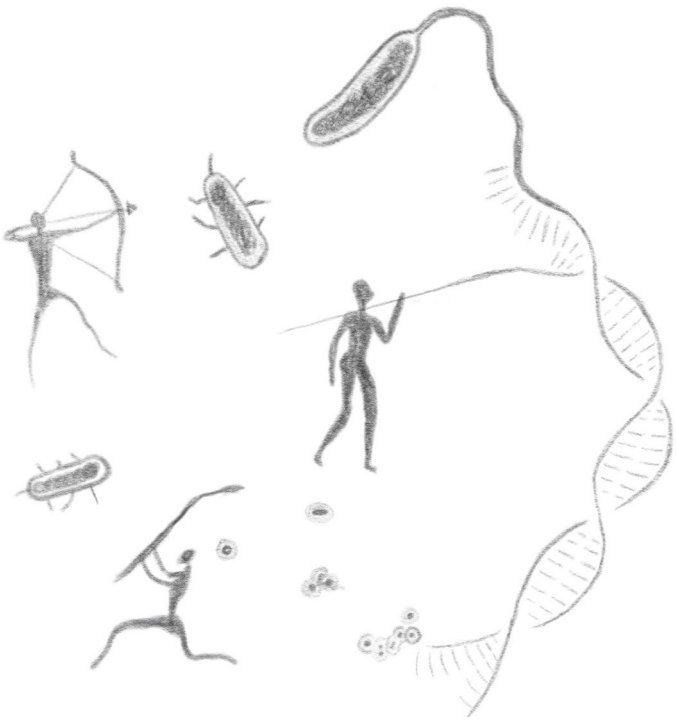
Stockholm University

Metagenomic analysis for detection of pathogenic microorganisms in prehistoric human populations

Nora Bergfeldt



Doctoral Thesis in Systematic Zoology at Stockholm University, Sweden 2025

Metagenomic analysis for detection of pathogenic microorganisms in prehistoric human populations

Nora Bergfeldt

Academic dissertation for the Degree of Doctor of Philosophy in Systematic Zoology at Stockholm University to be publicly defended on Friday 21 March 2025 at 10.00 in Vivi Täckholmsalen (Q-salen), NPQ-huset, Svante Arrhenius väg 20.

Abstract

Disease and pathogens have affected human populations throughout history, something that the global pandemics of the 21st century can attest for. With the development of methods for DNA extraction and sequencing during the last decade, it is now possible to study ancient pathogen evolution and transmission more in depth, within the field of ancient metagenomics. However, a long-standing challenge in ancient metagenomics has been high error rates and false positive identifications. In this thesis, I have aimed to initially improve the methods for analysing ancient DNA data, and further to study the presence and evolution of pathogens in populations across human prehistory. In **chapter I**, I present aMeta, an accurate ancient metagenomics profiling workflow that has been designed to minimize the number of false positive identifications, as well as to streamline computer memory usage. Using simulated as well as authentic ancient DNA data, aMeta was benchmarked against an existing workflow, and its superior sensitivity and specificity in both microbial detection and authentication was demonstrated. Further, we could show its substantially lower usage of computer memory. In **chapter II**, the aMeta workflow was applied on a dataset consisting of 38 individuals from four Mesolithic and Neolithic Scandinavian human cultural complexes. Several species of bacteria were identified in the dataset, for example the bacterium *Salmonella enterica* in two individuals from the Battle Axe cultural complex. Since osteological examination did not present any physical damage to the bones, this disease may have been the cause of death for the infected individuals. Several species of the bacterial genus *Yersinia* were identified in individuals from the Funnel Beaker culture context, and denser populations in an agricultural context may have facilitated the transmission of these pathogens. Further, in Mesolithic and Neolithic hunter-gatherers, two pathogenic species of the genus *Neisseria* were identified, representing the, to our knowledge, earliest findings of the species to date. In **chapter III**, aMeta was applied to a dataset from Mexico, consisting of 41 individuals dated between 900 – 1800 CE. In one individual, we identified DNA from the bacterium *Vibrio cholerae*, the causing agent of cholera. We created a phylogeny consisting of available, globally collected *Vibrio* genomes and concluded that our finding, the earliest of *V. cholerae* to date, likely belongs to a non-choleric strain and thus may not have been the cause of an epidemic. Further, the finding indicates that cholera may have arrived in the Americas decades earlier than previous research has shown. In **chapter IV**, we presented genomic data from 40 individuals in northeast Asia, dated between circa 16,900 and 550 years ago. Population demographics showed genetic affinity between the analysed individuals and present-day human populations in Asia and Native America. We further used the metagenomics tool Malt to identify *Yersinia pestis* reads in two individuals from 4,400 and 3,800 years ago respectively, representing the most northeastern ancient finding of the bacterium.

Keywords: *ancient DNA, pathogen evolution, metagenomics, Salmonella enterica, Vibrio cholerae, Yersinia pestis, Neolithic.*

Stockholm 2025
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-238718

ISBN 978-91-8107-098-9
ISBN 978-91-8107-099-6

Stockholm University

Department of Zoology

Stockholm University, 106 91 Stockholm

METAGENOMIC ANALYSIS FOR DETECTION OF PATHOGENIC MICROORGANISMS IN PREHISTORIC HUMAN POPULATIONS

Nora Bergfeldt

Stockholm University

Metagenomic analysis for detection of pathogenic microorganisms in prehistoric human populations

Nora Bergfeldt

ISBN print 978-91-8107-098-9
ISBN PDF 978-91-8107-099-6

Cover by Julia Höglund

Printed in Sweden by Universitetsservice US-AB, Stockholm 2025

If you do metagenomics, never map alone.
- E.K.

The thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I Pochon, Z.[#], **Bergfeldt, N.[#],** Kırdök, E., Vicente, M., Naidoo, T., van der Valk, T., Altınışık, E. N., Krzewińska, M., Dalén, L., Götherström, A.[#], Mirabello, C.[#], Unneberg, P.[#], Oskolkov, N.[#] (2023). aMeta: an accurate and memory-efficient ancient Metagenomic profiling workflow. *Genome Biology*, 24(1), 242.

II **Bergfeldt, N**., Kırdök, E., Oskolkov, N., Mirabello, C., Unneberg, P., Malmström, H., Fraser, M., Sanchez-Quinto, F., Jorgensen, R., Skar, B., Lidén, K., Jakobsson, M., Storå, J., Götherström, A. (2024). Identification of microbial pathogens in Neolithic Scandinavian humans. *Scientific Reports*, *14*(1), 5630.

III **Bergfeldt, N.**, Oskolkov, N., Rodriguez-Varela, R., Guinet, B., Larsson, P., Talavera González, J. A., Guevara Flores, S., Navarrete Linares, F., Götherström, A.[#], Valdiosera, C.[#]
Vibrio cholerae in 18th century Mexico. *Manuscript.*

IV Kılınç, G. M. [#], Kashuba, N. [#], Koptekin, D., **Bergfeldt, N.**, Dönertaş, H. M., Rodríguez-Varela, R., Shergin, D., Ivanov, G., Kichigin, D., Pestereva, K., Volkov, D., Mandryka, P., Kharinskii, A., Tishkin, A., Ineshin, E., Kovychev, E., Stepanov, A., Dalén, L., Günther, T., Kırdök, E., Jakobsson, M., Somel, M., Krzewińska, M., Storå, J., Götherström, A. (2021). Human population dynamics and *Yersinia pestis* in ancient northeast Asia. *Science Advances*, *7*(2), eabc4587.

[#]These authors contributed equally

Candidate contributions to thesis articles*

	I	II	III	IV
Conceived the study	Significant	Substantial	Significant	Minor
Designed the study	Substantial	Substantial	Substantial	Minor
Collected the data	Minor	Minor	Minor	Minor
Analysed the data	Significant	Substantial	Significant	Significant
Manuscript preparation	Significant	Substantial	Substantial	Significant

*** Contribution Explanation**

Minor: contributed in some way, but contribution was limited.
Significant: provided a significant contribution to the work.
Substantial: took the lead role and performed the majority of the work.

I am also a co-author to the following articles, which are not included in the thesis:

Dussex, N., **Bergfeldt, N.**, de Anca Prado, V., Dehasque, M., Díez-del-Molino, D., Ersmark, E., ... & Heintzman, P. D. (2021). Integrating multi-taxon palaeogenomes and sedimentary ancient DNA to study past ecosystem dynamics. *Proceedings of the Royal Society B*, *288*(1957), 20211252.

Larsson, P., von Seth, J., Hagen, I. J., Götherström, A., Androsov, S., Germonpré, M., **Bergfeldt, N.**, ... & Dalén, L. (2019). Consequences of past climate change and recent human persecution on mitogenomic diversity in the arctic fox. *Philosophical Transactions of the Royal Society B*, *374*(1788), 20190212.

Srigyan, M., Bolívar, H., Ureña, I., Santana, J., Petersen, A., Iriarte, E., Kırdök, E., **Bergfeldt, N.**, … & Valdiosera, C. (2022). Bioarchaeological evidence of one of the earliest Islamic burials in the Levant. *Communications biology*, *5*(1), 554.

This thesis builds partly upon the author's licentiate thesis (defended on October 25th, 2022). Of the papers included in this thesis, Paper I and II were part of the licentiate. By chapters, the contribution from the licentiate thesis is as follows:

The introductory chapter ("kappa"): This chapter was included in the licentiate; for this thesis it has been reviewed and updated and around 50% of the text and references are new.

Chapter 1: This chapter was included in the licentiate; for this thesis it has been reviewed and updated.

Chapter 2: This chapter was included in the licentiate; for this thesis it has been reviewed and updated.

Contents

Introduction

Ancient DNA

Studying the evolutionary history of prehistoric populations with the help of ancient DNA (aDNA) has been done for several decades (Higuchi et al., 1984; Pääbo, 1985). With the development of next-generation sequencing (NGS) methods, it is now possible to retrieve high-coverage mitochondrial and nuclear genomes from both extant and extinct species. With this high-quality genomic data, the genetic history and adaptations of these species can be studied. Furthermore, the age limit of a sample for DNA extraction and sequencing has increased dramatically during the last decade, from 560–780 thousand years old (Orlando et al., 2013), to circa 2 million years old (Kjær et al., 2022).

