UNIVERSITY OF VERONA

UNIVERSITY OF VERONA

DEPARTMENT OF NEUROSCIENCES, BIOMEDICINE AND MOVEMENT SCIENCES

GRADUATE SCHOOL OF

APPLIED LIFE AND HEALTH SCIENCES

DOCTORAL PROGRAM IN

LIFE AND HEALTH SCIENCES

31° CYCLE

S.S.D. MED/07

TITLE OF THE DOCTORAL THESIS

PALEOGENOMICS AND PALEOMICROBIOLOGY APPROACHES APPLIED TO THE UNDERSTANDING OF TRUE AND APOCRYPHAL PESTS IN DIFFERENT BIOARCHAEOLOGICAL CONTEXTS: THE STUDY OF ANCIENT INFECTIOUS DISEASES

Coordinator: Prof. Giovanni Malerba

Tutor: Prof. Giuseppe Cornaglia

Co-Tutor: Dott.ssa Elisabetta Cilli

Co-Tutor: Prof. Nicola Vitulo

Doctoral Student: Dott.ssa Alda Bazaj

Alda Bazaj

Paleogenomics and paleomicrobiology approaches applied to the understanding of true and apocryphal pests in different bioarchaeological contexts: the study of ancient infectious diseases – Alda Bazaj

PhD thesis

Verona, 8 April 2020

Abstract

Ph.D. Course in Applied Life and Health Sciences, 31th cycle

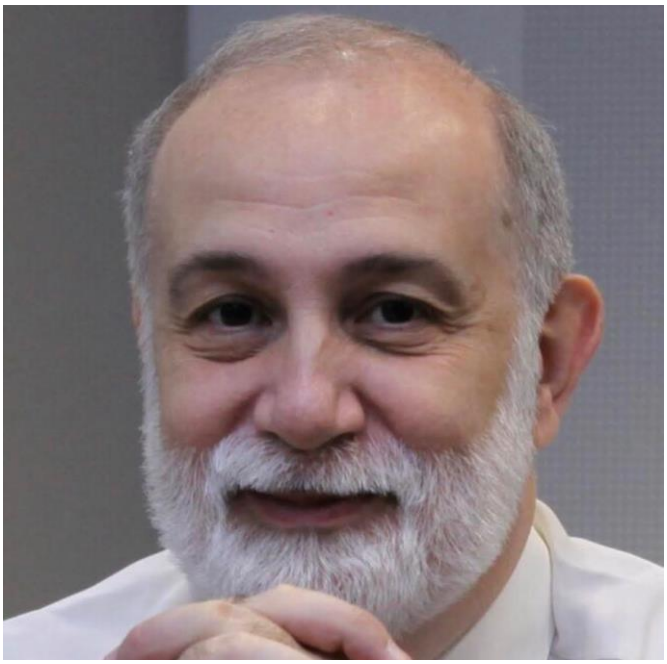
Doctor of Philosophy

Paleogenomics and paleomicrobiology approaches applied to the understanding of true and apocryphal pests in different bioarchaeological contexts: the study of Ancient infectious diseases

By Alda BAZAJ

Plague is a bacterial disease caused by *Yersinia pestis*, which primarily affects wild rodents. It is spread from one rodent to another by fleas. To date, Plague is probably the infectious pathology responsible of the largest amount of deaths among all human history. Unfortunately, it is still persistent in some area of the world, as on November 4, 2014 the Ministry of Health of Madagascar reported an outbreak of plague (https://www.epicentro.iss.it/peste/aggiornamenti) to the World Health Organization. The first case was identified on August 31, a male from Soamahatamana village in the district of Tsiroanomandidy, who died on September 3. As of November 16, a total of 119 cases of plague were confirmed, including 40 deaths.

Next-generation sequencing (NGS) and metagenomics has recently revolutionized genomic research, and its combination with high-throughput target enrichment method can be proficiently applied to the study of ancient DNA (aDNA), thus providing a powerful tool for understanding the evolution of pandemic infectious diseases like the plague.



(1958-2020)

Al professor ***Giuseppe Cornaglia***

Grazie per la fiducia che hai avuto in me e per avermi insegnato a non arrendermi mai…

INDEX

*To my husband **Valerio**, who fostered my love for learning, and my child **Daniele**.*

A.B.

Chapter 1

INTRODUCTION

1.1 Ancient DNA

Ancient DNA (aDNA) provides direct insights onto the past that modern DNA or paleontological studies alone cannot provide. It has been proven to address questions regarding history relationships, population dynamics and diversity through time. Although aDNA can be a very powerful tool, it should be handled with care [Fulton et al., 2012]. The first studies about aDNA began in 1984, when the DNA of an extinct Quagga, a relative of a zebra, were recovered [Higuchi et al., 1984]. History started changing from that moment, as technical and biological advances permitted to evolve Paleogenomics through time.

The state-of-the-art genomic technologies, through the combination of high-throughput sequencing and the most recent bioinformatics tools, has allowed the study of whole genome sequences of large population datasets, the identification of ancient pathogens and their evolution [Allentoft et al., 2015, Fu et al., 2016, Olalde et al., 2018]. Next-generation sequencing (NGS) has completely revolutionized aDNA research, when is well combined with high-throughput target enrichment methods. On the other side, aDNA present specific limitations that require careful consideration during data analysis.

1.1.1 aDNA specificities

Ancient DNA can be defined as any genomic sequence retrieved from dead organisms. The DNA of each living organism is frequently damaged through time, but such damages are repaired by mechanisms that preserve the integrity of the genetic material. While after death the damage of DNA still continues, the repairing mechanisms can no longer maintain the DNA intact. Therefore, most of the aDNA sequences found nowadays are in different stages of degradation.

1.1.2 aDNA degradation

DNA degradation is influenced mostly by atmospheric conditions such as temperature or humidity, and other variables linked with the burial environment, namely salt concentration, soil pH or the chemical composition of the ground.

The degradation process reduces the amount of endogenous aDNA present in the samples, usually accounting for less than 1% of the total sequenced reads [Fu et al., 2013]. On the other hand, in some exceptional cases, the retrieved endogenous DNA was found to exceed 70% [Meyer et al., 2012, Raghavan et al., 2014, Rasmussen et al., 2010, Gamba et al., 2014, Keller et al., 2012, Carpenter et al., 2013, Lazaridis et al., 2014, Olalde et al., 2014].

Nevertheless, environmental factors are still the biggest threat to DNA integrity. Upon exposure to climatic and environmental agents, DNA is subjected to chemical reactions such as deamination, depurination or hydrolysis that damage DNA structure (Figure 1), ultimately leading to its degradation [Höss et al., 1996]. In addition, oxidation can induce DNA lesions that impair Polymerase Chain Reaction

(PCR) by blocking the polymerase, resulting in low amplification and potentially chimeric artifacts, therefore called PCR inhibiting lesions. Due to oxidation, most of the "surviving" DNA sequences are very short, less than 100 base pairs (bp) [Poinar et al., 2016], and may contain damaged nucleotides.

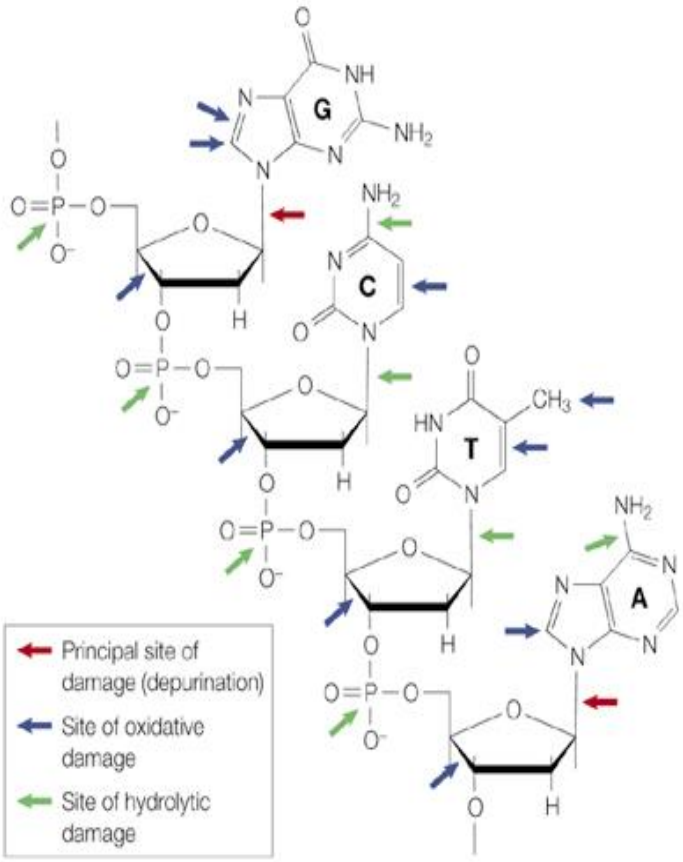


Figure 1. Principal sites where damage affects ancient DNA. *Oxidative damage modifies the nitrous bases and the sugar-phosphate backbone of the DNA, depurination causes breaks in the DNA chain, hydrolytic damage also leads to DNA chain breaks [Hofreiter et al., 2001].*

1.1.2.1 Cytosine Deamination

The most characteristic degradation of DNA molecules due to post-mortem damage is the hydrolytic deamination of Cytosine (C) to Uracil (U) [Gilbert et al., 2007], which codes as Thymine (T). During DNA replication, C deamination causes the misincorporation of an Adenine (A) instead of the original Guanine (G), which is followed by C to T substitutions in the 5' ends of the sequences [Hofreiter et al., 2001]. The C to T replacement at the 5' ends of the DNA fragments, results in a higher frequency of G to A substitutions at the 3' ends of the complementary strands [Briggs et al., 2007, Rasmussen et al., 2014]. It is unclear whether any other miscoding lesions may be relevantly frequent in aDNA molecules or their distribution along the sequence.

Cytosines deamination particularly affects aDNA reads ends, where the percentage of deaminated C can exceed 40% [Briggs et al., 2007]. Specific techniques have been implemented to minimize errors in sequencing and mapping procedures due to aDNA damage. Indeed, single stranded library building protocol can be an efficient method to analyse poor quality samples, characterized by a low rate of endogenous DNA and highly degraded strands [Gansauge et al., 2013]. On the other side, the presence of cytosine deamination can also be used to discriminate between reads obtained by real aDNA or modern contaminant [Rohland et al., 2009].

1.1.2.2 aDNA Depurination

DNA fragmentation is mainly caused by depurination, namely the disruption of an N-glucosyl bond between a purine and the sugar of the DNA chain, resulting in a chain with an abasic site. The chain is then fragmented through β elimination, leaving 3′-aldehydic and 5′-phosphate ends (Figure 2) [Briggs et l., 2007]. Depurination, considered the most critical chemical damage to aDNA structure, ultimately causes the underrepresentation of purines (G and A) usually at the 5' fragment ends [Briggs et al., 2007, Orlando et al., 2011 and Meyer et al., 2012].

A recent study demonstrated that depurination occurs at both ends of the aDNA fragments [Meyer et al., 2012] and suggested an explanation for the observation in previous publications of depurination only at the 5' ends. Such reason was hypothesized to reside in the pre-processing step required by most library building protocols, namely the blunt-ends repair, an enzymatic process that extends recessed and degrades overhanging 3′ ends of DNA fragments [Briggs et al., 2007]. Thus, only with the development of single-stranded DNA libraries, this process could be observed also at the 3' ends of the aDNA fragments [Meyer et al., 2012].

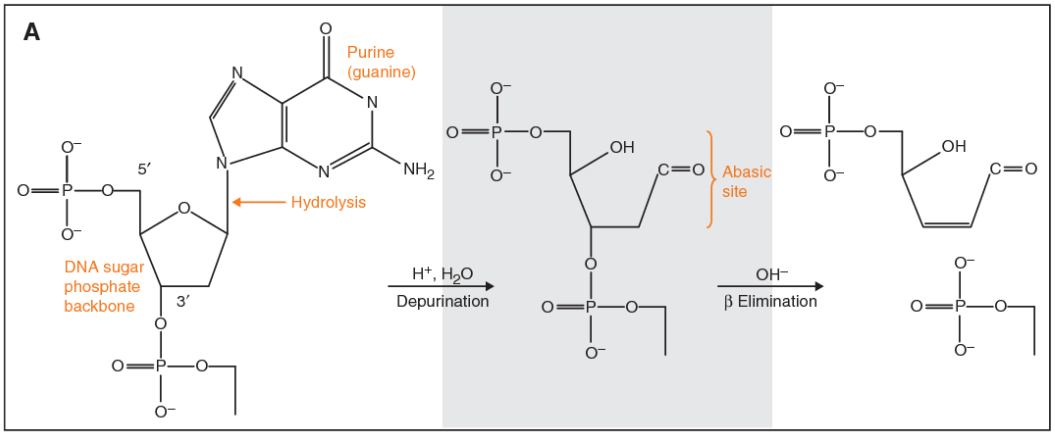


Figure 2. Depurination. *A chemical reaction in which an N-glycosyl bond is broken resulting in an abasic site. The abasic site is later removed, fragmenting the DNA through β elimination [Dabney et al., 2013].*

1.1.2.3 Fragment length

One of first published studies describing the aDNA peculiarities showed that even in well-preserved specimens, only very short fragments (50-150 bp) of aDNA could be retrieved. The entity of aDNA degradation is inversely correlated with the amplification efficiency and the length of the amplification product [Pääbo et al., 1989]. The same consideration was reported by the one of the first studies on aDNA involving high-throughput sequencing, describing the difficulties of both aDNA extraction and library preparation [Green et al., 2010].

1.1.3 Contamination

Research on endogenous aDNA of an extracted sample can result particularly challenging due to the small portion of survived copies of endogenous aDNA in an extract, compared with modern DNA present in the environment. The PCR amplifies not only those small portions of aDNA, but also DNA from different sources, which may contaminate the sample during different manipulation steps of aDNA specimen. Indeed, the sample itself can be contaminated by adhering microorganisms residing in the environment, as bones, ribs and teeth are extremely porous. The highest risk of contamination resides in the collection of specimens, particularly in human and microbials studies, yet another source of contamination is represented by sample handling procedures in the laboratory (DNA extraction and PCR). The air filtering system of the building may be contaminated by the presence of insects or other biological entities, although the major concern is represented by the contamination with exogenous DNA. Laboratory personnel and reagents may introduce exogenous human or animal DNA, as living organisms are constantly shedding DNA-bearing tissues in the form of skin cells, hair, saliva, and other secretions. Archaeological skeletal remains, for example (Figure 3), may be

contaminated by the DNA from organisms in the soil, microorganisms growing within the bones, excavators, curators, or even the DNA analysts themselves.

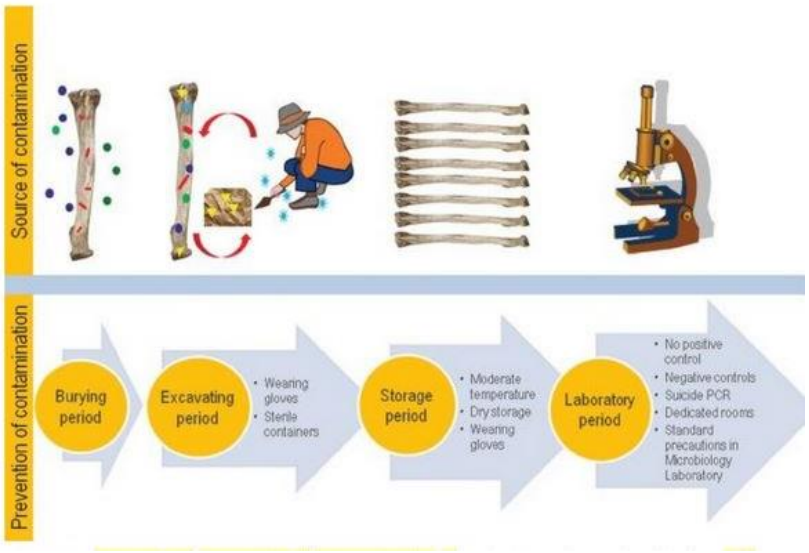


Figure 3. Example of source and prevention of contamination of ancient specimens. *Adapted from Drancourt and Raoult 2016.*

The problem of contamination is worsened by the nature of PCR, since the reaction preferentially amplifies well-preserved DNA molecules, which are more likely to be modern contaminants than actual aDNA. Since PCR produces large amounts of highly concentrated DNA, laboratories often encounter problems with previous PCR products contaminating current work. Contamination is critical when attempting to retrieve DNA from ancient human remains, as humans are also the main source of exogenous DNA, making contamination more difficult to detect. Several published DNA sequences from very ancient remains are now widely held to be inauthentic [Lindahl, 1997]. Moreover, the specimens from which the aDNA is extracted are often unique and the analysis is time consuming, therefore independent replication of results is not always carried out. A number of precautions have been developed to reduce the chances of contamination and to increase the likelihood of identifying contamination. Specifically, it is very important to have different isolated spaces in the laboratory where: i) to cut and classify specimens; ii) to perform DNA extraction (the main core of the lab) and iii) to run PCR.

1.2 Paleomicrobiology

Paleomicrobiology is defined as the study of microorganisms in ancient remains that were naturally present in healthy organisms as well as those that were responsible for infectious diseases [Rivera-Perez et al., 2016]. This discipline includes branches of medical microbiology, history, anthropology, and archaeozoology, and is aimed at evaluating the evolution of ancient pathogens (recently including also ancient microbiota) through their identification and the analysis of functional data such as antimicrobial resistance [Drancourt et al., 2016 chap. 5 Paleomicrobiology of Humans]. Molecular analysis of ancient pathogens can also provide useful information to reconstruct past epidemic trends and help refining even the most recent models of emerging infections, thus giving an important contribution to the development of adequate preventive measures. In addition, the combination of microbial and human metagenomics data has a huge potential to extend the paleomicrobiology field to a wider community of scientists and scholars [Bazaj et al., 2015].

Before the development of genomic techniques, the identification of infectious pathogens in ancient remains was restricted to the visual identification of bone injuries and its correlation with ancient written proofs [Kousoulis et al., 2012]. There are many pathogen casualties, though, for which it is impossible to visually identify the etiological cause of the reported mortality, for example Syphilis (caused by *Treponema pallidum*) can be confused with skeletal lesions [Rothschild et al., 2005], or Brucellosis (caused by *Brucella melitensis*) can be misclassified as Tuberculosis (caused by *Mycobacterium tuberculosis*) [Mutolo et al., 2012, Kay et al., 2014]. Another paradigmatic example of this approach is that of Medieval Black Death, which could be attributed to a viral pandemic with an aerosol transmission pattern, based on the descriptions from historical records [Bossak et al., 2007]. For such cases, the etiological cause can only be univocally identified through genetic markers.

The first genetic studies that targeted ancient pathogens were performed with PCR techniques [Kolman et al., 1999, Gernaey et al., 2001, Drancourt et al., 2003, Zink et al., 2003 and Nguyen-Hieu et al., 2010]. Such studies required a prior pathological diagnosis, as PCR relies on specific primers to amplify each possible pathogen [Willerslev et al., 2007], however as bacteria and virus are ubiquitous, the specificity of these tests is usually underrated and false positive results are frequent [Päabo et al., 2004, Gilbert et al., 2005, Gilbert et al., 2006].

The publication of the draft genome of *Y. pestis* from British individuals of the 14th century [Bos et al., 2011] is a milestone as it was the first draft genome of a pathogen obtained from ancient human remains. Other strains of *Y. pestis* were then characterized, from the Bronze Age [Rasmussen et al., 2015, Spyrou et al., 2018] to the 19th century pandemics [Stenseth et al., 2008] in Europe and from China [Cui et al., 2013]. The next step for paleomicrobiology is the retrieval of aDNA in unexplored environments, the development of protein-based approaches and their integration to unravel genetic adaptation through time.

The discoveries in the field have been always related with the ultimate technical improvements, indeed Second Generation Sequencing (SGS) provided a huge amount of data to analyse, whereas the advent of third generation sequencing methods might supply the basis for de-novo assembly of aDNA remains, the ultimate goal that was left unreached so far. Another alternative and promising approach is represented by metagenomics performed on aDNA, which could potentially lead to the recovery of complete extinct environments.

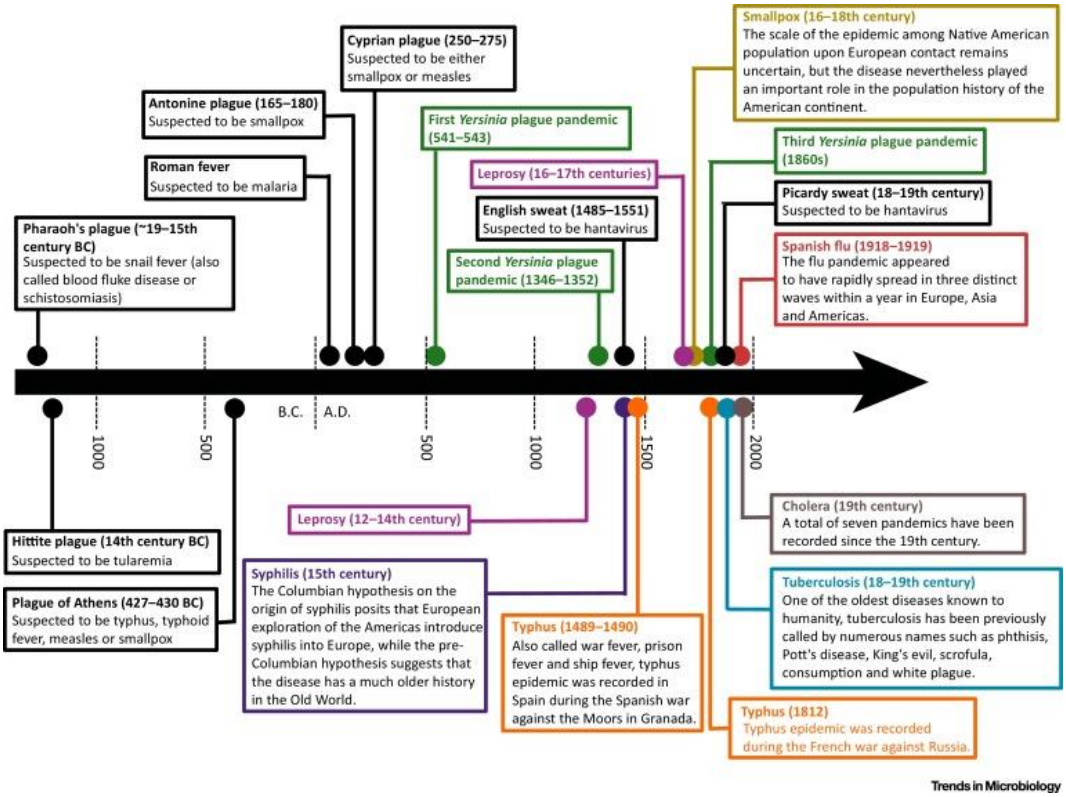


Figure 4. Overview and Timeline of Historically Notable Disease Outbreaks in Human History. *Adapted from Andam et al., 2016.*

1.3 The plague

Plague is an infectious disease of bacterial origin still present in many parts of the world, including some regions of industrialized countries. It is caused by *Yersinia pestis* bacterium, which is typically hosted by the parasitic fleas of rodents, rats, squirrels and in some cases also pets such as cats. Normally, *Y. pestis* has low mortality rates in these species which can therefore be considered as long-term infectious reserves, indeed it is still present in 22 rodent reservoirs. The origin of the plague is very ancient, and due to its destructive force, it has been collectively named “the black death”, a disease that has accompanied humanity over the centuries and was often present in the great literary works and art (Figure 5).

