Big Data, Small Microbes

Genomic analysis of the plague bacterium Yersinia pestis

A thesis submitted to the Department of Anthropology and the school of graduate studies of McMaster University in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

The web, pdf , and official versions of this thesis were generated from ktmeaton/obsidian-public@f0305af8 on January 28, 2022.

Author

- **Katherine Eaton (B.A. Hons)**
  0000-0001-6862-7756 · ktmeaton · eatonk3@mcmaster.ca
  McMaster Ancient DNA Center; Department of Anthropology, McMaster University

Supervisor

- **Hendrik Poinar (PhD)**
  0000-0002-0314-4160 · poinarh@mcmaster.ca
  McMaster Ancient DNA Center; Department of Anthropology, McMaster University

Lay Abstract

The Plague is a disease that has profoundly impacted human history and is responsible for some of the most fatal pandemics ever recorded. It may surprise many to know that this disease is not a bygone of a past era, but in fact is still present in many regions of the world. Although researchers have been studying plague for hundreds of years, there are many aspects of its epidemiology that are enigmatic. In this thesis, I focus on how DNA from the plague bacterium can be used to estimate where and when this disease appeared in the past. To do so, I reconstruct the evolutionary relationships between modern and ancient strains of plague, using publicly available data and new DNA sequences retrieved from the skeletal remains of plague victims in Denmark. This work offers a new methodological framework for large-scale genetic analysis, provides a critique on what questions DNA evidence can and cannot answer, and expands our knowledge of the global diversity of plague.

Abstract

Pandemics of plague have reemerged multiple times throughout human history with tremendous mortality and extensive geographic spread. The First Pandemic (6^{th} - 8^{th} century) devastated the Mediterranean world, the Second Pandemic (14^{th} - 19^{th} century) swept across much of Afro-Eurasia, and the Third Pandemic (19^{th} - 20^{th} century) reached every continent except for Antarctica, and continues to persist in various endemic foci around the world. Despite centuries of historical research, the epidemiology of these pandemics remains enigmatic. However, recent technological advancements have yielded a novel form of evidence: ancient DNA of the plague bacterium Yersinia pestis. In this thesis, I explore how genomic data can be used to unravel the mysteries of when and where this disease appeared in the past. In particular, I focus on phylogenetic approaches to studying this 'small microbe' with 'big data' (ie. 100s - 1000s of genomes). I begin by describing novel software I developed that supports the acquisition and curation of large amounts of DNA sequences in public databases. I then use this tool to create an updated global phylogeny of Y. pestis, which includes ~600 genomes with standardized metadata. I devise and validate a new approach for temporal modeling (ie. molecular clock) that produces robust divergence dates in pandemic lineages of Y. pestis. In addition, I critically examine the questions that genomic evidence can and cannot address in isolation, such as whether the timing and spread of short-term epidemics can be confidently reconstructed. Finally, I apply this theoretical and methodological insight to a case study in which I reconstruct the appearance, persistence, and disappearance of plague in Denmark during the Second Pandemic. The three papers enclosed in this sandwich-thesis contribute to a larger body of work on the anthropology of plague, which seeks to understand how disease exposure and experience change over time and between human populations. Furthermore, this dissertation more broadly impacts both prospective studies of infectious disease, such as environmental surveillance and outbreak monitoring, and retrospective studies, which seek to date the emergence and spread of past pandemics.

Acknowledgments

I'd like to thank my parents, Michelle and Michael Eaton. When I was little, I thought you knew everything. And now that I'm writing my doctoral dissertation... I realize you do know everything! I hope when I grow up, I turn out to be just like you <3 Thank you for your love, support, and encouragement over all these years.

To Miriam: Thank you for being my partner, my best friend, my everything. I hope that one day I have 1/10 of your intellect, kindness, and patience. (Maybe we won't hold our breath for that last one.)

To Hendrik Poinar: Thank you for your unending support and enthusiasm. Your mentorship and passion for research has been my rock during the hard times. Thank you for taking leaps of faith and trusting me when I proposed ridiculous project ideas. At least a few of those wound up in this thesis! And most importantly, thank you for traveling to Edmonton in 2013 to give a talk at the University of Alberta!

Thank you to members of my doctoral supervisory committee: Brian Golding, Tracy Prowse, and Nükhet Varlık. I am indebted to you for your generous support, careful guidance, and prompt feedback. Our affectionate motto of 'Keep It Simple Stupid' has played on loop in my head as I prepared this dissertation.

To John Silva, Marcia Furtado, and Delia Hutchinson: Thank you for guiding me through the labyrinth that is McMaster's administration. Your smiling faces up on the 5th floor were always so reassuring. I knew that if I ever had a problem, you would be there to investigate and advocate on my behalf.

To the Plague Team: Jennifer Klunk, what would I do without you? I think everything I've ever known and ever will know comes back to you. Thank you for being a dedicated mentor, a brilliant scientist, and the best companion for dancing in the lab. Madeline Tapson, I dearly miss sharing a desk with you. Your warm and friendly spirit was always comforting, and you opened my eyes to so many new avenues of plague research. Ravneet Sidhu, I learned so much from training you and I loved that you questioned everything. I'm so excited to be collaborating with you on current and future projects. Michael Klowak and Julianna Stangroom, thank you for your HARD work in screening a dizzying number of plague samples and making the lab such a fun and exciting place to be.