A major challenge when working with aDNA is that the DNA is highly degraded. The aDNA fragments are typically shorter than 100 base pairs in length (Pääbo, 1989) and are characterized by post-mortem damage patterns. Deamination of the nucleobase cytosine to uracil, which occurs in degrading DNA, leads to the misincorporation of thymine in the DNA sequence (Briggs et al., 2007; Stiller et al., 2006; Gilbert et al., 2007). These C-to-T transitions are more frequent at the 5'-end of the DNA molecule (Briggs et al., 2007), which makes it possible to use this pattern for authenticating the reads as ancient.

Another issue when studying aDNA is that the endogenous DNA in the sample often is vastly exceeded by DNA from other organisms (Handt et al., 1994). This exogenous DNA can come from modern contaminants, but may also be part of the individual's microbiome, or an infectious agent. With NGS methods, high quality data from this exogenous DNA can be retrieved, and likewise from environmental samples, which has led to the development of ancient metagenomics.

Ancient metagenomics and pathogens

Diseases in prehistoric populations have been studied using an interdisciplinary perspective, combining research subjects such as osteology and anthropology with medicine and microbiology. The prevalence of a disease has been possible to determine with the help of lesions some diseases cause in the skeleton of a sick individual, for example dental damage making it possible to

identify caries (W. J. Moore & Corbett, 1973; Meng et al., 2011) and periodontal disease (Clarke et al., 1986) in ancient humans. Further, tuberculosis infections can cause lesions in bones and joints, which occur in circa 1% of infected individuals (Davidson & Horowitz, 1970). With molecular data now being available for analysis, previous osteological diagnoses can be confirmed using molecular methods, and historical sources that tell of disease can be investigated. Moreover, aDNA makes it possible to study the evolution and transmission of pathogenic microbes in depth.

There are several early examples of claimed pathogen identification in ancient samples using PCR based methods for DNA analysis, for example tuberculosis in medieval humans being diagnosed by the presence of *Mycobacterium tuberculosis* DNA (Gernaey et al., 2001). With time, more reliable results have been presented. A well-studied pathogen in an ancient context is *Yersinia pestis*, the bacterial species that causes plague. The presence of the bacterium has been confirmed with aDNA in samples from the Black Death (Bos et al., 2011), as well as from other medieval, Bronze age and Neolithic sources (Rasmussen et al., 2015; Rascovan et al., 2019).

In addition to identifying pathogens in ancient samples, ancient metagenomics can be used to study changes in the human microbiome over time. In a study on early Neolithic humans, a shift in oral microbiota between hunter-gatherers and farmers has been seen, showing that earlier hunter-gatherers had fewer taxa associated to caries and periodontal disease than farmers (Adler et al., 2013). Another study, on palaeofaeces from 1,000 to 2,000 years ago, showed variations in human gut microbiome diversity between pre- and post-industrialized populations (Wibowo et al., 2021).

New insights on the transmission of diseases are also constantly developed using aDNA. A 2013 study suggested the spread of tuberculosis to the Americas from the African continent via the slave trade (Jaeger et al., 2013), after finding DNA evidence of *M. tuberculosis* in African slaves buried in Brazil. Another suggestion is that the bacterium spread to the Americas as a possible zoonosis from seals (Bos et al., 2014).

With the development of ancient metagenomic methods, a more comprehensive overview of the life history of ancient populations can be formed.

The Neolithic transition in Scandinavia

The transition from hunting and gathering to agriculture and farming is one of the most important processes in human history and is known as neolithization. This transition did not only mean a change in diet, but also resulted in a new lifestyle that has been described as the origin of civilization as we know it (Weisdorf, 2005). This shift in lifestyle may also have led to infectious diseases being spread more easily, due to human populations becoming larger and more dense (Armelagos & Dewey, 1970).

The shift from hunting and gathering to farming was largely driven across Europe by migration from Anatolia in present-day Turkey (Skoglund et al., 2012; Lazaridis et al., 2014; Kılınç et al., 2016). In Scandinavia, the Neolithic period started circa 6,000 years before present (BP) (Knutsson & Knutsson, 2003). The first farmer culture was the Funnel Beaker culture (FBC), which persisted until circa 4,800 years BP and was succeeded by the Battle Axe culture (BAC) (Larsson, 2008). However, an important hunter-gatherer cultural complex, the Pitted Ware culture (PWC), co-existed with these two farmer cultures between 5,400–4,400 years BP. The genetic ancestry and gene flow between these cultures has been studied previously and results have shown that the FBC has a major ancestry component deriving from Anatolian Neolithic farmers, while the PWC context share more ancestry with chronologically older Scandinavian hunter-gatherers (SHG) (Günther et al., 2018; Sánchez-Quinto et al., 2019; Skoglund et al., 2012; Skoglund, Malmström, et al., 2014). Individuals from a BAC context however seem to share most genetic ancestry with other groups belonging to the Corded Ware culture (CWC) on the European continent (Malmström et al., 2019). Further, low levels of admixture between individuals of FBC and PWC contexts have been detected (Skoglund, Malmström, et al., 2014), but while there is evidence that social and cultural interactions occurred between individuals from the PWC and BAC complexes in Scandinavia, there is limited evidence for genetic interactions between the groups (Coutinho et al., 2020).

Transatlantic migration in the 15th century

Human migration implies a movement of microbial communities associated with the population. In the 15th century Common Era (CE), Europeans arrived in the Americas, and the infectious diseases they brought with them, such as smallpox and measles (Merbs, 1992), have been described as a key factor in the depopulation of indigenous people (Howell, 2002). Today's México City had an important cultural role during the collision between the Eurasian and American worlds, being the seat of the viceroyalty of New Spain. The city was affected by multiple epidemics, such as the smallpox epidemic in 1779 that caused thousands of deaths (Cooper, 1965), and by using aDNA, several pathogenic bacterial species have been identified in Mexico in the centuries after the Europeans arrived. In samples from a cemetery in southern Mexico, *Salmonella enterica* DNA has been found, linking it to an epidemic in the 16th century and suggesting the introduction of *S. enterica* by Europeans (Vågene et al., 2018). aDNA has shown that *S. enterica* was widespread in Medieval Europe (Haller et al., 2021; Zhou et al., 2018), supporting the hypothesis of a spread from Europe to America at this time. Furthermore, aDNA determined to be from human parvovirus B and human hepatitis B virus has been found

in human remains from 16th century Mexico, suggesting the introduction of these viruses by the slave trade (Guzmán-Solís et al., 2021).

The peopling of northeast Asia

Humans arrived in northeastern Siberia circa 40,000 years ago (Sikora et al., 2019), before the Last Glacial Maximum (LGM) during the Last Glacial Period, and it has been suggested that a population from the area later dispersed into the Americas (Goebel et al., 2008). Lithic technologies and pre-Neolithic pottery show that several cultural complexes have been present in northeast Siberia since then (Derevianko et al., 2004; Goebel et al., 2008). Lately, studies have been published that show substantial changes in the genetic structure of the human populations, connecting these with migrations across the region (Wong et al., 2017). However, to fully understand the population dynamics of this vast area across the millennia since the LGM, further investigation is required. By analysing genetic data from the Trans-Baikal area – east of Lake Baikal – which carry the earliest traces of post-LGM human presence, and from sites in Yakutia associated with the ancestors of the Paleo-Inuit Saqqaq cultural complex, new insights can be developed.

Aims

The overall purpose of this thesis was to develop a time- and cost-efficient metagenomics workflow to expand the possibility of identifying microbial species related to ancient remains, and to use this workflow to investigate pathogen prevalence related to several prehistoric human populations from different places and time periods. In **chapter I**, we combined existing software for metagenomic analysis of DNA sequence data, with the aim to achieve a streamlined and automated process for species identification. Further, we compared the efficiency of our workflow to previously developed and established metagenomic workflows, with the purpose of benchmarking our workflow against existing tools. In **chapter II**, we applied our method on Scandinavian hunter-gatherers and farmers from the Mesolithic and Neolithic eras, to explore the prevalence and consequences of infectious diseases, and to gain insight in whether infectious agents may have been transmitted between populations from different cultural complexes. In **chapter III**, the workflow was used to analyse a dataset of human genomes from present-day Mexico, with the aim to investigate whether any infectious diseases were introduced by Europeans who arrived in the 15th century. Further analyses were made with the purpose of finding the phylogenetic relationships between our identified bacteria and other strains of the species. In **chapter IV**, we aimed to gain further understanding in how mobility, admixture processes, and disease may have

affected the population history of northeast Asia after the Last Glacial Maximum, by studying a dataset of ancient human genomes and using metagenomic tools to detect pathogenic microorganisms in the genetic material.

Materials and methods

Samples

For **chapter I**, we used simulated ancient metagenomic data to evaluate the power of our workflow. The method was then further evaluated using authentic aDNA data that had previously shown presence of *Yersinia pestis* (Rascovan et al., 2019). The human sample was collected in Gökhem, Sweden and was also included in **chapter II**.