Figure 5. The Triumph of Death (P. Bruegel, 1562). *Oil panel painting showing an allegory of the Last Judgment influenced by the medieval plague scenes.*

Three main *Y. pestis* epidemics have affected Europe in historical times: The First Pandemic started with the Plague of Justinian (541-543 AD) and continued until

~750 AD [Russell et al., 1968]. The Second Pandemic, named the Black Death, was probably the most famous and was responsible for killing up to 40% of the European population [Rasmussen et al., 2015] during the 14th century (1346-1352 AD). The second wave of the Black Death was named the Great Plague (1665-1666 AD), which caused the death of almost one third of the European population and literally infected all the countries from the Mediterranean to Scandinavia and Russia within five years. In Europe it remained endemic, bouncing back into cycles of 10-12 years at least for the next three centuries until the 18th century [Zietz et al., 2004]. Finally, the Third Pandemic started in China in 1860s, with the outbreak of a serious epidemy in 1894, before spreading all over the world as series of pestilences between the 18th and 19th centuries [Cohn et al., 2008; Stenseth et al., 2008] until the end of the 20th century [Bos et al., 2011; Cui et al., 2013; Drancourt et al., 1998; Wagner et al., 2014].

Y. pestis strains responsible for all three major epidemics were identified, namely the strains from the Justinian outbreak [Wagner et al., 2014], the 14th century Black Death strains [Bos et al., 2011, Schuenemann et al., 2014] and 18th century pandemic [Bos et al., 2016]. Based on literature records, earlier *Y. pestis* outbreaks may have occurred in Europe before the Justinian Plague, such as the Plague of Athens (427-430 BC) and Antonine Plague (165-180 AD). The lack of DNA evidence, though, does not allow either the confirmation of such events or the identification of the pathogen linked with the historical records [Drancourt et al., 2002; Drancourt and Raoult 2002]. The earliest evidence of *Y. pestis* DNA presence in human remains was detected in Late Neolithic and Bronze Age individuals from the steppe and eastern Europe (5000-3500 BP) [Rasmussen et al., 2015]. The analysis of these strains revealed that most recent common ancestor of all European *Y. pestis* strains lived up to ~6000 years BC, indicating that the Black Death causal strain was present in Europe at least since the Bronze Age.

Although as of 2018 [WHO 2018] it is still endemic in only 17 countries (Figure 6), the plague has a remarkable evocative power and immediately brings back images of horror and devastation.

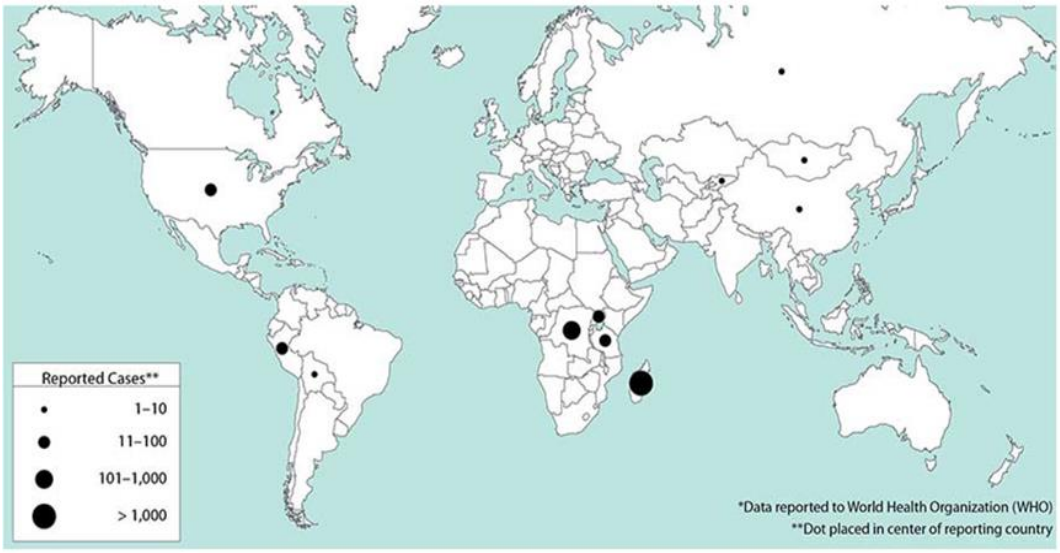


Figure 6. Reported Plague Cases by Country (2013-2018 WHO)

1.4 *Yersinia pestis*

The genus *Yersinia*, member of the family *Enterobacteriaceae*, consists of 11 species, 3 of them are human pathogens (*Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica*).

Yersinia pestis was discovered in 1894 and named *Pasteurella pestis* by Alexandre Yersin, a French/Swiss physician and bacteriologist from the Pasteur Institute, during an endemic of plague in Hong Kong, and was renamed *Yersinia pestis* in 1944 after his discoverer. Yersin noted that rats were affected by plague even before epidemics in humans and that plague was regarded by many locals as a disease of rats; indeed, villagers in China and India asserted that when large numbers of rats were found dead, plague outbreaks soon followed [Yersin and Treill 1894].

Y. pestis is a Gram-negative, rod-shaped, non-motile, non-lactose fermenting, non-spore forming, facultative anaerobic coccobacillus presenting cell wall, lipid composition and antigens typical of enterobacteria. The plague coccobacillus is a mandatory parasite growing at temperatures from 4°C to 40°C, optimally growing between 28°C and 30°C, with a remarkably stable and vigorous virulence, ability to multiply in the tissues of its host and cause death. Its lipopolysaccharide is characterized rough, there is no true capsule, however there is a carbohydrate-protein envelope, named capsular antigen or fraction 1 (F1), which forms during growth above 33°C and confers antiphagocytic properties [Perry and Fetherston 1997].

Human plague has three clinical forms: pneumonic, bubonic and septicemic. Bubonic plague is the most-known form in popular lore, it constitutes about three-fourths of plague cases. It is also the least dangerous form of plague, accounting today for virtually no deaths and in the past killing only half of its victims (at a time when contracting the other forms of plague brought almost certainly to death).

Typically, bubonic plague appears two-to-six days after *Yersinia* infection with symptoms like shivering, vomiting, headache, giddiness, light-susceptibility, back pain, limbs pain and sleeplessness with apathy or delirium in complicated cases. The most characteristic sign, however, is the appearance of one or more tender, swollen lymph nodes, or buboes, usually distributed in the groin area and armpits. The temperature rises rapidly above 40 °C and frequently falls slightly on the second or third day, with marked fatigue. Bubonic plague is not directly infectious from person to person; the bacillus is carried from rodent to person or from person to person by infected fleas (*Xenopsylla cheopis*). Once ingested by the flea, it multiplies until the insect's digestive tract is blocked. When the flea bites another rodent or a human, bacilli are released into the new host and migrate through the lymphatic system to lymph nodes, where they produce proteins that impair the normal inflammatory response by preventing the intervention of infection-fighting macrophages. After weakening the host's immune response system, the bacilli quickly colonize the lymph nodes, producing a painful swelling and, eventually, destroying the tissue. Finally, they cause general septicaemia or blood poisoning by entering the blood stream either directly, or from the lymph nodes, where they can be found in abundance together with spleen, bone marrow and liver.

At odds with bubonic plague, pneumonic plague can directly be transmitted through human-to-human contact [Treille and Yersin 1894], as the bacillus can be passed to other people in droplets expelled by coughing or sneezing, therefore it is highly infectious. Pneumonic plague can also develop as a complication of bubonic plague and displays the same symptoms as severe pneumonia (fever, weakness, and shortness of breath) followed by pulmonary edema and eventually death in 3-4 days if not treated properly. Additional symptoms include insomnia, stupor, staggering gait, speech disorder, and loss of memory. Extensive control measures against rats and their fleas have eliminated plague from Europe, but it is still occurring in other regions of the world.

Yersinia pestis is one of the most studied pathogenic microorganisms in paleomicrobiology, having the largest number of published strains recovered from ancient remains, as well as the most complete timeline due to its famous outbreaks

[Bos et al., 2011, Wagner et al., 2014]. For such reasons *Y. pestis* can be considered as the paradigmatic pathogen in the research field of infectious diseases in past populations. *Y. pseudotuberculosis* is the much less pathogenic ancestor (Figure 7) of *Y. pestis* [Achtman et al., 1999].

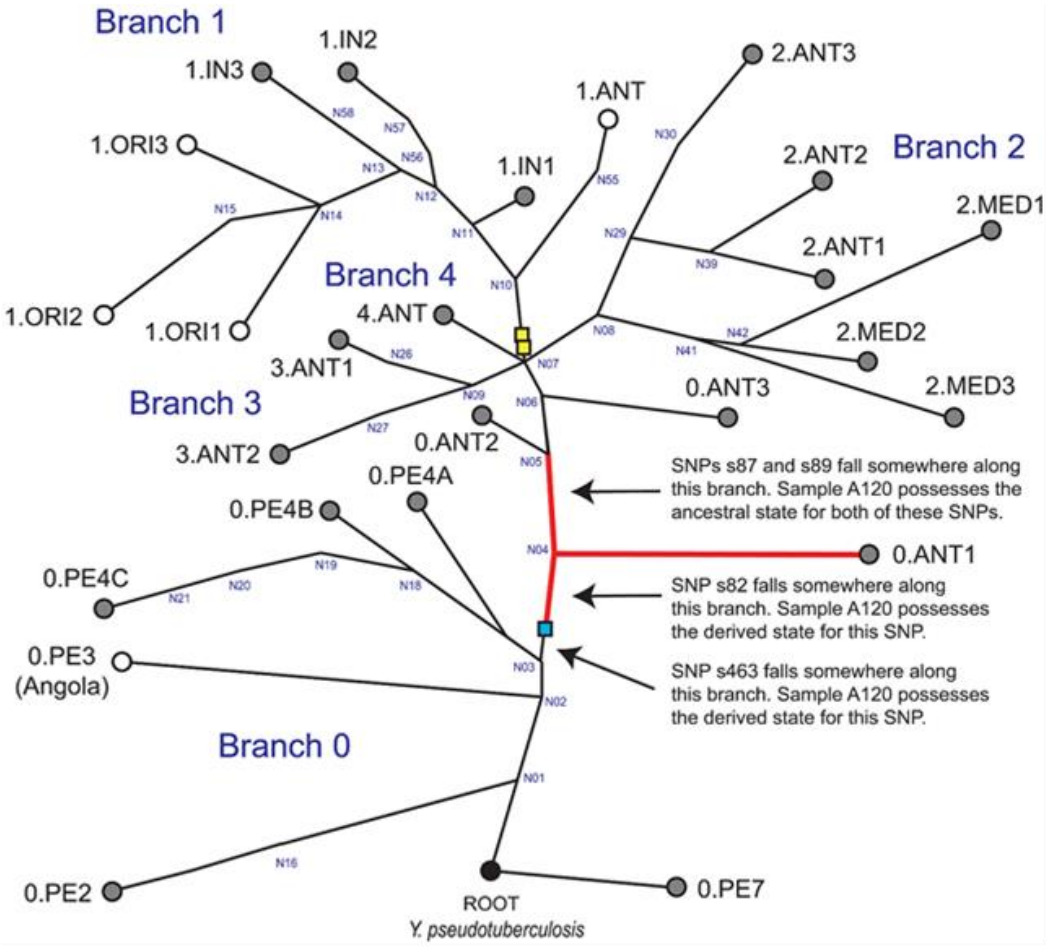


Figure 7. Global Phylogeny for Y. pestis [Harbeck et al., 2013]

Chapter 2

OBJECTIVES

The main objective of this thesis was to investigate ancient specimens from three different bioarchaeological contexts presumably related to *Yersinia pestis* and to discriminate between true and apocryphal plagues using a paleomicrobiological approach, through the application of state-of-the-art techniques such as NGS and Metagenomics.

This approach is quite expensive and technically demanding, but it can be proficiently applied to investigate the structure of microbial communities not only within the paleomicrobiological context, but it can also be applied to different fields like archaeology, anthropology and forensic medicine [Bazaj et al., 2015].

Unfortunately, no suitable aDNA laboratory where to properly perform the experiments was present at the University of Verona. Therefore, during the two years of my PhD program, I was involved in the design and realization of an aDNA laboratory at the Microbiology Section of the Department of Diagnostics and Public Health at the University of Verona.

In collaboration with the aDNA Laboratory at the Cultural Heritage Department of the University of Bologna, Ravenna-Campus, we had the possibility to work with teeth (most of which presenting decay) samples (Figures 12 and 13) collected from three different burial sites of presumed plague victims in Italy. Specifically, 3 samples related to the 1576 plague outbreak were collected from the Cemetery of Lazzaretto New Island (Venice); 9 samples belonged to the " Necropoli Tardoantica a Forum Semproni" in Fossombrone, which exhibited a peculiar burial method

typical of pandemic events like the plague; finally 13 samples were collected from “Necropoli Romana” in Modena - NOVISAD Park.

As previously stated, the broad goal of this project was to shed a light on infectious diseases of the past to predict the evolution of the potential pandemic events in the future. Such goal could only be reached upon achieving the following milestones:

- Building a dedicate aDNA laboratory in the Microbiology section
- Creation of a clean room to possibly isolate samples from modern DNA
- Setup of an “operator guideline” to avoid any kind of contamination
- Handling and proper manipulation of ancient specimens
- Optimization of the aDNA extraction protocols
- Setup of PCR reaction to maximize amplification of short fragments typical of aDNA (150-200 bp)
- NGS sequencing and Metagenomic analysis to identify the microorganism responsible for the presumed plague victims of the three different burial sites.

Chapter 3

METHODS

3.1 Setting up an “ad hoc” Paleomicrobiology laboratory

A critical requirement for this project was the presence of a suitable lab where to handle samples properly and to perform the analyses, because of the fragility of aDNA and of the high risk of external modern DNA contamination. Since paleomicrobiology specimens are non-reproducible and often of historical interest, it is fundamental to follow carefully each step of the analysis pipeline. The aDNA facility must be isolated from any location where PCR is routinely performed. After numbering, classifying, crashing, drilling, pulverizing (Figure 8) and sampling the

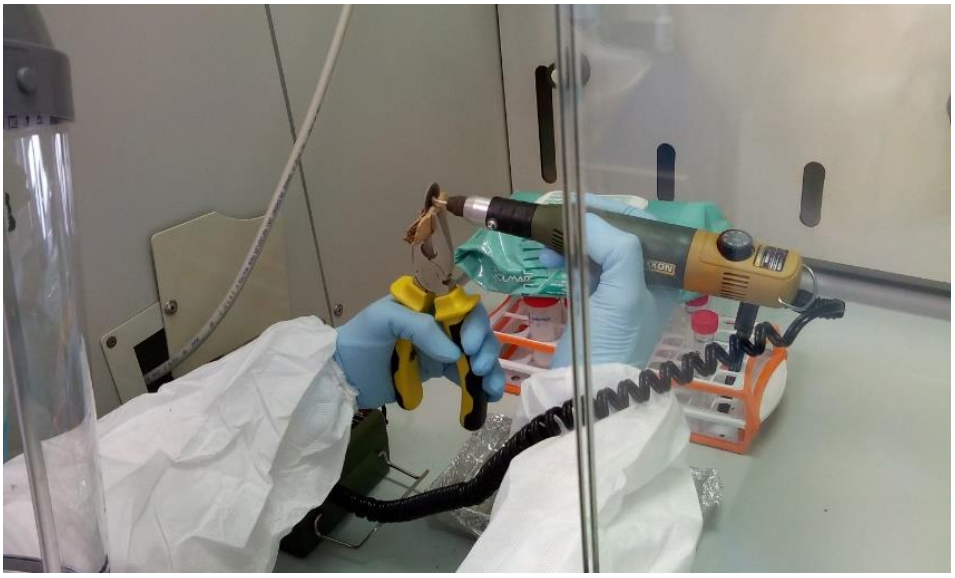


Figure 8. Drilling and pulverization of a rib specimen

specimens, the most important step that was the DNA extraction from the bones, paying the utmost attention to avoid contamination from external DNA.

For this reason, we introduced a negative control sample consisting of Hydroxyapatite (HA), which is a naturally occurring mineral form of calcium apatite ($Ca_5(PO_4)_3(OH)$). Hydroxyapatite is the hydroxyl endmember of the complex apatite group, where the OH^- ion can be replaced by fluoride, chloride or carbonate, producing fluorapatite or chlorapatite. Pure HA appears as a white powder constituted by hexagonal crystals, while its modified form, known as bone mineral, constitutes up to 50% volume and 70% weight of human bones. Moreover, carbonated calcium-deficient HA is the main mineral of which dental enamel and dentin are composed, making pure HA the best compound to use as a negative control. Pulverization of the specimens was performed under a laminar fumed hood under sterile conditions (Figure 8) and it is the step where the HA negative control was introduced. The sample (obtained from rib specimens) was then transferred for DNA extraction room to the core of the laboratory (named “clean lab”) (Figure 9) through a “pass-through cabinet” (Figure 10) fitted with UVC light in order to preserve sterility.

Figure 9. aDNA Extraction lab, called "Clean Laboratory”

Figure 10. Picture of the "pass-through cabinet" to preserve the sterility of the specimen during manipulation steps

In order to assess the suitability of the newly built "ad hoc" paleomicrobiology lab, we replicated the experiments performed in the established aDNA Laboratory at the Cultural Heritage Department of the University of Bologna, Ravenna-Campus. All the amplifications and sequencing reactions were replicated at least twice in each laboratory in order to authenticate the results and carefully check the mutations found. All the steps of the analysis were conducted under strict guidelines for contamination control and detection and reproducibility of data [Cooper & Poinar, 2000; Gilbert et al., 2005; Llamas et al., 2017].

Finally, we were able to reproduce their analysis in our laboratory and to evaluate the authenticity of the results obtained in all the samples included in the project [Cilli et al., 2020 article in press], meaning that our laboratory fulfilled all the requirements for proper aDNA manipulation.

The analysis of aDNA is time-consuming and very expensive, but it can provide a powerful tool for investigating evolutionary processes that cannot be approached using only modern data.

3.2 Biological specimens' collection and manipulation

The study of skeletal remains recovered in a catastrophic death assemblage can provide information about the health status of the population in different archaeological contexts, which may be useful to answer some questions about the plague, a still present disease in several parts of the world. This study, in collaboration with the aDNA Laboratory at the Cultural Heritage Department of the University of Bologna, Ravenna-Campus, was based on teeth samples (most of which with no signs of damage) collected from three different burial site of presumed plague victims in Italy. Three samples related to the Plague of 1576 (Table 1) were collected from the Cemetery of Lazzaretto New Island (Venice).

Figure 11. New Island Lazzaretto Venezia

The analysis of the composition of the samples allowed to reconstruct the diet of 16th century venetian population. Moreover, the ratio between trace elements indicated that the subjects belonged to different social classes, suggesting that the flagellum of *Y. pestis* didn't make any social distinction [M. Borrini, F. Bartoli et al., 2010].

Nine samples came from the "Necropoli Tardoantica a Forum Semproni" in Fossombrone (Figure 12), which exhibited a peculiar burial method.

Figure 12. Teeth samples collected from Fossombrone

Indeed, no individual burial was found at this site, instead all the corpses were buried together in a hasty, superficial and careless manner (see Figure 13), typical of pandemic events like the plague. In addition, based on historical events contemporary to the burial, a lethal epidemic event presumably occurred. Such event was characterized by a fairly large incubation period (up to 12 days for plague) and a few days period between the clinical manifestation and the death of the individual [M. Luni, O. Mei and P. Gobbi 2013].

Figure 13. Detail view of burial site of plague victims

Bodies are tossed from carts into hastily dug pits and covered with a layer of dirt thin enough that animals might dig up body parts [Ranson Riggs 2010].

Thirteen samples were collected from "Necropoli Romana" in Modena- NOVISAD Park (Figure 14). This burial site was found by chance in the city center of Modena during the construction of an underground parking area. The big cemetery was found to be constituted by several sepulchral nucleons of different ages, indeed the most ancient burial sites dated around the 2nd century BC, while the necropolis of the presumably plague victims dated back to the 17th century. The burial architecture was typical of the Christian Romans, the only population that took care of the sepulture of the individuals who died from the plague instead of incinerating them, the other treatment to which infected individuals were subjected. The graveyard consisted of 69 tombs arranged in several parallel rows each housing several bodies, often laid hurriedly one on top of the other, both wrapped in bandages and without

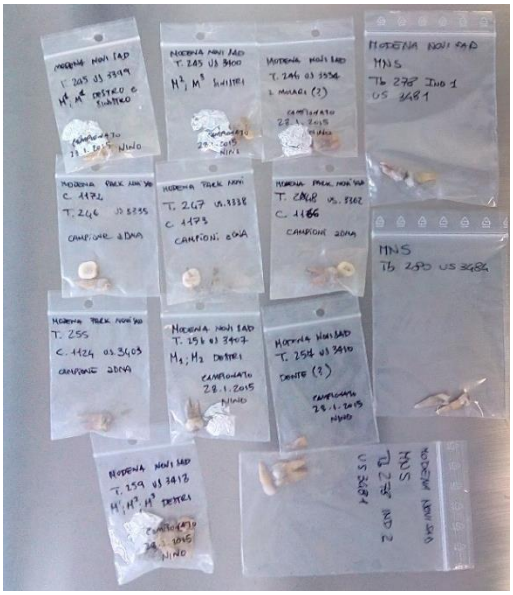


Figure 14. Teeth samples collected from Modena

clothing under a pile of quicklime. Some signs of burns on the bones were also found suggesting sterilization practices even before burial, probably to avoid the diffusion of the epidemic. The remains found are often incomplete and there are no signs of deposition according to religious canons, despite the land being religious property [Labate et al., 2010].