To Emil Karpinski: I always looked forward to our bus rides into campus together. You played such an important role in creating a welcoming atmosphere when I first arrived and throughout my whole degree. Also, you are lab notebooks goals (wow).

To Ana Duggan: You have been a role model for me in so many avenues of my academic, personal, and professional life! You were the first woman I met that was also passionate about computational analysis, and have made me feel more comfortable and confident in my own skin.

To Nathalie Mouttham: Thank you for being such a stellar trainer and friend! I have vivid (but positive) memories of long hours in a laboratory basement doing Phenol Chloroform extractions together and playing 20 questions.

Thank you to all past and present members of the McMaster Ancient DNA Centre. In particular: Melanie Kuch, Matthew Emery, Jess Hider, Samantha Price, Marie-Hélène B.-Hardy, and Dirk Hackenberger. You created such a unique sense of community, and left pretty big shoes to fill!

Thank you to all collaborators who have generously shared their time, energy, and resources with me. In particular, thank you to Rebecca Redfern for working with me on the Roman Londoners project!

To my colleagues at Red Lobster: I worked in many kitchens to fund my education, but working at Red Lobster was by far my favorite. Also to the managers, thank you for letting me have so many free Cheddar Bay Biscuits, they were a crucial component of my student diet (no joke).

Finally, I would like to acknowledge all individuals who financially supported me throughout my doctoral research. I thank Hendrik Poinar, the Department of Anthropology, the MacDATA Institute, the Sherman Centre for Digital Scholarship, McMaster University, the Province of Alberta, and the Social Sciences Research and Humanities Research Council.

List of Abbreviations and Symbols

aDNA	Ancient DNA
DNA	Deoxyribonucleic acid
MRCA	Most Recent Common Ancestor
NCBI	National Center for Biotechnology Information
SRA	Sequence Read Archive
tMRCA	Time to the most recent common ancestor

Declaration of Academic Achievement

I, Katherine Eaton, declare that this thesis titled, 'Big Data, Small Microbes: Genomic analysis of the plague bacterium Yersinia pestis' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at McMaster University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at McMaster University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

1 Introduction

In 2011, I learned about a researcher named Dr. Hendrik Poinar. His team had just published a seminal paper, in which they identified the causative agent of the infamous Black Death [1]. I discovered that this morbid term describes a pandemic that devastated the world in the 14th century, with unprecedented mortality and spread. In less than 10 years (1346-1353) the Black Death swept across Afro-Eurasia, killing 50% of the population [2]. Outbreaks of this new and mysterious disease, often referred to as The Plague, reoccurred every 10 years on average [3]. This epidemic cycling continued for 500 long years in Europe, but in Western Asia, the disease never truly disappeared [4]. The 10-year window of the Black Death alone has an estimated global mortality of 200 million people, making it the most fatal pandemic in human history [5], and also one of the most mysterious.

The cryptic nature of this medieval disease led to significant debate among contemporaries. The dominant theory of contagion at the time was miasma, in which diseases were spread through noxious air [6]. However, Ibn al-Khatib, a notable Islamic scholar, took issue with this theory. After studying outbreaks of The Plague in the 14th century, he proposed an alternative hypothesis in which minute bodies were transmissible between humans [7]. Like most controversial theories, this idea was not readily embraced. Some 400 years later, the British botanist Richard Bradley wrote a radical treatise on The Plague [8] where he similarly proposed that infectious diseases were caused by living, microscopic agents. Again, this theory was rejected. It was not until the 19th century that this “new” perspective would receive widespread acceptance [9]. It is quite remarkable that our modern conceptions of epidemiology and bacteriology can be traced back to diverse “founders” throughout history, who all happened to be grappling with the perplexing nature of The Plague.

After it was established that a living organism caused the Black Death, the intuitive next step was to precisely identify the organism. The symptoms described in historical texts seemed to incriminate bubonic plague [2], a bacterial pathogen that passes from rodents to humans, and leads to grotesquely swollen lymph nodes (buboes). On the other hand, the rapid spread of the Black Death suggests this was a contagion primarily driven by human to human transmission, which more closely fit the profile of an Ebola-like virus [10]. In the 1990s and 2000s, geneticists began contributing novel evidence to the debate, by retrieving pathogenic DNA from skeletal remains [11]. The plague bacterium, Yersinia pestis, played a central role in these molecular investigations, as researchers sought to either establish or refute its presence in medieval victims [12]. The competitive nature of this discourse fueled significant technological progress, and over the next decade, the study of ancient DNA became a well-established discipline. However, the origins of the Black Death remained unresolved, due to numerous controversies surrounding DNA contamination and scientific rigor [13].

As an undergraduate student of forensic anthropology, I was fascinated by the rapid pace at which the field of ancient DNA was developing. I attribute my developing academic obsession to two early-career experiences. First, was reading the highly entertaining back-and-forth commentaries in academic journals [14], where plague researchers would occasionally exchange personal insults [15]. It was clear that these researchers cared deeply about their work. Despite the occasional toxicity, I found these displays of passion to be engaging and refreshing, compared to the otherwise emotionally-sterile field of scientific publishing.

