Plagued by a cryptic clock

Insight and issues from the global phylogeny of *Y. pestis*.

This manuscript (permalink) was automatically generated from ktmeaton/obsidian-public@8c699f47 on September 11, 2021.

Authors

- **Katherine Eaton** 0000-0001-6862-7756
  McMaster Ancient DNA Center; Department of Anthropology, McMaster University
- **Leo Featherstone** 0000-0002-8878-1758
  The Peter Doherty Institute For Infection and Immunity , University of Melbourne
- **Sebastian Duchene** 0000-0002-2863-0907
  The Peter Doherty Institute For Infection and Immunity , University of Melbourne
- **Ann Carmichael**
  Indiana University Bloomington
- **Nükhet Varlık** 0000-0001-6870-5945
  University of South Carolina
- **Brian Golding** 0000-0002-7575-0282
  Department of Biology, McMaster University
- **Eddie Holmes** 0000-0001-9596-3552
  University of Sydney
- **Hendrik Poinar** 0000-0002-0314-4160
  McMaster Ancient DNA Center; Department of Anthropology, McMaster University

Introduction

- Plague has a powerful and lasting legacy. Tremendous societal change.
- Humanity has been visited by many waves of plague pandemics
- Bronze Age to current epidemics, has visited every continent, and remains endemic in x of those.

Results and Discussion

Population Structure

A critical step in reconstructing the evolutionary history of an organism is exploring the degree of population or genetic structure. When populations subdivide and diversify in isolation, the

can add nuance to spatiotemporal analyses, particularly when populations

through population-specific models [1]. For example… On the other hand, failing to account for how population subdivide and diversify in isolation can result in fundamentally different interpretations [2]. Such as… . We therefore began by examining how existing classification systems…

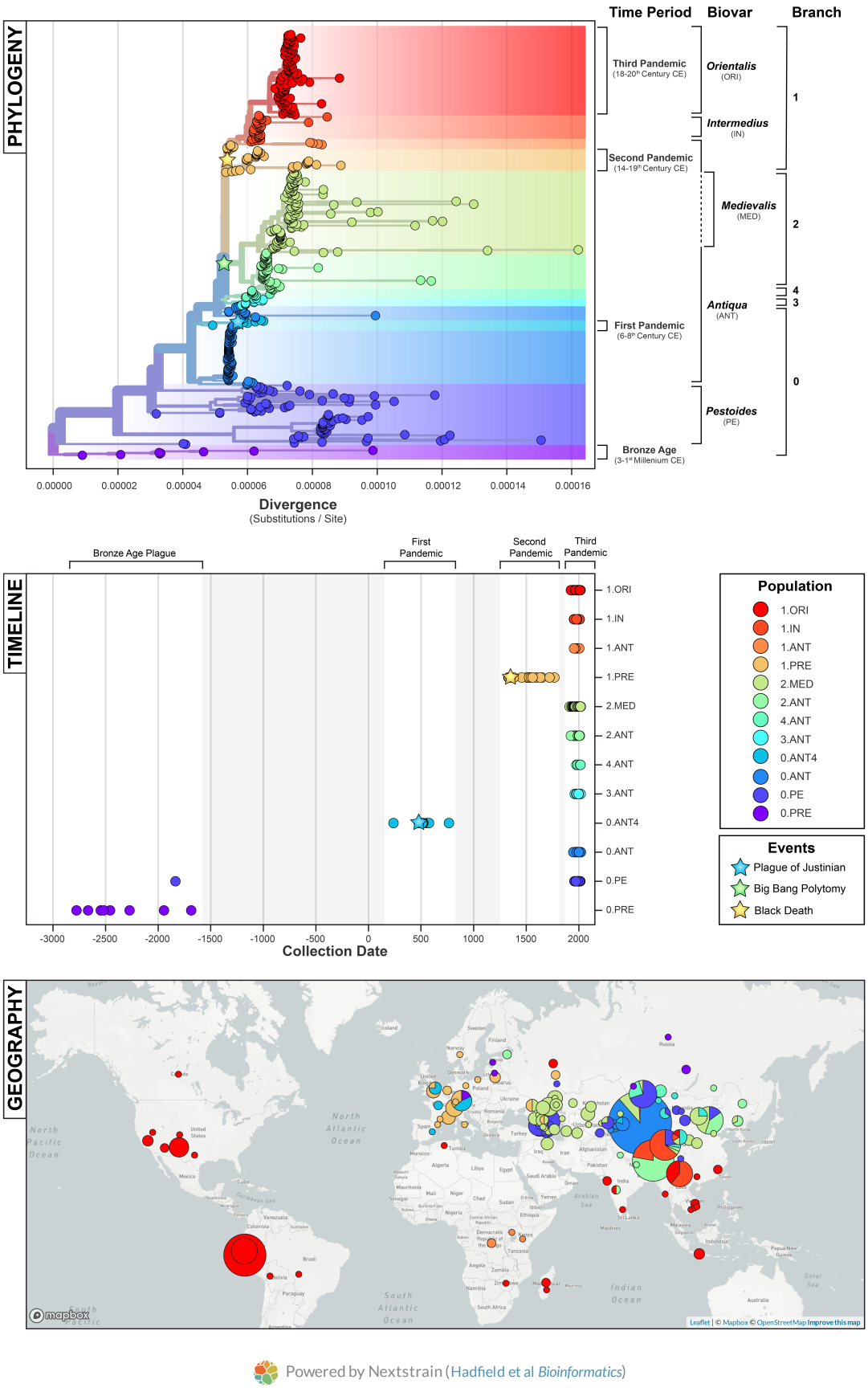
To explore the global population structure of *Y. pestis*, we estimated a maximum-likelihood phylogeny from 601 genomes, including 540 modern (89.9%) and 61 ancient (10.1%) strains. In addition, two genomes of the outgroup taxa *Yersinia pseudotuberculosis* were included to root the tree. The alignment consisted of 10,249 variant positions exclusive to *Y. pestis*, with 3,844 sites shared by at least two strains. Following phylogenetic estimation, we pruned the outgroup taxa *Y. pseudotuberculosis* from the tree to more closely examine the genetic diversity of *Y. pestis*.