Bones and dental pulp represent the best material for research on microbial pathogens, because their aDNA content better preserved. Before the analysis, each sample was subjected to a decontamination procedure consisting of 45 minutes of UVC irradiation on each side [Haensch et al., 2010]. Each lab operator had to wear a full body suit and a new overall dedicated only to aDNA laboratory before entering, and a second sterile overall on top of the first one every time the operator moved to a different room.

Handling of the samples followed the guidelines for contamination precautions described in the literature [Cooper & Pionar, 2000, Pääbo et al., 2004], consisting of changing gloves every time in between samples, sterilization of the bench with UVC for at least 45 minutes after sample manipulation, transfer of the samples using a pass-through cabinet (Figure 10) between the processing room and the DNA extraction room. Furthermore, all the equipment, benches, reagents, pipettes, scale, plasticware, masks, helmets, were incubated for 45 min under UVC light (250 nm) before and after usage and cleaned with bleach, which causes oxidative damage to DNA, producing chlorinated base products [Fulton et al., 2012].

All three rooms constituting the aDNA lab were irradiated with UVC light from 21 to 6 to ensure a contamination-free environment. At the end of the sample processing, pulverized bones were aliquoted (0.1-0.2 g) and stored at 4 °C until use. Unfortunately, since most of the specimens were collected by external archaeologists, we could not guarantee that the protocols for avoiding contamination were followed during the collection of the samples.

ID	SAMPLE_ID	ITEM_ID	DATA	NOTE	NOTE2	STANDARD	DATA ESTRAZIONE
		Lazzaretto Isola Nuova-Venezia					
LN 1	18082150	LN08 T37	21/08/2018	T: 0,047| P1C: 0,052| P2R:0,099	CAPO S. CIS 1 SETT IV	SI	23/08/2018
LN 2	18082151	LN08 T38	21/08/2018	T:0,051| P1C:0,080| P2C:0,077| P3R: 0,093	SET IV N C/O ID 15	SI	23/08/2018
LN 3	18082152	LN08 T48	21/08/2018	T:0,019| P1C:0,076 |P2R:0,093	CAMPO SANTO SETIV N CIS1	SI	23/08/2018
LN_CN							
		Modena NOVISAD					
NP 1	18070509	NP_TB_245_US 3399	24/07/2018	PTR: 71 mg	M^1; M^2 DESTRO E SINISTRO	SI	24/07/2018
NP 2	180705010	NP_TB_245_US 3400	24/07/2018	P: 36 mg	M^2; M^3 SINISTRI	SI	24/07/2018
NP 3	180705011	NP_TB_246_US 3334	24/07/2018	P: 77 mg | R: 124 mg	2 MOLARI	SI	24/07/2018
NP 4	180705012	NP_TB_246_US 3335	24/07/2018	P: 72 mg	C. 1172	SI	24/07/2018
NP 5	180705013	NP_TB_247 _US 3338	24/07/2018	PC: 30 mg	C. 1173 campioni aDNA	SI	24/07/2018
NP 6	180705014	NP_TB_248_US 3362	24/07/2018	R: 40 mg	C. 1166 campioni aDNA	SI	24/07/2018
NP 7	180705015	NP_TB_255_US 3403	24/07/2018	PR: 48 mg	C. 1124 campione aDNA	SI	24/07/2018
NP 8	180705016	NP_TB_256_US 3407	24/07/2018	R: 78 mg	M^1;M^2 DESTRI	SI	24/07/2018
NP 9	180705017	NP_TB_257_US 3410	24/07/2018	D: 242 mg	DENTE bambino (?)	SI	24/07/2018
NP 10	180705018	NP_TB_259_US 3418	24/07/2018	DR: 10 mg	M^1;M^2; M^3 DESTRI	SI	24/07/2018
NP 11	180705019	NP_TB_278_US 3481_IND 1	24/07/2018	PR: 37 mg	2 MOLARI	SI	24/07/2018
NP 12	180705020	NP_TB_278_US 3481_IND 2	24/07/2018	P: 77 mg	1 MOLARE	SI	24/07/2018
NP 13	18070521	NP_TB_280_US 3484	24/07/2018	P: 46 mg	2 PREMOLARI	SI	24/07/2018
NP_CN							
		Fossombrone					
FS1A; FS1I	18070501	FS_TB 3	23/07/2018	MR:_159mg | PR: 94MG_|	1 MOLARE | 1 PREMOLARE	SI	24/07/2018
FS 2	18070502	FS_TB 31	23/07/2018	R: 90 mg|	1 MOLARE	SI	24/07/2018
FS 3	18070503	FS_TB 34	23/07/2018	R: 189 mg|	1 MOLARE | 1 PREMOLARE	SI	24/07/2018
FS 4	18070504	FS_TB 35B	23/07/2018	R 142 mg	2 PREMOLARI	SI	24/07/2018
FS 5	18070505	FS_TB 114	23/07/2018	P: 103 mg	1 MOLARE | 1 PREMOLARE	SI	24/07/2018
FS 6	18070506	FS_TB121	23/07/2018	P: 127 mg	1 MOLARE | 1 PREMOLARE	SI	24/07/2018
FS 7	18070507	FS_TB 204_A	23/07/2018	P: 77 mg	1 MOLARE | 1 PREMOLARE	SI	24/07/2018
FS 8	18070508	FS_TB 204_B	23/07/2018	P: 116 mg	2 PREMOLARI	SI	24/07/2018
FS 9	18070509	FS_TB 212	23/07/2018	P: 59 mg	1 MOLARE | 1 PREMOLARE	SI	24/07/2018
FS_CN							

Table 1. Dataset of all samples received

3.3 DNA extraction

The most limiting factor in aDNA retrieval and analysis is the percentage of endogenous DNA found.

The content of endogenous aDNA is less than the 1% of the total sequenced reads, even though the rate of endogenous aDNA of different samples from the same individual can differ by orders of magnitude [Green et al. 2010]. In addition, the pathogens to be identified may require different extraction protocols depending on their life cycle and the specific tissue they target. For instance, aDNA extraction of *Y. pestis* or *Mycobacterium leprae* is often achieved from bone tissue, teeth [Schuenemann et al., 2013, Bos et al., 2011] or dental cementum, which exhibits a rate of endogenous DNA similar to that of the petrous bone [Hansen et al., 2017].

The petrous bone is the densest bone in the human skeleton, exhibiting the highest rate of aDNA [Gamba et al., 2014, Pinhasi et al., 2015], although in the study of pathogens from ancient remains its targeting is usually not an option [Margaryan et al., 2018]. On the other hand, the presence of *Mycobacterium tuberculosis* is usually assessed by extracting the DNA from ribs due to the pulmonary tuberculosis [Bowman et al., 2012]. Pathogens that do not infect bones, such as *Vibrio cholera*, can be particularly challenging to identify, as the DNA can be extracted only from soft tissues, which is restricted to rare mummified or dissected specimens [Devault et al., 2014].

DNA extraction protocols are typically divided in two main steps: the first is the solubilization of DNA-bearing tissues and the consequential release of the DNA molecules, the second consists in the purification of such DNA [Heintzman et al., 2015]. The release of the genetic material is achieved through compounds like proteinase K which hydrolyses collagen and Ethylenediaminetetraacetic acid (EDTA) which demineralizes hydroxyapatite [Rohland & Hofreiter, 2007] by chelating Ca^{2+} ions. DNA is then purified and separated from other organic and inorganic molecules by silica-based methods [Rohland & Hofreiter 2007, Dabney

et al., 2013, Allentoft et al., 2015], or alternatively by the less commonly used phenol-chloroform methods [Barnett et al., 2012].

Silica-based methods can be divided between in-solution based and column-based. In the in-solution based method the calcified tissue is digested, the DNA is electrostatically captured by a silica pellet [Rohland & Hofreiter 2007] which is finally washed to allow the release and recovery of the DNA. This method was recently improved by replacing the in-solution silica with silica columns [Dabney et al., 2013], which increased the recovery of extra-short endogenous aDNA fragments (<80 bp) [Gamba et al., 2016] representing the larger fraction of aDNA reads [Orlando et al., 2015]. Aliquots of 0.1-0.2 g were tested with a modified version of 3 aDNA extraction protocols [Rohland & Hofreiter 2007, Dabney et al., 2013, Allentoft et al., 2015]. The adapted protocol (described in detail in the supplementary data) was based on the "*The MinElute PCR Purification Kit"* commercial kit (Qiagen), which provides spin columns, buffers, and collection tubes for silica-membrane-based purification of PCR products (70 - 4000 bp in size), yielding large amounts of highly concentrated DNA in very small volumes (10-50 µl). DNA extraction was entirely performed under laminar hood (Figure 15) previously irradiated with UVC light for at least 30 min.

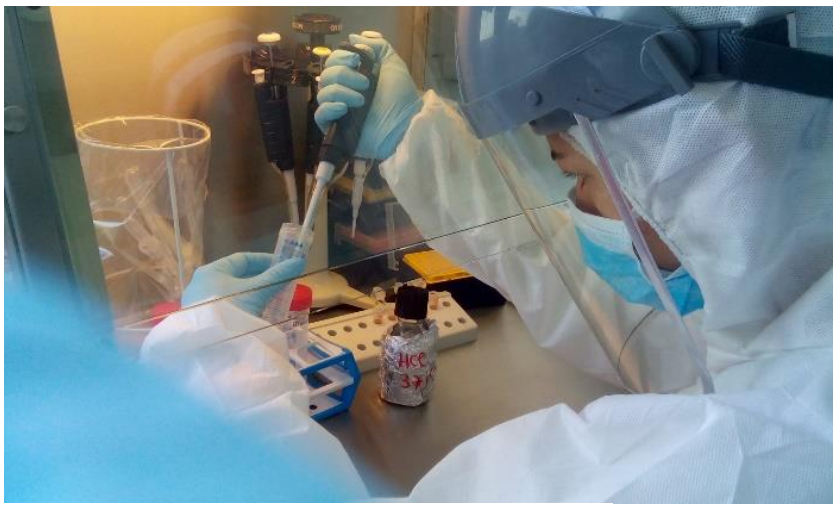


Figure 15. Extraction of aDNA from a rib

Quantification of 1 µL extracted aDNA sample was performed on a NanoDrop™ 2000 UV-Vis spectrophotometer (Thermo Scientific) according to the manufacturer instructions. The instrument is specifically intended for small volumes and measures the absorbance at different wavelengths to evaluate sample concentration (via Lambert-Beer's law) and to estimate the degree of contamination.

Indeed, the absorbance ratio at 260 and 280 nm (A_{260}/A_{280}) determines the presence of proteins in the sample; the optimal value for the study of mDNA is between 2.1 and 1.8 but in the case of aDNA such ratio ranges between 1.4 and 1.8, higher values determine the presence of protein contaminants. On the other hand, the absorbance ratio at 260 and 230 nm (A_{260}/A_{230}) indicates the presence of carbohydrates and phenols, values below 1.8 indicate the presence of contamination.

As the quantification method is absorbance-based, no distinction can be made between modern, human and microbial aDNA, moreover, the high level of fragmentation of aDNA makes its precise quantification particularly complex. Nevertheless, previous studies established this analysis of the fragments as the reference method for aDNA quantification [Brzobohatá et al., 2017].

3.4 Polymerase Chain Reaction (PCR): using "specific primers" to detect the presence of *Y. pestis*

A turning point for forensic science and DNA typing laboratories is represented by the Polymerase Chain Reaction (PCR), a technique described for the first time in 1985 by Kary Mullis, for which he was awarded with Nobel Prize in 1993. PCR has revolutionized molecular biology, as it allowed to generate hundreds of millions of copies of a specific sequence of DNA in only few hours. Specifically, due to the low number of surviving aDNA molecules, no aDNA analysis would be possible without PCR and therefore it would be impossible to reconstruct and have a clear view of the pandemic events in the past.

PCR is an enzymatic process in which a specific region of DNA is replicated over and over again to reproduce many copies of a particular sequence [Bulter J. M. 2012]. The amplification is done with two synthetic oligodeoxynucleotide primers, each about 25 bases long, a thermostable DNA polymerase, and the four deoxyribonucleotide triphosphates. PCR is an ideal tool to amplify a small number of intact aDNA sequences present in a vast excess of damaged molecules [Pääbo et al., 1989]. During enzymatic amplification, most damaged molecules will either not be replicated at all, e.g. due to interior intramolecular cross-links, or will be at a replicative disadvantage with respect to intact molecules, because lesions such as baseless sites, slow down the DNA polymerase. Moreover, the strong inverse correlation between amplification efficiency and the size of the amplification product observed for aDNA, but not for modern DNA, can be employed as another criteria for discriminating authentic aDNA from contaminating exogenous DNA. Another critical issue to assess is the specificity of the primers used to amplify the sequence of interest, as some doubts about the specificity of *pla* gene for *Y. pestis* were raised [Janse et al., 2013].

Here, three PCR analyses using SimpliAmp™ Thermal Cycler, were performed: the first, based on a paper by Hänsch and colleagues [Hänsch et al., 2015], to detect the plasminogen activator/coagulase (*pla)* gene, located on the pPCP1 plasmid, that was assumed to be specific for detecting *Y. pestis*, but was present also in *Citrobacter koseri* and *Escherichia coli.* The second pair of presumably specific primer used for detection of *Y. pestis* [Raoult et al., 2000], were used to perform a "suicide PCR", that is a PCR where the couple of primers is used only for the first cycle. In this PCR reaction there was not a positive control, so the amplicons obtained were sequenced to confirm the presence of *Y. pestis*. The third PCR reaction was performed using specific primers designed at the aDNA Laboratory of Ravenna (Table 2). As previously stated, the samples were independently analysed by two aDNA labs, using reagents manufactured by the same company, in detail: AmpliTaq Gold™ DNA Polymerase with Buffer II and $MgCl_2$ (Thermofisher) consisting in: 5 U/µl Ampli Taq™ Gold Hot Start; 10x Gold Buffer; 25 mM $MgCl_2$; 10 mM dNTP; 50 mg/ml Bovine Serum Albumin (BSA) as an enhancer; for each set of primer: 10 µM Primer Fw 10 µM Primer Rv and DNAse/RNAse free H_2O, with the following cycling conditions:

Pla_1 Fw and Pla_1 Rv [Hänsch et al., 2015]

Activation: 15 min at 95 °C

Amplification (50 cycles):

- Denaturation: 30 s at 94 °C
- Annealing: 30 s at 60 °C
- Elongation: 1 min at 72 °C

Final extension: 10 min at 72 °C

Cooling and storage: 8 °C until analysis

YP12D and YP11R [Didier and Raoult, 2000]

Activation: 5 min at 95 °C

Amplification (50 cycles):

- Denaturation: 45 s at 95 °C
- Annealing: 45 s at 55 °C
- Elongation: 1 min at 72 °C

Final extension: 10 min at 72 °C

Cooling and storage: 8 °C until analysis

pst_Fw and pst_Rv [Ravenna Campus]

Activation: 5 min at 95 °C

Amplification (50 cycles):

- Denaturation: 45 s at 95 °C
- Annealing: 45 s at 60 °C
- Elongation: 1 min at 72 °C

Final extension: 10 min at 72 °C

Cooling and storage: 8 °C until analysis

Name	Seq 5'→3'	Annealing temperature (°C)	Amplicon size (bp)	Ref.	Notes
pla_1 Fw	GACTGGGTTCGGGCACATG	60	70 (33)	Hänsch et al., 2015	qPCR (pla)
pla_1 Rv	CGGATGTCTTCTCACGGA	60	70 (33)	Hänsch et al., 2015	qPCR (pla)
YP12D	CAGCAGGATATCAGGAAACA	55	148 (106)	Raoult et al., 2000	pla gene
YP11R	GCAAGTCCAATATATGGCATAG	55	148 (106)	Raoult et al., 2000	pla gene
pst_Fw	CTGTGGGAGCAGTTCTGGAT	60	73 (33)	Ravenna Campus	
pst_Rv	TTGAGAACCCGTACAGCACT	60	73 (33)	Ravenna Campus	

Table 2. Primers used for PCR analyses

The amplicons were separated on a 1.5% agarose gel to evaluate the presence of real aDNA, providing optimal resolution for small DNA fragments (0.2 – 1 kbp).

Gel electrophoresis is a method for separation and analysis of biological macromolecules (DNA, RNA and proteins), based on their size and charge. DNA is a negatively charged molecule due to the presence of phosphate groups, therefore, when immersed in an electric field, it will move towards the anode. The mobility of the substance on a gel depends on its charge and mass, implying that a larger strand of DNA (potentially mDNA, or contaminant) will travel a shorter distance compared to a smaller fragment, which is what usually aDNA consists of. For a correct identification of the amplicons we used Gene Ruler 50 bp DNA Ladder (Thermofisher) which ensures high mass and size resolution for double-stranded DNA in the 50 to 1000 bp range. Each amplicon was visually quantified according to the DNA standard, finally the 70, 73 and 148 bp fragments were sent for sequencing using traditional Sanger method [Sanger et al., 1975].

The obtained sequences were used as query strings for BLAST (Basic Local Alignment Statistical Tool), a program that compares nucleotide or protein sequences to sequence databases to find regions of local similarity between sequences and calculates the statistical significance of each match. BLAST matches can be used to infer functional and evolutionary relationships between sequences as well as to help identifying members of gene families.

3.5 Next Generation Sequencing (NGS)

Next-generation DNA sequencing (NGS) involves rapid, high-throughput collection of short DNA sequences ranging from 25 to 250 bp. Recently, pathogen bacteria genome sequencing has been used as an epidemiological tool to trace contemporary outbreaks, like the cholera outbreak in Haiti in 2010 [Eppinger et al., 2014] and the Ebola epidemy in west Africa in 2014 [Carroll et al., 2015, Gire et al., 2014]. The majority of pathogen genome sequencing efforts have been focused on contemporary DNA, due to the easier handling and conservation and the higher availability of specimens with respect to aDNA. Indeed, the application of NGS to aDNA remains a big challenge due to the high degree of degradation caused by endonucleases and environmental factors that may result in modified bases or strand breaks [Briggs A. W, Heyn P. 2012]. Thus, obtaining high sequencing depth and accuracy from such samples is often difficult, as the combination of the presence of uracil (caused by cytosine deamination) and the low copy of number of aDNA can potentially lead to miscoding errors.

Despite all these drawbacks, NGS represented a promising technique for unravelling aDNA due to the similarity between the length of the fragments required for NGS and that available from aDNA samples, usually too short for traditional sequencing techniques. Indeed, recent studies showed successful applications of NGS to aDNA belonging to the Neanderthal mitochondrial genome [Green et al., 2008] and the extinct woolly mammoth [Miller et al., 2008, Poinar et al., 2006].

It is still unclear whether NGS techniques may be applied to traditional forensic DNA as proficiently as to aDNA analysis [Blow et al., 2008]. Currently, NGS techniques cannot accurately identify repetitive sequences and thus, unless future improvements are made, they cannot reliably deal with the short tandem repeat regions (STR) which forensic DNA analysis is based on [Hert et al., 2008]. Moreover, the amount of data produced by NGS approaches, consisting of millions of short reads, makes bioinformatics support crucial for forensic DNA laboratories, which might also require to switch the analysis of genetic markers from STRs to single nucleotide polymorphisms (SNPs) to fully exploit the potentiality of NGS.

3.6 Metagenomics

Metagenomics is a branch of genomics that simultaneously studies a complex community of microorganisms present in a sample, avoiding the growth on selective media. Indeed, growth on culture media it allows to identify only 1-3% of the microorganisms actually present in natural samples, losing 97-99% of the information [Gordon, 2012; Hugenholtz et al., 1998], because of their particular growth conditions, such as specific nutrients and anaerobic conditions.

The drawbacks of the classical microbiology techniques to identify microorganisms can be overcome by metagenomics approaches, consisting of extracting genomic DNA from the samples. Followed by sequencing of the 16S rRNA. In such way it is possible not only to identify simultaneously every single microorganism belonging to the community, but also to analyze their interaction with each other, with the environment (microbial ecology) or with the hosting organism.

As microbial communities are involved in a variety of complex biological processes, metagenomics can be fruitfully applied to numerous fields in order to unravel their specific function within each context.

A great contribution was provided by Carl Woese, who in 1967 separated the Archea and the Bacteria dominia, using molecular phylogeny techniques applied to ribosomal RNA 16S [Woese et al., 1990, 1978 and 1977]. The molecular analysis of the gene sequence that encodes the minor ribosomal subunit (16S) is still today considered the most relevant sequence for the classification of Bacteria and Archea.

3.7 16S rRNA Metagenomics and aDNA

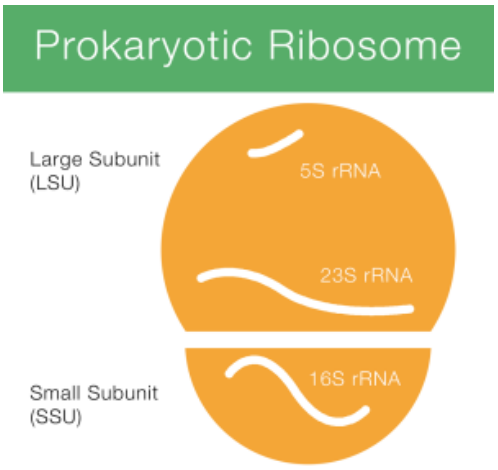


Figure 16. 16S rRNA subunit of prokaryotic

16S rRNA stands for 16S Ribosomal Ribonucleic Acid (rRNA) (Figure 16), where S (Svedberg) is a unit of measurement for the sedimentation rate.

The 16S rRNA encodes the small subunit of prokaryotic ribosomal RNA and contains nine hypervariable regions (V1-V9) separated by ten highly conserved regions (Figure 17).

The hypervariable regions characterize the diversity between species and allow bacterial taxonomy studies. Several studies showed that each of the 9 hypervariable regions provide specific information for bacterial classification [Petrosino et al., 2009].