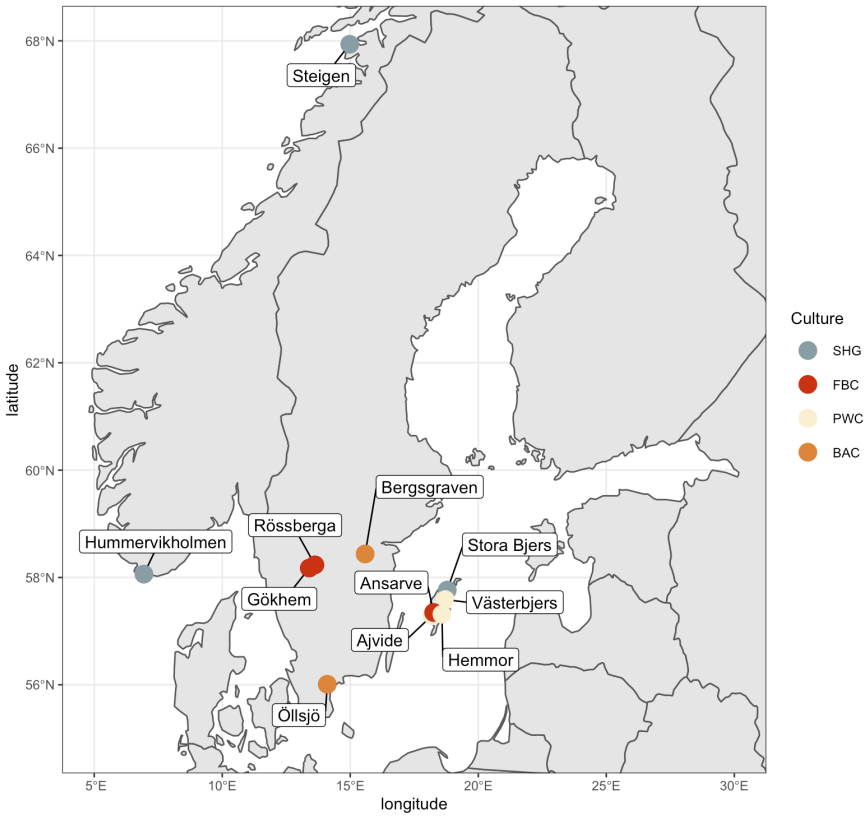


Figure 1. Map of Scandinavia showing the locations where samples for chapters I (Gökhem) and II were collected. SHG = Mesolithic Scandinavian hunter-gatherers, FBC = Funnel Beaker Culture, PWC = Pitted Ware Culture, BAC = Battle Axe Culture. (**chapter II**)

The 38 archaeological samples used in **chapter II** were collected from 11 sites in Sweden and Norway (Fig. 1). The human content of 36 of these samples had been previously investigated, while for two samples the DNA was sequenced for this study. The samples included have been radiocarbon dated between 9,700 to 4,400 years BP.

In **chapter III** we analysed human DNA from 41 individuals, collected from different archaeological sites in the Valley of Mexico and dated between 900–1800 CE.

Since bacterial DNA has been shown to be better preserved in teeth than in other skeletal parts (Margaryan et al., 2018), only tooth samples were used in **chapters II** and **III**.

In **chapter IV**, a genomic dataset consisting of 40 individuals from north-eastern Asia, dated between 16,900 to 550 years BP, was investigated. Samples were collected from five regions in Russia: Yakutia, Trans-Baikal, Cis-Baikal, Krasnoyarsk and Amur Oblast.

Laboratory and sequencing methods

DNA from the samples used in **chapter II**, as well as the authentic ancient data in **chapter I**, was extracted in aDNA laboratory facilities in Stockholm and Uppsala, Sweden. All samples were extracted using demineralization and a protein denaturing agent and were purified using silica binding protocols. aDNA from the two individuals in **chapter II** that hadn't previously been sequenced were extracted and sequenced according to the procedure described in (Günther et al., 2018). The samples analysed in **chapter III** and **chapter IV** were processed in the aDNA facilities at the Archaeological Research Laboratory, Stockholm University.

For **chapter III**, samples were decontaminated using ultraviolet radiation, and the root tips of the teeth were treated with predigestion buffer to further eliminate external contaminants. Extractions were performed according to the protocol in (Dabney et al., 2013). Libraries were prepared mainly according to (Meyer & Kircher, 2010). DNA sequencing was performed on an Illumina HiSeq X10 at SciLifeLab National Genomics Infrastructure in Stockholm.

Samples in **chapter IV** were cleaned mechanically and subjected to ultraviolet radiation. For some samples, libraries were prepared using USER treatment (NEB/BioNordika) (**chapter IV** table S1) to remove deaminated cytosines. Sequencing was performed on an Illumina HiSeq X platform at the SciLifeLab sequencing centre in Stockholm.

Bioinformatics

During the design of these studies, we identified the lack of a metagenomics tool to perform our desired analysis. After considering several software and existing workflows, we decided to create a workflow that includes both a screening step and an authentication step, resulting in the workflow aMeta (**chapter I**) (Fig. 2). The screening step was performed with KrakenUniq (Breitwieser et al., 2018), a tool using *k*-mers, short sequences of length *k*, to quickly classify the origin of a DNA sequence. To be as inclusive as possible, we built a custom extended NCBI non-redundant (NT) database, consisting of microbial (bacteria, viruses, archaea, fungi and parasitic worms) as well as human and available complete eukaryotic genomes. We performed global alignment of the pathogenic microbial reads using Bowtie 2 (Langmead & Salzberg, 2012) and deamination patterns were computed using MapDamage2 tool (Jónsson et al., 2013).

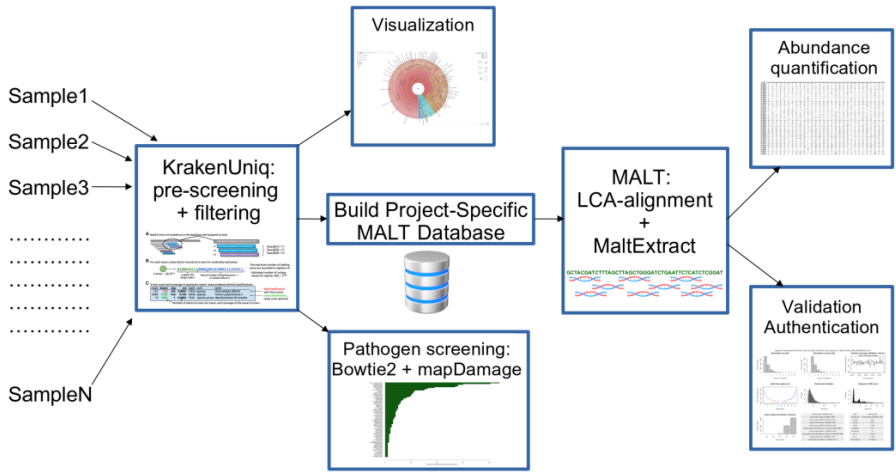


Figure 2. aMeta: ancient metagenomics profiling workflow overview. The workflow uses a combination of taxonomic classification and filtering steps with KrakenUniq, and the output is used for building a MALT database. Alignments are run against the database, and validation and authentication analysis are performed based on the alignments. (**chapter I**)

Furthermore, several previous studies have used MALT (MEGAN Alignment Tool) (Herbig et al., 2016) as an approach for alignment and analysis of aDNA sequencing data. MALT consists of two programs, where malt-build creates an index of reference sequences from the reference database GenBank (Benson et al., 2013), while malt-run aligns the samples, trimmed of adapters,

against the reference database (Herbig et al., 2016). We used MALT to validate the results from the KrakenUniq screening step, and to eliminate potential false positive identifications. We performed an alignment round with the Lowest Common Ancestor (LCA) algorithm, implemented in the MALT software (Herbig et al., 2016). MALT alignments were run using a custom reference database consisting of the microbial genomes from the species identified in the dataset by KrakenUniq.

To further authenticate the microbial organisms found in the human aDNA samples, the deamination pattern, read length distribution, and edit distance of the sequences are computed. Additionally, breadth/evenness of coverage of the reads aligned to each microbial organism is addressed using Samtools (Li et al., 2009). Lastly, histograms of post-mortem damage (PMD) scores are computed using PMDtools (Skoglund, Northoff, et al., 2014) to evaluate ancient status of the reads, and a graphical output is created (Fig. 3). Microbial abundance quantification from rma6 MALT alignments is performed using MaltExtract and rma2info wrapper script from MEGAN tool, as well as a custom awk script applied to sam-alignments from MALT.

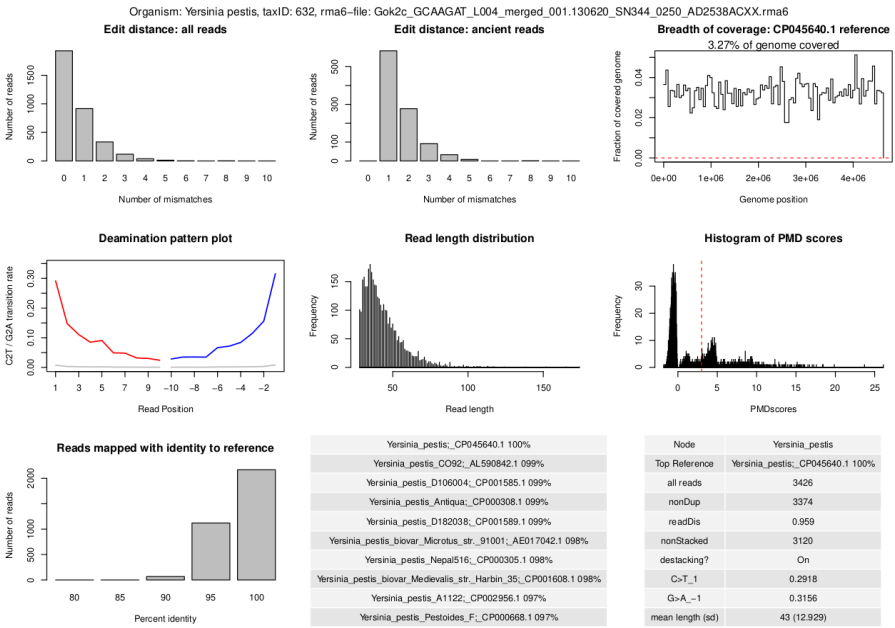


Figure 3. Final output of aMeta. Panels from left to right, top to bottom: a) edit distance computed on all assigned reads, b) edit distance computed on damaged reads, c) evenness / breadth of coverage, d) deamination pattern, e) read length distribution, f) PMD scores distribution, g) number of reads assigned with an identity to a reference, h) candidate reference sequences with percentages of mapped reads, i) MaltExtract statistics. (**chapter I**)

In **chapter II** we applied the metagenomics workflow we developed in **chapter I** to a dataset consisting of pre-historic Scandinavian humans, from hunter-gatherer and farmer cultural complexes. In **chapter III**, the workflow was applied to a dataset from present-day Mexico. A sample containing *Vibrio cholerae* DNA was further analysed. All DNA libraries from the sample were merged and *V. cholera* reads were extracted after alignment against a reference genome using bwa aln (Li & Durbin, 2009). Modern reference genomes were aligned using Sibeliaz (Minkin & Medvedev, 2020) and a phylogenetic tree was inferred using IQ-TREE (Nguyen et al., 2015), with a maximum likelihood framework.