In Figure 1, we contextualize the maximum-likelihood phylogeny using three nomenclature systems: the major branches, biovars, and time periods. In the following section, we compare how each system defines the population structure of *Y. pestis*, and the uncertainty surrounding these incongruent divisions.

–> Brief critique of the strengths and weaknesses of each.

Biovar

The oldest system to date is the biovar nomenclature, which uses phenotypic differences to define population structure. *Y. pestis* can be categorized into four classical biovars: *antiqua* (ANT), *medievalis* (MED), *orientalis* (ORI), and *microtus*/*pestoides* (PE) [3] [4] . Non-classical biovars have also been introduced, such as the *intermedium* biovar (IN), which reflects a transitional state from *antiqua* to *orientalis* [5]. The biovar system is simple and comprehensible as it largely focuses on two traits: the ability to ferment glycerol and reduce nitrate [4]. However, this simplicity is offset by the growing recognition of regional inconsistencies in metabolic profiles [6], which weakens its broader applicability. This is exacerbated by the sequencing of non-viable, 'extinct' *Y. pestis*, for which metabolic sub-typing is impossible [7]. Researchers have responded to this uncertainty in a variety of ways, by creating pseudo-biovars (PRE) [8] or extrapolating existing ones [9]. Other still have foregone the *biovar* nomenclature altogether in favor of locally-developed taxonomies [6]. Despite extensive research, it remains unclear which traits, if any, can be used to classify *Y. pestis* into distinct populations at a global scale.



Powered by Nextstrain (Hadfield et al *Bioinformatics*)

Figure 1: The phylogenetic, temporal, and geographic structure of *Yersinia pestis*. Top: The maximum-likelihood phylogeny. Middle: The timeline of collection dates. Bottom: The global geographic distribution.

Major Branch

In contrast to the biovar nomenclature which emphasizes phenotypes, the major branch nomenclature focuses on the evolutionary relationships between strains. This system divides the global phylogeny of *Y. pestis* into populations according to their relative position to the "Big Bang" polytomy [10]. All lineages that diverged prior to this polytomy are grouped into Branch 0 and those diverging after form Branches 1-4. Because this multifurcation plays such a central role in this system, there is great interest in estimating its timing and geographic origins [11]. However, the epidemiological significance of the "Big Bang" polytomy remains unclear, as no definitive phenotype has been identified that correlates with the observed branching pattern [**???**]. While the major branch system excels at reconstructing the evolutionary relationships between candidate populations, it struggles to connect these relationships to meaningful biological changes.

Time Period

As previously mentioned, the sequencing of ancient *Y. pestis* poses a problem for classification, as direct metabolic testing is impossible for these non-viable samples. Ancient DNA researchers thus use an alternative strategy, by incorporating contextual evidence such as the collection date or associated time period. The known genetic diversity of *Y. pestis* has been most commonly divided into four time periods: the Bronze Age (3rd - 1st millenium BCE) [8] , the First Pandemic, (6th - 8th century CE) [9], the Second Pandemic (14th - 18th century CE) [12], and the Third Pandemic (19th - 20th century CE) [10].