The second defining experience, was the perplexing and often frustrating task of diagnosing infectious diseases from skeletal remains. I was intrigued by the idea of reconstructing an individual’s life story from their skeleton, and using this information to solve the mysteries of the dead. However, while some forms of trauma leave diagnostic marks on bone (ex. sharp force), acute infectious diseases rarely manifest in the skeleton [16,17] and thus are ‘invisible’ to an anthropologist. Because of this, I

found the new field of ancient DNA to be extremely appealing, as it offered a novel solution to this problem. Anthropologists could now retrieve the precise pathogen that had infected an individual, and contribute new insight regarding disease exposure and experience throughout human history. These experiences suggested to me that studying the ancient DNA of pathogens would be an exciting, dynamic, and productive line of research for a graduate degree. I'm happy to say that 10 years later, I still agree with this statement, and by writing this dissertation I hope to convince you, the reader, as well.

Which brings us back to Dr. Hendrik Poinar and his team's seminal work on the mysterious Black Death. The study, led by first author Kirsten Bos, had found DNA evidence of the plague bacterium Y. pestis in several Black Death victims buried in a mass grave in London [1]. However, they did not just retrieve a few strands of DNA, they captured millions of molecules (10.5 million to be precise) which allowed them to reconstruct the entire Y. pestis genome, comprising four million DNA bases. The amount of molecular evidence was staggering, and offered irrefutable proof that the plague bacterium was present during the time of Black Death. However, with a sample size of N=1, the genetic link between Y. pestis and this ancient pandemic was tentative at best.

Armed with the proposal of finding more evidence of Y. pestis in the archaeological record, I applied to work for Dr. Hendrik Poinar at the McMaster Ancient DNA Centre. In 2014, I had the delight and privilege of being accepted into the graduate program at McMaster University. Alongside other members of the "McMaster Plague Team", I set about the daunting task of screening the skeletal remains of more than 1000 individuals for molecular evidence of Y. pestis. This material was generously provided by archaeological collaborators, who were similarly invested in the idea that ancient DNA techniques could identify infectious diseases in their sites. These archaeological remains reflected nearly a millennium of human history, with sampling ages ranging from the 9th to the 19th century CE. The geographic diversity was also immense, with individuals sampled across Europe, Africa, and Asia.

Of the 1000+ individuals screened, approximately 30% originated in Denmark. Due to this large sample size, we, the "Plague Team", had the greatest success in identifying ancient Y. pestis in this region. Over a period of 5 years, we retrieved Y. pestis DNA from 13 Danish individuals dated to the medieval and early modern periods. To contextualize these plague isolates, we reconstructed their evolutionary relationships using a large comparative dataset of global Y. pestis. In Chapter 4, I present the results of this collaborative study, which marks the first longitudinal analysis of an ancient pathogen in a single region. I explore whether the genetic evidence of Y. pestis aligns with the historical narrative of the Black Death, and whether or not subsequent epidemics can be attributed to long-distance reintroductions. However, while this high-throughput study was the first one I embarked on, as the chapter numbering indicates, it would be the last project I completed due to several unforeseen complications.

While the McMaster Plague Team was busy screening for Y. pestis, so too were other ancient DNA centres throughout the world. Between 2011 and 2021, more than 100 ancient Y. pestis genomes were published, making plague the most intensively sequenced historical disease. The sequencing of modern isolates accelerated in tandem, with over 1500 genomes produced from culture collections of 20th and 21st century plague outbreaks [18]. Because of this influx of evidence, the research questions changed accordingly. Geneticists were no longer interested in just establishing the presence of Y. pestis during the short time frame of the Black Death (1346-1353), they wanted to know how it behaved and spread throughout the long 500 years of this pandemic. The longitudinal study design of Chapter 4 was therefore well-positioned to address these nuanced epidemiological questions. However, this novel genetic evidence also introduced new complexities.

It quickly became clear that isolates of Y. pestis sampled during epidemic periods were highly similar in terms of genetic content, if not indistinguishable clones [19]. This called into question the resolution of genomic evidence, and whether the geographic origins and spread of the Black Death could be accurately inferred using ancient DNA studies. This was further confounded by the finding that the rate of evolutionary change in Y. pestis could vary tremendously [20] which led to the discovery that previously published temporal models were erroneous [21]. It became increasingly uncertain whether genetic evidence could be used to produce informative estimates of the timing of plague's frequent reemergences [22]. As I read these critical studies, I began developing an idea to address the substantial gaps in our evolutionary theory and methodology concerning the plague bacterium Y. pestis. This idea culminated in Chapter 3, where I curated and contextualized the largest global data set of plague genomes. I critiqued the existing spatiotemporal models of plague's evolutionary history, and with the assistance of my co-authors, devised a new methodological approach. This method would then be repurposed for Chapter 4, so that I could infer the emergence and disappearance of Y. pestis in Denmark with greater accuracy. However, as the chapter numbering once again reflects, there was one final obstacle.

Synthesizing the largest genomic data set was a lofty ambition, especially considering that there were few software tools available to perform such a task. New plague genomes of Y. pestis were being published monthly, and at times even weekly, with such volume that manual tracking became impossible. My excel spreadsheet of genetic metadata became riddled with errors and fields with missing data. The era of "Big Data" had arrived, and I was woefully unequipped to effectively manage this deluge of information. In response, I ventured into the tumultuous waters of software development. In Chapter 2, I describe my original software that automates the acquisition and organization of genetic metadata. Academic publishing in the field of software was a unique experience, as I had to both produce a scholarly manuscript and demonstrate expertise as a service-provider. This database tool has continually proven to be indispensable, and is the backbone upon which the studies in Chapter 3 and Chapter 4 would be rebuilt upon.