In **chapter IV**, we investigated the demography and relatedness of the 40 pre-historic humans in the dataset. The authenticity of the sequence reads was assessed by evaluating C-to-T transitions near the 5'-ends of the reads, and by estimating contamination using mitochondrial DNA and – for male samples – X chromosome DNA. The program lcMLkin (Lipatov et al., 2015) was used to detect related individuals. Principal components analysis was performed on the dataset using smartpca tool (v.1600) of the EIGENSOFT (Patterson et al., 2006), with principal components calculated on present-day populations from Asia, Eurasia, Oceania and America. To estimate the ancestry components of ancient and present-day individuals, unsupervised clustering was performed using ADMIXTURE (Alexander et al., 2009).

To investigate the presence of *Yersinia pestis* DNA in the dataset, the DNA libraries were first mapped to a hybrid genome of *Y. pestis* and the closely related *Y. pseudotuberculosis* (Rasmussen et al., 2015), and reads mapping to *Y. pestis* CO92 were extracted, using MALT (Herbig et al., 2016). Edit distance distribution, length distribution, deamination patterns and breadth of coverage were used to assess the taxonomical origin of *Y. pestis*. Deamination patterns and deamination rate of *Y. pestis* and human DNA sequences were calculated using MapDamage (Jónsson et al., 2013).

Results and discussion

Chapter I

The results of our evaluation using simulated ancient DNA data showed that our workflow could detect present microbes in ancient samples. Compared to HOPS (Hübler et al., 2019), one of the most frequently used metagenomics workflows, aMeta showed a better sensitivity vs. specificity balance for both microbial detection and authentication (Fig. 4). We could also conclude that aMeta had a lower authentication error than HOPS. Further, we could report that by using a KrakenUniq-based selection of microbial candidates before inclusion in the MALT database, we dramatically reduced the resource usage of aMeta compared to metagenomic profiling with MALT alone.

When running the workflow on an authentic ancient sample that had previously shown presence of *Y. pestis* DNA, we could confirm that the workflow can detect microbial species in real ancient data.

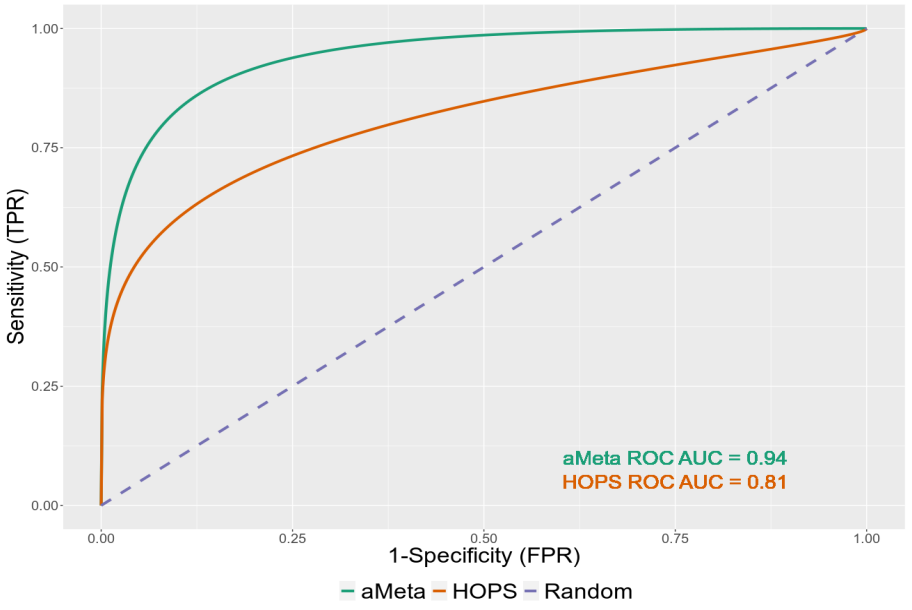


Figure 4. ROC-curve comparison of authentication scoring implemented in aMeta and HOPS. (**chapter I**)

With aMeta, we present a more precise as well as more resource efficient workflow than has previously been available in the field of metagenomics, and we believe that aMeta can become the new standard tool for classifying microorganisms in ancient material.

Chapter II

In **chapter II**, we detected the presence of several bacterial species, both commensal and potentially pathogenic, in the prehistoric human populations studied. We identified several species of microorganisms in the first screening with KrakenUniq, though after performing MALT and the authentication steps of the workflow, we could only confirm DNA belonging to bacterial species. Both viral and fungal DNA has been identified in ancient samples previously (G. Moore et al., 2020; Nielsen et al., 2021), and we therefore concluded that no fungi nor DNA viruses were present in the samples, at least not on a detectable level. Moreover, the human microbiome consists of over 99% bacteria (Qin et al., 2010), which can further explain why the majority of the DNA reads have bacterial origin. We can however not exclude that RNA viruses were present in the samples.

In the Mesolithic hunter-gatherers (SHG), the most prominent finding was *Neisseria meningitidis*, known to cause meningococcal disease, in one individual. This infection is commonly transmitted via saliva, for example through coughing, sneezing, and kissing. Common symptoms of meningitis include headache and fever, and though it is easily treated with the help of modern healthcare, the infection can be fatal. In one Neolithic hunter-gatherer (PWC) individual, we found another bacterium of the *Neisseria* genus, *N. gonorrhoeae*, which causes the sexually transmitted infection gonorrhoea. Although *N. gonorrhoeae* was only found in one individual, the effect on the individual and the population may have been substantial. Gonorrhoea can lead to sepsis, which is potentially fatal, but more commonly it can lead to infertility. Untreated gonorrhoea in mothers has also been connected to several foetal conditions, and the infection can be transmitted to the baby during delivery. Furthermore, considering the transmission of the infection it is likely that other individuals in the population were infected. *N. meningitidis* has been identified using aDNA from human samples from California, North America, from 1250–1640 CE (Eerkens et al., 2018), and both *N. meningitidis* and *N. gonnorrhoeae* have been identified in samples from Germany c. 950–1200 CE (Warinner et al., 2014). Our findings may thus be the oldest evidence of pathogenic *Neisseria* so far.

In FBC individuals, we identified several species of the genus *Yersinia*. In one of the individuals, Gok2, we found *Y. pestis*, corresponding to previous results (Rascovan et al., 2019). The sample is dated to 5,000 years BP and is

one of the earliest findings of *Y. pestis* to date. The bacterium is transmitted to humans from infected rodents via flea bites, however, the variant of the *ymt* gene that allows flea transmission was likely not acquired until around 1,500 years later (Rasmussen et al., 2015). The severeness of the disease may thus have been limited in the population.

In another individual, from Ansarve on Gotland, we identified another species from the same genus, *Y. enterocolitica*. *Y. enterocolitica* causes yersiniosis, a possibly lethal infection with symptoms including diarrhoea and fever. Yersinosis is transmitted by ingestion of contaminated food and water, although there are also signs of some transmission between humans as well as zoonotic transmission (Sabina et al., 2011).

In two individuals from a BAC context, excavated from Bergsgraven in Östergötland, Sweden, we found *Salmonella enterica*, the causing agent of the disease salmonellosis (Fig. 5). The symptoms of salmonellosis include stomach cramps, fever, and diarrhoea. *S. enterica* is typically transmitted to humans through ingestion of contaminated meat, eggs, or milk, though zoonotic transmissions can also occur. Interestingly, these two individuals were recovered from the same grave (Malmström et al., 2019) and since the osteological examination of the bones didn't show any signs of damage, this finding could imply the cause of death for these individuals. A third individual, an infant, was discovered after the initial excavation (Malmström et al., 2019), and it would be interesting to investigate in future studies whether *S. enterica* is present in these remains as well. DNA from *S. enterica* has previously been identified in Neolithic samples and been linked to an agricultural lifestyle (Key et al., 2020). This study has further shown that the oldest strains of the bacterium likely were host generalists and not adapted to humans (Key et al., 2020). Our findings could be integrated in studies using previously recovered data to gain further knowledge of the evolutionary history of *S. enterica*.

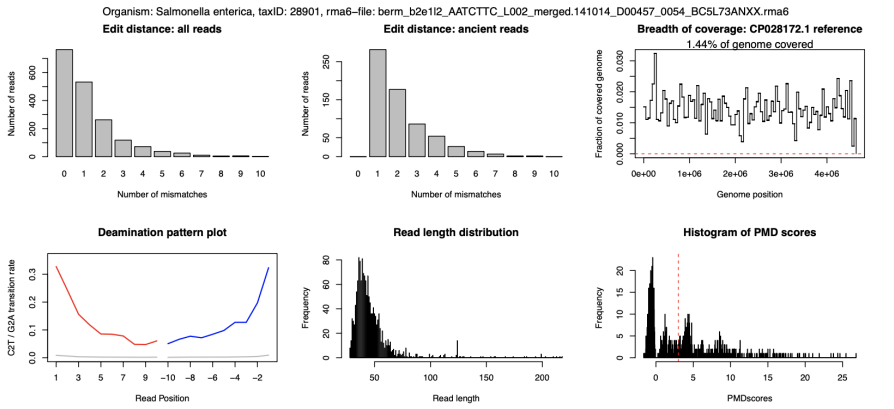


Figure 5. Edit distance, breadth of coverage, deamination plot and read length distribution of *Salmonella enterica* in individual ber001. (**chapter II**)

Chapter III

In **chapter III**, we applied the metagenomics workflow on a dataset of humans from present-day Mexico. The samples had been radiocarbon dated to different time periods, before and after the 15th century, when Europeans arrived in the Americas. We were able to distinguish two species with high authentication scores that were only present in post-contact individuals: *Parvimonas micra* and *Vibrio cholerae*. *P. micra* is strongly connected to periodontal disease and was found in three individuals from the late 18th century. The bacterium was prevalent in Europe at the same time (Willmann et al., 2018), and may have spread to the Americas through European expeditions.

