoS One **12**(2): e0172562.

LALUEZA-FOX, C., M. L. SAMPIETRO, D. CARAMELLI, Y. PUDER, M. LARI, F. CALAFELL, C. MARTINEZ-MAZA, M. BASTIR, J. FORTEA AND M. D. L. RASILLA (2005). "Neandertal evolutionary genetics: mitochondrial DNA data from the Iberian Peninsula." Molecular biology and evolution **22**(4): 1077-1081.

LATHAM, K. E. AND J. J. MILLER (2019). "DNA recovery and analysis from skeletal material in modern forensic contexts." Forensic sciences research **4**(1): 51-59.

LINDAHL, T. (1993). "Instability and decay of the primary structure of DNA." Nature **362**(6422): 709-715.

LLAMAS, B., G. VALVERDE, L. FEHREN-SCHMITZ, L. S. WEYRICH, A. COOPER AND W. HAAC (2017). "From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era." STAR: Science & Technology of Archaeological Research **3**(1): 1-14.

- MALMSTRÖM, H., J. STORÅ, L. DALÉN, G. HOLMLUND AND A. GÖTHERSTRÖM (2005). "Extensive human DNA contamination in extracts from ancient dog bones and teeth." Molecular biology and evolution **22**(10): 2040-2047.
- MALMSTRÖM, H., E. M. SVENSSON, M. T. P. GILBERT, E. WILLERSLEV, A. GÖTHERSTRÖM AND G. HOLMLUND (2007). "More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA." Molecular biology and evolution **24**(4): 998-1004.
- MANN, A. E., J. A. F. YATES, Z. FAGERNÄS, R. M. AUSTIN, E. A. NELSON AND C. A. HOFMAN (2020). "Do I have something in my teeth? The trouble with genetic analyses of diet from archaeological dental calculus." Quaternary International.
- MASUYAMA, K., H. SHOJO, H. NAKANISHI, S. INOKUCHI AND N. ADACHI (2017). "Sex determination from fragmented and degenerated DNA by amplified product-length polymorphism bidirectional SNP analysis of amelogenin and SRY genes." PLoS One **12**(1): e0169348.
- MATHISEN, R. W. AND D. SHANZER (2011). Romans, barbarians, and the transformation of the Roman world: cultural interaction and the creation of identity in late antiquity, Ashgate Publishing, Ltd.
- MEYER, M., M. KIRCHER, M.-T. GANSAUGE, H. LI, F. RACIMO, S. MALICK, J. G. SCHRAIBER, F. JAY, K. PRÜFER AND C. DE FILIPPO (2012). "A high-coverage genome sequence from an archaic Denisovan individual." Science **338**(6104): 222-226.
- MILELLA, M., C. GERLING, T. DOPPLER, T. KUHN, M. COOPER, V. MARIOTTI, M. G. BELCASTRO, M. S. P. DE LEON AND C. P. ZOLLIKOFER (2019). "Different in death: Different in life? Diet and mobility correlates of irregular burials in a Roman necropolis from Bologna (Northern Italy, 1st–4th century CE)." Journal of Archaeological Science: Reports **27**: 101926.
- MOHANDESAN, E., S. J. MOWLA, A. HOJABRI NOOBARI, M. M. YAGHOUBI AND S. A. MESBAH-NAMIN (2004). "Extraction and analysis of ancient DNA from human remains of Masjede Kabood burial site." Iranian Journal of Biotechnology **2**(4): 236-242.
- MONROY KUHN, J. M., M. JAKOBSSON AND T. GÜNTHER (2018). "Estimating genetic kin relationships in prehistoric populations." PLoS One **13**(4): e0195491.
- ORLANDO, L., A. GINOLHAC, M. RAGHAVAN, J. VILSTRUP, M. RASMUSSEN, K. MAGNUSSEN, K. E. STEINMANN, P. KAPRANOV, J. F. THOMPSON AND G. ZAZULA (2011). "True single-molecule DNA sequencing of a pleistocene horse bone." Genome research **21**(10): 1705-1719.
- PÄÄBO, S., H. POINAR, D. SERRE, V. JAENICKE-DESPRÉS, J. HEBLER, N. ROHLAND, M. KUCH, J. KRAUSE, L. VIGILANT AND M. HOFREITER (2004). "Genetic analyses from ancient DNA." Annu Rev Genet **38**(1): 645-679.
- PEARCE, J. (2014). "The Interpretation of Provincial Roman Grave Good Assemblages." Death and Changing Rituals: Function and meaning in ancient funerary practices **7**: 223.
- PEYRÉNE, S. AND B. M. PETER (2020). "AuthentiCT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination." bioRxiv.