The key strength of this nomenclature is that it provides an excellent foundation for interdisciplinary discourse. However, this system runs the risk of grouping unrelated populations, as contemporaneous strains have been observed to have distinct evolutionary histories [13]. Furthermore, there is growing awareness of the temporal overlap of the Second and the Third Pandemics. Previously, the temporal extents of these events were mutually exclusive, dating from the 14th to 18th century, and the late 19th to mid-20th century respectively 14. Recent historical scholarship has contested this claim, and demonstrated that these constraints are a product of a Eurocentric view of plague [15]. The Second Pandemic is now recognized to have extended into at least the 19th Century in the Ottoman Empire [16; 17. Similarly, the Third Pandemic is now hypothesized to have began as early as the 18th century in southern China [18]. Unfortunately, this period of overlap remains unsampled, thus it is unclear where exactly to draw a genetic distinction, if it even exists, between these pandemic events.

In addition, some populations are curiously excluded from the time period/pandemic nomenclature. In particular, Branch 2 populations emerged at the same time as, but separate from, the Second Pandemic and have been associated with high mortality epidemics [19]. In particular, the *medievalis* population (2.MED) has spread throughout Asia (Figure 1) and was observed to have the fastest spread velocity of any *Y. pestis* lineage [18]. Given this epidemiological significance, it is surprising that Branch 2 populations have been largely overlooked in the pandemic taxonomy of *Y. pestis*. As ancient DNA sampling strategies expand out of Europe, it will be important to consider how to adapt and expand the three pandemic nomenclature to encompass this new diversity.

Uncertainty

In light of this uncertainty and inconsistencies, no classification system comprehensively represents the global population structure of *Y. pestis*. However, a combined approach has been previously used in large comparative studies of *Y. pestis* [10,20]. We therefore take the intersection of the three taxonomic systems discussed previously, to define 12 populations for statistical analysis (Figure 1). In

the following sections, we highlight the novel insight that comes from explicitly incorporating this population structure, and the key areas of uncertainty that remain.

Phylodynamics

A long-standing line of inquiry in plague phylogenetics has been estimating evolutionary rates in order to date internal nodes. Key areas of the phylogeny that have been intensively researched are the first emergence of *Y. pestis* in human populations [8], the "Big Bang" polytomy [11], and the onset of past pandemics [7,9,10]. Recent technological advancements, such as ancient DNA sequencing and new molecular clock methods, have enabled researchers to reach further back in time with increasingly complex models. But despite this intensive interest and methodological advancement, *Y. pestis* remains notoriously difficult to model using a molecular clock approach.

This difficulty can largely be attributed to the substantial rate variation that has been documented across the phylogeny of *Y. pestis* [10,12]. As a result, considerable debate has emerged over whether *Y. pestis* has absolutely no temporal signal [9], or if populations have such distinct rates that a species-wide signal is obscured [12,21]. This uncertainty has produced radically different temporal models between studies, with node dates shifted by as much as several millennia [8,10]. Thus a comprehensive understanding of plague's molecular clock, or lack thereof, is necessary before we can begin to untangle when and where this disease appeared in the past.

Rate Variation

In light of this evolutionary debate, we began by exploring the extent of rate variation present in our updated genomic dataset, which is notably larger and more diverse than those used in previous studies [12]. Given this expanded diversity, it is unsurprising that a root-to-tip regression on collection date reproduces the finding that substitution rates in *Y. pestis* are poorly represented by a simple linear model or "strict clock" (Figure 2). While there is a statistically significant relationship between date and genetic distance (P-value=4.959×10^{-13}), an extremely low coefficient of determination (R^2=0.09) indicates there is tremendous variation that is not accounted for.

Thus far, the hypotheses proposed to explain this variation have primarily focused on biological processes, such as the cycling between endemic and epidemic phases [**???**] and adaptations to new environments [**???**]. However, we argue that several methodological factors must first be taken into account, before investigating more complex ecological factors such as host and landscape.

Time Dependency

One striking factor that is methodological in nature is the time dependency of molecular rates. In Figure 2, we show how rate variation in *Yersinia pestis* correlates with the sampling time frame, in which populations sampled over several millennia (Bronze Age) have less variation than those sampled over multiple centuries (Second Pandemic) or decades (*medievalis*). This correlation is a well-known and widely-documented phenomenon in many organisms [22] and occurs due to two conflicting signals: a slower, long-term substitution rate combined with a higher, short-term mutation rate.

Separating out these signals can be extremely challenging and failure to do so can have significant consequences when estimating and interpreting a molecular clock. Of particular concern for epidemiological investigations is the risk of artificially inflating the substitution rate due to transient mutations, which will lead to underestimating internal node dates. With regards to plague genomics, this may result in incorrect molecular dates linked to key historical events, such as the emergence of pandemic populations. Because of this risk, we first evaluate the presence of spurious mutations in our dataset before attempting to estimate a molecular clock model.

