Big Data, Small Microbes: Genomic analysis of the plague bacterium *Yersinia pestis*

Katherine Eaton

Abstract here (no more than 300 words).

Table of Contents

# 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](#ref-bos2011DraftGenomeYersinia). 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[**cite?**](#ref-cite). Outbreaks of this new and mysterious disease, often referred to as *the Plague*, reoccurred every 10 years on average[**cite?**](#ref-cite). This epidemic cycling continued for 500 long years in Europe, but in Western Asia, the disease never truly disappeared[**cite?**](#ref-cite). 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[2](#X7d4172177775123404348b58dd121e929c44cf0), and it remains 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[3](#ref-ober1982PlagueGranada13481349). Ibn al-Khatib, a notable Islamic scholar, took issue with this theory. After studying outbreaks of *Plague* in the 14th century, he proposed an alternative hypothesis in which *minute bodies* were transmissible between humans[4](#ref-syed1981IslamicMedicine1000). Like most controversial theories, this idea was not readily embraced. Some 400 years later, the British botanist Richard Bradley wrote a radical treatise on *Plague*[5](#ref-bradley1721PlagueMarseillesConsider) 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, with the experiments of Louis Pasteur and Robert Koch, that this “new” perspective would receive widespread acceptance[**cite?**](#ref-cite).

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 seem to incriminate bubonic plague[**cite?**](#ref-cite), 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 fits the profile of an Ebola-like virus[6](#ref-scott2001BiologyPlaguesEvidence). In the 2000s, geneticists began contributing novel evidence to the debate, by retrieving pathogenic DNA from skeletal remains[**cite?**](#ref-cite). 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. The competitive nature of this discourse fueled significant technological progress, and over the next decade, the study of ancient DNA (aDNA) became its own discipline. However, the origins of the Black Death remained unresolved, due to numerous controversies surrounding contamination and scientific rigor[**cite?**](#ref-cite).

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, where plague researchers would exchanges snide and personal insults[**cite?**](#ref-cite). It was clear that these researchers cared *deeply* about their work. Despite the toxicity, I found these displays of passion to be engaging and refreshing, as compared to the otherwise emotionally-sterile field of scientific publishing.

The second defining experience, was the perplexing and often frustrating task of trying to diagnose 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[7](#Xdebc45c6f972e8ed8e66aa67d1e8d8645017850),[8](#X3e92459bec36978eff10e15852ff1eb9dd9eb31) 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 confirmed 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](#ref-bos2011DraftGenomeYersinia). But 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. Admittedly, any inference of causality would be mere speculation, but given the extremely high case-fatality rates of plague in modern times[**cite?**](#ref-cite), there was strong support for the hypothesis that the causative agent of the Black Death was plague.

In the following decade (2011-2021) more than 100 ancient *Y. pestis* genomes have been retrieved from skeletal remains across Europe, making plague the *most intensively sequenced historical disease*. The sequencing of modern isolates has accelerated in tandem, with over 1000 genomes produced from 20th and 21st century plague outbreaks[**cite?**](#ref-cite). Because of this influx of evidence, the research questions changed accordingly. Geneticists were no longer interested in *just* establishing the presence *Y. pestis* during the short time frame of the Black Death (1346-1353), they wanted to know *how it behaved* throughout the long 500 years of this pandemic. These questions included lines of inquiry such as: How did this ancient plague pandemic spread? Did it take the rare form of pneumonic plague, where the bacterium could be rapidly transmitted between humans along maritime routes? Or were outbreaks highly localized cases of bubonic plague, where disease exposure in humans occurs via spillover from nearby wild rodent populations? And perhaps most importantly of all, why have we never experienced another pandemic quite like the Black Death?

These nuanced epidemiological questions, particularly the last one, were what motivated and resonated with me. In 2013, I had the delight and privilege of being accepted into the graduate program at McMaster University, to investigate these questions under the supervision of Dr. Poinar at the McMaster Ancient DNA Centre.

Previously, I referred to one of the ultimate questions in plague research: why have we never experienced another pandemic quite like the Black Death?

In 2019, my relationship with infectious diseases transformed from an intellectual curiosity to a lived experience. The emergence of the novel coronavirus (SARS-CoV-2) triggered a global pandemic, operating on a scale that had not been seen for a 100 years. For years, I had written grants to fund my plague research under the auspice of *we have to know the past to understand the present*[**cite?**](#ref-cite).

* In 2019, pandemics changed from an intellectual passion to a live experience.
* The novel coronavirus (SARS-CoV-2) pandemic.

## 1.1 References

1. Bos, K. I. *et al.* [A draft genome of *Yersinia* *Pestis* from victims of the Black Death](https://doi.org/10.1038/nature10549). *Nature* **478**, 506–510 (2011).

2. Sampath, S. *et al.* [Pandemics Throughout the History](https://doi.org/10.7759/cureus.18136). *Cureus* **13**, e18136 (2021).