- PEYRÉGNE, S. AND K. PRÜFER (2020). "Present-Day DNA Contamination in Ancient DNA Datasets." BioEssays **42**(9): 2000081.
- PILLI, E., A. MODI, C. SERPICO, A. ACHILLI, H. LANCIONI, B. LIPPI, F. BERTOLDI, S. GELICHI, M. LARI AND D. CARAMELLI (2013). "Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains." PLoS One **8**(1): e52524.
- POWER, R. C., D. C. SALAZAR-GARCÍA, M. RUBINI, A. DARLAS, K. HARVATI, M. WALKER, J.-J. HUBLIN AND A. G. HENRY (2018). "Dental calculus indicates widespread plant use within the stable Neanderthal dietary niche." Journal of Human Evolution **119**: 27-41.
- PROWSE, T., J. BARTA, T. VON HUNNIUS AND A. SMALL (2010). "Stable isotope and ancient DNA evidence for geographic origins at the site of Vagnari (2nd-4th centuries AD), Italy." H. Eckhart (a cura di), Roman Diasporas: Archaeological Approaches to Mobility and Diversity in the Roman Empire. Journal of Roman Archaeology Supplement(78): 175-197.
- R_CORE_TEAM (2020). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.
- R_STUDIO_TEAM (2020). RStudio: Integrated Development Environment for R. Boston, MA, RStudio, PBC.
- RAVEN, P., G. JOHNSON, J. LOSOS AND S. SINGER (2007). Biologie. de Boeck, Bruxelles.
- RENAUD, G., M. SCHUBERT, S. SAWYER AND L. ORLANDO (2019). Authentication and assessment of contamination in ancient DNA. Ancient DNA, Springer: 163-194.
- RICHARDS, M. B., B. C. SYKES AND R. E. HEDGES (1995). "Authenticating DNA extracted from ancient skeletal remains." Journal of Archaeological Science **22**(2): 291-299.
- SAMPIETRO, M., D. CARAMELLI, O. LAO, F. CALAFELL, D. COMAS, M. LARI, B. AGUSTI, J. BERTRANPETIT AND C. LALUEZA-FOX (2005). "The genetics of the pre-Roman Iberian Peninsula: A mtDNA study of ancient Iberians." Annals of human genetics **69**(5): 535-548.
- SAMPIETRO, M. L., M. T. P. GILBERT, O. LAO, D. CARAMELLI, M. LARI, J. BERTRANPETIT AND C. LALUEZA-FOX (2006). "Tracking down human contamination in ancient human teeth." Molecular biology and evolution **23**(9): 1801-1807.
- SCHWEISSING, M. M. AND G. GRUPE (2003). "Stable strontium isotopes in human teeth and bone: a key to migration events of the late Roman period in Bavaria." Journal of Archaeological Science **30**(11): 1373-1383.
- SERRE, D., A. LANGANEY, M. CHECH, M. TESCHLER-NICOLA, M. PAUNOVIC, P. MENNECIER, M. HOFREITER, G. POSSNERT AND S. PÄÄBO (2006). No evidence of Neandertal mtDNA contribution to early modern humans. Early modern humans at the Moravian gate, Springer: 491-503.
- "SNP." (2014b). Retrieved 14-12-2020, from <https://www.nature.com/scitable/definition/snp-295/>.
- STARK, R. J., M. V. EMERY, H. SCHWARCZ, A. SPERDUTI, L. BONDIOLI, O. E. CRAIG AND T. PROWSE (2020). "Imperial Roman mobility and migration at Velia (1st to 2nd c. CE) in southern Italy." Journal of Archaeological Science: Reports **30**: 102217.