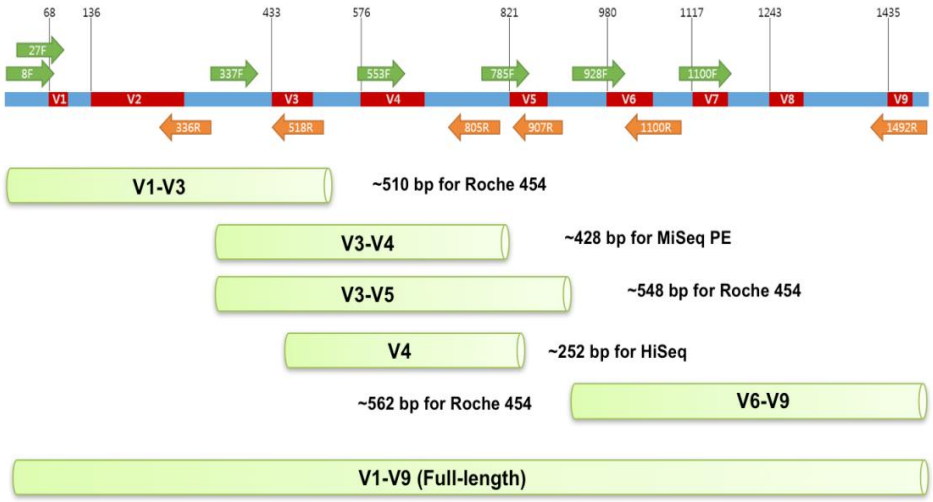


Figure 17. Full length of 16S rRNA gene. Adapted from https://help.ezbiocloud.net/wp/wp-content/uploads/2017/05/16s_var_pcr.png

The 16S rRNA gene is the most used gene in microbial metataxonomic analysis because it is conserved across members of the paraphyletic prokaryotic domains Bacteria and Archaea [Ziesemer et al., 2015], therefore allowing the design of "universal" primers for microbial PCR amplification, yet also sufficiently variable to allow the classification at an approximate species level [Roh et al., 2010]. The full-length 16S gene is usually amplified using 27F and 1492R primers, although multiple primers on both strands are required for accurate DNA sequencing.

Archaea-specific primer sequences typically lack in specificity. The V3 region (primers F333 5'- TCCTACGGGAGGCAGCAG-3' and U592R 5'-ACCGCGGCKGCTGGC-3'), was used for first time for the preparation of 16S libraries from thermophile communities [Baker et al., 2004]. Such libraries suggested that the V3 region was an excellent candidate for aDNA amplification for two main reasons: i) it is the shortest 16S rRNA region (~100 bp shorter than V4 region), ii) it exhibits high sequence heterogeneity, resulting in a good taxonomic variability [Ziesemer et al., 2015].

Region V4 (universal primers 515F 5'-GTGYCAGCMGCCGCGGTAA-3' and 806R28 5'-GGACTACNVGGGTWTCTAAT-3') was considered not a suitable choice for aDNA studies, as the highly fragmented aDNA molecules, rarely exceeding 200 bp in length, were shorter than the entire V4 sequence (~292 bp including primers).

To guarantee the exclusive analysis of aDNA in our samples we focused also on the V5 region. Indeed, the V5 hypervariable region (primers F784 5'-AGGATTAGATACCCTGGTA-3' and R934 5'-TGTGCGGGCCCCCGTCAATT-3') performs well on a number of metrics: it has very good predicted taxonomic coverage and is relatively short (144–148 bp) with little amplicon length variation. The sequence encompassing V3 and V5 regions (primers F333 and R934) was used for the construction of the libraries of bacterial 16S rRNA gene [Weil et al., 2017].

In addition, since the V3-V5 amplicon would be much longer than the expected sequences for aDNA, the two regions were amplified individually, as the combined analysis of such regions was found to be a suitable discriminant for microbial

research. We analysed eight different samples from the three burial sites presumably associated with the presence of *Y. pestis* (see figures in supplementary information), and two additional samples, from a different bio-archaeological context (Forlì) in which bisome burials were found, were considered as negative control because they cannot be attributed to pestilential periods (see Material and Methods). PCR reactions were performed with the same reagents as previously described for conventional PCR, with the following thermal cycling profiles:

Amplification for V3 region of 16S gene with F333/U529R primer [Baker et al., 2004]

Activation: 10 min at 95 °C
Amplification (25-35 cycles):

- Denaturation: 45 s at 94 °C
- Annealing: 1 min at 56 °C
- Elongation: 1 min at 72 °C

Final extension: 10 min at 72 °C
Cooling and storage: 4 °C until analysis

Amplification for V5 region of 16S gene with F784/R934 primer

Activation: 10 min at 94 °C
Amplification (35cycles):

- Denaturation: 45 s at 94 °C
- Annealing: 1 min at 55 °C
- Elongation: 1 min at 72 °C

Final extension: 10 min at 72 °C
Cooling and storage: 4 °C until analysis

3.8 Preparation of NGS Libraries from aDNA

Library preparation consists of all the chemical reactions and procedures that modify DNA fragments to meet the experimental requirements for NGS sequencing. In detail, the DNA fragments to be sequenced are end-repaired and ligated with universal sequencing-adaptors.

Adaptor-ligated libraries are convenient for aDNA studies because they can help overcoming the main inconvenience of NGS which is the short reads length. Moreover, adaptor primers outside the aDNA sequence allow the recovery of information from molecules too short for traditional PCR. Finally, universal adaptors provide higher amplification of the entire library before any downstream experiments. Even though NGS presents advantages for aDNA studies, post-mortem aDNA damage, consisting of strand-breaks and base modification (see aDNA degradation), still represent a challenge to face.

There are no standard protocols for the preparation and NGS sequencing of aDNA libraries, they are usually tailored on the specific requirements of each project, but they follow the same pipeline. The first step usually consists of DNA fragmentation, but it can be considered unnecessary for aDNA due to its intrinsic fragmentation. Then, the ends of aDNA fragments have to be repaired and ligated with short sequences (adapters) which will be eventually recognized by NGS platforms (Figure 18). The reparation reaction basically consists in the degradation of the overhanging 3' ends and the filling of 5' overhanging ends. Unfortunately, fragment ends are frequently affected by one of the most common miscoding lesions in aDNA molecules, namely cytosine deamination to uracil (resulting in C to T transition). The elimination of deamination products on one hand reduces the number of mismatches and the genotyping errors, but on the other hand prevents the recognition of specific patterns used to validate the presence of aDNA.

The following step is common for both treated and non-treated reads and represented by the adapter ligation. This process can be accomplished either using

two different adapters (blunt-end ligation) to the read ends, or a single Y-shaped adapter with a T-overhang to both ends of DNA. However, the Y-shaped adapter requires an additional pre-processing step consisting of the addition of A-overhangs through the so-called A-tailed ligation [Willmann et al., 2018], and may lead to the misincorporation of T in the read ends [Seguin-Orlando 2013]. It is also recommended to include a negative (blank) library control during library preparation to assess the quality of the library.

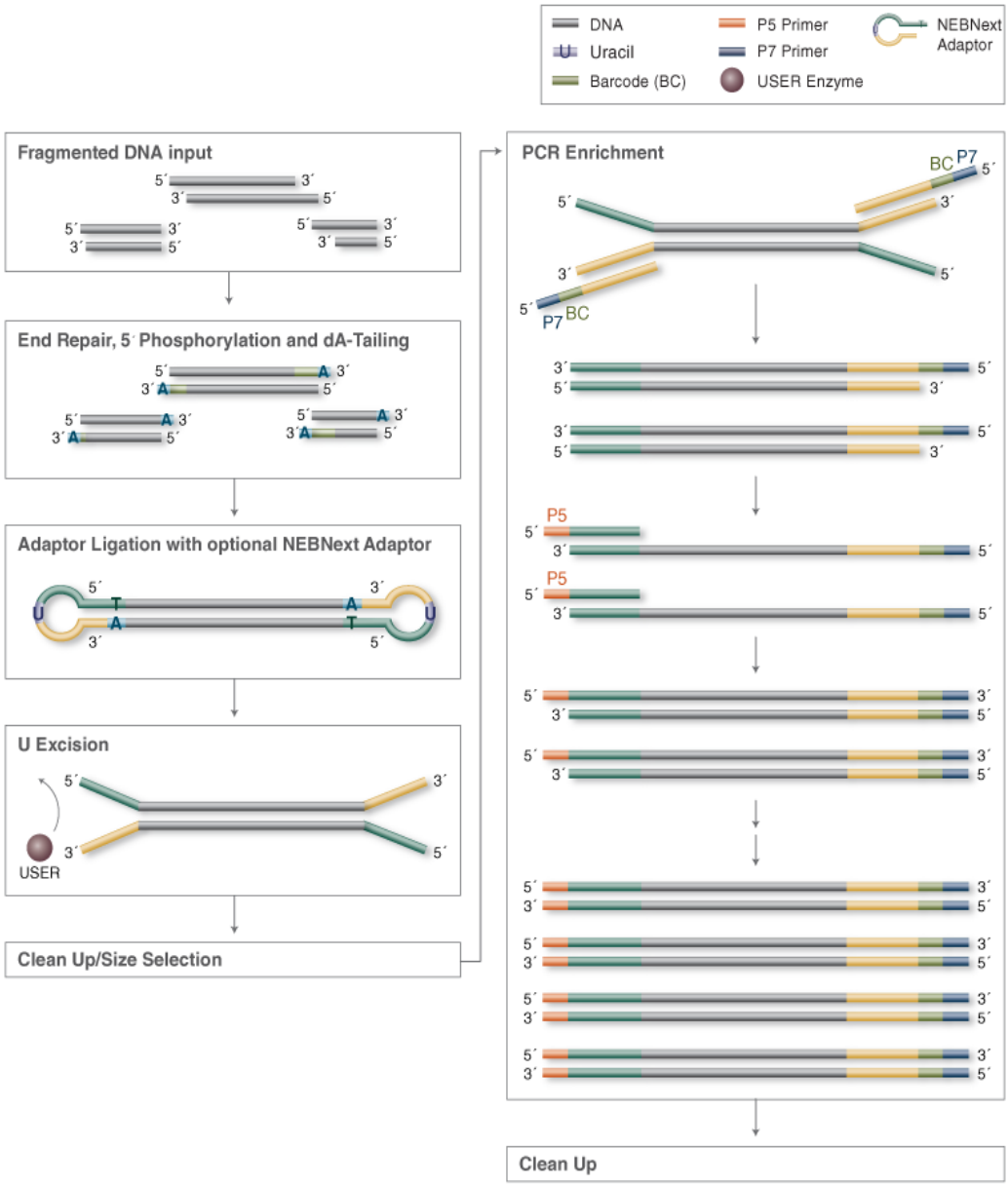


Figure 18. Library preparation method for double strand libraries. *The DNA is fragmented and the reads ends are repaired by adding A bases, then the adapters are ligated to the repaired ends. Finally reads are amplified with PCR technique (Ilumina).*

Finally, the DNA sequences obtained are amplified through a limited number of PCR cycles, in order to preserve the variability of the library, moreover, the choice of an adequate PCR polymerase is also crucial to avoid GC and read length biases [Dabney and Meyer 2012]. To differentiate several libraries sequenced in the same run, barcodes are attached to the adapters during the amplification process [Craig et al., 2008, Knapp et al., 2012].

The amplicon obtained from the two PCR reactions were sent to the Paleogenetic laboratory at University of Firenze for the library preparation and DNA sequencing [Modi et al., 2017].

3.9 Microbial taxonomic profiling

Bioinformatic analyses and microbial taxonomic profiling were conducted using MALT (MEGAN [Huson et al., 2007] ALignment Tool) software, by aligning the sequences obtained from teeth samples against SILVA database (https://www.arb-silva.de/).

3.10 Phylogenetic trees

Phylogenetic trees are one of the most common representations of the biological diversity, displaying the relations of biological samples or species using branches, nodes and taxa. The branching pattern of a tree is define its topology, while the group of the descendants of a node defines a clade. Clades composed by all the descendants of a common ancestor are called monophyletic, otherwise they are named paraphyletic.

Phylogenetic trees can be built based on genetic or morphological diversity. One of the most used criteria to build a genetic phylogenetic tree is called Maximum Likelihood (ML), which is a statistical method for estimating unknown parameters in a probability model. In terms of phylogeny, the likelihood represents the probability of an observed sequence on a particular tree assuming a specific evolutionary model (see figure 19). If the ancestor of all the taxa present in the phylogeny is unknown, the phylogenetic tree is considered unrooted, otherwise it is called rooted and the branches length can be interpreted as time estimates. Moreover, if an outgroup is added, it is possible to identify the ancestral node of the phylogeny and root it.

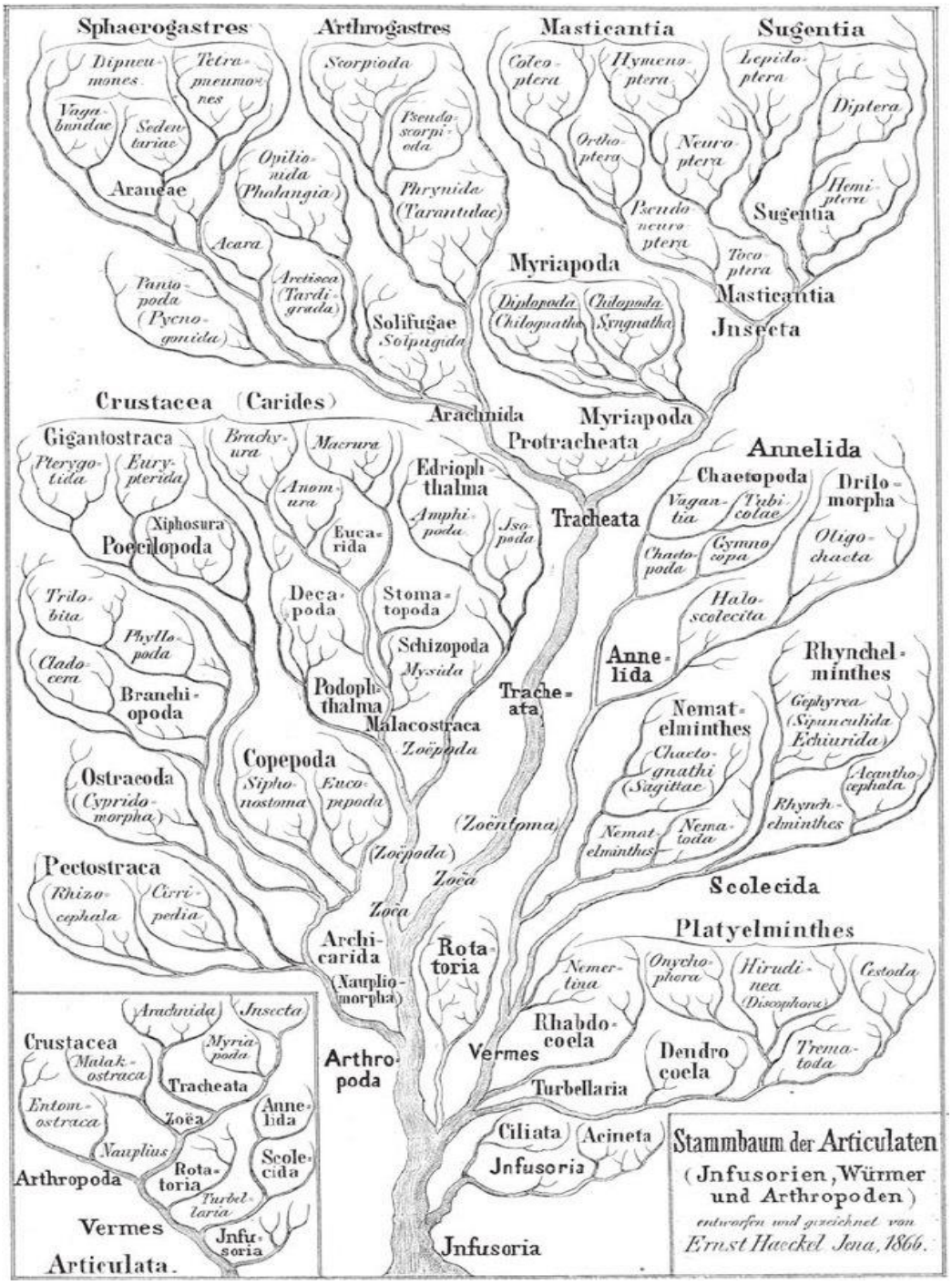


Figure 19. Phylogenetic Tree. *Adapted from Hackel 1866*

Chapter 4

RESULTS

4.1 aDNA extraction and purity

Three different bioarchaeological context from Italy (Venice, Fossombrone, Modena) were taken under analyses for research of *Y. pestis*, the bacterium causing the plague. The specimens were treated with extreme caution, as detailed in the methods section, to prevent any cross-contamination with modern DNA. Extraction and purification of aDNA was performed with MinElute PCR purification Kit (Qiagen GmbH, Hilden, Germany), for each extraction reaction, a negative control based on hydroxyapatite powder was introduced to detect possible DNA contamination from exogenous sources during both pulverization and extraction procedures.

Quantification results showed an overestimation of the amount of DNA present in all samples, with a measured concentration range from 12.7 to 43.4 ng/µL, with an average A_{260}/A_{230} ratio ranging from 0.79 to 3.03 and an average A_{260}/A_{280} ratio range between 1.38 and 1.55. Overall, our results suggest that the samples do not exhibit a high degree of purity, most probably due to the internal degradation of aDNA and the exposition to environmental factors. Nevertheless, such values can still be considered satisfactory, as they are comparable with previously reported data on aDNA samples [Brzobohatá et al., 2017].

4.2 PCR and Sanger Sequencing

All the aDNA samples were subjected to conventional PCR reactions using three sets of supposedly "specific primers" to assess primer specificity (Figure Sx). The first set was used detect the plasminogen activator/coagulase (*pla)* gene (pla_1_FW 5'-GACTGGGTTCGGGCACATG-3', pla_1_RV 5'-CGGATGTCTTCTCACGGA-3') located on the pPCP1 plasmid, that was assumed to be specific for detecting *Y. pestis* [Hänsch et al., 2015], but was found in *Citrobacter koseri* and *Escherichia coli.* The second set of primers was employed for a "suicide PCR" (YP12D 5'-CAGCAGGATATCAGGAAACA-3', YP11R 5'-GCAAGTCCAATATATGGCATAG-3'), where the primers were used only for the first cycle [Raoult et al., 2000].

The third couple of primers (pst_FW 5'-CTGTGGGAGCAGTTCTGGAT, pst_RV 5'- TTGAGAACCCGTACAGCACT-3') was specifically designed at the aDNA Laboratory of Ravenna for this research. The amplification products were purified with the MinElute PCR purification Kit (Qiagen GmbH, Hilden, Germany), and sent for Sanger sequencing on both strands.

The obtained sequences were used as query strings for BLAST (Basic Local Alignment Statistical Tool) against all non-redundant nucleotide sequences.

Unfortunately, due to shortness of the query sequences, ascribable to the high degree of fragmentation of aDNA, the detection of *Y. pestis* with traditional PCR amplification was not successful. Indeed, most of the sequenced samples consisted only of short fragments, not informative enough to retrieve reliable information from the databases available online. Indeed, the most frequently identified species from BLAST analysis were environmental bacteria such as *Streptomyces bingchenggensis* (strain BCW-1) and *Rhodococcus hoagii* (strain 103S, also called *Rhodococcus equi*).

S. bingchenggensis (strain BCW-1) belongs to the Streptomyces genus, consisting of soil and water Gram positive filamentous bacteria, well known for their ability to produce complex secondary metabolites including many antibiotics. It was isolated in Harbin, China (China General Microbiology Culture Collection Center CGMCC1734) and has one of the largest sequenced bacterial genomes at almost 12 Mb.

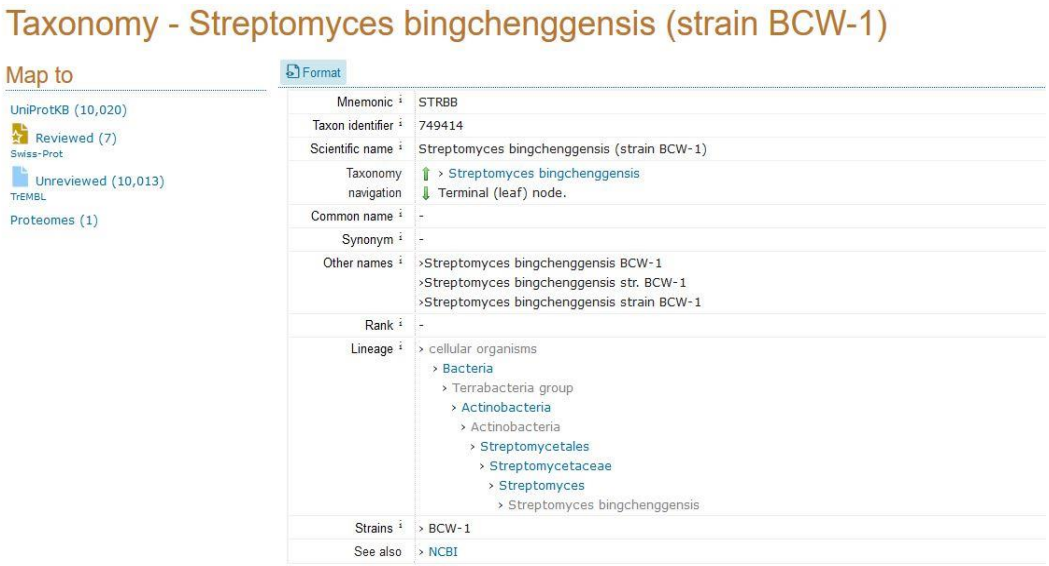


Figure 20. Streptomyces bingchenggensis (strain BCW-1) taxonomy adapted from https://www.uniprot.org/taxonomy/749414

R. hoagii (strain 103S, or *R. equi*) belongs to the Rhodococci genus, consisting of aerobic, Gram positive actinomycetes characterized by a high G/C content and by an environmental-dependent morphological differentiation (e.g., cocci or filaments). Moreover, Rhodococci display long-term survival in soil, an exceptional tolerance for high levels of heavy metals and a metabolic propensity towards hydrophobic pollutants even in the presence of more readily assimilable carbon sources, which make Rhodococci particularly suitable for bioremediation applications [Kämpfer et al., 2014]. *R. equi* is an important pathogen causing pneumonia in foals, wild boars, domestic pigs and immunocompromised humans such as HIV-AIDS patients or transplant recipients, whose infections symptoms resemble clinical and pathological signs of pulmonary tuberculosis.