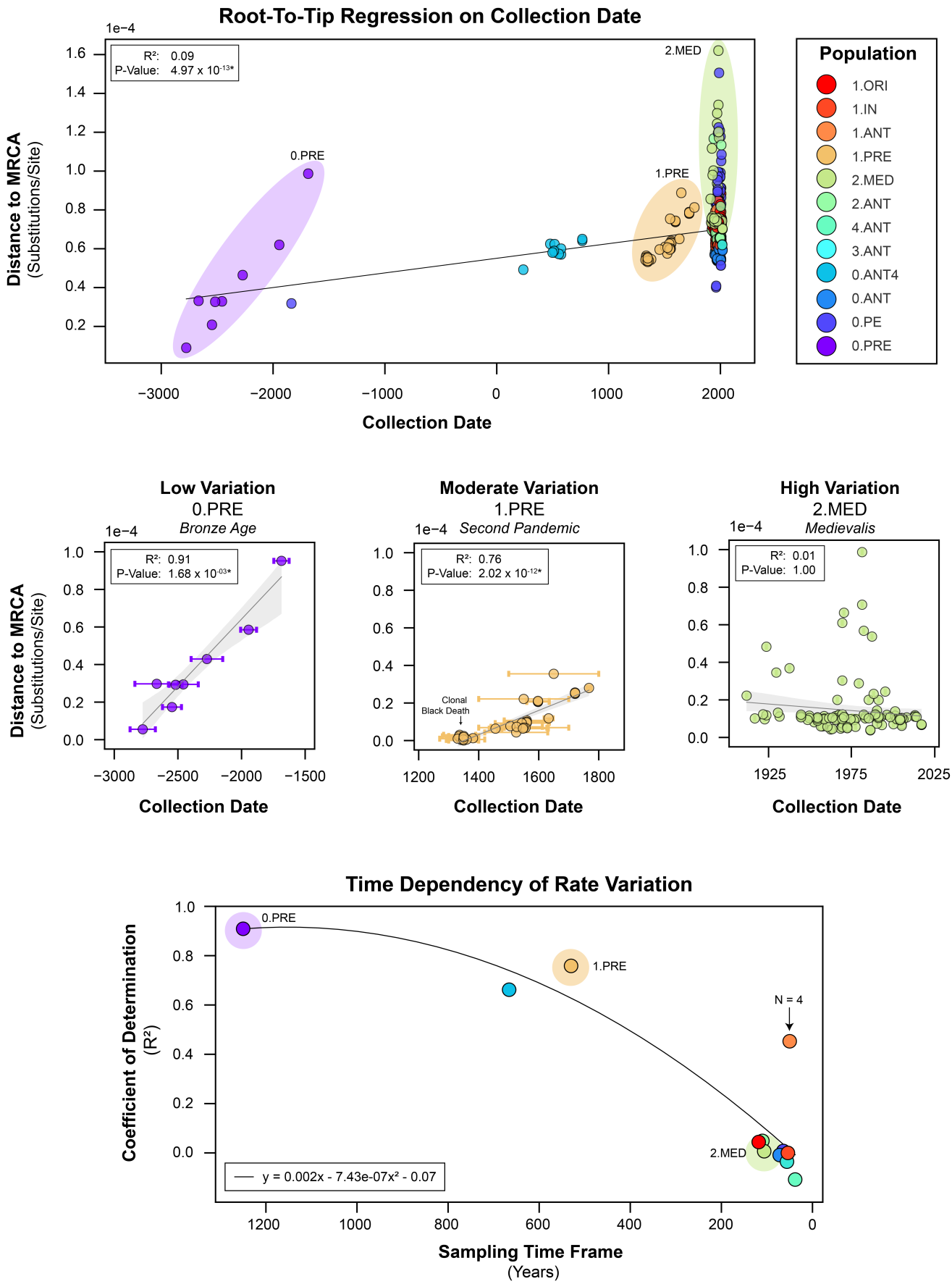


Figure 2: Rate variation in *Yersinia pestis* as observed through a regression of root-to-tip distances on collection date. Top: A species-wide model using all genomes from the maximum-likelihood phylogeny. Middle: Population-specific models based on extracted subtrees from the phylogeny. Bottom: The time-dependency of population-specific rate variation on the sampling time frame.

Slow, Long-Term Substitution Rate

Y. pestis is particularly susceptible to the time-dependency of molecular rates, as it has one of the slowest substitution rates observed among bacterial pathogens [21]. The substitution rate of *Y. pestis* has previously been estimated to range from 1×10^{-8} to 2×10^{-8} substitutions/site/year, [10,12] or 1 substitution every 10 - 25 years. In application, this means that *Y. pestis* lineages often cannot be differentiated until multiple decades have passed, a concept generally referred to as the phylodynamic threshold [23].