- STEVENSON, E., S. ACIEGO, P. CHUTCHARAVAN, I. PARKINSON, K. BURTON, M. BLAKOWSKI AND C. ARENDT (2016). "Insights into combined radiogenic and stable strontium isotopes as tracers for weathering processes in subglacial environments." Chemical Geology **429**: 33-43.
- TANG, S. AND T. HUANG (2010). "Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system." Biotechniques **48**(4): 287-296.
- UNIVERSITÉ-PIERRE-ET-MARIE-CURRIE (s. d.). "Origine des polymorphismes : SNP." Retrieved 14/08/2017, from http://www.edu.upmc.fr/sdv/masselot_05001/polymorphisme/snp.html.
- VAI, S., C. E. G. AMORIM, M. LARI AND D. CARAMELLI (2020). "Kinship Determination in Archeological Contexts Through DNA Analysis." Frontiers in Ecology and Evolution **8**: 83.
- VAN OVEN, M. AND M. KAYSER (2009). "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation." Human mutation **30**(2): E386-E394.
- VANEK, D., L. SASKOVA AND H. KOCH (2009). "Kinship and Y-chromosome analysis of 7th century human remains: novel DNA extraction and typing procedure for ancient material." Croatian medical journal **50**(3): 286-295.
- VERSLYPE, L. AND E. WEINKAUF (n.d.). Personnal communication about the archaeological studies of the Gallo-Roman necropolis of the Later Roman Empire on the Rue Perdue in Tournai.
- WANDELER, P., S. SMITH, P. A. MORIN, R. PETTIFOR AND S. FUNK (2003). "Patterns of nuclear DNA degeneration over time—a case study in historic teeth samples." Molecular Ecology **12**(4): 1087-1093.
- WEYRICH, L. S., S. DUCHENE, J. SOUBRIER, L. ARRIOLA, B. LLAMAS, J. BREEN, A. G. MORRIS, K. W. ALT, D. CARAMELLI AND V. DRESELY (2017). "Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus." Nature **544**(7650): 357-361.
- WEYRICH, L. S., A. G. FARRER, R. EISENHOFER, L. A. ARRIOLA, J. YOUNG, C. A. SELWAY, M. HANDSLEY-DAVIS, C. J. ADLER, J. BREEN AND A. COOPER (2019). "Laboratory contamination over time during low-biomass sample analysis." Molecular ecology resources **19**(4): 982-996.
- WILLMES, M., L. MCMORROW, L. KINSLEY, R. A. ARMSTRONG, M. AUBERT, S. EGGINS, C. FALGUERES, B. MAUREILLE, I. MOFFAT AND R. GRÜN (2014). "The IRHUM (Isotopic Reconstruction of Human Migration) database-bioavailable strontium isotope ratios for geochemical fingerprinting in France."

