Plague Denmark Paper

This manuscript (permalink) was automatically generated from ktmeaton/obsidian-public@d90f1c60 on July 13, 2021.

Authors

- **Katherine Eaton** 0000-0001-6862-7756 · ktmeaton
  McMaster University
- **Ravneet Sidhu**
  McMaster University
- **Jennifer Klunk** 0000-0002-6521-8516
  Daicel Arbor Biosciences
- **Julia Gamble** 0000-0001-7486-757X
  University of Manitoba
- **Jesper Boldsen** 0000-0002-2850-0934
  University of Southern Denmark; ADBOU
- **Ann Carmichael**
  Indian University Bloomington
- **Nükhet Varlık** 0000-0001-6870-5945
  University of South Carolina
- **Sebastian Duchene** 0000-0002-2863-0907 · sebastianduchene
  University of Melbourne
- **Leo Featherstone** 0000-0002-8878-1758
  University of Melbourne
- **Vaughan Grimes** 0000-0002-2177-3147
  Memorial University
- **G. Brian Golding** 0000-0002-7575-0282
  McMaster University
- **Hendrik N. Poinar** 0000-0002-0314-4160
  McMaster University

Abstract

> This is a placeholder from the original paper proposal. To be re-written once results are finalized.

Background

The sequencing of ancient pathogen genomes has resulted in vast advancements to our understanding of the second plague pandemic. However, due to limited sample availability, debate remains about the Plague's origins, routes of dissemination, genomic diversity, and persistence. Specifically, Scandinavia has a unique history with regards to plague persistence as it is home to the oldest known strain of plague to date and has been ravaged by historic epidemics, only for the plague to have disappeared from this region in the modern era.

Problem

It is unknown to what extent local plague reservoirs fed the recurring epidemics in Scandinavia as compared to the continual introduction of globally circulated strains. Previously studied historical records primarily derive from large commercial centres such as London, which are contrasted by countries such as Denmark where the archives have retained limited information about the spread of the plague and its impact on society.

Objectives

To confidently identify and sequence ancient *Yersinia pestis* from Danish archaeological sites across a wide geographic and temporal range. With the aim of performing genomic analyses to estimate the timing, spread, and evolutionary changes occurring within Danish plague strains as compared to neighboring regions.

Significance

First, there have been relatively few studies that explore the genetics of plague in Scandinavia across time and geography, particularly in Denmark. Second, this paper contributes to a larger body of epidemiological literature that considers the patterns and mechanisms by which diseases emerge, propagate, and go extinct.

Introduction

1. When does plague appear? The presence of LNBA plague in Sweden raises the possibility that plague may be an endemic disease for the region. However, most European. However, most evidence of plague comes from the pandemic times (Justinian, and Black Death). We want to know if Denmark is the same (novel disease) or different.

2. How is plague in Northern/Western Europe connected to Modern plague. There's the theory of a back migration out of Europe (Netherlands –> Russia). If this is true, then Denmark may also be involved. At 'worst' we expand the relationship of European regions connected to this lineage. At 'best' we capture where that transition from the Black Death clonal lineage to the post-Black Death lineage emerges.

3. Longitudinal study: The possibility to capture how plague affected the same people over time. (Highly regional, with 1 Danish strain?), or

4. Document the *pla* depletion/restoration events.

Results

Sites and Samples

326 individuals were sampled across 6 regions from 14 archaeological sites (Table 1). The site occupation dates span from the 10th to 18th centuries which encompasses the Viking Age (8th - 10th century), the Medieval Period (11th - 16th century) and the Early Modern Period (16th - 19th century) in Denmark.

Table 1: Summary of archaeological sites sampled in this study.