At this point, I re-introduce the dissertation as a collection of three hierarchical, but independently published, studies. I first describe an original piece of software in Chapter 2, which automates the retrieval and organization of publicly available sequence data. In Chapter 3, I outline how this tool was used to generate an updated and curated phylogeny of Y. pestis, which yielded novel insight regarding the timing and origins of past pandemics. In this chapter, I also conduct a critical examination of the historical questions that genomic evidence can, or cannot, address. In Chapter 4, I use these theories and methods to reconstruct the emergence and continuity of plague in Denmark over a period of 400 years. I conclude in Chapter 5 with a discussion of the contributions of each study, with a particular focus on their significance within the broader field of anthropology.

2 NCBImeta: Efficient and comprehensive metadata retrieval from NCBI databases

Published 03 February 2020 in
The Journal of Open Source Software, 5(46), 1990.
https://doi.org/10.21105/joss.01990

Katherine Eaton[1,2]

[1] McMaster Ancient DNA Centre, McMaster University
[2] Department of Anthropology, McMaster University

Eaton_2020_NCBImeta_dissertation.md

3 Plagued by a cryptic clock: Insight and issues from the global phylogeny of Yersinia pestis

Submitted 17 December 2021 to
Communications Biology.
https://www.researchsquare.com/article/rs-1146895

Katherine Eaton[1,2], Leo Featherstone[3], Sebastian Duchene[3], Ann G. Carmichael[4], Nükhet Varlık[5], G. Brian Golding[6], Edward C. Holmes[7], Hendrik N. Poinar[1,2,8,9,10]

[1]McMaster Ancient DNA Centre, McMaster University, Hamilton, Canada.
[2]Department of Anthropology, McMaster University, Hamilton, Canada.
[3]The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia.
[4]Department of History, Indiana University Bloomington, Bloomington, USA.
[5]Department of History, Rutgers University-Newark, Newark, USA.
[6]Department of Biology, McMaster University, Hamilton, Canada.
[7]Sydney Institute for Infectious Diseases, School of Life & Environmental Sciences and School of Medical Sciences, University of Sydney, Sydney, Australia.
[8]Department of Biochemistry, McMaster University, Hamilton, Canada.
[9]Michael G. DeGroote Institute of Infectious Disease Research, McMaster University, Hamilton, Canada.
[10]Canadian Institute for Advanced Research, Toronto, Canada.

`Eaton_et_al._2021_Plagued_by_a_cryptic_clock.md`

4 Plague in Denmark (1000-1800): A longitudinal study of Yersinia pestis

Prepared 08 December 2021 for submission to
The Proceedings of the National Academy of Sciences

Katherine Eaton*[1,2], Ravneet Sidhu*[1,3], Jennifer Klunk[1,4], Julia Gamble[5], Jesper Boldsen[6], Ann G. Carmichael[7], Nükhet Varlık[8], Sebastian Duchene[9], Leo Featherstone[9], Vaughan Grimes[10], G. Brian Golding[3], Sharon DeWitte[11], Hendrik N. Poinar[1,2,12,13,14]

*Contributed equally.

[1]McMaster Ancient DNA Centre, McMaster University, Hamilton, Canada.
[2]Department of Anthropology, McMaster University, Hamilton, Canada.
[3]Department of Biology, McMaster University, Hamilton, Canada.
[4]Daicel Arbor Biosciences, Ann Arbor, USA.
[5]Department of Anthropology, University of Manitoba, Winnipeg, Canada.
[6]Department of Forensic Medicine, Unit of Anthropology (ADBOU), University of Southern Denmark, Odense, Denmark.
[7]Department of History, Indiana University Bloomington, Bloomington, USA.
[8]Department of History, Rutgers University-Newark, Newark, USA.
[9]The Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Australia.
[10]Department of Archaeology, Memorial University of Newfoundland, St. Johns, Canada.
[11]Department of Anthropology, University of South Carolina, Columbia, USA.
[12]Department of Biochemistry, McMaster University, Hamilton, Canada.
[13]Michael G. DeGroote Institute of Infectious Disease Research, McMaster University, Hamilton, Canada.
[14]Canadian Institute for Advanced Research, Toronto, Canada.

`Eaton_et_al._2021_Plague_in_Denmark_1000-1800.md`

5 Conclusion

5.1 Main Findings and Contributions

In this dissertation, I developed computational methods for genomics research and used them to reconstruct past and present pandemics of plague. In Chapter 2, I resented a novel software called `NCBImeta` that facilitates the acquisition of sequence data and metadata from the NCBI repository. This specialized tool supports genomics research in the era of big data, where manual processing of abundant sequence records (10,000+) is impossible. As a paper on software development, its contributions and significance to the field of anthropology are understandably unclear. I targeted this

article exclusively towards computational biologists because, at the time, few anthropologists had expressed interested in the issue of collecting and curating sequencing data. Reflecting this, `NCBImeta` has mainly been cited across biological fields including studies of the human microbiome [23], plant-associated bacteria in agriculture [24], and emerging infectious diseases in public health (Matthew Gopez & Philip Mabon, personal communication, https://github.com/ktmeaton/NCBImeta/pull/9).