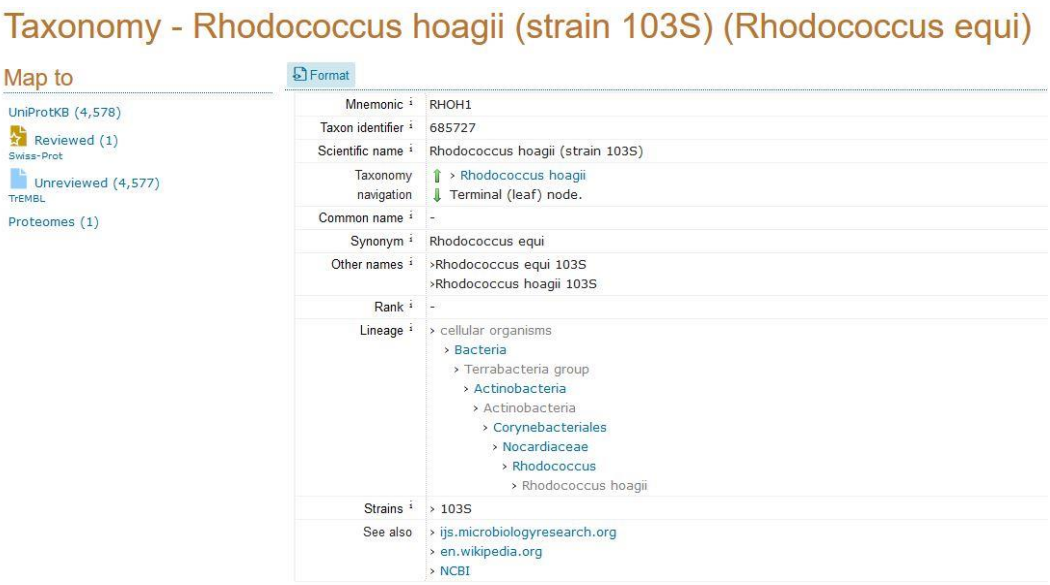


Figure 21. Rhodococcus hoagii (strain 103S) taxonomy adapted from https://www.uniprot.org/taxonomy/685727

4.3 Metagenomics

As conventional PCR allowed the identification of only environmental bacteria, a large number of questions still remained unanswered, such as which bacteria/pathogens were present during particular historical periods and if the diversity of commensal microorganisms was affected by modern diet, lifestyle and especially environmental and climatic changes over time. We therefore analyzed taxonomic profiles generated by amplicon sequencing in temporally and geographically diverse archaeological teeth specimens.

We analyzed aDNA samples from 8 different subjects belonging to 3 burial sites: 2 from Modena (NP), 3 from Venice (LN), 3 from Fossombrone (FS). Two samples collected in a burial site in Forlì (FODI) not dated back to a plague outbreak were added to the analysis and considered as the “negative control” for *Y. pestis* and the “positive control” for environmental/soil bacteria. The authenticity of aDNA was confirmed by the presence of typical molecular signatures of post-mortem DNA damage, such as fragmentation patterns consistent with depurination and mis-incorporation patterns compatible with cytosine deamination within overhangs.

The relative phylum abundance analysis of both V3 and V5 regions highlighted a considerable presence of Proteobacteria (Figures 22 and 23), a major phylum of Gram-negative bacteria including a wide variety of pathogenic genera, such as Escherichia, Salmonella, Vibrio, Helicobacter, Legionellales, Yersinia and many others. Due to its large variety of genera, the Proteobacteria phylum was named after Proteus, a Greek god of the sea capable of assuming many different shapes, while its classification, informally called the "purple bacteria and their relatives", was established in 1987 [Woese, 1987]. Most of the other identified phyla, such as Acidobacteria, Actinobacteria, Chloroflexi and others, include soil bacteria commonly found in buried specimens.

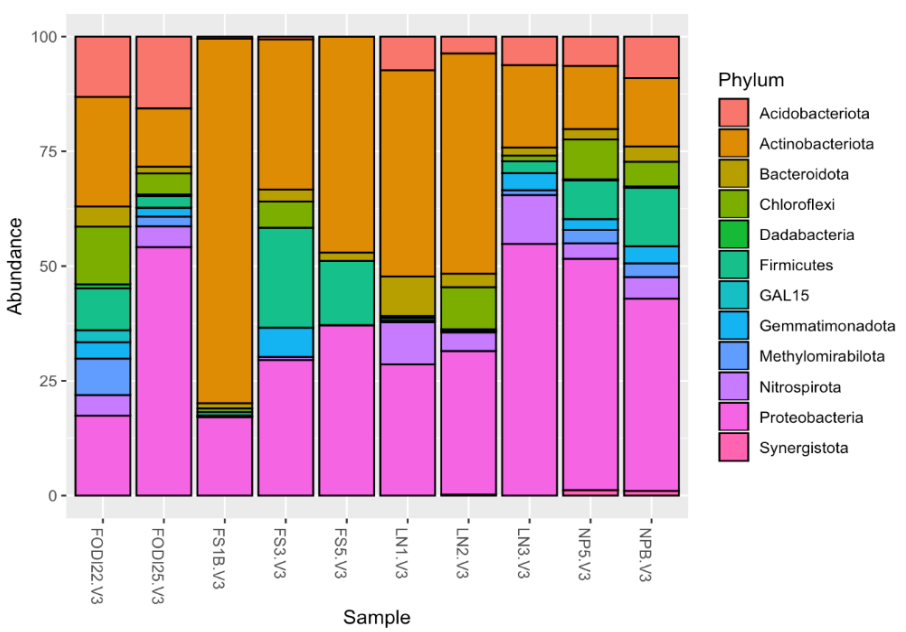


Figure 22. Relative phylum abundance according to V3 region

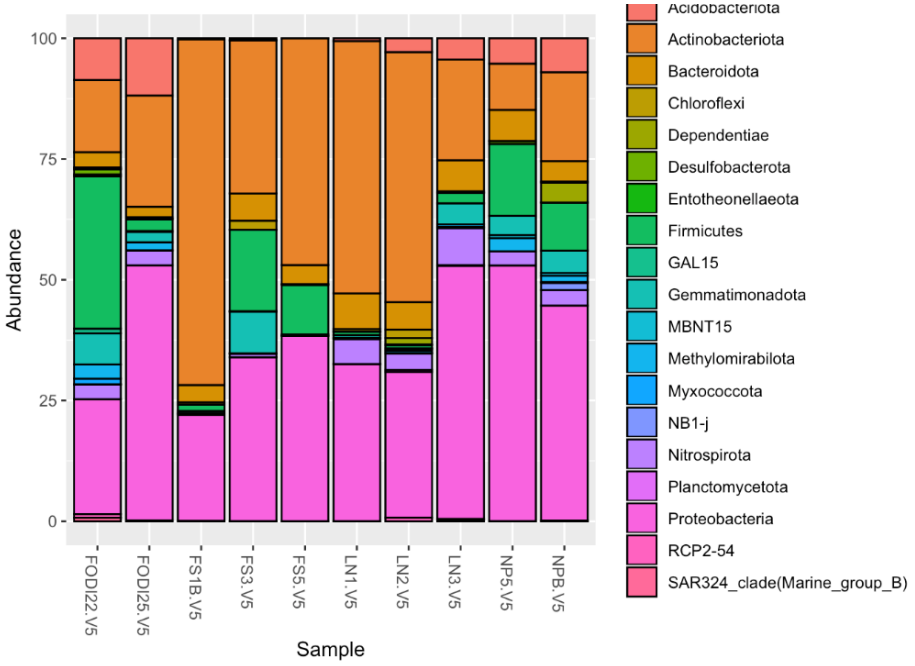


Figure 23. Relative phylum abundance according to V5 region

4.3.1 Paleomicrobiome analysis on V3 region

In 16S metagenomics approaches, OTU (Operational Taxonomic Unit) are cluster of similar sequence variants of the 16S rDNA marker gene sequence. Each of these cluster represents a taxonomic unit of a bacteria species or genus, depending on the sequence similarity threshold. Typically, OTU clusters are defined by a 97% identity threshold of the 16S gene sequences to distinguish bacteria at the genus level. Here, after deep sequencing of 16S rRNA gene V3 we described for each specimen we the microbiome profile in terms of the different OTUs identified at the Phylum, Class, Family and Genus taxonomic levels.

The analysis of relative abundances (Figures 22 and 23) revealed that, at phylum level, the most representative OTUs belong to the Actinobacteria (17.3 %), Proteobacteria (16.5 %) and Chloroflexi (13.1 %), followed by Acidobacteriota (9.7 %), Bacteroidota (5.8 %), Firmicutes (5 %), Gemmatimonadota (4 %) and Patescibacteria (3.1 %). Cyanobacteria, Dadabacteria, Deinococcota, Dependentiae, Desulfobacterota, Elusimicrobiota, Entotheonellaeota, Euryarchaeota, Fibrobacterota, Verrucomicrobiota, Nitrospira, Dependentiae, Elusimicrobiota, Spirochaetota and other non-classifiable were found in very lower frequency (0.1%-2.7%).

A closer inspection of the prevalence at class level (Figure 24) highlighted that the most abundant classes were Actinobacteria, Alphaproteobacteria, Bacilli, Acidobacteriae, together with Clostridia and Gammaproteobacteria, particularly interesting as they include several potential pathogens.

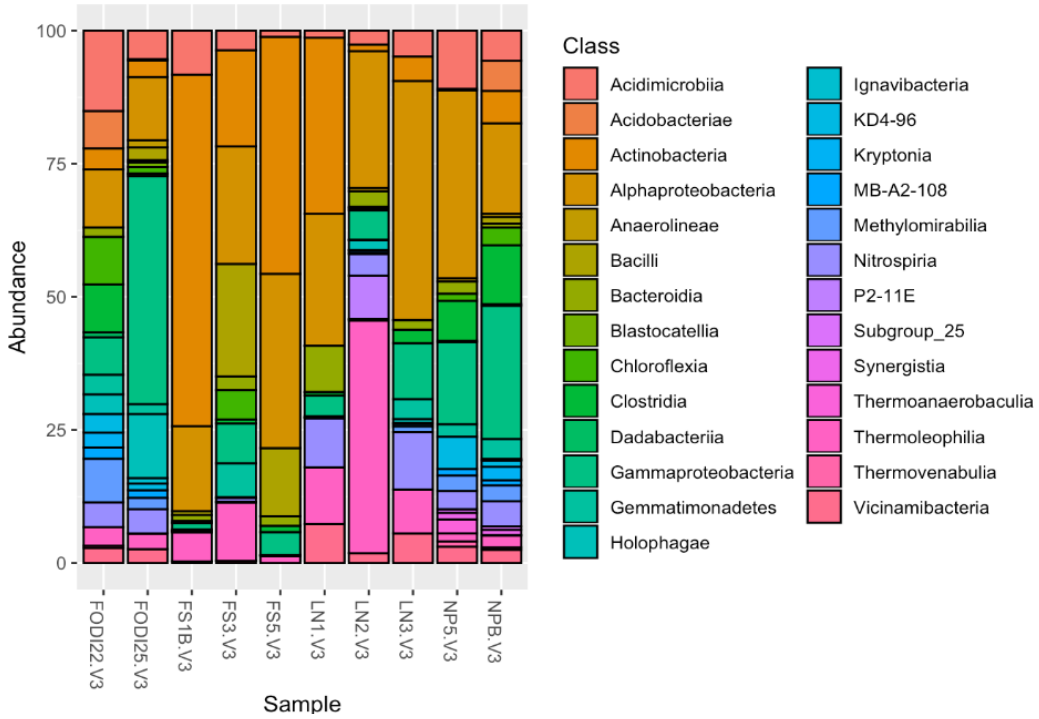


Figure 24. Relative class abundance according to V3 region

Indeed, Clostridia is a class of ubiquitous strictly anaerobic to aerotolerant spore-forming bacilli, usually found in soil as well as in normal intestinal flora of humans and animals. This class includes both Gram-positive and Gram-negative species, although the majority of isolates are Gram-positive. Clostridial wound infections are typically polymicrobic, where the clostridial species, including *C. perfringens*, *C. novyi*, *C. septicum*, and others [Wells and Wilkins 1996], represent the primary pathogens.

Gammaproteobacteria is the class of *Vibrionales*, inhabitants of fresh or saltwater, which include several pathogenic species such as *Vibrio cholerae*, the agent responsible for cholera. Moreover, most bioluminescent bacteria belong to this family, typically found as symbionts of deep-sea animals [Devault et al., 2014].

Gammaproteobacteria also the class of *Legionellales*, consisting of the *Legionella* and *Coxiella* families, both of which include notable pathogens. Indeed, legionellosis (acute pneumonia), is usually caused by *Legionella pneumophila*, although potentially any *Legionella* species may be responsible for such disease. Less often, legionellosis presents as a non-pneumonic, epidemic, influenza-like

illness called Pontiac fever, while extrapulmonary *Legionella* infections (e.g., pericarditis and endocarditis) are rare. Legionella was first recovered from the blood of a soldier more than 50 years ago, but its importance as a human pathogen was not recognized until 1976, when a mysterious epidemic of pneumonia struck members of the Pennsylvania American Legion [Winn WC Jr. 1996].

Finally, Gammaproteobacteria includes also the class of *Yersina pestis* (formerly *Pasteurella pestis*), the cocco-bacillus responsible for plague outbreaks, which was isolated and described by Yersin in 1894 [Yersin and Treille. 1894, Collins. 1996]. These results encouraged us to proceed with a deeper examination in order to classify also family and genus.

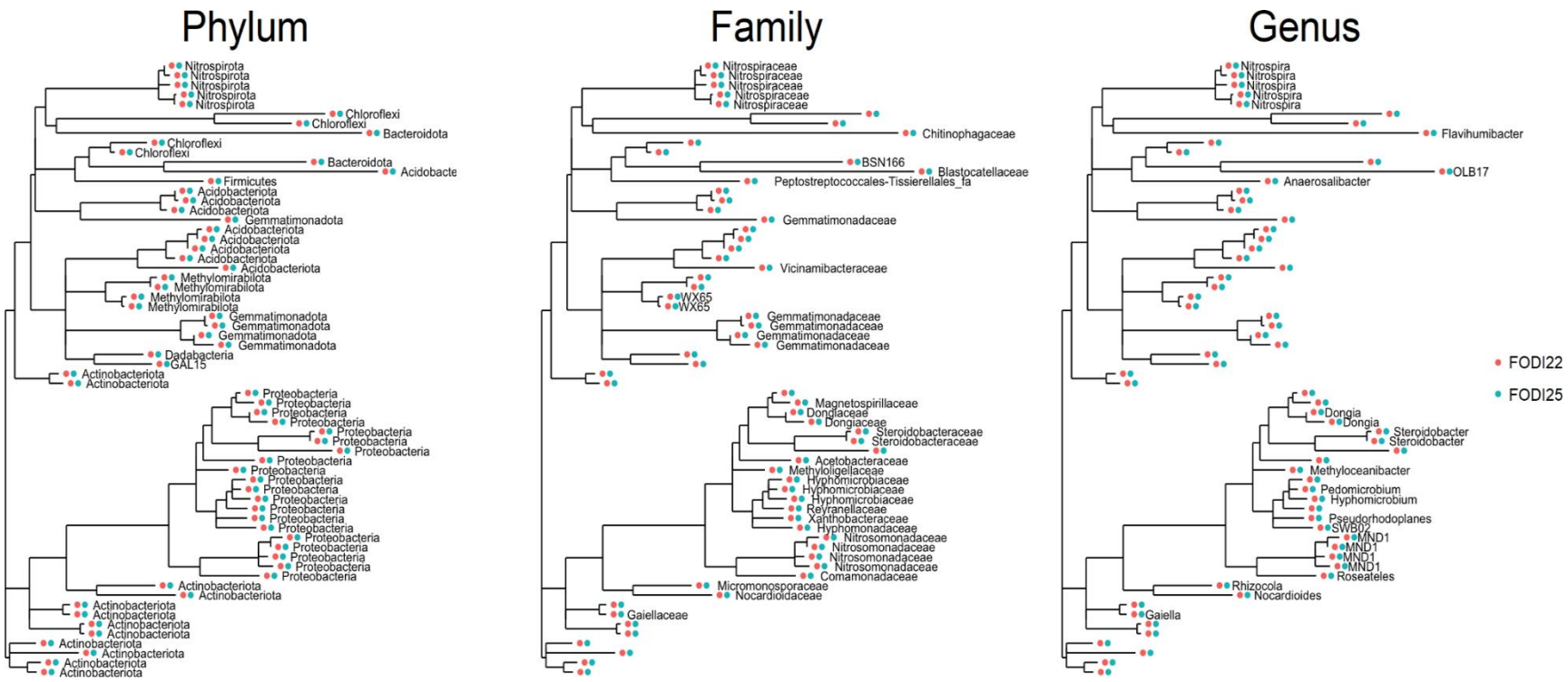


Figure 25. Phylogenetic tree based on V3 region of Forlì (FODI) cemetery at phylum, family and genus levels

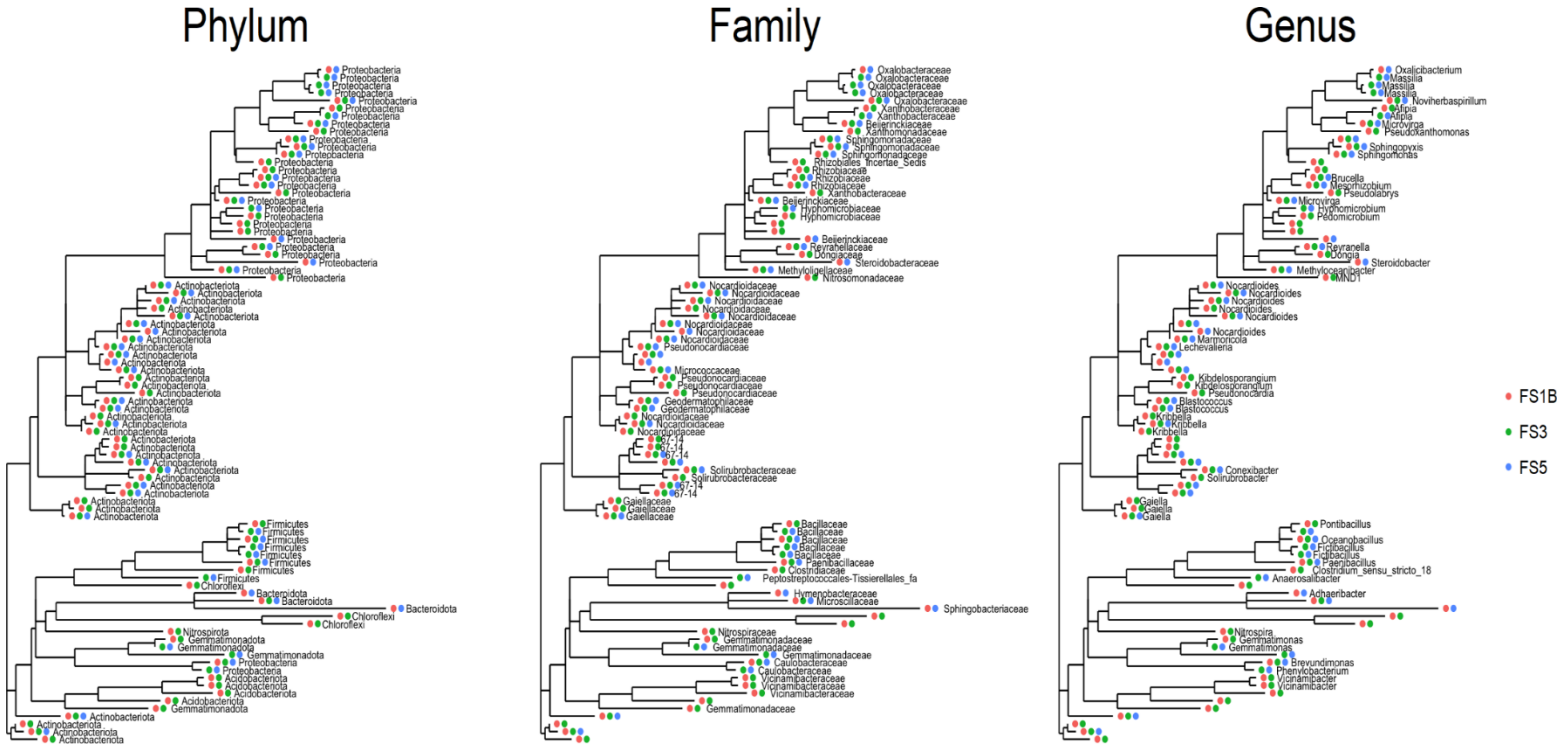


Figure 26. Phylogenetic tree based on V3 region of Fossombrone (FS) burial site at phylum, family and genus levels

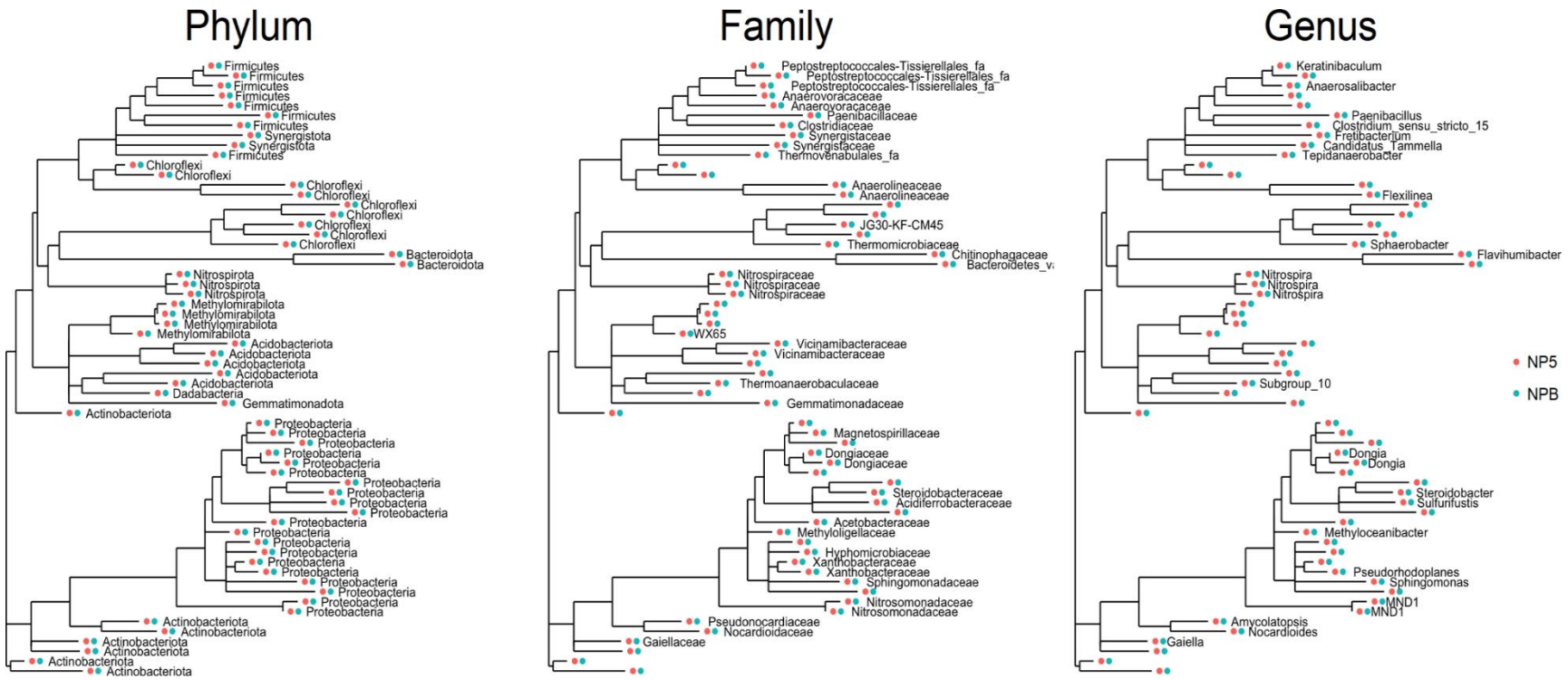


Figure 27. Phylogenetic tree based on V3 region of Modena cemetery (NP) at phylum, family and genus levels