The *V. cholerae* DNA, which was found in remains dated to the late 18th century, was further investigated, and a phylogeny including several globally collected strains of *V. cholerae* and the closely related *V. paracholerae* was built (Fig. 6). Cholera is typically caused by *V. cholerae* strains belonging to the O1 serogroup (Kaper et al., 1995; Ramamurthy et al., 2022), although strains belonging to the O139 serogroup also carry the cholera toxin (*ctx*) gene (Ramamurthy et al., 2022) and are considered epidemic (Bhadra et al., 1995). Moreover, non-O1, non-139 strains can cause infections with severe diarrhoea (Dutta et al., 2013; Engel et al., 2016; Onifade et al., 2011). The phylogenetic analysis placed the identified strain with environmental, non-choleric strains, collected from North America. Although the phylogeny implies that the strain is non-choleric, such strains have been isolated from the environment in epidemic settings previously (Kaper et al., 1995; Ramirez et al., 2021). It can therefore not be ruled out that the sample is connected to an epidemic outbreak of cholera in México City in the late 18th century. Another possibility is that the region of the genome containing the *ctx* gene is not covered by the sequence data due to post-mortem fragmentation and degradation of the DNA, and the strain could in fact belong to an epidemic serogroup.

The first global cholera pandemic emerged in India in 1817 (Cockburn & Cassanos, 1960), although historical sources describe the disease as early as 1503 (Moukassa et al., 2015). Our results, regardless of which serogroup the strain belongs to, represents the, to our knowledge, earliest finding of *V. cholerae* in the Americas, and imply a spread of *V. cholera* decades earlier than the first known pandemic. This obscure spread of the disease could be explained by the strain lacking the *ctx* gene, since small, non-epidemic outbreaks with mild symptoms would likely go mainly unnoticed.

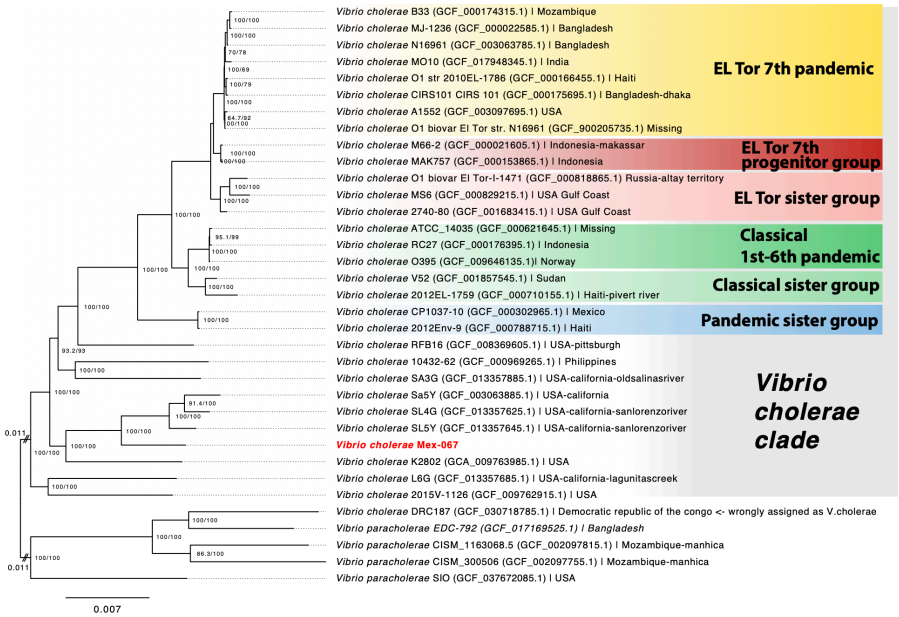


Figure 6. Phylogenetic tree of *Vibrio cholerae* strains. The strain identified in sample mex067 is shown in red text. (**chapter III**)

Chapter IV

In **chapter IV**, we used a human genomic analysis to show that the oldest studied individual, a 16,900-year-old female from the Khaiyrgas Cave, had a genetic affinity towards a present-day north-Siberian population, as well as to indigenous populations of Native America. The results further indicate that the Khaiyrgas-1 individual represents the first major genetic shift throughout the investigated region after the LGM. The genetic legacy of the group Khaiyrgas-1 represents can still be seen in populations in the area ~6,000 years later.

In two individuals, we identified *Y. pestis* specific DNA. The results from one of the individuals, from Cis-Baikal and dated to 4,400 years BP, showed the expected pattern of C-to-T transmission close to the 5'-end of the DNA read (Fig. 7), and the coverage of the *Y. pestis* genome was 3.66%. In the other individual, a 3,800-year-old individual from Yakutia, we found *Y. pestis* DNA with 1.65% genome coverage, representing the most northeastern ancient finding of the bacterium. Our results are in accordance with previous findings of *Y. pestis* in human remains from the Baikal region dated to circa 4,500 years BP (Yu et al., 2020). Although further research on additional data is necessary, our results suggest that a plague outbreak may have been affecting the population of Siberia during this time period.

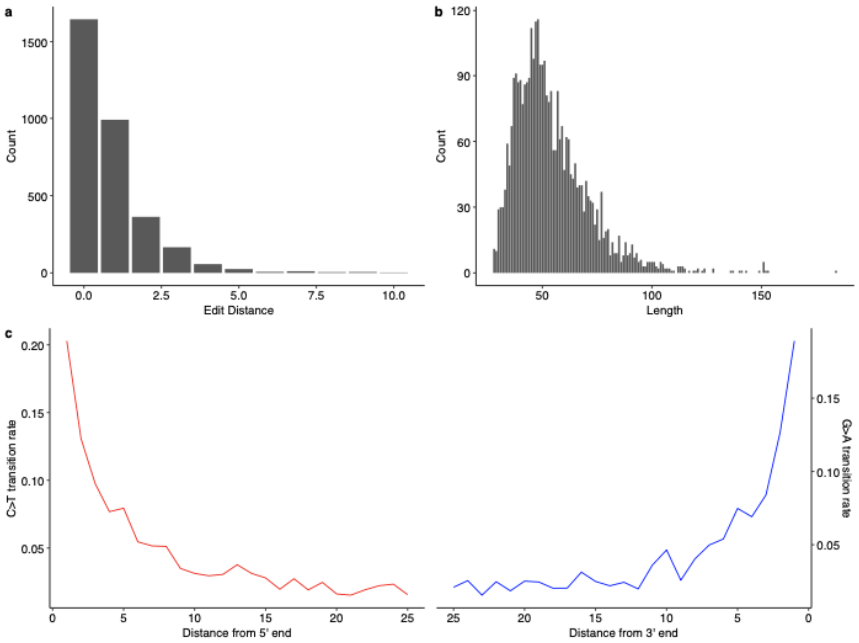


Figure 7. The edit distance (a), length distribution (b) and deamination plot (c) for reads aligned to *Y. pestis* in individual Anosovo-1. (**chapter IV, fig. S24**)

Concluding remarks and future work

With this thesis, I have demonstrated how a novel metagenomics workflow can contribute to the research on ancient populations as a time and resource efficient complement to demographic studies, making it possible to easily see additional aspects of the life history of a population. Further, I have shown an example of how the reference database used for metagenomic species assignment can be adjusted according to research objective, and in that way further streamline the analysis.

In **chapter I**, I presented an ancient metagenomics workflow that combines available software in a time and resource efficient manner. The interest in human-associated microbes has increased in the past years, and the effect of host microbiome on the response to diseases has been shown in multiple studies, for example the response to immunotherapy in cancer patients (Gopalakrishnan et al., 2018). Ancient microbes from prehistoric populations will likely emerge as an important complement to other types of information on historic

and prehistoric health. With more thorough studies on the microbiome of prehistoric humans, more details about their resilience to diseases can be investigated. In addition to diagnosing prehistoric humans with various diseases, integrating ancient human and microbial DNA data may illuminate the process of host-pathogen interactions and co-evolution (Key et al., 2017). It is my view that the aMeta workflow can be of considerable importance in the study of ancient diseases and the evolution of pathogens, not only in humans but in other animals as well.

Moreover, a de novo assembly module is currently being tested, and we plan to add it to the aMeta workflow in a future release. This way, aMeta will leverage all the power of classification, alignment and de novo assembly that can be used complementary to each other and provide a more informative overview of microbial composition in ancient samples. Another planned extension is to develop the workflow for working with sedimentary ancient DNA, by fine-tuning the workflow to deal with large eukaryotic reference genomes.

A weakness with the workflow when working with ancient pathogens is that it is developed for DNA sequence data and thus cannot be used to identify RNA viruses. Since more than 70% of viruses have RNA as genetic material (Domingo, 1997), there is an obvious risk that pathogenic viruses in a sample are not detected using aMeta. A future prospect could be to develop the workflow for work on RNA sequence data.

Another important aspect to consider in future research is the treatment of samples before sequencing. UV-radiation and predigestion steps remove contamination from the surface of bone fragments but may also inadvertently result in removing ancient bacteria and other pathogens. Similarly, ancient DNA samples are commonly treated with USER enzyme to remove uracil residues. This can lead to ancient microorganisms erroneously being classified as modern contaminants, since the deamination pattern close to the 5'-end of the fragment is masked. For DNA samples extracted from teeth this should be a minor problem, as the DNA in the pulp of the tooth shouldn't be affected by surface treatments. However, when analysing bones with lesions potentially caused by disease, treatments like these should be made with care.

In **chapter II**, I demonstrated how the metagenomics workflow can be applied to investigate diseases in a prehistoric human population. We could demonstrate the presence of several bacterial species that can cause severe infections. Some of these represent the earliest findings of the respective species, and some have possibly been the cause of death for the individuals in which the pathogens were detected. We have thus shown the potential of a high-precision, resource efficient metagenomics classifier, and how it can be incorporated in future demographic studies to give a broader perspective of the life history of human populations.

