Plagued by a cryptic clock

New insights from the global phylogeny of *Yersinia pestis*

This manuscript (permalink) was automatically generated from ktmeaton/plague-global-phylo-paper@8ad2585 on June 17, 2021.

Authors

- **Katherine Eaton**
  0000-0001-6862-7756 · ktmeaton
  Ancient DNA Centre, McMaster University; Department of Anthropology, McMaster University · Funded by Social Sciences and Humanities Research Council XXXXXXXX

- **Leo Featherstone**
  0000-0002-8878-1758
  The Peter Doherty Institute For Infection and Immunity, University of Melbourne

- **Sebastian Duchene**
  0000-0002-2863-0907 · sebastianduchene
  The Peter Doherty Institute For Infection and Immunity, University of Melbourne

- **Hendrik Poinar**
  0000-0002-0314-4160
  Ancient DNA Centre, McMaster University; Department of Anthropology, McMaster University

Abstract

Introduction

Plague has an impressively long and expansive history as a human disease. The earliest evidence of the plague bacterium, *Yersinia pestis*, comes from ancient DNA studies, dating its emergence to at least the Neolithic [andradesvaltuena2017StoneAgePlague?,rascovan2019EmergenceSpreadBasal?]. Since then, *Y. pestis* has traveled extensively due to ever-expanding global trade networks and the ability to infect a wide variety of mammalian hosts [yue2017TradeRoutesPlague?,perry1997YersiniaPestisEtiologic?]. Few regions of the ancient and modern world remain untouched by this disease, as plague has an established presence on every continent except Oceania [who2017Plague?].

Accompanying this prolific global presence is unnervingly high mortality. The infamous medieval Black Death is estimated to have killed more than half of Europe's population [benedictow2004BlackDeath13461353?]. This virulence can still be observed in the post-antibiotic era, where case fatality rates range from 22-71% [bertheratPlagueWorld2019?]. As a result, plague maintains its status as a disease that is of vital importance to current public health initiatives.

This high priority disease status is unsurprising given that *Y. pestis* is a member of the Enterobacteriaceae family. This family includes other notable pathogens such as *Escherichia coli* and *Salmonella typhi* that are commonly transmitted by contaminated food and water. In comparison, the plague bacterium is unique among this family due to a striking difference in host habitat and transmission. *Y. pestis* commonly resides in the blood of its mammalian hosts and can be transmitted to new hosts through an infectious fleabite [gage2006RecentTrendsPlague?]. Furthermore, this unique mechanism evolved relatively recently, possibly around the 1st millennium BC [rasmussen2015EarlyDivergentStrains?], long after *Y. pestis* emerged as a monomorphic clone of the enteric pathogen *Yersinia pseudotuberculosis* [achtman1999YersiniaPestisCause?].

While the population structure of *Y. pseudotuberculosis* is well-defined, [laukkanen-ninios2011PopulationStructureYersinia?,seecharran2017PhylogeographicSeparationFormation?], the phylogenetic patterning of *Y. pestis* remains cryptic. Populations of *Y. pestis* have been historically categorized according to a a vast array of historical, ecological, biochemical, and molecular characteristics. As a result, disparate sub-typing systems have emerged over the years to differentiate lineages of plague [qi2016TaxonomyYersiniaPestis?]. It has thus been argued that the taxonomy of *Y. pestis* should be revised and consolidated according to the latest global phylogenetic analysis [kutyrev2018PhylogenyClassificationYersinia?].

Unfortunately, there are a number of obstacles that have stalled large-scale phylogenetic analysis. The first challenge is data availability, both in terms of the genomic sequences, as well as the metadata required for interpretation. Genomic sampling of *Y. pestis* has recently intensified [zhou2020EnteroBaseUserGuide?], thus providing exceptional new datasets for statistical inference. This intensive sampling has produced over 1000 *Y. pestis* genomes that are now publicly available, with tremendous diversity spanning five continents and 5000 years of human history. Unfortunately, the majority of these genomic records lack curated metadata, such as sampling date and location, which are crucial variables in testing population structure.

The second major obstacle is an apparent lack of temporal structure in *Y. pestis*. Detecting temporal signal and estimating a molecular clock model are general pre-requisites for sophisticated methods of quantifying population structure [volz2020IdentificationHiddenPopulation?]. However, there has been significant debate concerning whether *Y. pestis* can be appropriately modeled using the available clock methods [cui2013HistoricalVariationsMutation?,wagner2014YersiniaPestisPlague?,spyrou2019Phylogeog

raphySecondPlague?]. To some extent, this debate can be explained by different *Y. pestis* datasets, which have been shown to produce dramatically different patterns of temporal signal [duchene2016GenomescaleRatesEvolutionary?]. Therefore, it is uncertain whether the new intensively sampled genomes will bring clarity or greater uncertainty.

In response to these debates and obstacles, this paper proposes a theoretical and methodological shift in plague genomics. Rather than conceptualizing *Y. pestis* as a conglomerate species, we highlight how novel insight emerges through analyzing *Y. pestis* sub-populations in isolation. To accomplish this shift in discourse, we focus on four objectives, specifically to:

1. Curate and contextualize the most recent *Y. pestis* genomic metadata.
2. Review and critique our current understanding of *Y. pestis* population structure.
3. Conduct robust and nuanced molecular clock analyses.
4. Identify key areas of phylogenetic uncertainty to be expanded on in future research.

Progress towards these key objectives is anticipated to benefit both prospective studies of plague, such as environmental surveillance and outbreak monitoring, and retrospective studies, which seek to date emergence and spread of past pandemics.

References

Appendix