A historical example of this can be seen during the Second Pandemic, where isolates dated to the medieval Black Death (1348-1353) are nearly indistinguishable clones (Figure 2). A modern example is the *medievalis* population, where the youngest samples (2010s) have diverged little compared to those from a century prior (1910s). This highlights a significant limitation and cautionary note for *Y. pestis* phylogenetics, as comparisons over short time scales (10 to 100 years) may have limited resolving power.

High, Short Term Mutation Rate

Furthermore, the little phylogenetic signal (ie. substitutions) that accumulates in the population may be easily obscured by spurious mutations in a single sample. The *medievalis* population is an extreme example of this, as several samples collected between 1970 and 1980 are exceptionally divergent (SI Figure 2). This short-term diversity is largely due to mutations observed in only a single-isolate, which manifests as long terminal branches in the maximum-likelihood phylogeny (Figure 1). The presence of transient mutations, or long branches, is not isolated to the *medievalis* population and also strongly impacts the *pestoides* 0.PE, *intermedium* 1.IN, and *orientalis* (1.ORI) populations (Figure 4 SI). Given the extensive presence of these apparent outliers, inclusion or exclusion of these samples may have profound impacts on the models used to estimate a molecular clock.

> So what?

Bayesian Evaluation of Temporal Signal

To investigate the degree of temporal signal in the genomic data, we performed a Bayesian evaluation of temporal signal (BETS) test. In brief, this method compares the likelihoods of two different phylogenetic models, one where the true collection dates are used and the other where all collection dates are assumed to be contemporaneous. A comparison of the model likelihoods, or Bayes Factors, is then used to assess the degree of temporal signal in the dataset.

Unfortunately, a BETS test was inconclusive when attempting to fit a clock model to all *Y. pestis* populations combined. The Bayesian analysis exhibited poor sampling of the clock parameters, for both a strict and relaxed clock, even when attempting to reduce sources of variation such as down-sampling the number of genomes, using fixed tip dates, and fixing the tree topology. This observation is consistent with previous analyses [9] where robust estimates of model parameters could not be estimated, thus leading to the conclusion that *Y. pestis* lacks temporal signal. However, as other studies have suggested data composition is a strong determinant of temporal signal [21] we then repeated the BETS test by modeling each of the 12 populations of *Y. pestis* in isolation.

In contrast to the species-wide model, separating the genomic dataset by population dramatically stabilized the Bayesian analysis. All clock model parameters were well-sampled with effective sample

sizes (ESS) greater than 200 and temporal signal was detected in 9 out of 12 populations (Figure 3, SI Table 1). We can draw several conclusions from these nuanced models...