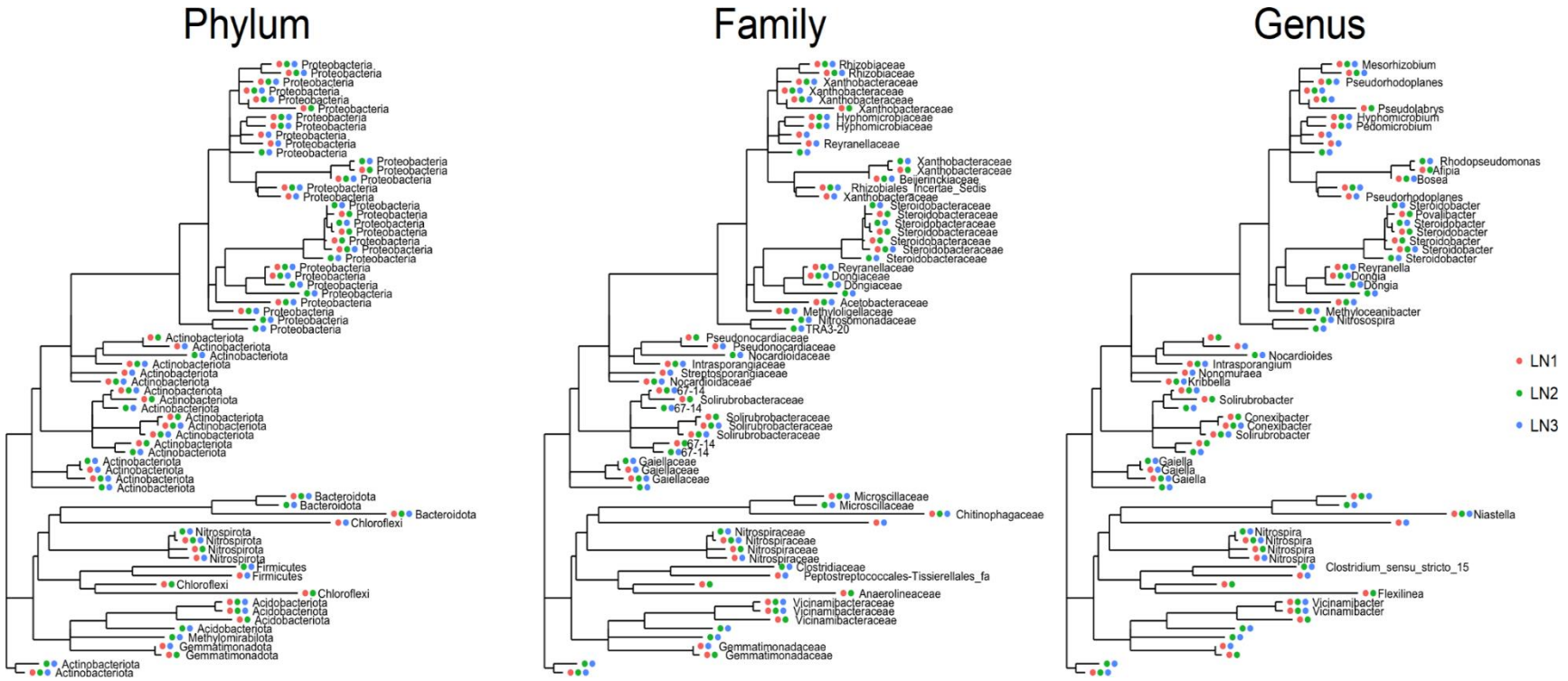


Figure 28. Phylogenetic tree based on V3 region of New Lazarette in Venice (LN) at phylum, family and genus levels

We reconstituted the phylogenetic tree for each burial site independently, focusing on potential pathogenic bacteria that could represent the cause or co-cause of death.

As previously described, the cemetery in Forlì (FODI, Figure 25) was considered as a negative control for *Y. pestis* identification but a positive control for aDNA analyses and soil bacteria investigation. Results showed that the most abundant phyla present in FODI samples were *Proteobacteriota* (31.3%), *Actinobacteriota* (18.8%), *Acidobacteriota* (14.1%), *Nitrospirota* (7.8%), *Gemmatimonadota* (7.8%), *Chloroflexi* (6.3%) and *Methylomirabilota* (6.3%), all normally found in soil and water. Unfortunately, due to the degradation and fragmentation of aDNA, most of the samples family and genus could not be identified (Figure 25). Nonetheless, the most prevalent family of the Nitrospiriota phylum was that of the *Nitrospiraceae* (13.5%), whereas the most frequent families belonging to Proteobacteria phylum were *Gemmatimonadaceae* (13.5%) and *Nitrosomonadaceae* (10.8%), while other families like *Hyphomicrobiaceae*, *Xanthobacteriaceae*, *Steroidobacteriaceae*, *Dongiaceae* displayed a very low frequency. Finally, also at the genus level, the most prevalent genera identified were *Nitrospira*, *Dongia*, *Steroidobacter* and *Hyphomicrobium*, all microorganisms normally present in soil and water.

The results for the samples belonging to Fossombrone (FS) burial site (Figure 26) show more homogenic phyla, with the most prevalent being Actinobacteriota (40.9%), Proteobacteria (34.1%), and Firmicutes (9.1%), followed by Acidobacteriota, Gemmatimonadota, Chloroflexi and Bacteroidota, all showing a much lower frequency (3.4% to 4.5%).

Among the Actinobacteriota the *Nocardiodaceae*, *Micrococcaceae*, *Pseudonocardaceae* families were identified, and in particular the *Kribbella* [Urzì et al., 2008] and *Blastococcus* [Hezbri et al., 2016] genera, whose first discovery was related to catacombs. Interestingly, the *Bacillaceae* and *Clostridiaceae* families, belonging to the phylum of Firmicutes were also identified.

The Proteobacteria phylum included families of nitrogen-fixing (diazotrophic) bacteria such as *Oxalobacteraceae*, *Rhizobiaceae*, *Sphingomonadaceae*, *Nitrosomonadaceae*, present in various habitats like water, soil, and plant-

associated environments. Some species/strains belonging to the *Oxalobacteraceae* family are mild plant pathogens or are claimed to be opportunistic human pathogens [Baldani et al., 2014].

One of the most interesting genera found in all samples from Fossombrone was *Brucella*, whose infection results in *Brucellosis* (also known as undulant fever, Malta fever, or Mediterranean fever [Di Pierdomenico et al., 2011]), a highly contagious zoonosis caused by ingestion of unpasteurized milk or undercooked meat from infected animals, or by close contact with their secretions.

Another potential pathogen identified was the *Clostridium_sensu_stricto_18* genus, which was also recently found in the microbiome of HIV-positive patient [Ahmed et al., 2020]. Unfortunately, no deeper investigation could be performed about the species of such potential pathogens due to aDNA degradation. We can though hypothesize that Fossombrone subjects died because of *Brucella* infection, rather than from a *Y. pestis* infection, as it was not identified, moreover we can also exclude an infection caused by *Clostridium*, because it was found in only 2 out of the 3 samples analysed.

As far as Modena (NP) samples are concerned (see figure 27), the Proteobacteria, Chloroflexi, Firmicutes and Actinobacteria were the four most abundant phyla, accounting for 32.8 %, 14.8 %, 13.1 % and 11.5% respectively. Again, a deeper investigation at family and genus level of Proteobacteria, Chloroflexi and Actinobacteria confirmed the presence of several groups normally found in soil and water, such as *Nocardioides, Nitrospira*, *Dongia*, *Flavihumbacterium.*

Although both samples belonging to Modena burial site displayed the presence of *Clostridium_sensu_stricto_15* of the *Clostridiaceae* family, no conclusions could be drawn regarding potential diseases attributed to this genus as the fundamental information about the specie was missing. Modena samples also presented *Fretibacterium* genus, isolated from subgingival plaque, suggesting that the samples were overall handled properly.

Finally, the most relevant phyla found in samples belonging to the New Lazarette (LN) of New Island in Venice (Figure 28) were Proteobacteria (44.3%) and

Actinobacteriota (28.3%), while all other phyla showed very low frequency. Concerning family classification, *Burkholderiaceae* and *Clostridiaceae* were the most interesting families, detected in 3 and 2 samples, respectively, while at genus level the presence of the *Nonomurieae* suggested that the samples underwent several climatic and atmospheric variations, as a generic abundance of such genus was found to be dependent on geographical changes [Sungthong and Nakaew 2015]. Similarly, to Fossombrone, also the catacombs-associated genus *Kribbella* was detected in 2 out of 3 samples. Unfortunately, none of the identified genera could be associated with any possible cause of death.

4.3.2 Paleomicrobiome analysis on V5 region

The same analysis presented for V3 region was performed on the other informative region of the 16S-rRNA sequence, that is the V5 region. The analysis of relative abundances revealed that the most abundant phyla were *Proteobacteria* (27 %), *Actinobacteria* (13.7 %), *Acidobacteriota* (7.2 %), *Bacteroidota* (6.7%), *Plancomycetota* (6.6%) and *Firmicutes* (6.3%), while all other phyla, such as *Verrucomicrobiota* displayed an abundance lower than 5%.

The prevalence analysis at class level (Figure 29) highlighted the presence of several Gram-positive and Gram-negative bacteria particularly relevant from a clinical microbiology standpoint. Indeed, the largest fraction of classes was related to Gram-negative bacteria, represented by *Gammaproteobacteria* and *Alphaproteobacteria*, while the Gram-positive bacteria were represented by *Clostridia* and *Bacilli* classes. It is worth mentioning, though, that the *Bacilli* class includes not only Gram-positive families like *Staphylococcaceae,*

Streptococcaceae and Listeria, but also for example *Escherichia coli*, which is a Gram-negative bacteria.

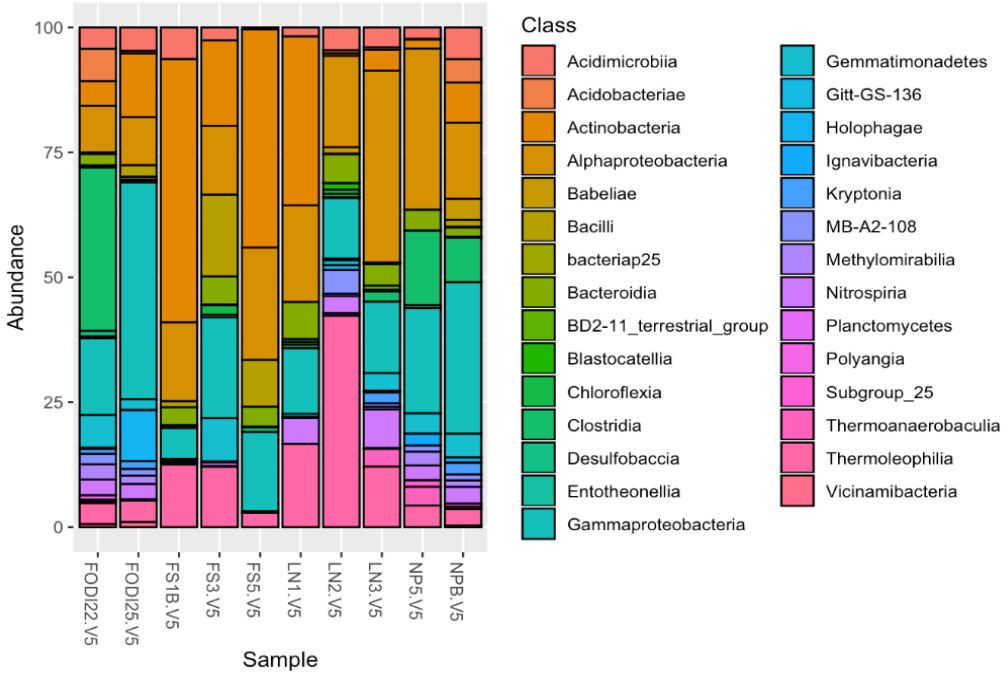


Figure 29. Relative phylum abundance according to V5 region

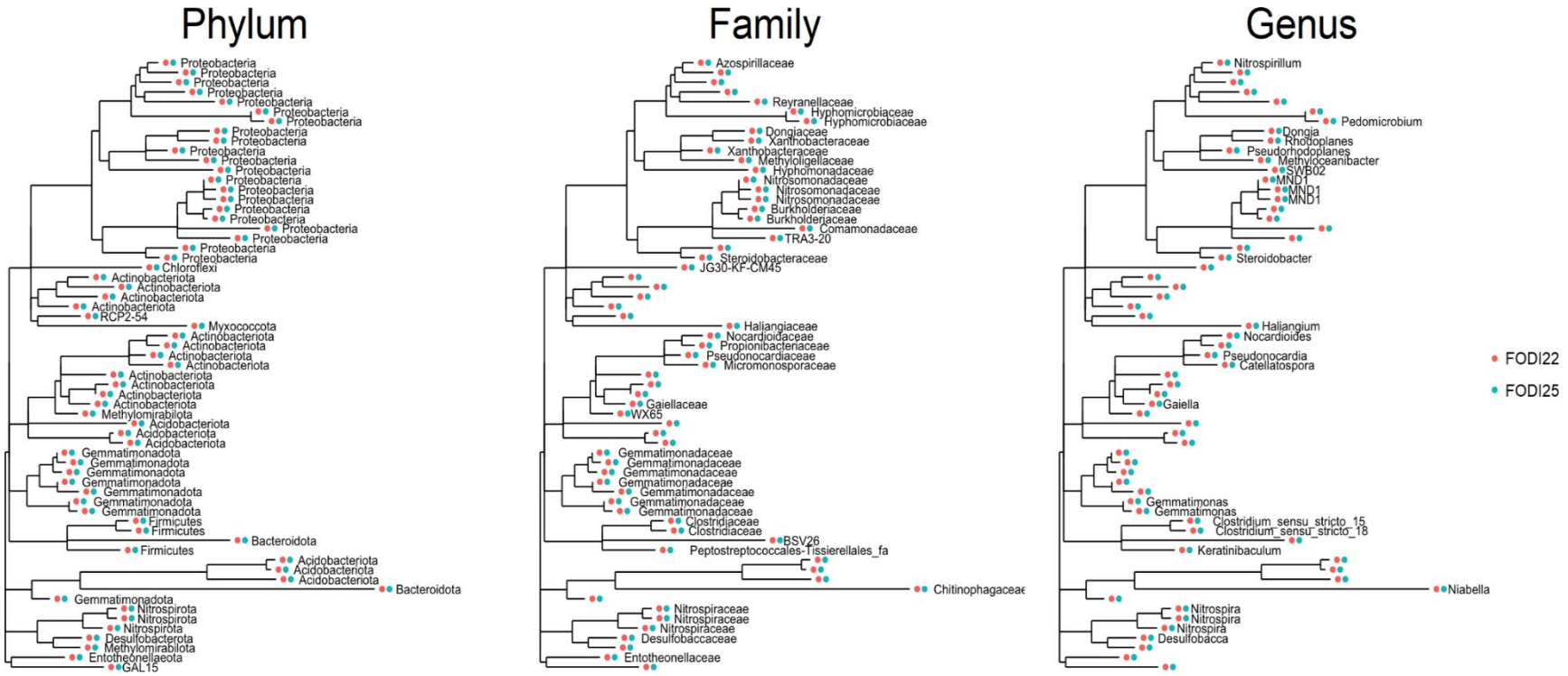


Figure 30. Phylogenetic tree based on V5 region of Forlì (FODI) cemetery at phylum, family and genus levels

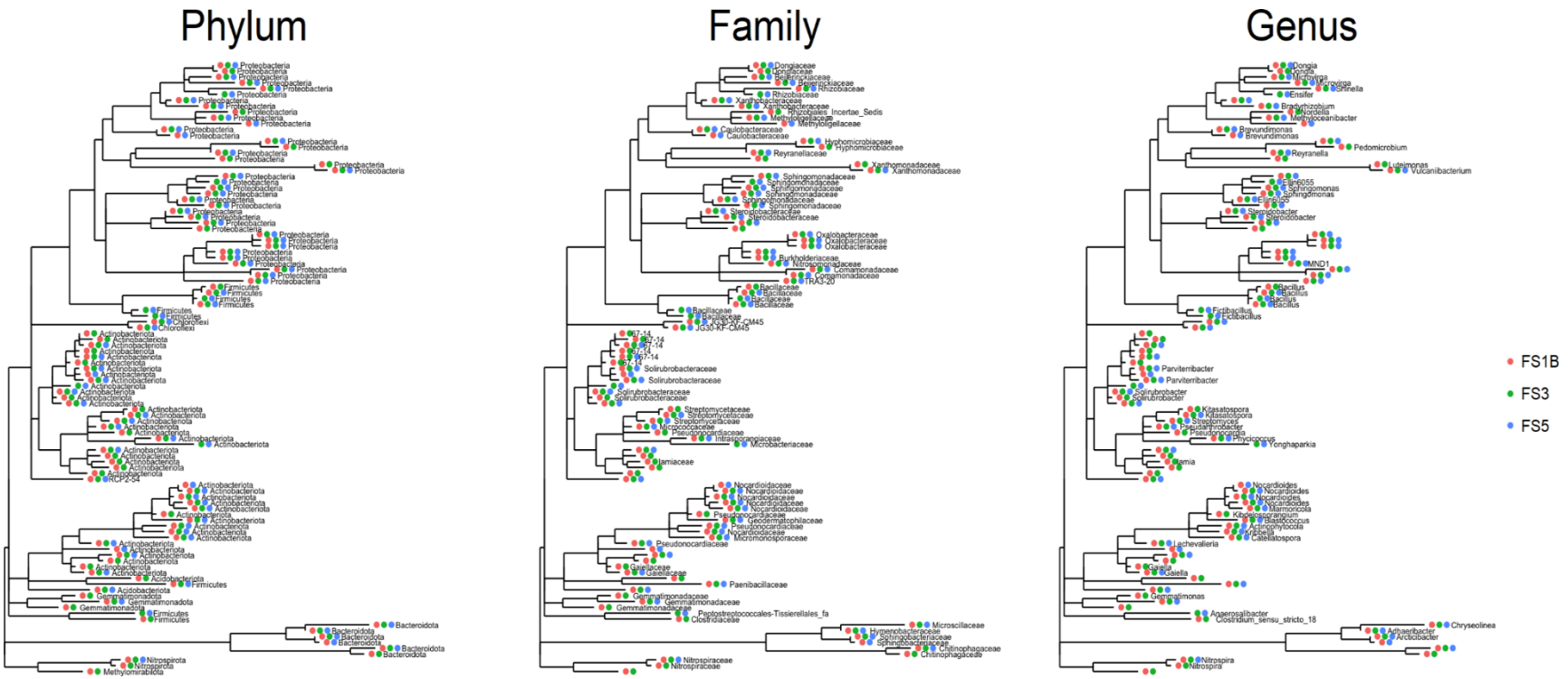


Figure 31. Phylogenetic tree based on V5 region of Fossombrone (FS) burial site at phylum, family and genus levels

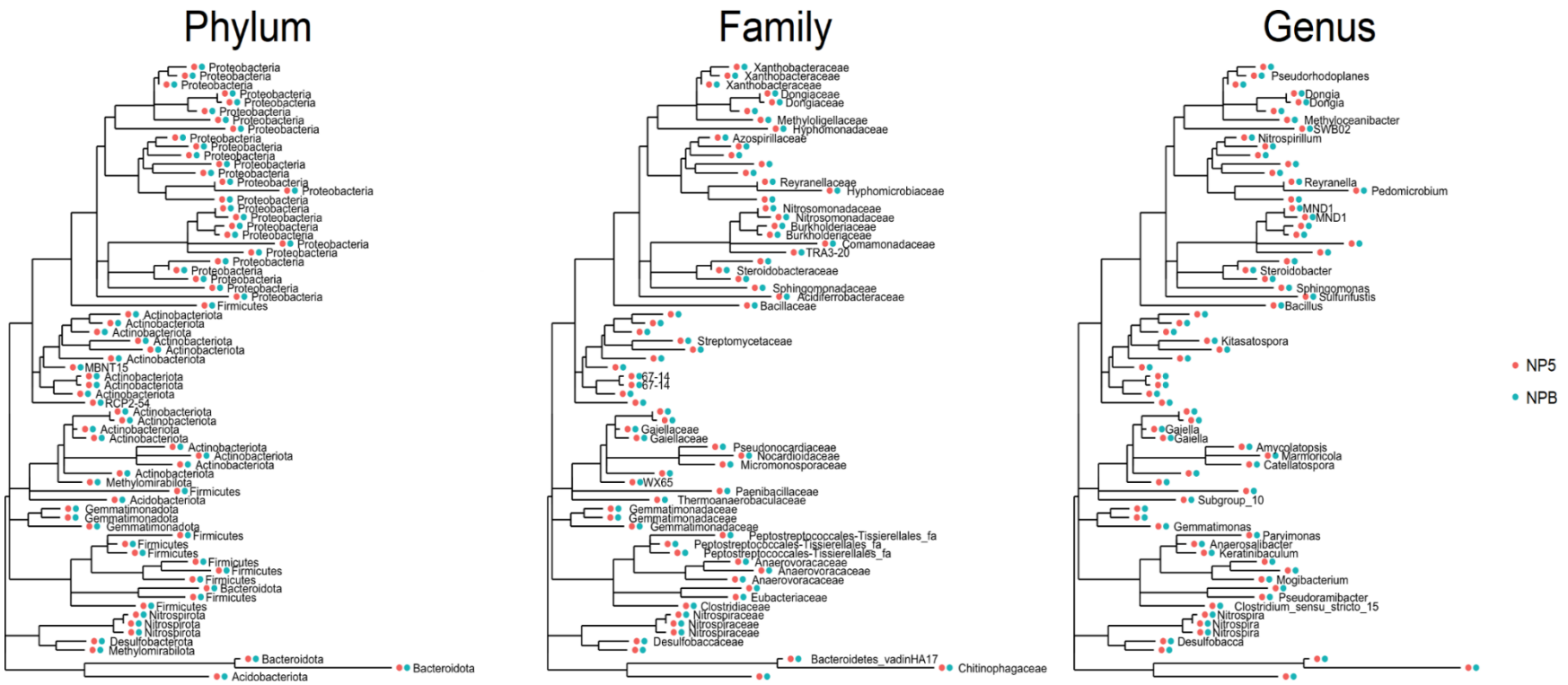


Figure 32. Phylogenetic tree based on V5 region of Modena cemetery (NP) at phylum, family and genus levels