3. Ober, W. B. & Aloush, N. [The plague at Granada, 1348-1349: Ibn Al-Khatib and ideas of contagion.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1808550) *Bulletin of the New York Academy of Medicine* **58**, 418–424 (1982).

4. Syed, I. Islamic medicine: 1000 years ahead of its times. *Journal of the International Society for the History of Islamic Medicine* **13**, 2–9 (1981).

5. Bradley, R. *The Plague at Marseilles: Consider’d with Remarks Upon the Plague in General*. (W. Mears, 1721).

6. Scott, S. & Duncan, C. J. *Biology of Plagues: Evidence from Historical Populations*. (Cambridge University Press, 2001). doi:[10.1017/CBO9780511542527](https://doi.org/10.1017/CBO9780511542527).

7. Ortner, D. J. Differential Diagnosis of Skeletal Lesions in Infectious Disease. in *Advances in Human Palaeopathology* 189–214 (John Wiley & Sons, Ltd, 2007). doi:[10.1002/9780470724187.ch10](https://doi.org/10.1002/9780470724187.ch10).

# 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>  
Licensed under a Creative Commons Attribution 4.0 International License.

Katherine Eaton1,2

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

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

Submitted 06 December 2021 to  
*Nature Communications*.  
<https://www.researchsquare.com/article/rs-1146895>  
Licensed under a Creative Commons Attribution 4.0 International License

Katherine Eaton1,2, Leo Featherstone3, Sebastian Duchene3, Ann G. Carmichael4, Nükhet Varlık5, G. Brian Golding6, Edward C. Holmes7, Hendrik N. Poinar1,2,8,9,10

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

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

Prepared 07 December 2021 for submission as a Brief Report to  
*The Proceedings of the National Academy of Sciences*  
Licensed under a Creative Commons Attribution 4.0 International License

Katherine Eaton1,2, Leo Featherstone3, Sebastian Duchene3, Ann G. Carmichael4, Nükhet Varlık5, G. Brian Golding6, Hendrik N. Poinar1,2,8,9,10

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

# 5 Conclusion

As a paper on software development, its contributions and significance to the field of anthropology are understandably unclear. I admittedly targeted this article exclusively towards computational biologists because, at the time, few anthropologists had expressed interested in the issue of collecting and curating sequence data from online repositories. However, since its publication, my software has been used to support several bodies of anthropological research.

The database software NCBImeta was recently used to support an environmental reconstruction of Beringia[9](#ref-murchieInPrepNoBonesIt), 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).

This tool was also recently used to curate sequence data in a case study of the zoonotic disease brucellosis in the 14th century[10](#ref-hiderInPrepExaminingPathogenDNA). 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.

## 5.1 References

1. Murchie, T. *et al.* No bones about it: Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost. (In Prep).

2. Hider, J. *et al.* Examining pathogen DNA recovery across the remains of a 14th century Italian monk (St. Brancorsini) infected with Brucella melitensis. (In Prep).

# References

1. Bos, K. I. *et al.* [A draft genome of *Yersinia* *Pestis* from victims of the Black Death](https://doi.org/10.1038/nature10549). *Nature* **478**, 506–510 (2011).

2. Sampath, S. *et al.* [Pandemics Throughout the History](https://doi.org/10.7759/cureus.18136). *Cureus* **13**, e18136 (2021).

3. Ober, W. B. & Aloush, N. [The plague at Granada, 1348-1349: Ibn Al-Khatib and ideas of contagion.](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1808550) *Bulletin of the New York Academy of Medicine* **58**, 418–424 (1982).

4. Syed, I. Islamic medicine: 1000 years ahead of its times. *Journal of the International Society for the History of Islamic Medicine* **13**, 2–9 (1981).

5. Bradley, R. *The Plague at Marseilles: Consider’d with Remarks Upon the Plague in General*. (W. Mears, 1721).

6. Scott, S. & Duncan, C. J. *Biology of Plagues: Evidence from Historical Populations*. (Cambridge University Press, 2001). doi:[10.1017/CBO9780511542527](https://doi.org/10.1017/CBO9780511542527).

7. Ortner, D. J. Differential Diagnosis of Skeletal Lesions in Infectious Disease. in *Advances in Human Palaeopathology* 189–214 (John Wiley & Sons, Ltd, 2007). doi:[10.1002/9780470724187.ch10](https://doi.org/10.1002/9780470724187.ch10).

8. Brown, P. J. & Inhorn, M. C. *The Anthropology of Infectious Disease: International Health Perspectives*. (Routledge, 2013).

9. Murchie, T. *et al.* No bones about it: Pleistocene mitogenomes reconstructed from the environmental DNA of permafrost. (In Prep).

10. Hider, J. *et al.* Examining pathogen DNA recovery across the remains of a 14th century Italian monk (St. Brancorsini) infected with Brucella melitensis. (In Prep).