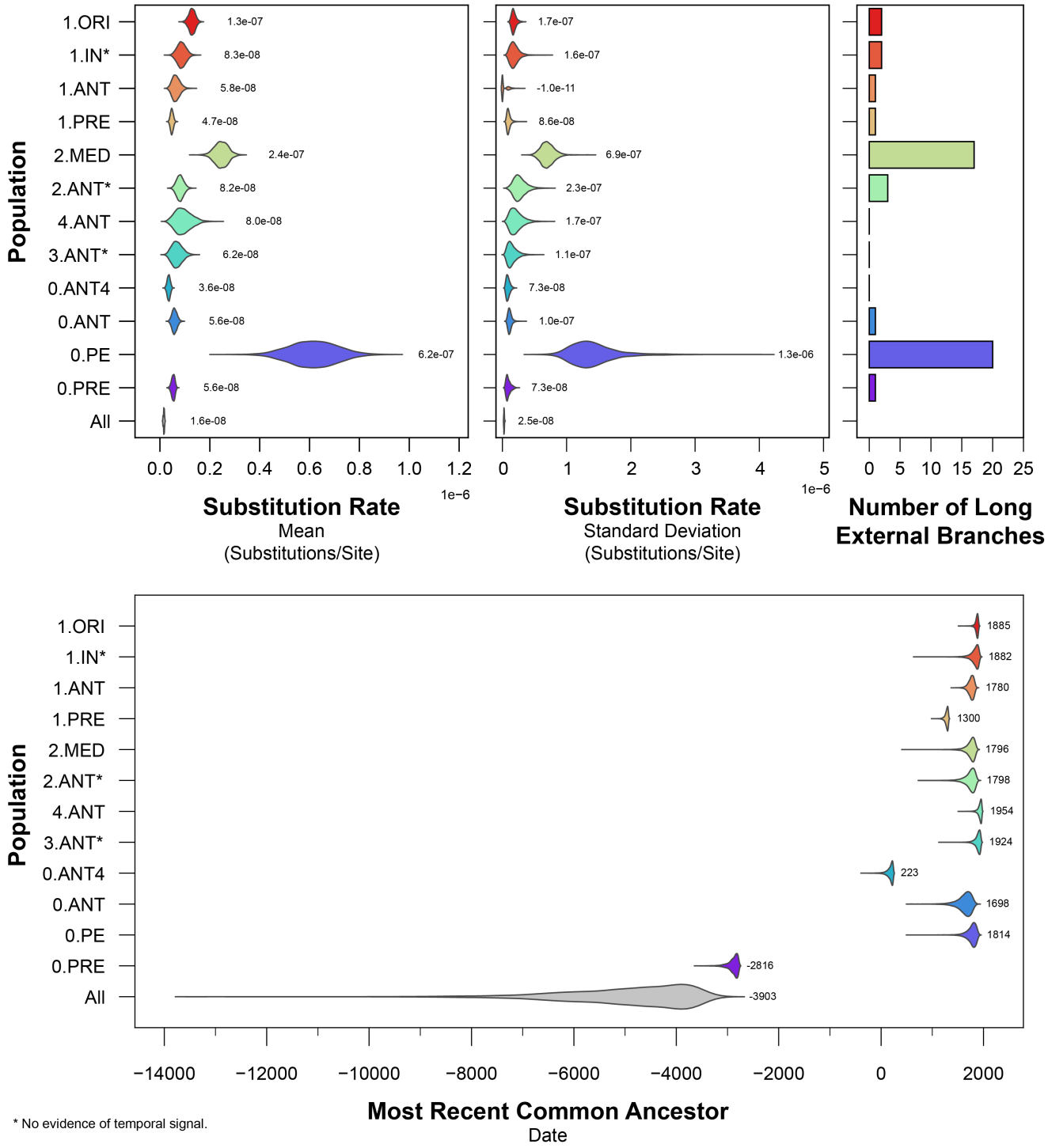


Figure 3: The 95% HPD estimates of the substitution rate and the date of the most recent common ancestor (MRCA) by population. Each distribution is annotated with the peak value.

A New Global Rate

The first finding is that substitution rates in *Y. pestis* have been considerably underestimated. The BETS analysis on the non-segregated data, which was highly unstable, fell within the published range of 1-2 x 10^{-8} at 1.6 x 10^{-8}. However, no *Y. pestis* population was observed to have a rate this slow, and instead ranged from 3.6 x 10^{-8} during the First Pandemic (0.ANT4 to 6.16 x 10^{-7} in the *pestoides* biovar (0.PE). This study therefore reports the substitution rate of *Y. pestis* to be much higher than previously thought and more comparable to bacteria such as *Mycobacterium tuberulcosis* [21].

As mentioned previously, inaccurate substitution rates have important consequences for node-dating. Specifically, underestimating the global rate leads to overestimating the age of the MRCA.

Unreliable Outliers

The second finding is the outlier populations. Long external branches correlate with increased substitution rates, and in turn, younger MRCA estimates. The tMRCA of the *pestoides* biovar at 1814 is highly incongruent with the maximum-likelihood phylogeny in which the *pestoides* population is ancestral to the First Pandemic (6th to 8th century). Similarly, the estimate rates and dates associated with *medievalis* are likely untrustworthy?

> Should be very cautious trying to date inter-population nodes, ie. the Big Bang Polytomy. Because we don't fully understand how the rate changes in between them.

Phylogeography

Conclusions

Methods

References

1. **Genomic Insights into a Sustained National Outbreak of Yersinia pseudotuberculosis**
   Deborah A. Williamson, Sarah L. Baines, Glen P. Carter, Anders Gonçalves da Silva, Xiaoyun Ren, Jill Sherwood, Muriel Dufour, Mark B. Schultz, Nigel P. French, Torsten Seemann, … Benjamin P. Howden
   Genome Biology and Evolution (2016-12-01) https://doi.org/10.1093/gbe/evw285
   DOI: 10.1093/gbe/evw285

2. **Did Our Species Evolve in Subdivided Populations across Africa, and Why Does It Matter?**
   Eleanor M. L. Scerri, Mark G. Thomas, Andrea Manica, Philipp Gunz, Jay T. Stock, Chris Stringer, Matt Grove, Huw S. Groucutt, Axel Timmermann, G. Philip Rightmire, … Lounès Chikhi
   Trends in Ecology & Evolution (2018-08) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6092560/
   DOI: 10.1016/j.tree.2018.05.005 · PMID: 30007846 · PMCID: PMC6092560

3. **Variétés de l'espèce Pasteurella pestis**
   R. Devignat
   Bulletin of the World Health Organization (1951) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2554099/
   PMID: 14859080 · PMCID: PMC2554099

4. **Comparative and evolutionary genomics of *Yersinia pestis***
   Dongsheng Zhou, Yanping Han, Yajun Song, Peitang Huang, Ruifu Yang
   Microbes and Infection (2004-11-01) http://www.sciencedirect.com/science/article/pii/S1286457904002357
   DOI: 10.1016/j.micinf.2004.08.002