In 2021, I took a more active approach in my discipline and used this software to support several bodies of anthropological research. `NCBImeta` was recently used in an environmental reconstruction of Beringia [25], the former land-bridge that facilitated early human migrations into North America from northeast Asia. The study by Murchie et al. furthers our understanding of the peopling of the Americas, and the possible interactions between early human populations and large animals (ie. megafauna) before the Last Glacial Period (~12,000 years ago). `NCBImeta` was also recently used to curate sequence data in a case study of the zoonotic disease brucellosis in the 14th century [26]. The pioneering work by Hider et al. demonstrates how pathogen DNA preserves differently throughout the body, ranging from being the dominant microorganism in several tissues while being completely absent in others. It raises an important cautionary note for ancient DNA analysis and the anthropology of disease, by empirically demonstrating how sampling strategies can bias our understanding of what diseases were present in past populations.

In Chapter 3, I explored the challenges in estimating where and when plague appeared in the past, and why these estimates are often not reproducible between studies. I used the software tool from Chapter 2 to collect all publicly available Y. pestis genomes, and carefully curated their collection dates, locations, and hosts. My co-authors and I then used this data set for phylodynamic analysis, and devised a new approach for modeling the rates of evolutionary change (ie. molecular clock). We used these results to explain why divergence dates varied between studies, and outlined a critical framework for identifying which divergence dates should be considered non-informative. In addition, we found that past pandemics of plague may have emerged decades, or even centuries, before they were historically documented in European sources. These early dates are in agreement with recent historical work that examines more diverse (ie. non-European) sources. Through this finding, we demonstrated how genomic dating plays an important role in expanding the timelines of past pandemics to make space for more diverse narratives.

In contrast to our claims of the 'power' of genomic evidence, a prominent takeway from Chapter 3 was our discussion of the limitations of DNA. In particular, we found that Y. pestis genomes in isolation are not suitable for reconstructing evolutionary relationships during short-term epidemics. This is because the evolutionary rate of past pandemic lineages is approximately 1 substitution every 10 years. Isolates collected within this time frame (<10 years) are often identical, which means that the directionality of spread cannot be confidently inferred. To mitigate this weakness, complementary evidence is needed that has a higher temporal resolution. Historical case records are an excellent candidate, where plague cases are recorded annually if not weekly [27]. Based on initial comments from readers of the preprint, this conclusion was particularly exciting as it provided guidance on how to avoid over-interpreting ancient DNA evidence, and suggested a new avenue for inter-disciplinary collaboration (Boris Schmidt, personal communication).

In Chapter 4, I applied this updated genomic dataset and molecular clock method to a new problem. While in Chapter 3 we were concerned with estimating the first emergence of pandemic lineages, in Chapter 4 we reconstructed the persistence or continuity of ancient pandemics. We designed a unique longitudinal study, where we sampled skeletal remains spanning 800 years (1000 - 1800 CE) dated to before, during, and after the Second Pandemic (14thth - 18th century). Our sampling strategy focused on Scandinavia, particularly Denmark, as this region is underrepresented in the historical

narrative and because the Anthropological DataBase Odense University collection (ADBOU, University of Southern Denmark) has exquisitely curated over 17,000 skeletal remains dated from the Viking Age (10^{th} century) to the Early Modern Period (18^{th} century). Using ancient DNA techniques, we recovered evidence of Y. pestis throughout the 14^{th} to 17^{th} centuries, which perfectly aligns with the historical narrative, limited as it is. Furthermore, our positivity rate for Y. pestis (3.3% - 14.3%) overlaps with mortality estimates from several historical outbreaks during the Second Pandemic. Our results strengthen the argument that Y. pestis was the causative agent of the Second Pandemic, and suggests that plague was a relatively new disease in medieval Denmark. These findings are being expanded on in two upcoming studies. The first, is an examination of how Danish populations responded to this new disease with regards to changes in the human immune system [28]. The second, is a reconstruction of how and when virulence in Y. pestis became attenuated during the Second Pandemic. Taken together, we anticipate these studies will deepen our understanding of disease exposure and experience in Denmark and across Europe.

5.2 Future Directions

5.2.1 Same 'Plague', New Problems

A reoccurring problem in plague research is how best to integrate multidisciplinary sources, as there is great interest in combining genetic, historical, and environmental records to better understand past pandemics of plague [29,30]. An approach that is commonly used in ancient DNA studies of Y. pestis involves two steps: (1) reconstructing the relationships between epidemics using genetic evidence, and then (2) interpreting those relationships using historical records [19,31,32]. However, a major limitation of this method is that multidisciplinary sources are only integrated during the final interpretation phase. This runs the risk that errors and uncertainty associated with the genetic analysis will propagate, leading to high levels of ambiguity when attempting to provide historical context for this genetic 'noise'.

An alternative method, is to leverage the strengths and mitigate the weaknesses of interdisciplinary sources in a joint phylogenetic analysis. This novel approach treats historical records (ex. location and date of an outbreak) as special 'sequence-free' samples. These records are then combined with DNA evidence to jointly infer a phylogeny, which can then be used to estimate the timing and location of historical events. Recent studies have demonstrated how critical this approach is, as case occurrence records can effectively correct for sampling biases in sparse genomic datasets [33,34]. However, incorporating sequence-free datasets is still a relatively recent method, and to date has only been applied to the study of viruses. Furthermore, it has only been tested on outbreaks occurring over a relatively small geographic area and time range. It remains unknown whether this approach is feasible for bacterial genomics, let alone ancient DNA, where genomes are larger and sampled across greater temporal and geographic scopes. This presents a key line of inquiry for future research, for which the plague bacterium Y. pestis would be an excellent case study.