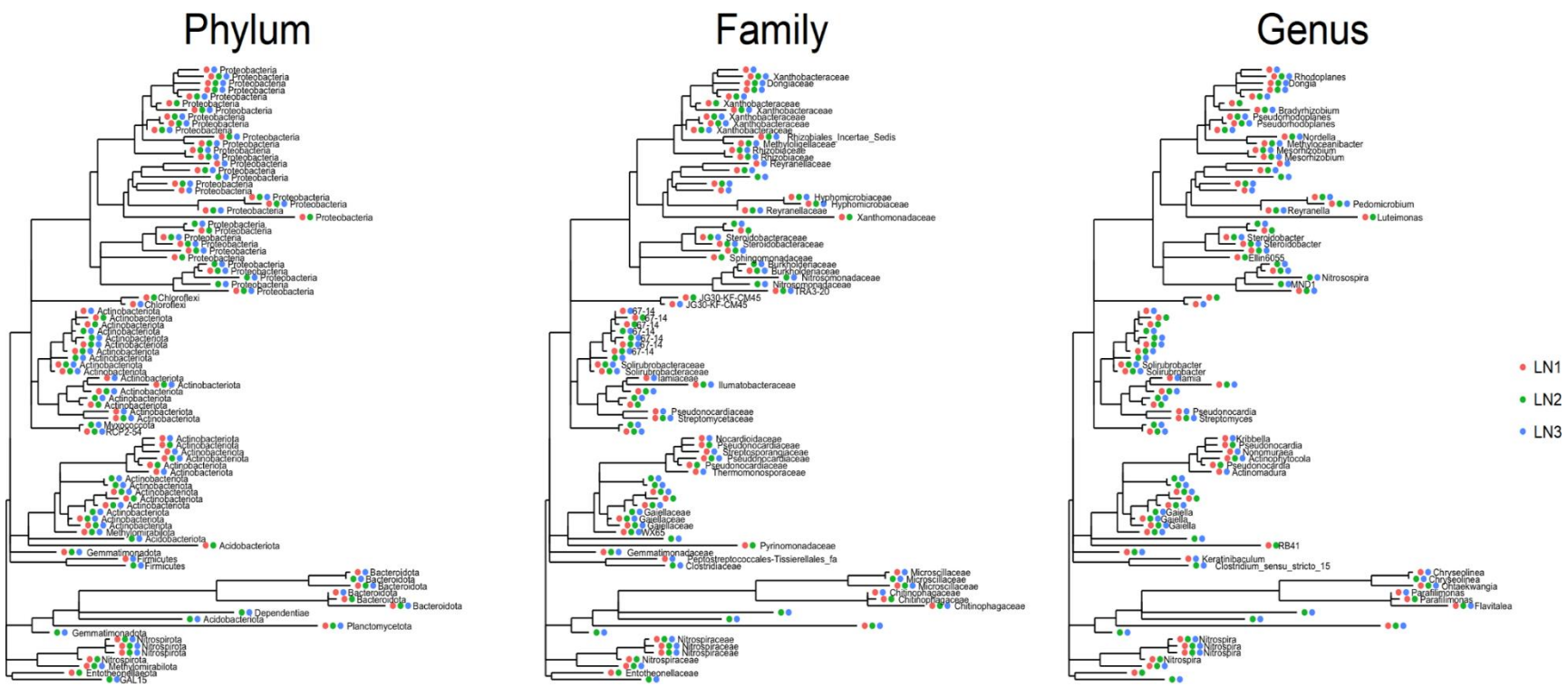


Figure 33. Phylogenetic tree based on V5 region of New Lazarette in Venice (LN) at phylum, family and genus levels

The phylogenetic tree concerning Forlì burial site (Figure 30), our "negative control" for infectious diseases research, showed that the most prevalent phyla *Proteobacteria* (33.3%), *Actinobacteria* (19 %) and *Gemmatimonadota* (12.7%), while *Firmicutes*, *Acidobacteriota* and others were detected in a much lower frequency. At a deeper level, the only potential pathogenic families found were *Burkholderiaceae*, *Nocardiaceae* and *Clostridiaceae*, but none of the identified genera could be associated with infectious diseases.

The most prevalent phyla from Fossombrone (Figure 31) were *Actinobacteria* (39.8%), followed by *Proteobacteria* (36.9%) and *Firmicutes* (6.8%), all other phyla displayed a frequency lower than 5%. Again, the potentially pathogenic families found were *Bacillaceae*, *Burkholdeiaceae*, *Clostridiaceae*, *Nocardiodaceae*, while at genus level we detected again several genera typically found in catacombs such as *Kribella* and *Blastococcus*. Interestingly, *Bacillus* was the most prevalent genus, including both free-living (nonparasitic) species and two parasitic pathogenic species, namely *B. anthracis*, responsible for anthrax, and *B. cereus*, associated with food poisoning.

Modena burial site presented a highly conserved phylogenetic tree in terms of grade/branches of phyla (Figure 32), with the most frequent being *Proteobacteria* (42.9%), *Actinobacteria* (27 %) and *Firmicutes* (15.9 %). At family level, several branches of *Bacillaceae* were detected, followed by *Burkholderiaceae*, *Nocardiodaceae* and *Clostridiaceae*, while at genus level, the most frequently found was the *Bacillus* genus, similarly to what was observed in Fossombrone samples.

In line with all other burial sites, *Proteobacteria* (37%), *Actinobacteria* (33%) and *Bacteroidota* (6.5%) were the most prevalent phyla found in the lazarette of New Island in Venice (Figure 33). The investigation at family level highlighted the presence of only the *Clostridiaceae* family, but no more precise information about genus could be retrieved from aDNA samples, implying that no potential cause of an infectious disease could be identified.

Chapter 5

CONCLUSIONS AND FUTURE PERSPECTIVES

Ancient DNA research is a complicate and expensive discipline to approach, but the continue development of sophisticated analytical techniques such as NGS, Metagenomics and WGS allowed to achieve a considerable success in studying ancient pathogens. Although aDNA analysis was established as the gold standard for unravelling the evolution of bacteria, such approach has to deal with critical problems like the retrieval, detection and characterization of aDNA molecules. Moreover, the overwhelmingly higher presence of modern DNA and PCR products in all the steps of the specimen handling can lead to potential contaminations. However, when all possible precautions are taken, the constitution of an aDNA repository can also provide important additional information on post-mortem DNA base modification, particularly on the interpretation and update of ancient genomes datasets [Der Sarkissian C et al., 2015].

Metagenomic analysis is facing a wide range of challenges, in particular the comparison of huge amounts of sequencing data against a steadily increasing number of reference sequences. While the overall composition can be characterized even without precise alignments, more sophisticated analyses require the assessment of complete alignments of the metagenomic sequencing reads against a comprehensive reference database.

Indeed, 16S rRNA amplicon sequencing has been the primary tool for characterizing ancient microbiome samples since 1998, yet no study has systematically investigated the effect of aDNA fragmentation on the fidelity of amplicon-based ancient microbiome reconstructions. Such issue results in an

unclear interpretation of the differences observed between modern and ancient microbial communities. Overamplification of archaeal taxa and altered microbial diversity estimates are predictable artifacts observed in poorly preserved (highly fragmented) but relatively uncontaminated aDNA samples.

We demonstrated that although amplicon-based 16S rRNA gene sequencing may be a useful high-throughput screening tool for qualitative characterization of the preservation and contamination of ancient microbiome samples, it cannot be used to reliably reconstruct qualitative information about microbial diversity. In addition, always due to the high fragmentation of DNA, no quantification of the taxonomic frequency in ancient microbial communities cannot be obtained, resulting in the impossibility to identify the precise species of microorganisms that could have been the cause of death.

The most abundant taxon in all burial sites resulting from the investigation of both V3 and V5 regions of the 16S rRNA was *Proteobacteria,* more specifically the class of *Gammaproteobacteria,* which includes several clinically relevant microbial genera such as *Legionella*, *Vibrio* and *Yersinia*.

For the samples belonging to Fossombrone, a possible cause of death was found in Brucellosis [Kay GL et al., 2014; Mutolo MJ et al., 2012], as *Brucella* was detected in this burial site, moreover also aDNA related to *Bacillus* genus was found in both V3 and V5 analysis, suggesting that probably this cemetery was erroneously associated with the plague.

Unfortunately, no traces of *Y. pestis* DNA were found in the teeth samples either in V3 or in V5 region. We also showed that the putatively specific primers were not as specific as expected, since the atmospheric and soil condition may have contributed to the accumulation of mutations in the targeted regions, leading to their a-specificity. Besides, the evolutionary and epidemic history of this bacterium could not be easily addressed due to the degradation of endogenous DNA of *Y. pestis*, which may have been present at that time but lost over time.

As reported in *Y. pestis* related literature, a single small genetic mutation fundamentally influenced the evolution of the deadly pathogen, and thus the course

of human history, and such single variant may have not survived the passing of time.

On the other hand, the metagenomic analysis revealed to be successful in identifying the variability of bacteria found in “a closed box” as the tooth is defined. Finally, teeth samples were found to be much more informative than any other bone belonging to the same individual, in terms of microbiome analysis.

Chapter 6

PUBLICATIONS LIST

Publications related to aDNA research

- Bazaj A, Turrina S, De Leo D, Cornaglia G. Palaeomicrobiology meets forensic medicine: time as a fourth-dimension for the crime scene. *New Microbes New Infect*. 2015;4:5–6. Published 2015 Jan 12. doi:10.1016/j.nmni.2014.12.006

- Saegeman V, Cohen MC, Alberola J, et al. How is post-mortem microbiology appraised by pathologists? Results from a practice survey conducted by ESGFOR. *Eur J Clin Microbiol Infect Dis*. 2017;36(8):1381–1385. doi:10.1007/s10096-017-2943-6

- Drancourt M, Barbieri R, Cilli E, et al. Did Caravaggio die of Staphylococcus aureus sepsis?. Lancet Infect Dis. 2018;18(11):1178. doi:10.1016/S1473-3099(18)30571-1

- Cilli E, Gabanini G, Ciucani M M, et al. 2020 Investigating childbirth deaths in double burials: Anthropology, paleopathology and ancient DNA. In press in Journal of Archaeological Science

Other publications

- Fan X, Wu Y, Xiao M, et al. Diverse Genetic Background of Multidrug-Resistant Pseudomonas aeruginosa from Mainland China, and Emergence of an Extensively Drug-Resistant ST292 Clone in Kunming. *Sci Rep*. 2016;6:26522. Published 2016 May 20. doi:10.1038/srep26522

- Mazzariol A, Bazaj A, Cornaglia G. Multi-drug-resistant Gram-negative bacteria causing urinary tract infections: a review. *J Chemother*. 2017;29(sup1):2–9. doi:10.1080/1120009X.2017.1380395

- Bazaj A, Bombiero E, Naso L D, Lo Cascio G, Cornaglia G. Fighting antibiotic resistance, it's in your hands: mobile phones are a fertile ground for microorganisms' growth. Published 2019 Journal Current trends in Microbiology

- Brandi J, Di Carlo C, Manfredi M, et al. Investigating the Proteomic Profile of HT-29 Colon Cancer Cells After Lactobacillus kefiri SGL 13 Exposure Using the SWATH Method. *J Am Soc Mass Spectrom*. 2019;30(9):1690–1699. doi:10.1007/s13361-019-02268-6

- Piccirilli A, Perilli M, Piccirilli V, et al. Molecular characterization of carbapenem-resistant Klebsiella pneumoniae ST14 and ST512 causing bloodstream infections in ICU and surgery wards of a tertiary university hospital of Verona (northern Italy): co-production of KPC-3, OXA-48, and CTX-M-15 β-lactamases. *Diagn Microbiol Infect Dis*. 2020;96(3):114968. doi:10.1016/j.diagmicrobio.2019.114968

- Brandi J, Cheri S, Manfredi M, Di Carlo C, Federici F, Bombiero E, Bazaj A, Rizzi E, Manna L, Cornaglia G, Marini U, Valenti M T, Marengo E, Cecconi D. Exploring the Wound Healing, Anti-Inflammatory, Anti-Pathogenic and Proteomic Effects of Lactic Acid Bacteria on Keratinocytes. Submitted to Scientific Reports

Bibliography

Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E, et al. Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proc Natl Acad Sci USA. 1999;96(24):14043–14048.

Allentoft ME, Sikora M, Sjögren KG, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. Nature. 2015;522(7555):167-172.

Andam CP, Worby CJ, Chang Q, Campana MG. Microbial Genomics of Ancient Plagues and Outbreaks. Trends Microbiol. 2016;24(12):978–990. doi:10.1016/j.tim.2016.08.004

Baker GC, Cowan DA. 16 S rDNA primers and the unbiased assessment of thermophile diversity. Biochem Soc Trans. 2004;32(Pt 2):218–221. doi:10.1042/bst0320218

Barnett R, Larson G. A phenol-chloroform protocol for extracting DNA from ancient samples. Methods Mol Biol. 2012;840:13–19. doi:10.1007/978-1-61779-516-9_2

Bazaj A, Turrina S, De Leo D, Cornaglia G. Palaeomicrobiology meets forensic medicine: time as a fourth-dimension for the crime scene. New Microbes New Infect. 2015;4:5–6. Published 2015 Jan 12. doi:10.1016/j.nmni.2014.12.006

Bennett EA, Massilani D, Lizzo G, Daligault J,Geigl EM, Grange T. Library construction for ancient genomics: single strand or double strand? BioTechniques. 2014;56:289–298.

Blow MJ, Zhang T, Woyke T, et al. Identification of ancient remains through genomic sequencing [published correction appears in Genome Res. 2008 Nov;18(11):1859]. Genome Res. 2008;18(8):1347–1353. doi:10.1101/gr.076091.108

Borrini M, Bartoli F. Analisi paleonutrizionale su alcuni campioni dalla mass grave dell'Isola del Lazzaretto Nuovo (Venezia) Archivio per l'Antropologia e la Etnologia Vol. CXL 2010

Bos KI, Herbig A, Sahl J, et al. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. Elife. 2016;5:e12994. Published 2016 Jan 21. doi:10.7554/eLife.12994

Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee JB, DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJ, Herring DA, Bauer P, Poinar HN, Krause J. A draft genome of Yersinia pestis from victims of the Black Death. Nature. 2011 Oct 12;478(7370):506-10. doi: 10.1038/nature10549. Erratum in: Nature. 2011 Dec 8;480(7376):278.

Bouwman AS, Kennedy SL, Müller R, Stephens RH, Holst M, Caffell AC, et al. Genotype of a historic strain of Mycobacterium tuberculosis. Proc Natl Acad Sci U S A. 2012;109(45):18511-18516.

Briggs AW, Heyn P. Preparation of next-generation sequencing libraries from damaged DNA. Methods Mol Biol. 2012;840:143–154. doi:10.1007/978-1-61779-516-9_18

Briggs AW, Stenzel U, Johnson PL, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A. 2007;104(37):14616–14621. doi:10.1073/pnas.0704665104

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, et al. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. Am J Hum Genet. 2013;93(5): 852–864.

Carroll MW, Matthews DA, Hiscox JA, et al. Temporal and spatial analysis of the 2014-2015 Ebola virus outbreak in West Africa. Nature. 2015;524(7563):97–101. doi:10.1038/nature14594

Cilli E, Gabanini G, Ciucani M.M, De Fanti S, Serventi P, Bazaj A, Sarno S, Ferri G, Fregnani A, Cornaglia G, Gruppioni G, Luiselli D, Traversari M. A multifaceted approach towards investigating childbirth deaths in double burials: Anthropology, paleopathology and ancient DNA. Submitted to Journal of Archaeological Science.

Cohn SK Jr. Epidemiology of the Black Death and successive waves of plague. Med Hist Suppl. 2008;(27):74-100

Collins FM. Pasteurella, Yersinia, and Francisella. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 29. Available from: https://www.ncbi.nlm.nih.gov/books/NBK7798/

Cooper A, Poinar HN. Ancient DNA: do it right or not at all. Science. 2000;289(5482):1139. doi:10.1126/science.289.5482.1139b

Cowan DA, Arslanoglu A, Burton SG, et al. Metagenomics, gene discovery and the ideal biocatalyst. Biochem Soc Trans. 2004;32(Pt 2):298–302. doi:10.1042/bst0320298

Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, et al. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. Proc Natl Acad Sci USA. 2013; 110(2):577–582.

Dabney J, Meyer M, Pääbo S. Ancient DNA damage. Cold Spring Harb Perspect Biol. 2013;5(7):1–7.

Der Sarkissian C, Allentoft ME, Ávila-Arcos MC, et al. Ancient genomics. Philos Trans R Soc Lond B Biol Sci. 2015;370(1660):20130387. doi:10.1098/rstb.2013.0387.

Devault AM, Golding GB, Waglechner N, Enk JM, Kuch M, Tien JH, et al. Second-pandemic strain of Vibrio cholerae from the Philadelphia cholera outbreak of 1849. N Engl J Med. 2014;370(4):334-340.

Di Pierdomenico A, Borgia SM, Richardson D, Baqi M (2011). "Brucellosis in a returned traveller". CMAJ. 183 (10): E690-2. doi:10.1503/cmaj.091752. PMC 3134761. PMID 21398234

Drancourt M and Raoult D. Paleomicrobiology of Humans. ISBN-13: 978-1555819163. First publication 2016

Drancourt M, Aboudharam G, Signoli M, Dutour O, Raoult D. Detection of 400-year-old Yersinia pestis DNA in human dental pulp: an approach to the diagnosis of ancient septicemia. Proc Natl Acad Sci USA. 1998;95(21):12637–12640.

Drancourt M, and Raoult D. Cause of Black Death. Lancet Infect Dis. 2002 Aug;2(8):459.

Drancourt M, and Raoult D. Palaeomicrobiology: current issues and perspectives. Nature Reviews Microbiology. 2005;3(1):23-55.

Eppinger M, Pearson T, Koenig SS, et al. Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. mBio. 2014;5(6):e01721. Published 2014 Nov 4. doi:10.1128/mBio.01721-14

Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, et al. DNA analysis of an early modern human from Tianyuan Cave, China. PNAS. 2013;110(6): 2223-2227.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, et al. The genetic history of Ice Age Europe. Nature. 2016;534(7606):200-205.

Fulton TL, Stiller M. PCR amplification, cloning, and sequencing of ancient DNA. Methods Mol Biol. 2012;840:111–119. doi:10.1007/978-1-61779-516-9_15

Gamba C, Hanghøj K, Gaunitz C, Alfarhan AH, Alquraishi SA, Khaled AS, et al. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. Mol Ecol Resour. 2016;16:459–469.

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, et al. Genome flux and stasis in a five millennium transect of European prehistory. Nat Commun. 2014;5:5257.

Gansauge MT, Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat Protoc. 2013;8(4):737–748. doi:10.1038/nprot.2013.038

Gernaey AM, Minnikin DE, Copley MS, Dixon RA, Middleton JC, Roberts CA. Mycolic acids and ancient DNA confirm an osteological diagnosis of tuberculosis. Tuberculosis. 2001;81:259–265.

Gilbert MTP, Hansen AJ, Willerslev E, Turner-Walker G, Collins M. Insights into the processes behind the contamination of degraded human teeth and bone samples with exogenous sources of DNA. Int J Osteoarchaeol. 2006;16:156–164.

Gilbert MTP, Rudbeck L, Willerslev E, Hansen AJ, Smith C, Penkman KEH, et al. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. J Arch Sci. 2005;32:785–793.

Gire SK, Goba A, Andersen KG, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. Science. 2014;345(6202):1369–1372. doi:10.1126/science.1259657

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328(5979):710-722.

Green RE, Malaspinas AS, Krause J, Briggs AW, Johnson PLF, Uhler C, et al. A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. Cell. 2008; 134(3):416–426.

Hänsch S, Cilli E, Catalano G, et al. The pla gene, encoding plasminogen activator, is not specific to Yersinia pestis. BMC Res Notes. 2015;8:535. Published 2015 Oct 5. doi:10.1186/s13104-015-1525-x

Hansen HB, Damgaard PB, Margaryan A, et al. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. PLoS One. 2017;12(1):e0170940. Published 2017 Jan 27. doi:10.1371/journal.pone.0170940

Harbeck M, Seifert L, Hänsch S, et al. Yersinia pestis DNA from skeletal remains from the 6(th) century AD reveals insights into Justinianic Plague. PLoS Pathog. 2013;9(5):e1003349. doi:10.1371/journal.ppat.1003349

Heintzman PD, Zazula GD, Cahill JA, Reyes AV, MacPhee RDE, Shapiro B, et al. Genomic Data from Extinct North American Camelops Revise Camel Evolutionary History. Molecular Biology and Evolution. 2015;32(9):2433–2440.

Hert DG, Fredlake CP, Barron AE. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. Electrophoresis. 2008;29(23):4618–4626. doi:10.1002/elps.200800456

Hezbri, K; Louati, M; Nouioui, I; Gtari, M; Rohde, M; Spröer, C; Schumann, P; Klenk, HP; Ghodhbane-Gtari, F; Montero-Calasanz, MD (November 2016). "Blastococcus capsensis sp. nov., isolated from an archaeological Roman pool and emended description of the genus Blastococcus, B. aggregatus, B. saxobsidens, B. jejuensis and B. endophyticus". International Journal of Systematic and Evolutionary Microbiology. 66 (11): 4864–4872. doi:10.1099/ijsem.0.001443. PMID 27553620.

Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC.DNA sequences from the quagga, an extinct member of the horse family. Nature. 1984;312(5991):282-284.

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S. Ancient DNA. Nat Rev Genet. 2001;2(5):353–359. doi:10.1038/35072071

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, Pääbo S. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Res. 1996;24(7): 1304–1307.

Höss M, Pääbo S. DNA extraction from Pleistocene bones by a silica-based purification method. Nucleic Acids Res. 1993; 21:3913–3914.

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome research*, *17*(3), 377-386.