In **chapter III**, we used the workflow to investigate the diseases in a historic human population from Mexico. Here, I further showed how the workflow can be used to obtain an accurate overview of the microbial content of a dataset, and that the results can subsequently be used to perform more in-depth analyses of the phylogenetic relationships of the identified species. From these results, we have showcased the possibilities of studying the evolutionary history of bacterial species, as well as the spread of pathogens, using aDNA.

In **chapter IV**, we used metagenomic tools to investigate the presence of DNA from one particular bacterial species in an ancient dataset, demonstrating how metagenomic analysis can be performed quickly as a complement to demographic studies, by customizing the reference database to the species of interest.

References

Adler, C. J., Dobney, K., Weyrich, L. S., Kaidonis, J., Walker, A. W., Haak, W., Bradshaw, C. J. A., Townsend, G., Sołtysiak, A., Alt, K. W., Parkhill, J., & Cooper, A. (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. *Nature Genetics*, *45*(4), 450–455. https://doi.org/10.1038/ng.2536

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Armelagos, G. J., & Dewey, J. R. (1970). Evolutionary Response to Human Infectious Diseases. *BioScience*, *20*(5), 271–275. https://doi.org/10.2307/1295204

Benson, D. A., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *42*(Database issue), D32.

Bhadra, R. K., Roychoudhury, S., Banerjee, R. K., Kar, S., Majumdar, R., Sengupta, S., Chatterjee, S., Khetawat, G., & Das, J. (1995). Cholera toxin (CTX) genetic element in Vibrio cholerae O139. *Microbiology*, *141*(8), 1977–1983. https://doi.org/10.1099/13500872-141-8-1977

Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S. A., Bryant, J. M., Harris, S. R., Schuenemann, V. J., Campbell, T. J., Majander, K., Wilbur, A. K., Guichon, R. A., Wolfe Steadman, D. L., Cook, D. C., Niemann, S., Behr, M. A., Zumarraga, M., … Krause, J. (2014). Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, *514*(7523), 494–497. https://doi.org/10.1038/nature13591

Bos, K. I., Schuenemann, V. J., Golding, G. B., Burbano, H. A., Waglechner, N., Coombes, B. K., McPhee, J. B., DeWitte, S. N., Meyer, M., Schmedes, S., Wood, J., Earn, D. J. D., Herring, D. A., Bauer, P., Poinar, H. N., & Krause, J. (2011). A draft genome of Yersinia pestis from victims of the Black Death. *Nature*, *478*(7370), 506–510. https://doi.org/10.1038/nature10549

Breitwieser, F. P., Baker, D. N., & Salzberg, S. L. (2018). KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biology*, *19*(1), 198. https://doi.org/10.1186/s13059-018-1568-0

Briggs, A. W., Stenzel, U., Johnson, P. L. F., Green, R. E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M. T., Lachmann, M., & Pääbo, S. (2007).

Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, *104*(37), 14616–14621. https://doi.org/10.1073/pnas.0704665104

Clarke, N. G., Carey, S. E., Srikandi, W., Hirsch, R. S., & Leppard, P. I. (1986). Periodontal disease in ancient populations. *American Journal of Physical Anthropology*, *71*(2), 173–183. https://doi.org/10.1002/ajpa.1330710205

Cockburn, T. A., & Cassanos, J. G. (1960). Epidemiology of endemic cholera. *Public Health Reports*, *75*(9), 791.

Cooper, D. B. (1965). Epidemic Disease in Mexico City, 1761–1813. In *An Administrative, Social, and Medical Study* (pp. 86–156). University of Texas Press. https://doi.org/10.7560/732285-008

Coutinho, A., Günther, T., Munters, A. R., Svensson, E. M., Götherström, A., Storå, J., Malmström, H., & Jakobsson, M. (2020). The Neolithic Pitted Ware culture foragers were culturally but not genetically influenced by the Battle Axe culture herders. *American Journal of Physical Anthropology*, *172*(4), 638–649. https://doi.org/10.1002/ajpa.24079

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., & Arsuaga, J.-L. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763.

Davidson, P. T., & Horowitz, I. (1970). Skeletal tuberculosis: A review with patient presentations and discussion. *The American Journal of Medicine*, *48*(1), 77–84. https://doi.org/10.1016/0002-9343(70)90101-4

Derevianko, A. P., Kuzmin, Y. V., Burr, G. S., Jull, A. J. T., & Kim, J. C. (2004). AMS 14C age of the earliest pottery from the Russian Far East: 1996–2002 results. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, *223–224*, 735–739. https://doi.org/10.1016/j.nimb.2004.04.136

Domingo, E. (1997). Rapid Evolution of Viral RNA Genomes12. *The Journal of Nutrition*, *127*(5), 958S-961S. https://doi.org/10.1093/jn/127.5.958S

Dutta, D., Chowdhury, G., Pazhani, G. P., Guin, S., Dutta, S., Ghosh, S., Rajendran, K., Nandy, R. K., Mukhopadhyay, A. K., Bhattacharya, M. K., Mitra, U., Takeda, Y., Nair, G. B., & Ramamurthy, T. (2013). Vibrio cholerae Non-O1, Non-O139 Serogroups and Cholera-like Diarrhea, Kolkata, India. *Emerging Infectious Diseases*, *19*(3), 464. https://doi.org/10.3201/eid1903.121156

Eerkens, J. W., Nichols, R. V., Murray, G. G. R., Perez, K., Murga, E., Kaijankoski, P., Rosenthal, J. S., Engbring, L., & Shapiro, B. (2018). A probable prehistoric case of meningococcal disease from San Francisco Bay: Next generation sequencing of Neisseria meningitidis from dental calculus and osteological evidence. *International Journal of Paleopathology*, *22*, 173–180. https://doi.org/10.1016/j.ijpp.2018.05.001

Engel, M. F., Muijsken, M. A., Mooi-Kokenberg, E., Kuijper, E. J., & Westerloo, D. J. van. (2016). Vibrio cholerae non-O1 bacteraemia: Description of three cases in the Netherlands and a literature review. *Eurosurveillance*, *21*(15), 30197. https://doi.org/10.2807/1560-7917.ES.2016.21.15.30197

Gernaey, A. M., Minnikin, D. E., Copley, M. S., Dixon, R. A., Middleton, J. C., & Roberts, C. A. (2001). Mycolic acids and ancient DNA confirm an osteological diagnosis of tuberculosis. *Tuberculosis*, *81*(4), 259–265. https://doi.org/10.1054/tube.2001.0295

Gilbert, M. T. P., Binladen, J., Miller, W., Wiuf, C., Willerslev, E., Poinar, H., Carlson, J. E., Leebens-Mack, J. H., & Schuster, S. C. (2007). Recharacterization of ancient DNA miscoding lesions: Insights in the era of sequencing-by-synthesis. *Nucleic Acids Research*, *35*(1), 1–10. https://doi.org/10.1093/nar/gkl483

Goebel, T., Waters, M. R., & O'Rourke, D. H. (2008). The Late Pleistocene Dispersal of Modern Humans in the Americas. *Science*, *319*(5869), 1497–1502. https://doi.org/10.1126/science.1153569

Günther, T., Malmström, H., Svensson, E. M., Omrak, A., Sánchez-Quinto, F., Kılınç, G. M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., Munters, A. R., Coutinho, A., Simões, L. G., Vicente, M., Sjölander, A., Jansen Sellevold, B., Jørgensen, R., Claes, P., Shriver, M. D., … Jakobsson, M. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biology*, *16*(1). https://doi.org/10.1371/journal.pbio.2003703

Guzmán-Solís, A. A., Villa-Islas, V., Bravo-López, M. J., Sandoval-Velasco, M., Wesp, J. K., Gómez-Valdés, J. A., Moreno-Cabrera, M. de la L., Meraz, A., Solís-Pichardo, G., Schaaf, P., TenOever, B. R., Blanco-Melo, D., & Ávila Arcos, M. C. (2021). Ancient viral genomes reveal introduction of human pathogenic viruses into Mexico during the transatlantic slave trade. *eLife*, *10*, e68612. https://doi.org/10.7554/eLife.68612

Haller, M., Callan, K., Susat, J., Flux, A. L., Immel, A., Franke, A., Herbig, A., Krause, J., Kupczok, A., Fouquet, G., Hummel, S., Rieger, D., Nebel, A., & Krause-Kyora, B. (2021). Mass burial genomics reveals outbreak of enteric paratyphoid fever in the Late Medieval trade city Lübeck. *iScience*, *24*(5), 102419. https://doi.org/10.1016/j.isci.2021.102419

Handt, O., Höss, M., Krings, M., & Pääbo, S. (1994). Ancient DNA: Methodological challenges. *Experientia*, *50*(6), 524–529. https://doi.org/10.1007/BF01921720

Hawass, Z. (2010). Ancestry and Pathology in King Tutankhamun's Family. *JAMA*, *303*(7), 638. https://doi.org/10.1001/jama.2010.121

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

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

Howell, K. W. (2002). In the Wake of Conquest: A global perspective on the depopulation of indigenous peoples of Latin America. *Diálogos Latinoamericanos*, *5*, 58–72.

Hübler, R., Key, F. M., Warinner, C., Bos, K. I., Krause, J., & Herbig, A. (2019). HOPS: Automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biology*, *20*(1), 280. https://doi.org/10.1186/s13059-019-1903-0

Jaeger, L. H., Souza, S. M. de, Dias, O. F., & Iñiguez, A. M. (2013). Mycobacterium tuberculosis Complex in Remains of 18th–19th Century Slaves, Brazil. *Emerging Infectious Diseases*, *19*(5), 837. https://doi.org/10.3201/eid1905.120193

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kaper, J. B., J G Morris, J., & Levine, M. M. (1995). Cholera. *Clinical Microbiology Reviews*, *8*(1), 48. https://doi.org/10.1128/cmr.8.1.48

Key, F. M., Posth, C., Esquivel-Gomez, L. R., Hübler, R., Spyrou, M. A., Neumann, G. U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., Lankapalli, A. K., Vågene, Å. J., Meyer, M., Nagel, S., Tukhbatova, R., Khokhlov, A., Chizhevsky, A., Hansen, S., Belinsky, A. B., … Krause, J. (2020). Emergence of human-adapted Salmonella enterica is linked to the Neolithization process. *Nature Ecology & Evolution*, *4*(3), 324–333. https://doi.org/10.1038/s41559-020-1106-9

Kılınç, G. M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A. M., Bıçakçı, E., Baird, D., Dönertaş, H. M., Ghalichi, A., Yaka, R., Koptekin, D., Açan, S. C., Parvizi, P., Krzewińska, M., Daskalaki, E. A., Yüncü, E., Dağtaş, N. D., Fairbairn, A., Pearson, J., … Götherström, A. (2016). The Demographic Development of the First Farmers in Anatolia. *Current Biology*, *26*(19), 2659–2666. https://doi.org/10.1016/j.cub.2016.07.057

Kjær, K. H., Winther Pedersen, M., De Sanctis, B., De Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavić, S., Ruter, A. H., & Schmidt, A. M. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature*, *612*(7939), 283–291.