5.2.2 New 'Plague', Same Problems

During the course of this dissertation, my interest in global pandemics turned from an academic curiosity to a lived experience. In 2019, the novel coronavirus SARS-CoV-2 emerged to cause a global pandemic, with over 270 million cases recorded worldwide. While there are many unique aspects of this pandemic, one that has captured my attention is that it is the first pandemic to be monitored with real-time genomic surveillance [35]. Over two million genomic sequences have been deposited in public repositories, which can be used to inform public health responses [36]. However, this avalanche of data has also caused numerous problems, as researchers are struggling to manage this

information and utilize it effectively [37]. As a result, database tools such as `NCBImeta` presented in Chapter 2, are playing an important role in information management.

One field of ongoing research involves improving the scalability of these tools. For example, `NCBImeta` was developed for a data set of 'only' 15,000 records, and in its current implementation, cannot process the 1+ million SARS-CoV-2 records on NCBI. A second critical avenue is integrating information from multiple repositories, as surveillance data is inconsistently being deposited in national and international databases [38,39,40]. Progress towards these two objectives will result in more diverse genomic data being analyzed (geographically and temporally), which may improve of our understanding of transmission and spread between and within countries.

Another parallel between this dissertation and the ongoing pandemic involves spatiotemporal modeling. In Chapter 3, we discovered that in our expanded genomic data set, Y. pestis' rate of spread tends to outpace its rate of evolutionary change. This leads to identical Y. pestis isolates found across multiple countries, such as the case throughout the Black Death (1346-1353). However, we sporadically observed the opposite trend, in which Y. pestis strains collected in a short time frame (<10 years) were extremely different. This tremendous diversity in evolutionary rates meant that we were unable to estimate a single molecular clock for Y. pestis. These issues, clonal spread and rate variation, were also recently documented in SARS-CoV-2 [41]. Ferreira et al. describe this as a paradox in which we "become increasingly uncertain about the relationships among specific lineages as we collect greater amounts of data". This runs counterintuitive to the general expectation in scientific studies that the more data we collect, the closer we get to the 'truth'. Overall, this presents a complex theoretical problem that is becoming increasingly prevalent across various disciplines moving into the era of 'big data'.

6 References

1. **A draft genome of Yersinia pestis from victims of the Black Death**
   Kirsten I Bos, Verena J Schuenemann, GBrian Golding, Hernán A Burbano, Nicholas Waglechner, Brian K Coombes, Joseph B McPhee, Sharon N DeWitte, Matthias Meyer, Sarah Schmedes, ... Johannes Krause
   Nature (2011) http://www.nature.com/articles/nature10549
   DOI: 10.1038/nature10549

2. **The Black Death, 1346-1353: The Complete History**
   OJ Benedictow
   Boydell Press (2004)
   ISBN: 0-85115-943-5

3. **"In these perilous times": Plague and plague policies in early modern Denmark**
   Peter Christensen
   Medical History (2003-10) https://www.cambridge.org/core/journals/medical-history/article/in-these-perilous-times-plague-and-plague-policies-in-early-modern-denmark/EFF71835DE9EBFB610E35451FD7A0A86
   DOI: 10.1017/s0025727300057331

4. **The plague that never left: restoring the Second Pandemic to Ottoman and Turkish history in the time of COVID-19**
   Nükhet Varlık

New Perspectives on Turkey (2020-11) https://www.cambridge.org/core/journals/new-perspectives-on-turkey/article/plague-that-never-left-restoring-the-second-pandemic-to-ottoman-and-turkish-history-in-the-time-of-covid19/AE81F48AD6FEB71C42AAA2115A0307A9
DOI: 10.1017/npt.2020.27

5. **Pandemics Throughout the History**
Shrikanth Sampath, Anwar Khedr, Shahraz Qamar, Aysun Tekin, Romil Singh, Ronya Green, Rahul Kashyap
Cureus (2021) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8525686/
DOI: 10.7759/cureus.18136 · PMID: 34692344 · PMCID: PMC8525686

6. **The plague at Granada, 1348-1349: Ibn Al-Khatib and ideas of contagion.**
WB Ober, N Aloush
Bulletin of the New York Academy of Medicine (1982-05) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1808550/
PMID: 7052179 · PMCID: PMC1808550

7. **Islamic medicine: 1000 years ahead of its times**
Ibrahim Syed
Journal of the International Society for the History of Islamic Medicine (1981) https://jima.imana.org/article/view/11925

8. **The Plague at Marseilles: Consider'd with Remarks Upon the Plague in General**
Richard Bradley
W. Mears (1721) https://books.google.ca/books?id=qQYAmMH1nS4C

9. **Richard Bradley: A Unified, Living Agent Theory of the Cause of Infectious Diseases of Plants, Animals, and Humans in the First Decades of the 18th Century**
Melvin Santer
Perspectives in Biology and Medicine (2009) http://muse.jhu.edu/article/362917
DOI: 10.1353/pbm.0.0124

10. **Biology of Plagues: Evidence from Historical Populations**
Susan Scott, Christopher J Duncan
Cambridge University Press (2001) https://www.cambridge.org/core/books/biology-of-plagues/9AE9838FB1BBE059455F24C75DDB9DE2
DOI: 10.1017/cbo9780511542527 · ISBN: 978-0-521-80150-8