5. **Genotyping and Phylogenetic Analysis of Yersinia pestis by MLVA: Insights into the Worldwide Expansion of Central Asia Plague Foci**
   Yanjun Li, Yujun Cui, Yolande Hauck, Mikhail E. Platonov, Erhei Dai, Yajun Song, Zhaobiao Guo, Christine Pourcel, Svetlana V. Dentovskaya, Andrey P. Anisimov, … Gilles Vergnaud
   PLOS ONE (2009-06-22) https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006000
   DOI: 10.1371/journal.pone.0006000

6. **Phylogeny and classification of *Yersinia pestis* through the lens of strains From the plague foci of Commonwealth of Independent States**
   Vladimir V. Kutyrev, Galina A. Eroshenko, Vladimir L. Motin, Nikita Y. Nosov, Jaroslav M. Krasnov, Lyubov M. Kukleva, Konstantin A. Nikiforov, Zhanna V. Al'khova, Eugene G. Oglodin, Natalia P. Guseva
   Frontiers in Microbiology (2018-05-25) https://www.frontiersin.org/article/10.3389/fmicb.2018.01106/full
   DOI: 10.3389/fmicb.2018.01106

7. **A draft genome of *Yersinia pestis* from victims of the Black Death**
   Kirsten I. Bos, Verena J. Schuenemann, G. Brian Golding, Hernán A. Burbano, Nicholas Waglechner, Brian K. Coombes, Joseph B. McPhee, Sharon N. DeWitte, Matthias Meyer, Sarah Schmedes, … Johannes Krause
   Nature (2011-10) http://www.nature.com/articles/nature10549
   DOI: 10.1038/nature10549

8. **Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago**
Simon Rasmussen, Morten Erik Allentoft, Kasper Nielsen, Ludovic Orlando, Martin Sikora, Karl-Göran Sjögren, Anders Gorm Pedersen, Mikkel Schubert, Alex Van Dam, Christian Moliin Outzen Kapel, … Eske Willerslev
Cell (2015-10-22) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4644222/
DOI: 10.1016/j.cell.2015.10.009 · PMID: 26496604 · PMCID: PMC4644222

9. ***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

10. **Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis***
Y. Cui, C. Yu, Y. Yan, D. Li, Y. Li, T. Jombart, L. A. 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

11. **How a microbe becomes a pandemic: a new story of the Black Death**
Monica H. Green
The Lancet Microbe (2020-12-01) https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(20)30176-2/abstract
DOI: 10.1016/s2666-5247(20)30176-2

12. **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

13. **Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague**
Maria A. Spyrou, Rezeda I. Tukhbatova, Chuan-Chao Wang, Aida Andrades Valtueña, Aditya K. Lankapalli, Vitaly V. Kondrashin, Victor A. Tsybin, Aleksandr Khokhlov, Denise Kühnert, Alexander Herbig, … Johannes Krause
Nature Communications (2018-06-08) https://www.nature.com/articles/s41467-018-04550-9
DOI: 10.1038/s41467-018-04550-9

14. **Natural history of plague: perspectives from more than a century of research**
Kenneth L. Gage, Michael Y. Kosoy
Annual Review of Entomology (2005)
DOI: 10.1146/annurev.ento.50.071803.130337 · PMID: 15471529

15. **New Science and Old Sources: Why the Ottoman Experience of Plague Matters**
Nükhet Varlık
The Medieval Globe (2014) https://scholarworks.wmich.edu/tmg/vol1/iss1/9/

16. **The Ottomans During the Global Crises of Cholera and Plague: The View from Iraq and the Gulf.**
Isacar A. Bolaños
International Journal of Middle East Studies (2019-11) http://www.cambridge.org/core/journals/international-journal-of-middle-east-studies/article/ottomans-during-the-global-crises-of-cholera-and-plague-the-view-from-iraq-and-the-gulf/82A6A0E7CF8BD2138FA230AE4189057E
DOI: 10.1017/s0020743819000667

17. **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

18. **Wet climate and transportation routes accelerate spread of human plague**
Lei Xu, Leif Chr. Stige, Kyrre Linné Kausrud, Tamara Ben Ari, Shuchun Wang, Xiye Fang, Boris V. Schmid, Qiyong Liu, Nils Chr. Stenseth, Zhibin Zhang
Proceedings of the Royal Society B: Biological Sciences (2014-04-07) https://royalsocietypublishing.org/doi/10.1098/rspb.2013.3159
DOI: 10.1098/rspb.2013.3159

19. **Evolution and circulation of Yersinia pestis in the Northern Caspian and Northern Aral Sea regions in the 20th-21st centuries**
Galina A. Eroshenko, Nikolay V. Popov, Zhanna V. Al'khova, Lyubov M. Kukleva, Alina N. Balykova, Nadezhda S. Chervyakova, Ekaterina A. Naryshkina, Vladimir V. Kutyrev
PLOS ONE (2021-02-11) https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0244615
DOI: 10.1371/journal.pone.0244615