Knutsson, H., & Knutsson, K. (2003). Stone age transitions. Neolithisation in central Scandinavia. *Documenta Praehistorica*, *30*, 48–78. https://doi.org/10.4312/dp.30.2

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

Larsson, Å. M. (2008). Taking out the trash: On excavating settlements in general, and houses of the Battle Axe Culture in particular. *Current Swedish Archaeology*, 111–136.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P. H., Schraiber, J. G., Castellano, S., Lipson, M., Berger, B., Economou, C., Bollongino, R., Fu, Q., Bos, K. I., Nordenfelt, S., Li, H., De Filippo, C., Prüfer, K., … Krause, J. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, *513*(7518), 409–413. https://doi.org/10.1038/nature13673

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

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

Malmström, H., Günther, T., Svensson, E. M., Juras, A., Fraser, M., Munters, A. R., Pospieszny, Ł., Tõrv, M., Lindström, J., Götherström, A., Storå, J., & Jakobsson, M. (2019). The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1912), 20191528. https://doi.org/10.1098/rspb.2019.1528

Margaryan, A., Hansen, H. B., Rasmussen, S., Sikora, M., Moiseyev, V., Khoklov, A., Epimakhov, A., Yepiskoposyan, L., Kriiska, A., Varul, L., Saag, L., Lynnerup, N., Willerslev, E., & Allentoft, M. E. (2018). Ancient pathogen DNA in human teeth and petrous bones. *Ecology and Evolution*, *8*(6), 3534–3542. https://doi.org/10.1002/ece3.3924

Meng, Y., Zhang, H.-Q., Pan, F., He, Z.-D., Shao, J.-L., & Ding, Y. (2011). Prevalence of dental caries and tooth wear in a Neolithic population (6700–5600 years BP) from northern China. *Archives of Oral Biology*, *56*(11), 1424–1435. https://doi.org/10.1016/j.archoralbio.2011.04.003

Merbs, C. F. (1992). A new world of infectious disease. *American Journal of Physical Anthropology*, *35*(S15), 3–42. https://doi.org/10.1002/ajpa.1330350603

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb-prot5448.

Minkin, I., & Medvedev, P. (2020). Scalable multiple whole-genome alignment and locally collinear block construction with SibeliaZ. *Nature Communications*, *11*(1), 6327. https://doi.org/10.1038/s41467-020-19777-8

Moore, G., Tessler, M., Cunningham, S. W., Betancourt, J., & Harbert, R. (2020). Paleo-metagenomics of North American fossil packrat middens: Past biodiversity revealed by ancient DNA. *Ecology and Evolution*, *10*(5), 2530–2544. https://doi.org/10.1002/ece3.6082

Moore, W. J., & Corbett, E. (1973). The Distribution of Dental Caries in Ancient British Populations. *Caries Research*, *7*(2), 139–153. https://doi.org/10.1159/000259838

Moukassa, D., Pointe-Noire, C., & Ibara, J. (2015). Natural History of Cholera. *Sleep Medicine: A Comprehensive Guide to Its Development, Clinical Milestones, and Advances in Treatment*, 143.

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Nielsen, S. H., Van Dorp, L., Houldcroft, C. J., Pedersen, A. G., Allentoft, M. E., Vinner, L., Margaryan, A., Pavlova, E., Chasnyk, V., Nikolskiy, P., Pitulko, V., Pimenoff, V. N., Balloux, F., & Sikora, M. (2021). *31,600-year-old human virus genomes support a Pleistocene origin for common childhood infections*. Evolutionary Biology. https://doi.org/10.1101/2021.06.28.450199

Onifade, T. M., Hutchinson, R., Zile, K. V., Bodager, D., Baker, R., & Blackmore, C. (2011). Toxin producing Vibrio cholerae O75 outbreak, United States, March to April 2011. *Eurosurveillance*, *16*(20), 19870. https://doi.org/10.2807/ese.16.20.19870-en

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., Johnson, P. L. F., Fumagalli, M., Vilstrup, J. T., Raghavan, M., Korneliussen, T., Malaspinas, A.-S., Vogt, J., Szklarczyk, D., Kelstrup, C. D., … Willerslev, E. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*(7456), 74–78. https://doi.org/10.1038/nature12323

Pääbo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA. *Nature*, *314*(6012), 644–645. https://doi.org/10.1038/314644a0

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

Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLOS Genetics*, *2*(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Pochon, Z., Bergfeldt, N., Kırdök, E., Vicente, M., Naidoo, T., Van Der Valk, T., Altınışık, N. E., Krzewińska, M., Dalén, L., Götherström, A., Mirabello, C., Unneberg, P., & Oskolkov, N. (2023). aMeta: An accurate and memory-efficient ancient metagenomic profiling workflow. *Genome Biology*, *24*(1), 242. https://doi.org/10.1186/s13059-023-03083-9

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D. R., Li, J., Xu, J., Li, S., Li, D.,

Cao, J., Wang, B., Liang, H., Zheng, H., … Zoetendal, E. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*(7285), 59–65. https://doi.org/10.1038/nature08821

Ramamurthy, T., Pragasam, A. K., Taylor-Brown, A., Will, R. C., Vasudevan, K., Das, B., Srivastava, S. K., Chowdhury, G., Mukhopadhyay, A. K., Dutta, S., Veeraraghavan, B., Thomson, N. R., Sharma, N. C., Nair, G. B., Takeda, Y., Ghosh, A., Dougan, G., & Mutreja, A. (2022). Vibrio cholerae O139 genomes provide a clue to why it may have failed to usher in the eighth cholera pandemic. *Nature Communications*, *13*(1), 3864. https://doi.org/10.1038/s41467-022-31391-4

Ramirez, D. A., Saka, H. A., & Nores, R. (2021). Detection of Vibrio cholerae aDNA in human burials from the fifth cholera pandemic in Argentina (1886–1887 AD). *International Journal of Paleopathology*, *32*, 74–79. https://doi.org/10.1016/j.ijpp.2020.12.004

Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., & Rasmussen, S. (2019). Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. *Cell*, *176*(1–2), 295-305.e10. https://doi.org/10.1016/j.cell.2018.11.005

Rasmussen, S., Allentoft, M. E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G., Pedersen, A. G., Schubert, M., Van Dam, A., Kapel, C. M. O., Nielsen, H. B., Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M. V., Gnuni, A., Kriiska, A., Lasak, I., Metspalu, M., … Willerslev, E. (2015). Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. *Cell*, *163*(3), 571–582. https://doi.org/10.1016/j.cell.2015.10.009

Sabina, Y., Rahman, A., Ray, R. C., & Montet, D. (2011). Yersinia enterocolitica: Mode of Transmission, Molecular Insights of Virulence, and Pathogenesis of Infection. *Journal of Pathogens*, *2011*, 1–10. https://doi.org/10.4061/2011/429069

Sánchez-Quinto, F., Malmström, H., Fraser, M., Girdland-Flink, L., Svensson, E. M., Simões, L. G., George, R., Hollfelder, N., Burenhult, G., Noble, G., Britton, K., Talamo, S., Curtis, N., Brzobohata, H., Sumberova, R., Götherström, A., Storå, J., & Jakobsson, M. (2019). Megalithic tombs in western and northern Neolithic Europe were linked to a kindred society. *Proceedings of the National Academy of Sciences*, *116*(19), 9469–9474. https://doi.org/10.1073/pnas.1818037116

Sikora, M., Pitulko, V. V., Sousa, V. C., Allentoft, M. E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., Renaud, G., Yang, M. A., Fu, Q., Dupanloup, I., Giampoudakis, K., Nogués-Bravo, D., Rahbek, C., Kroonen, G., Peyrot, M., McColl, H., … Willerslev, E. (2019). The population history of northeastern Siberia since the Pleistocene. *Nature*, *570*(7760), 182–188. https://doi.org/10.1038/s41586-019-1279-z

Skoglund, P., Malmström, H., Omrak, A., Raghavan, M., Valdiosera, C., Günther, T., Hall, P., Tambets, K., Parik, J., Sjögren, K.-G., Apel, J., Willerslev, E., Storå,

J., Götherström, A., & Jakobsson, M. (2014). Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science*, *344*(6185), 747–750. https://doi.org/10.1126/science.1253448

Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M. T. P., Götherström, A., & Jakobsson, M. (2012). Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science*, *336*(6080), 466–469. https://doi.org/10.1126/science.1216304

Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(6), 2229–2234. https://doi.org/10.1073/pnas.1318934111

Stiller, M., Green, R. E., Ronan, M., Simons, J. F., Du, L., He, W., Egholm, M., Rothberg, J. M., Keates, S. G., Ovodov, N. D., Antipina, E. E., Baryshnikov, G. F., Kuzmin, Y. V., Vasilevski, A. A., Wuenschell, G. E., Termini, J., Hofreiter, M., Jaenicke-Després, V., & Pääbo, S. (2006). Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences*, *103*(37), 13578–13584. https://doi.org/10.1073/pnas.0605327103