11. **Detection of 400-year-old Yersinia pestis DNA in human dental pulp: An approach to the diagnosis of ancient septicemia**
Michel Drancourt, Gérard Aboudharam, Michel Signoli, Olivier Dutour, Didier Raoult
Proceedings of the National Academy of Sciences (1998-10-13) https://www.pnas.org/content/95/21/12637
DOI: 10.1073/pnas.95.21.12637 · PMID: 9770538

12. **Absence of Yersinia pestis-specific DNA in human teeth from five European excavations of putative plague victims**
MThomas P Gilbert, Jon Cuccui, William White, Niels Lynnerup, Richard W Titball, Alan Cooper, Michael BYR 2004 Prentice
Microbiology (2004) https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.26594-0
DOI: 10.1099/mic.0.26594-0

13. **Ancient DNA: do it right or not at all**

A Cooper, HN Poinar
Science (New York, N.Y.) (2000-08-18)
DOI: 10.1126/science.289.5482.1139b · PMID: 10970224

14. **Response to Drancourt and Raoult**
MThomas P Gilbert, Jon Cuccui, William White, Niels Lynnerup, Richard W Titball, Alan Cooper, Michael B Prentice
Microbiology (2004-02-01) https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.26959-0
DOI: 10.1099/mic.0.26959-0

15. **Was the Black Death yersinial plague?**
Didier Raoult
The Lancet Infectious Diseases (2003-06) https://linkinghub.elsevier.com/retrieve/pii/S1473309903006522
DOI: 10.1016/s1473-3099(03)00652-2

16. **Differential Diagnosis of Skeletal Lesions in Infectious Disease**
Donald J Ortner
Advances in Human Palaeopathology (2007) https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470724187.ch10
DOI: 10.1002/9780470724187.ch10 · ISBN: 978-0-470-72418-7

17. **The Anthropology of Infectious Disease: International Health Perspectives**
Peter J Brown, Marcia C Inhorn
Routledge (2013-11-05)
ISBN: 978-1-134-38642-0

18. **The EnteroBase user's guide, with case studies on Salmonella transmissions, Yersinia pestis phylogeny, and Escherichia core genomic diversity**
Zhemin Zhou, Nabil-Fareed Alikhan, Khaled Mohamed, Yulei Fan, Mark Achtman
Genome Research (2020-01) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6961584/
DOI: 10.1101/gr.251678.119 · PMID: 31809257 · PMCID: PMC6961584

19. **Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes**
Maria A Spyrou, Marcel Keller, Rezeda I Tukhbatova, Christiana L Scheib, Elizabeth A Nelson, Aida Andrades Valtueña, Gunnar U Neumann, Don Walker, Amelie Alterauge, Niamh Carty, … Johannes Krause
Nature Communications (2019-10-02) https://www.nature.com/articles/s41467-019-12154-0
DOI: 10.1038/s41467-019-12154-0

20. **Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis**
Y Cui, C Yu, Y Yan, D Li, Y Li, T Jombart, LA Weinert, Z Wang, Z Guo, L Xu, … R Yang
Proceedings of the National Academy of Sciences (2013-01-08) http://www.pnas.org/cgi/doi/10.1073/pnas.1205750110
DOI: 10.1073/pnas.1205750110

21. **Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic analysis**
David M Wagner, Jennifer Klunk, Michaela Harbeck, Alison Devault, Nicholas Waglechner, Jason W Sahl, Jacob Enk, Dawn N Birdsell, Melanie Kuch, Candice Lumibao, … Hendrik Poinar
The Lancet Infectious Diseases (2014-04) https://linkinghub.elsevier.com/retrieve/pii/S1473309913703232

DOI: 10.1016/s1473-3099(13)70323-2

22. **Genome-scale rates of evolutionary change in bacteria**
Sebastian Duchêne, Kathryn E Holt, François-Xavier Weill, Simon Le Hello, Jane Hawkey, David J Edwards, Mathieu Fourment, Edward C Holmes
Microbial Genomics (2016-11-30) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5320706/
DOI: 10.1099/mgen.0.000094 · PMID: 28348834 · PMCID: PMC5320706

23. **SKIOME Project: a curated collection of skin microbiome datasets enriched with study-related metadata**
G Agostinetto, D Bozzi, D Porro, M Casiraghi, M Labra, A Bruno
Cold Spring Harbor Laboratory (2021-08-18) https://www.biorxiv.org/content/10.1101/2021.08.17.456635v1
DOI: 10.1101/2021.08.17.456635

24. **Comparative Genomics and In Vitro Plant Growth Promotion and Biocontrol Traits of Lactic Acid Bacteria from the Wheat Rhizosphere**
Sabrina Strafella, David J Simpson, Mohammad Yaghoubi Khanghahi, Maria De Angelis, Michael Gänzle, Fabio Minervini, Carmine Crecchio
Microorganisms (2021-01) https://www.mdpi.com/2076-2607/9/1/78
DOI: 10.3390/microorganisms9010078

25. **Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost**
Tyler Murchie, Emil Karpinski, Katherine Eaton, Ana T Duggan, Sina Baleka, Grant Zazula, Ross DE MacPhee, Duane Froese, Hendrik N Poinar
Current Biology (Accepted, 2021)

26. **Examining pathogen DNA recovery across the remains of a 14th century Italian monk (St. Brancorsini) infected with Brucella melitensis**
Jessica Hider, Ana T Duggan, Jennifer Klunk, Katherine Eaton, Long, George Sam, Emil Karpinski, GBrian Golding, Tracy L Prowse, Hendrik N Poinar, Gino Fornaciari
(In Prep)