Janse I, Hamidjaja RA, Reusken C. Yersinia pestis plasminogen activator gene homolog in rat tissues. Emerg Infect Dis. 2013;19(2):342–344. doi:10.3201/eid1902.120659

Kämpfer P, Dott W, Martin K, Glaeser SP. Rhodococcus defluvii sp. nov., isolated from wastewater of a bioreactor and formal proposal to reclassify [Corynebacterium

hoagii] and Rhodococcus equi as Rhodococcus hoagii comb. nov. Int J Syst Evol Microbiol. 2014;64(Pt 3):755–761. doi:10.1099/ijs.0.053322-0

Kay GL, Sergeanta MJ, Giuffra V, Bandiera P, Milanesed M, Bramanti B, et al. Recovery of a Medieval Brucella melitensis Genome Using Shotgun Metagenomics. mBio. 2014;5 (4): e01337-14. Published 2014 Jul 15. Doi: 10.1128/mBio.01337-14

Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. Nat Commun. 2012;3(698).

Kolman CJ, Centurion-Lara A, Lukehart SA, Owsley DA, Tuross N. Identification of Treponema pallidum subspecies pallidum in a 200-year-old skeletal specimen. J Infect Dis. 1999;180(6):2060–2063.

Kousoulis AA, Economopoulos KP, Poulakou-Rebelakou E, Androutsos G, Tsiodras S. The plague of Thebes, a historical epidemic in Sophocles' Oedipus Rex. Emerg Infect Dis. 2012;18(1):153–157. doi:10.3201/eid1801.AD1801

Labate D, Malnati L, et al., 2017 PARCO NOVI SAD DI MODENA: DALLO SCAVO AL PARCO ARCHEOLOGICO. Archeologia, antropologia, storia e ambiente di un insediamento periurbano di età romana e medievale. Ministero dei Beni Culturali e il Turismo.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature. 2014;513(7518):409-413.

Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993;362(6422):709-715.

Llamas B, Willerslev E, Orlando L. Human evolution: a tale from ancient genomes. Philos Trans R Soc Lond B Biol Sci. 2017;372(1713):20150484. doi:10.1098/rstb.2015.0484

Luni M, Mei O, Gobbi P. Necropoli Tardoantica a Forum Sempronii cum bibl. 2013

Margaryan A, Hansen HB, Rasmussen S, Sikora M, Moiseyev V, Khoklov A, et al. Ancient pathogen DNA in human teeth and petrous bones. Ecol Evol. 2018;8(6):3534-3542.

Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science. 2012;338(6104):222-226.

Miller W, Drautz DI, Ratan A, Pusey B, Qi J, Lesk AM, et al. Sequencing the nuclear genome of the extinct woolly mammoth. Nature. 2008;456(7220):387-90.

Modi A, Tassi F, Susca RR, et al. Complete mitochondrial sequences from Mesolithic Sardinia. Sci Rep. 2017;7:42869. Published 2017 Mar 3. doi:10.1038/srep42869

Mutolo MJ, Jenny LL, Buszek AR, Fenton TW, Foran DR. Osteological and molecular identification of brucellosis in ancient Butrint, Albania. Am J Phys Anthropol. 2012;147:254–263.

Nguyen-Hieu T, Aboudharam G, Signoli M, Rigeade C, Drancourt M, Raoult D. Evidence of a louse-born outbreak involving typhus in Douai, 1710–1712 during the War of Spanish Succession. PLoS One. 2010;5:e15405.

Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CW, DeGiorgio M, et al. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature. 2014;507(7491):225-8.

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, et al. The Beaker phenomenon and the genomic transformation of northwest Europe. Nature. 2018;555(7695):190-196.

Orlando L, Gilbert MT, Willerslev E. Reconstructing ancient genomes and epigenomes. Nat Rev Genet. 2015;16(7):395–408. doi:10.1038/nrg3935

Pääbo S, Higuchi RG, Wilson AC. Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. J Biol Chem. 1989;264(17):9709–9712.

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, et al. Genetic analyses from ancient DNA. Annu Rev Genet. 2004;38:645–679.

Pääbo S. Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci USA. 1989;86(6):1939-1943.

Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome Biol. 2016;17:60.

Perry RD, Fetherston JD. Yersinia pestis--etiologic agent of plague. Clin Microbiol Rev. 1997 Jan;10(1):35-66.

Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, et al. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science. 2006;311(5759):392-394.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature. 2014;505(7481):87-91.

Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, Stafford TW Jr, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature. 2014;506(7487):225-229.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen, MoltkeI, et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature. 2010; 463(7282):757-762.

Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren KG, et al. Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. Cell. 2015;163(3):571-582.

Rivera-Perez JI, Santiago-Rodriguez TM, Toranzos GA. Paleomicrobiology: a Snapshot of Ancient Microbes and Approaches to Forensic Microbiology. Microbiol Spectr. 2016;4(4):10.1128/microbiolspec.EMF-0006-2015. doi:10.1128/microbiolspec.EMF-0006-2015

Rohland N, Hofreiter M. Ancient DNA extraction from bones and teeth. Nature Protocols. 2007; 2:1756–1762.

Rothschild BM. History of syphilis. Clin Infect Dis. 2005;40(10):1454–1463. doi:10.1086/429626

Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94(3):441-448.

Schuenemann VJ, Avanzi C, Krause-Kyora B5, Seitz A, Herbig A, Inskip S, et al. Ancient genomes reveal a high diversity of Mycobacterium leprae in medieval Europe. PLoS Pathog. 2018;14(5):e1006997.

Schuenemann VJ, Bos K, DeWitte S, Schmedes S, Jamieson J, Mittnik A, et al. Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of Yersinia pestis from victims of the Black Death. Proc Natl Acad. Sci USA. 2011;108(38):E746-52.

Schuenemann VJ, Peltzer A, Welte B, van Pelt WP, Molak M, Wang CC, et al. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. Nat Commun. 2017;8:15694.

Schuenemann VJ, Singh P, Mendum TA, et al. Genome-wide comparison of medieval and modern Mycobacterium leprae. Science. 2013;341(6142):179–183. doi:10.1126/science.1238286

Seguin-Orlando A, Korneliussen TS, Sikora M, et al. Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. Science. 2014;346(6213):1113–1118. doi:10.1126/science.aaa0114

Seguin-Orlando A, Schubert M, Clary J, et al. Ligation bias in illumina next-generation DNA libraries: implications for sequencing ancient genomes. PLoS One. 2013;8(10):e78575. Published 2013 Oct 29. doi:10.1371/journal.pone.0078575

Spyrou MA, Tukhbatova RI, Feldman M, Drath J, Kacki S, Beltrán de Heredia J, et al. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. Cell Host Microbe. 2016;19(6):874-881.

Spyrou MA, Tukhbatova RI, Wang CC, Valtueña AA, Lankapalli AK, Kondrashin VV, et al. Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. Nat Commun. 2018;9(1):2234.

Stenseth NC, Atshabar BB, Begon M, Belmain SR, Bertherat E, Carniel E, Gage KL, Leirs H, Rahalison L. Plague: past, present, and future. PLoS Med. 2008 Jan 15;5(1):e3. doi: 10.1371/journal.pmed.0050003

Stiller M, Fulton TL. Multiplex PCR amplification of ancient DNA. Methods Mol Biol. 2012;840:133–141. doi:10.1007/978-1-61779-516-9_17

Sungthong R, Nakaew N. The genus Nonomuraea: A review of a rare actinomycete taxon for novel metabolites. J Basic Microbiol. 2015;55(5):554–565. doi:10.1002/jobm.201300691

Treille G-F, and Yersin A. La peste bubonique à Hong-Kong. Ann. Inst. Pastur. 2, 428-430, 1894

Urzì C; De Leo, F; Schumann, P; et al. (2008). "Kribbella catacumbae sp. nov. and Kribbella sancticallisti sp. nov., isolated from whitish-grey patinas in the catacombs of St Callistus in Rome, Italy". International Journal of Systematic and Evolutionary Microbiology. 58 (Pt 9): 2090–2097. doi:10.1099/ijs.0.65613-0. PMID 18768610.

Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, et al. Yersinia pestis and the plague of Justinian 541-543 AD: a genomic analysis. Lancet Infect Dis. 2014 Apr;14(4):319-26. doi: 10.1016/S1473-3099(13)70323-2. Epub 2014 Jan 28.

Wagner S, Lagane F, Seguin-Orlando A, Schubert M, Leroy T, Guichoux E, et al. High-Throughput DNA sequencing of ancient wood. Mol Ecol. 2018;27(5):1138-1154.

Weil T, De Filippo C, Albanese D, et al. Legal immigrants: invasion of alien microbial communities during winter occurring desert dust storms. Microbiome. 2017;5(1):32. Published 2017 Mar 10. doi:10.1186/s40168-017-0249-7

Wells CL, Wilkins TD. Clostridia: Sporeforming Anaerobic Bacilli. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 18. Available from: https://www.ncbi.nlm.nih.gov/books/NBK8219/

Willerslev E, Cappellini C, Boomsma W, Nielsen R, Hebsgaard MB,1 Brand TB, et al. Ancient biomolecules from deep ice cores reveal a forested southern Greenland. Science. 2007;317(5834):111–114.

Willmann C, Mata X, Hanghoej K, et al. Oral health status in historic population: Macroscopic and metagenomic evidence. PLoS One. 2018;13(5):e0196482. Published 2018 May 16. doi:10.1371/journal.pone.0196482

Winn WC Jr. Legionella. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 40. Available from: https://www.ncbi.nlm.nih.gov/books/NBK7619/

Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87(12):4576–4579. doi:10.1073/pnas.87.12.4576

Woese CR. Bacterial evolution. Microbiol Rev. 1987;51(2):221–271.

World Health organization (WHO). Plague report 2018 Available from https://www.who.int/health-topics/plague#tab=tab_1

Zaremba-Niedźwiedzka K, Andersson SG. No ancient DNA damage in Actinobacteria from the Neanderthal bone. PLoS One. 2013;8(5):e62799. Published 2013 May 3. doi:10.1371/journal.pone.0062799

Zhou B, Wen S, Wang L, Jin L, Li H, Zhang H. Ant Caller: an accurate variant caller incorporating ancient DNA damage. Molecular Genetics and Genomics. 2017;292(6):1419–1430

Ziesemer KA, Mann AE, Sankaranarayanan K, et al. Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification [published correction appears in Sci Rep. 2016 Jun 02;6:27163]. Sci Rep. 2015;5:16498. Published 2015 Nov 13. doi:10.1038/srep16498

Zietz BP, Dunkelberg H. The history of the plague and the research on the causative agent Yersinia pestis. Int J Hyg Environ Health. 2004 Feb;207(2):165-78.

Zink AR, Sola C, Reischl U, Grabner W, Rastogi N, Wolf H, Nerlich AG. Characterization of Mycobacterium tuberculosis complex DNAs from Egyptian mummies by spoligotyping. J Clin Microbiol. 2003;41(1):359–367.

Supplementary Information

Ancient DNA extraction (Rohland & Hofreiter, 2007; Dabney et al., 2013; Allentoft et al., 2015)

MATERIAL

Gloves
Micropipettes
Tips
Serological pipettes
Germicidal wipes
NaClO
NaCl solution 5M in H_2O
H_2O_2
Falcon 15/50 ml
Eppendorf 1.5 / 2 ml
Heater
Stove
Centrifuge 14,000 rpm
MiniElute PCR® purification kit (Qiagen) Buffer TE and BP
Ethanol at 80%
HPLC water
QG buffer Qiagen
TE Buffer (Tris-EDTA pH 8.0)
EDTA (Ethylenediaminetetraacetic acid 0.5 M)

PROTOCOL

DAY 1

Sterilize reagents and UV materials for at least 60 minutes

1. Calculate the number of samples to be extracted plus a negative control

2. In a 15 ml falcon, add at maximum 100 mg of bone or tooth powder and 480µl of the extraction Buffer as it is shown in Table below

3. Ensure that the tubes are closed with Parafilm

4. Incubate overnight at 37 ° C in continuous agitation

Preparation of the extraction Buffer

Reagents	**Stock Conc.**	**Final Conc.**	**Q.ty for 1 sample (µl)**
EDTA* pH8	0,5 M	0,45 M	432
Proteinase K*	20 mg/ml	0.25mg/ml	6
H_2O			42
Final Volume			480

*EDTA 0.5 M (chelating agent for Ca ions)
*Proteinase K (enzyme that digests proteins)

DAY 2

Sterilize all reagents and materials under UVC rays for at least 60 minutes

1. Check the personal rates of Buffer PE and PB, if they are sufficient for extraction;

2. Remove the samples from the stove and centrifuge at 10,000 rpm for 3 min;

3. Transfer the supernatant to a new Eppendorf tube

4. Add 960 µl of Buffer PB and mix for 10 s with a vortex and a short spin of centrifuge;

5. Transfer 750 µl of liquid into the MinElute® Qiagen Vials and centrifuge at 14,000 rpm for 2 min. (in this way the liquid passes through the membrane in the lower part, while instead the DNA will be linked to the siliceous membrane), then throw the eluate, dab the tube and put it back under the column.

6. Repeat the operation until the solution is exhausted in each Eppendorf;

7. Add 750 µl of Buffer PE (washing buffer for DNA cleaning), then centrifuge at 6,000 rpm for 2 min, the liquid passes through the membrane in the tube, then throw the eluate and put the tube back under the column; Repeat the operation;

8. Centrifuge at 14,000 rpm for one minute, to dry the filter, and then throw tube;

9. Place the silica column in a new 1.5 ml Eppendorf, cut the caps and elute with 20 µl of Buffer TE or water and then incubate the whole at 37 ° C or room Temperature for 5 minutes;

10. Finally centrifuge at 14,000 rpm for 1 min;

11. elute with another 20 µl of Buffer TE or water and then incubate the whole at 37 ° C or room T for 5 minutes;

12. Centrifuge at 14,000 rpm for 1 min. Now throw the silica column and keep the Eppendorf with the eluate in the freezer at - 20 ° C for further use.

N.B. It is also advisable to make aliquots of the sample in 0.5 ml Eppendorf to avoid contamination and frequent thawing / freezing.

After the procedure:

- Clean any instrument / material used with sodium hypochlorite
- Rinse / clean with H_2O_d and possibly germicidal wipes
- Turn on UVC in working spaces and under the hoods

Electrophoresis results of traditional PCR:

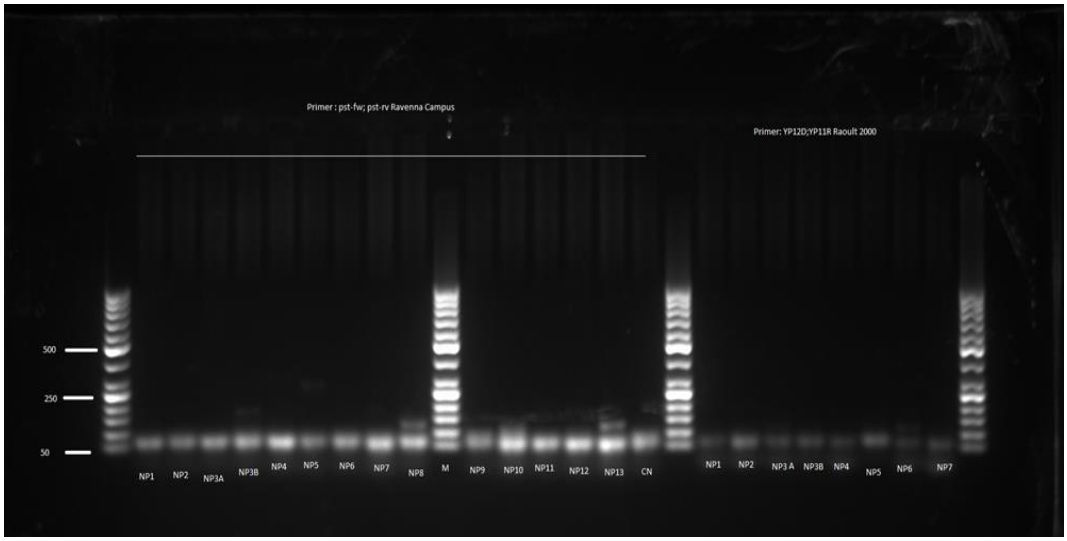


Figure 1. Traditional PCR performed for Modena burial site.

On the left primers from Ravenna campus. On the right "Single suicide PCR" primer [Raoult et al., 2000].

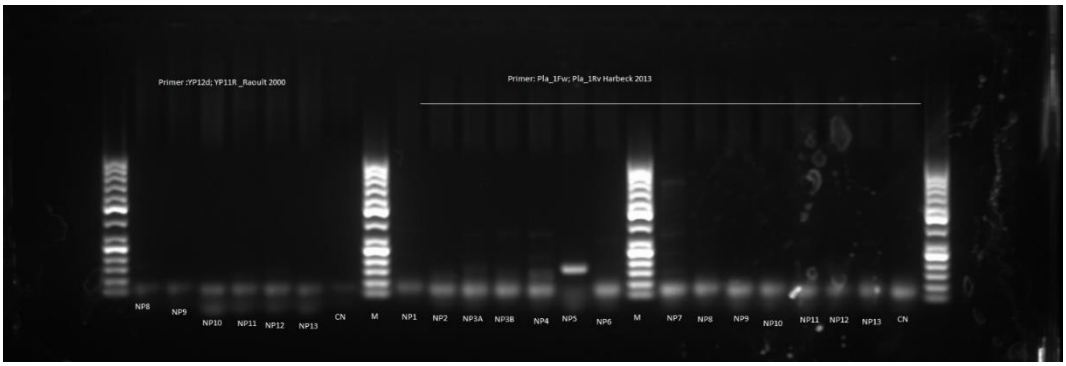


Figure 2. Traditional PCR performed for Modena burial site.

On the left "Single suicide PCR" primer[Raoult et al., 2000] On the right the "specific" pla gene primer [Hänsch et al., 2015].

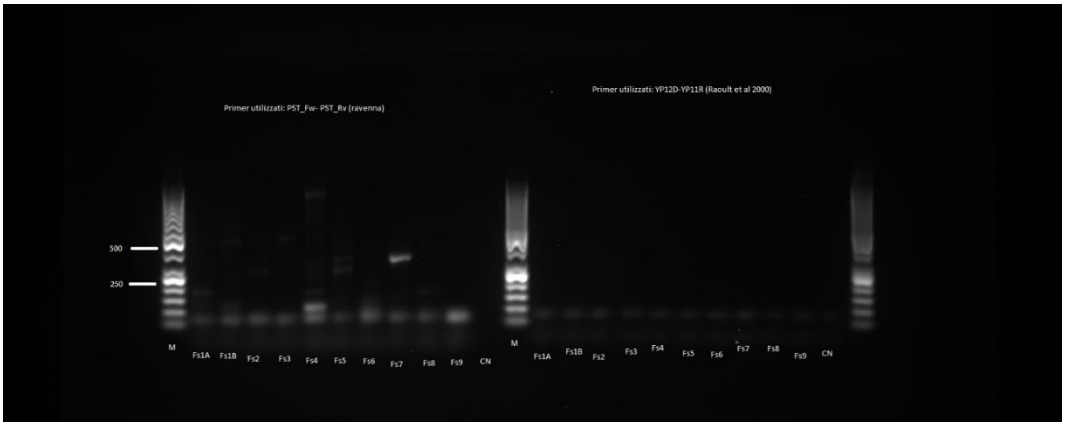


Figure 3. Electrophoresis of traditional PCR for Fossombrone burial site.

On the left primers from Ravenna campus. On the right "Single suicide PCR" primer [Raoult et al., 2000].

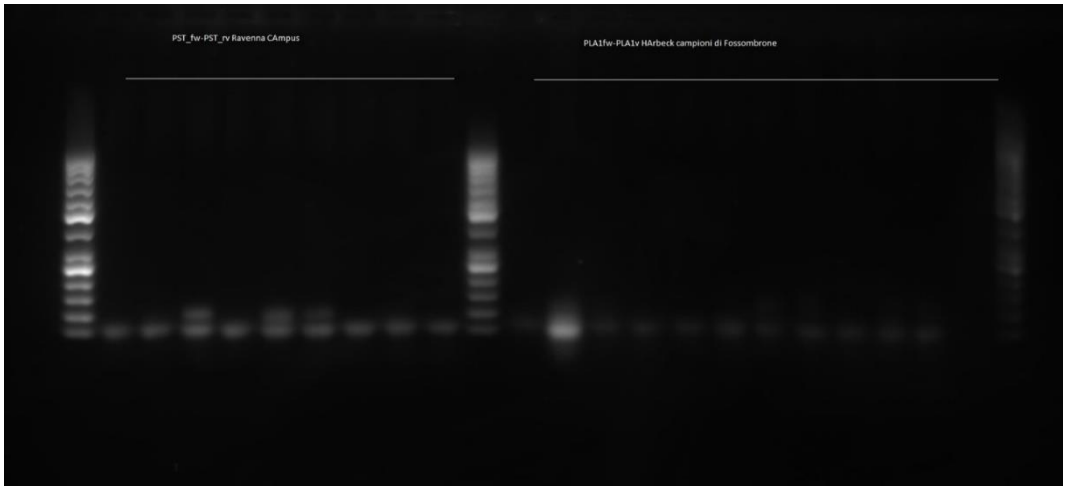


Figure 4. Electrophoresis of traditional PCR for Venezia on the left and Fossombrone on the right burial sites.

On the left with Primer from Ravenna campus, and on the right with "specific" pla gene primer [Hänsch et al., 2015].

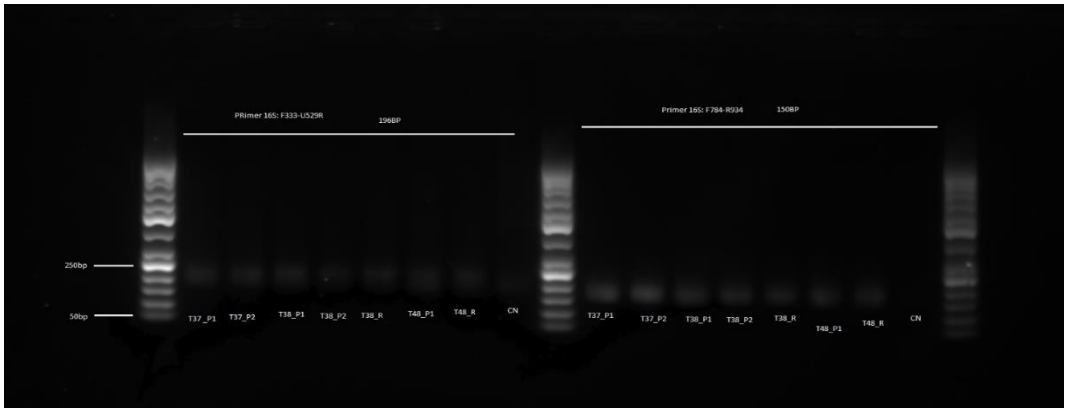


Figure 5. 16s rRNA V3-V5 PCR amplification of specimens from Venezia Burial site.