Vågene, Å. J., Herbig, A., Campana, M. G., Robles García, N. M., Warinner, C., Sabin, S., Spyrou, M. A., Andrades Valtueña, A., Huson, D., Tuross, N., Bos, K. I., & Krause, J. (2018). Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution*, *2*(3), 520–528. https://doi.org/10.1038/s41559-017-0446-6

Warinner, C., Rodrigues, J. F. M., Vyas, R., Trachsel, C., Shved, N., Grossmann, J., Radini, A., Hancock, Y., Tito, R. Y., Fiddyment, S., Speller, C., Hendy, J., Charlton, S., Luder, H. U., Salazar-García, D. C., Eppler, E., Seiler, R., Hansen, L. H., Castruita, J. A. S., … Cappellini, E. (2014). Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics*, *46*(4), 336–344. https://doi.org/10.1038/ng.2906

Weisdorf, J. L. (2005). From Foraging To Farming: Explaining The Neolithic Revolution. *Journal of Economic Surveys*, *19*(4), 561–586. https://doi.org/10.1111/j.0950-0804.2005.00259.x

Wibowo, M. C., Yang, Z., Borry, M., Hübner, A., Huang, K. D., Tierney, B. T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., Martínez-Hernández, A., Luber, J. M., Kirstahler, P., Blohm, T., Smiley, F. E., Arnold, R., Ballal, S. A., Pamp, S. J., Russ, J., … Kostic, A. D. (2021). Reconstruction of ancient microbial genomes from the human gut. *Nature*, *594*(7862), 234–239. https://doi.org/10.1038/s41586-021-03532-0

Willmann, C., Mata, X., Hanghoej, K., Tonasso, L., Tisseyre, L., Jeziorski, C., Cabot, E., Chevet, P., Crubézy, E., Orlando, L., Esclassan, R., & Thèves, C. (2018).

Oral health status in historic population: Macroscopic and metagenomic evidence. *PLOS ONE*, *13*(5), e0196482. https://doi.org/10.1371/journal.pone.0196482

Wong, E. H. M., Khrunin, A., Nichols, L., Pushkarev, D., Khokhrin, D., Verbenko, D., Evgrafov, O., Knowles, J., Novembre, J., Limborska, S., & Valouev, A. (2017). Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Research*, *27*(1), 1–14. https://doi.org/10.1101/gr.202945.115

Yu, H., Spyrou, M. A., Karapetian, M., Shnaider, S., Radzevičiūtė, R., Nägele, K., Neumann, G. U., Penske, S., Zech, J., Lucas, M., LeRoux, P., Roberts, P., Pavlenok, G., Buzhilova, A., Posth, C., Jeong, C., & Krause, J. (2020). Paleolithic to Bronze Age Siberians Reveal Connections with First Americans and across Eurasia. *Cell*, *181*(6), 1232-1245.e20. https://doi.org/10.1016/j.cell.2020.04.037

Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M. J., Langridge, G., Fotakis, A. K., Nair, S., Stenøien, H. K., Hamre, S. S., Casjens, S., Christophersen, A., Quince, C., Thomson, N. R., Weill, F.-X., Ho, S. Y. W., Gilbert, M. T. P., & Achtman, M. (2018). Pan-genome Analysis of Ancient and Modern Salmonella enterica Demonstrates Genomic Stability of the Invasive Para C Lineage for Millennia. *Current Biology*, *28*(15), 2420-2428.e10. https://doi.org/10.1016/j.cub.2018.05.058

Svensk sammanfattning

I takt med att metoderna för DNA-extrahering och -sekvensering har utvecklats de senaste åren har det blivit möjligt att studera förhistoriska mikroorganismer i större detalj än tidigare. Evolution och spridning av sjukdomar är ett viktigt ämne att forska på, inte minst med tanke på de senaste årens pandemier. Ett problem inom metagenomiken har dock varit felaktiga identifieringar av mikroorganismer i form av bland annat falska positiva resultat.

I **kapitel I** av den här avhandlingen presenteras aMeta, ett strömlinjeformat verktyg som utvecklats för att identifiera förhistoriska organismer och som utformats specifikt för att minimera antalet falska positiva identifikationer. Med hjälp av simulerat och autentiskt förhistoriskt DNA-data har vi utvärderat aMeta i relation till de verktyg som använts i tidigare studier. Vi kunde påvisa hur vårt verktyg presterar överlägset jämfört med andra verktyg, dels genom känslighet när det kommer till att identifiera arter och att autentisera resultaten, dels i att effektivisera resursanvändningen.

I **kapitel II** applicerades verktyget på individer från fyra olika stenålderskulturer i Skandinavien. I dessa prover kunde flera bakteriearter identifieras. Hos två individer från stridsyxekulturen identifierade vi *Salmonella enterica*, som vi med tanke på avsaknaden av fysiska skador vid de osteologiska undersökningarna tror kan ha varit dödsorsaken för dessa individer. Flera arter ur släktet *Yersinia* upptäcktes också i proverna och det är möjligt att dessa bakterier lättare kunde spridas i populationen i och med de förändringar i livsstil som jordbruksrevolutionen innebar. Även två arter ur släktet *Neisseria* identifierades, vilka representerar de hittills tidigaste fynden av bakterier från detta släkte.

Även i **kapitel III** identifierade vi med hjälp av aMeta förekomsten av sjukdomsorsakande bakterier, denna gång i ett DNA-dataset med prover från mänskliga kvarlevor insamlade i Mexiko. Vi identifierade DNA från bakterien *Parvimonas micra*, som är starkt kopplad till inflammation som leder till tandlossning, i flera indivier från sent 1700-tal. I en individ från 1700-talet hittade vi DNA från *Vibrio cholerae*, bakterien som orsakar sjukdomen kolera, vilket representerar det hittills tidigaste fyndet av denna bakterie. Vi undersökte även släktskapet mellan det identifierade DNA:t och andra stammar av bakterien och kunde dra slutsatsen att vårt fynd troligen inte orsakat någon epidemi, utan istället tillhör en stam som är vanligen förekommande i naturen.

I **kapitel IV** presenterade vi genomiska data från 40 individer i nordöstra Asien, mellan 16900 och 550 år gamla. Vi redovisade ursprunget av och släktskapet mellan individerna och undersökte vidare om det fanns DNA från sjukdomsorsakande bakterier i datat. Vi hittade bakterien *Yersinia pestis*, som orsakar sjukdomen pest, i två individer, 3800 respektive 4400 år gamla, från olika områden i Sibirien. Våra resultat överensstämmer med tidigare studier som identifierat bakterien i närliggande områden, vilket väcker frågan om en pestepidemi härjade i Sibirien och påverkade människorna som levde där för 4000 år sedan.

Acknowledgements

Accomplishing anything in science is impossible without people around you that support and help you. I want to direct my greatest thanks to the following people:

Anders Götherström, my supervisor. You've guided me through the past years research-wise, but more importantly I've had your full support during the ups and downs that are part of life outside work and that means a lot more than I think you realise. You've helped me see my own potential and strengths as a researcher, for which I am very grateful.

Love Dalén, a great mentor and (co-)supervisor, first during my undergrad studies and lately during my research studies. I am ever so grateful for getting the opportunity to be a part of your research group and becoming a better researcher with your support, and I've truly enjoyed bossing you around during the choir's Lucia performances.

If you'd have told me twelve years ago, when I took a genetics course as a first-year biology student at Högskolan på Gotland, that I'd be defending my PhD thesis with the support of those strange professors who talked about cloning mammoths, I probably wouldn't have believed it, but then as now the two of you have been an inspiration.

Lars Arvestad, my co-supervisor, thank you for keeping an interest in my work although the planned paired PhD project didn't go as intended.

Karin Norén and Per Ericson, my follow-up group at the department of Zoology, thank you for bringing up interesting thoughts and helping me see the bigger picture of my projects, as well as encouraging me to be a lot more confident in my work.

Past and current members of the Götherström group – Emrah, Zoé, Mário, Reyhan, Maja, Ricardo, TJ – you have helped me with all kinds of problems during the past years and I've enjoyed working with you every day.

CPG staff of past and present – especially Johanna, Edana, Marianne, Johannes, Isabelle, Ernst, Alicia, Camilo, David, Julia, Erik, Vendela, Petter, Dave, Ioana, Pete, Benjamin, Anna. You've all helped me a lot, but mostly I will miss all the lunches, fika, games and pub nights that made the workplace a truly great environment. Julia – thank you for fixing my boring front cover!

Nikolay Oskolkov, Claudio Mirabello, and Per Unneberg at NBIS – thank you for teaching me so much and for putting more time than was required of you into helping me. This thesis couldn't have been written without you!

Other collaborators: Gülşah, Björn, Ezgi, Mehmet, Mattias, Jan, Charlotte to mention a few – you have all been so generous in sharing your knowledge, thank you!

Anders Andersson was the opponent of my licentiate thesis and introduced a lot of new perspectives for which I am very grateful!

Gunilla, Sarah and Maike at The County Administrative Board in Norrbotten – thank you for bearing with me (see what I did there?) and making it possible for me to be in two places at the same time during the past year and a half.

I want to thank my father, Kjell, for asking a lot of annoying questions about how my research is going, and my mother, Susanna, for not asking as many annoying questions about how my research is going. Thank you for hosting and feeding me and everything else. Vendela, Anton and the niblings Mika, Ottillia and Emelia for always being there to cheer me up and reminding me of the importance of life outside the workplace. Sören, Johanna and Birgitta for being a part of the bigger little family and all the support, always. Linus, thank you for taking care of everything else in our lives and for, together with Lucas and Khepri, making my life a lot more entertaining.

Friends – you've kept me company, kept me fed, been great study partners, and listened to me complain about work, life and everything in between. Thank you, Catharina, Amanda, Camilla, Cecilia, Alice, Anna, Susanna, Linnea.

My warmest thanks go to Emma, Ina and Mia for creating a community that kept me company during lockdown and helped me find the best people I didn't know existed. Jolina, Emina, Line, Linus, Moa, Maria, Alexandra, Ulrica, Sorina, Björn, Hanna, Elin, Christofer, Daniel, Daniel, Matilda and all other FF-friends – you made PhD studies during a pandemic bearable and I'm so grateful to have you in my life.