27. **Dangers of noncritical use of historical plague data**
Joris Roosen, Daniel R Curtis
Emerging Infectious Diseases (2018-01) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5749453/
DOI: 10.3201/eid2401.170477 · PMID: null · PMCID: PMC5749453

28. **Black Death shaped the evolution of immune genes**
Jennifer Klunk, Tauras Vilgalys, Christian Demeure, Mari Cobb, Derek Elli, Rebecca Redfern, Sharon N DeWitte, Julia Gamble, Jesper L Boldsen, Ann G Carmichael, … Luis Barreiro
(In Prep, 2021)

29. **Climate-driven introduction of the Black Death and successive plague reintroductions into Europe**
Boris V Schmid, Ulf Büntgen, WRyan Easterday, Christian Ginzler, Lars Walløe, Barbara Bramanti, Nils Chr Stenseth
Proceedings of the National Academy of Sciences (2015-03-10) http://www.pnas.org/content/112/10/3020
DOI: 10.1073/pnas.1412887112 · PMID: 25713390

30. **Human ectoparasites and the spread of plague in Europe during the Second Pandemic**

Katharine R Dean, Fabienne Krauer, Lars Walløe, Ole Christian Lingjærde, Barbara Bramanti, Nils Chr Stenseth, Boris V Schmid
Proceedings of the National Academy of Sciences of the United States of America (2018-02-06) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5819418/
DOI: 10.1073/pnas.1715640115 · PMID: 29339508 · PMCID: PMC5819418

31. **Integrative approach using Yersinia pestis genomes to revisit the historical landscape of plague during the Medieval Period**
Amine Namouchi, Meriam Guellil, Oliver Kersten, Stephanie Hänsch, Claudio Ottoni, Boris V Schmid, Elsa Pacciani, Luisa Quaglia, Marco Vermunt, Egil L Bauer, … Barbara Bramanti
Proceedings of the National Academy of Sciences (2018-12-11) http://www.pnas.org/lookup/doi/10.1073/pnas.1812865115
DOI: 10.1073/pnas.1812865115

32. **A genomic and historical synthesis of plague in 18th century Eurasia**
Meriam Guellil, Oliver Kersten, Amine Namouchi, Stefania Luciani, Isolina Marota, Caroline A Arcini, Elisabeth Iregren, Robert A Lindemann, Gunnar Warfvinge, Lela Bakanidze, … Barbara Bramanti
Proceedings of the National Academy of Sciences (2020-11-10) https://www.pnas.org/content/117/45/28328
DOI: 10.1073/pnas.2009677117 · PMID: 33106412

33. **Sampling bias and model choice in continuous phylogeography: getting lost on a random walk**
Antanas Kalkauskas, Umberto Perron, Yuxuan Sun, Nick Goldman, Guy Baele, Stephane Guindon, Nicola De Maio
PLOS Computational Biology (2021-01-06) https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008561
DOI: 10.1371/journal.pcbi.1008561

34. **Infectious disease phylodynamics with occurrence data**
Leo A Featherstone, Francesca Di Giallonardo, Edward C Holmes, Timothy G Vaughan, Sebastián Duchêne
Methods in Ecology and Evolution (2021) https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13620
DOI: 10.1111/2041-210x.13620

35. **The next phase of SARS-CoV-2 surveillance: real-time molecular epidemiology**
Bas B Oude Munnink, Nathalie Worp, David F Nieuwenhuijse, Reina S Sikkema, Bart Haagmans, Ron AM Fouchier, Marion Koopmans
Nature Medicine (2021-09) https://www.nature.com/articles/s41591-021-01472-w
DOI: 10.1038/s41591-021-01472-w

36. **SARS-CoV-2 Whole Genome Sequencing in Ontario, December 14, 2021**
Public Health Ontario
(2021-12-14) https://www.publichealthontario.ca/-/media/documents/ncov/epi/covid-19-sars-cov2-whole-genome-sequencing-epi-summary.pdf

37. **Phylogenetic Analysis of SARS-CoV-2 Data Is Difficult**
Benoit Morel, Pierre Barbera, Lucas Czech, Ben Bettisworth, Lukas Hübner, Sarah Lutteropp, Dora Serdari, Evangelia-Georgia Kostaki, Ioannis Mamais, Alexey M Kozlov, … Alexandros Stamatakis
Molecular Biology and Evolution (2021-05-01) https://doi.org/10.1093/molbev/msaa314

DOI: 10.1093/molbev/msaa314

38. **National Center for Biotechnology Information**
NCBI
https://www.ncbi.nlm.nih.gov/

39. **GISAID - Initiative**
GISAID
https://www.gisaid.org/

40. **VirusSeq Portal**
CanCOGeN
https://virusseq-dataportal.ca/

41. **CoVizu: Rapid analysis and visualization of the global diversity of SARS-CoV-2 genomes**
Roux-Cil Ferreira, Emmanuel Wong, Gopi Gugan, Kaitlyn Wade, Molly Liu, Laura Muñoz Baena, Connor Chato, Bonnie Lu, Abayomi S Olabode, Art FY Poon
Virus Evolution (2021-12-01) https://doi.org/10.1093/ve/veab092
DOI: 10.1093/ve/veab092