20. ***Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity**
Giovanna Morelli, Yajun Song, Camila J. Mazzoni, Mark Eppinger, Philippe Roumagnac, David M. Wagner, Mirjam Feldkamp, Barica Kusecek, Amy J. Vogler, Yanjun Li, … Mark Achtman
Nature Genetics (2010-12)
DOI: 10.1038/ng.705 · PMID: 21037571 · PMCID: PMC2999892

21. **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

22. **Time Dependency of Molecular Rate Estimates and Systematic Overestimation of Recent Divergence Times**
Simon Y. W. Ho, Matthew J. Phillips, Alan Cooper, Alexei J. Drummond
Molecular Biology and Evolution (2005-07-01) https://doi.org/10.1093/molbev/msi145
DOI: 10.1093/molbev/msi145

23. **Temporal signal and the phylodynamic threshold of SARS-CoV-2**
Sebastian Duchene, Leo Featherstone, Melina Haritopoulou-Sinanidou, Andrew Rambaut, Philippe

Lemey, Guy Baele
Virus Evolution (2020-07-01) https://doi.org/10.1093/ve/veaa061
DOI: 10.1093/ve/veaa061

Supplementary Information

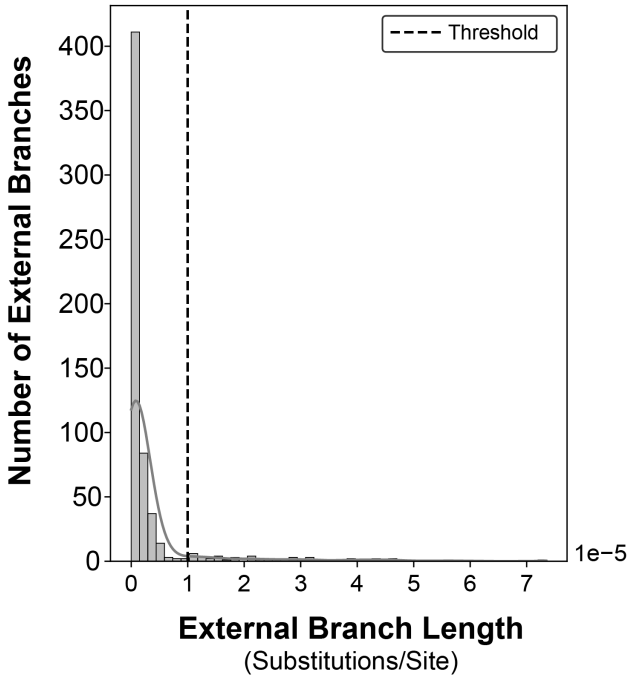


Figure 4: The distribution of external branch lengths across the maximum-likelihood phylogeny. The threshold to be considered a long external branch is set at 1e-5 substitutions/site.

Table 1: Model selection and log marginal likelihoods obtained from a Bayesian evaluation of temporal signal (BETS) test. SC: strict clock. UCLN: uncorrelated lognormal relaxed clock. *0.PRE had temporal signal according to a SC, although the ultrametric UCLN model had the highest likelihood.

Population	N	Time Span (Years)	Best Model	Bayes Factor	Strict Clock Ultrametric	Strict Clock Dates	Relaxed Clock Ultrametric	Relaxed Clock Dates
1.ORI	117	118	UCLN	35.670	-5899691.115	-5899661.493	-5899601.413	-5899565.743
1.IN	39	54	–	-10.331	-5891399.168	-5891402.697	-5891344.183	-5891354.514
1.ANT	4	50	UCLN	12.691	-5882596.155	-5882586.874	-5882594.555	-5882581.864
1.PRE	40	530	UCLN	44.082	-5888139.985	-5888129.886	-5888082.134	-5888038.053
2.MED	116	106	UCLN	3.902	-5920837.35	-5920732.774	-5919662.038	-5919658.136
2.ANT	54	110	–	-13.385	-5892876.227	-5892894.924	-5892791.269	-5892804.654
4.ANT	11	38	UCLN	3.609	-5886031.423	-5886034.116	-5886025.578	-5886021.969
3.ANT	11	56	–	-11.172	-5887496.544	-5887506.036	-5887494.669	-5887505.841
0.ANT4	12	666	UCLN	5.921	-5889525.703	-5889520.445	-5889501.725	-5889495.805
0.ANT	103	72	UCLN	13297.716	-5896014.089	-5896016.472	-5895879.702	-5882581.985
0.PE	85	64	UCLN	12.378	-5945602.843	-5945574.023	-5944626.698	-5944614.32
0.PRE*	8	1250	SC	83.002	-5892925.901	-5892842.899	-5892738.563	-5892741.377