7. Supplementary data

S1. R code for Figure 12 and 13.

```
#-----  
# Définition du répertoire de travail et installation des packages  
#-----  
setwd("C:/Users/laura/Documents/UCL/Mémoire")  
  
#install.packages("readxl") # pour lire des fichier Excel  
#install.packages("ade4") # Permet d'installer le package "ade4".  
#install.packages("pander") # Librairie pour afficher des tableaux  
#install.packages("FactoMineR") # librairie Factominer développé par URennes  
#install.packages("pastecs") # librairie qui inclut la fonction de stat descriptive stat.desc  
#install.packages("cluster") # librairie pour faire du clustering  
#install.packages("lme4") # Librairie pour estimer des modèles mixtes  
#install.packages("lmerTest") # Librairie pour faire des tests dans les modèles mixtes  
  
#-----  
# Chargement des packages  
#-----  
#require(pander) # Librairie pour afficher des tableaux  
#require(FactoMineR) # librairie Factominer développé par URennes  
#require(pastecs) # librairie qui inclut la fonction de stat descriptive stat.desc  
#require(cluster) # librairie pour faire du clustering  
#require(lme4) # Librairie pour estimer des modèles mixtes  
#require(lmerTest) # Librairie pour faire des tests dans les modèles mixtes  
#require(tinytex) # librairie pour pouvoir créer un fichier pdf à partir de RMD  
library(ggplot2) # pour faire des plot de type ggplot  
library(ggpubr) # Wrapper au dessus de ggplot avec des chouettes facilités supplémentaires  
#library(gridExtra) # package pour arranger des ggplot en grilles  
#require(patchwork) # pour mettre plusieurs ggplots en patwork sur un graphique  
#library(pastecs) # Package pour l'analyse de "Space-Time Ecological Series" mais aussi des  
stat descriptives sympa tout simplement  
require(readxl) # pour lire des fichier Excel  
#require(plotly) # pour pouvoir faire des graphiques interactifs  
#require(GGally) # Wrapper de ggplot 2 pour faire des belles matrices de graphiques  
#require(arm)# Fonction display (plutôt que summary)  
require(tidyverse) # Fonction tibble pour contruire un data frame à implémenter en ggplot.  
require(ade4)  
require(car)  
#require(VCA)  
#require(lmerTest)  
#require(ggeffects)  
#require(pbkrtest)  
library(readxl)
```

```

#library(lsmmeans)
#library(VCA)
#library(lmerTest)
#library(cowplot)
#library(reshape)      # melt()

#-----
# importation des donnees
#-----

Mitochondrial<-read.table("DataDNAa.csv", sep=";", dec=".", header=T)
Mitochondrial[, "PCR"]=as.factor(Mitochondrial[, "PCR"])
Mitochondrial[, "Target"]=as.factor(Mitochondrial[, "Target"])

alldata<-read.table("DataAlIDNAa.csv", sep=";", dec=".", header=T)

#-----
#Mitochondrial
#-----

ggplot(Mitochondrial) +
  aes(x =Size, y =SuccessRate) +
  geom_point(colour = "blue", alpha = .2) +
  xlab("Size") +
  ylab("SuccessRate")

# modele lineaire pour ln(SuccessRate) en fonction de Size
# sans les observations t.q. SuccessRate == 0
# Fonction log(): logarithme neperien

Mitochondrial_In <- filter(Mitochondrial, SuccessRate > 0) %>%
  mutate(lnSuccessRate = log(SuccessRate))
display(lm(lnSuccessRate ~ Size, data = Mitochondrial_In), digits = 3)

ggplot(Mitochondrial, aes(x = Size, y = SuccessRate)) +
  geom_point(size = 2) +
  geom_line(data = Mitochondrial_In,
    mapping = aes(x = Size,
      y = exp(fitted.values(lm(lnSuccessRate ~ Size, Mitochondrial_In)))),
    colour = "blue", size = 0.5) +
  labs(title = "Success Rate of mitochondrial amplification by size (bp)")

#---- QQ-plot des residus standardizes-----

```

```

lm_fitted <- lm(SuccessRate ~ Size, data = Mitochondrial)
plot(lm_fitted, which = 2)

## ----homoskedasticity, dependson = "QQ", cache = TRUE, out.width = "80%"----
# figure: les residus standardises en fonction des valeurs predites
r <- rstandard(lm_fitted)
ggplot(tibble(`standardized residual` = r,
              `fitted value` = fitted(lm_fitted)),
       aes(x = `fitted value`, y = `standardized residual`)) +
  geom_point(size = 3) +
  geom_hline(yintercept = 0, color = "grey", size = 1.5)

## ----leverage, cache = TRUE, out.width = "70%"-----
# figure: les leviers
ggplot(tibble(`observation` = 1:nrow(Mitochondrial),
              `leverage` = lm.influence(lm_fitted)$hat),
       aes(x = observation, y = `leverage`)) +
  geom_point(size = 2)

## ----Cook, dependson = "data", cache = TRUE, echo = TRUE, out.width = "70%"----
# figure: distance de Cook
plot(lm(SuccessRate ~ Size, data = Mitochondrial), which = 4)

## ----res_lev_Cook, dependson = "data", cache = TRUE, echo = TRUE-----
# figure: residus standardises et leviers
plot(lm(SuccessRate ~ Size, data = Mitochondrial),
     which = 5, add.smooth = FALSE)

#-----
#Individuals
#-----

ggplot(alldata) +
  aes(x = Individuals, y = SuccessRateIndividual) +
  geom_point(colour = "black", alpha = 2) +
  xlab("Individuals") +
  ylab("SuccessRate")

```