Region	Site Name	Site Code	Site Occupation	Samples	Plague Positive
Ribe	Ribe Gräbrødre	ASR 1015	1200 - 1560	53	5
	Ribe Lindegärden	ASR 2391	900 - 1000	5	0
		ASR 13/13II	900 - 1000	15	0
		ASR 13II	1200 - 1560	28	1
Viby	Nordby	FHM 3970	1050 - 1250	36	0
Horsens	Monastery Church	HOM 1272	1600 - 1800	50	0
	Ole Wormsgade	HOM 1649	1100 - 1500	17	2
	Sejet	HOM 1046	1150 - 1574	25	1
	Tirup	VKH 1201	1150 - 1350	12	1
Hågerup	Hågerup	ØHM 1247	1100 - 1555	7	1
Refshale	Refshale	Refshale	1100 - 1350	19	0
Viborg	Sct. Mikkel	JAH 1-77	1000 - 1529	4	0
	The Catholic Church	VSM 09264	1100 - 1529	6	0
	Sct. Mathias	VSM 855F/906F	1100 - 1529	23	0
	Sct. Drotten	VSM 902F	1100 - 1529	8	0
	Faldborg	VSM 29F	1100 - 1600	17	2
Total				325	13

Figure 1: Geographic map of archaeological sites by region. Stars indicate sites where *Yersinia pestis* was detected.

Plague Detection

13 individuals were identified as plague-positive based on a combination of PCR assays, shotgun sequencing, and targeted enrichment for the *Y. pestis* whole genome (Tables 2 and 3). Of the 13 individuals, 9 had chromosomal coverage sufficient for phylogenetic analyses.

Overall, plague was detected in 4% (13/325) of all individuals in this study (Table 1). When excluding plague-negative archaeological sites, this estimate rises to 8.2% (13/159) of individuals. *Y. pestis* was observed primarily in the Medieval Period, with no evidence of plague in the Viking Age settlements at Ribe Lindegärden or the Early Modern cemetery at Horsens.

> - Absence of plague in Viking/Early Modern could easily be a false negative.
> - These periods are sparsely sampled, with fewer sites and individuals sampled.
> - Are G25A and G25B two individuals from the same grave?

Table 2: Plague positive summary of high coverage genomes. Human (%) and plague (%) derive from shotgun estimates. The mean depth of coverage for the chromosome (CHROM) and the plasmids (pCD1, pMT1, pPCP1) are reported after targeted enrichment.

Arch ID	Project ID	Site	PCR	Human (%)	Plague (%)	CHROM	pCD1	pMT1	pPCP1
G16	D71	Ribe Gräbrødre	6/6	5.95	0.18	22.6	39.4	14.7	4.6
G861 x1035	D75	Ribe Gräbrødre	6/6	4.42	0.23	17.4	40.2	16.5	3.4
G25B x98	R36	Ribe Gräbrødre	6/6	8.41	0.25	24.0	51.8	14.9	5.8
G25A	D62	Ribe Gräbrødre	6/6	1.12	0.10	3.8	10.5	2.5	0.9
G207	D72	Ribe Gräbrødre	6/6	12.94	0.04	6.0	13.5	5.8	2.2
A146 x3011	P187	Sejet	6/6	0.68	0.01	4.9	18.4	6.6	52.2
G371	P212	Tirup	6/6	0.61	0.04	6.7	26.3	8.5	56.6
Gr GC 15	D51	Faldborg	6/6	0.67	0.05	9.0	25.4	8.1	2.0
A1480 x1480	P387	Ole Wormsgade	6/6	0.04	0.01	6.5	21.7	5.0	75.0

Table 3: Plague positive summary of low coverage genomes. Human (%) and plague (%) derive from shotgun estimates. The mean depth of coverage for the chromosome (CHROM) and the plasmids (pCD1, pMT1, pPCP1) are reported after targeted enrichment.

Arch ID	Project ID	Site	PCR	Human (%)	Plague (%)	CHROM	pCD1	pMT1	pPCP1
A1155 x1155	P384	Ole Wormsgade	4/6	0.11	0.01	1.1	4.8	1.4	19.6
Gr ID 319	R21	Faldborg	6/6	0.85	0.01	2.6	3.8	2.3	0.4
A19 X21	D24	Hågerup	6/6	0.55	0.01	2.6	6.1	1.9	0.7
X1265	P246	Ribe Lindegärden	6/6	0.03	0.01	0.1	0.1	0.1	3.2

Dating

I'm relying heavily on the discussion in Boldsen (2009) [1], as quoted here:

"The dating of individual skeletons is a fundamental problem in historical studies like this, and even the period of usage of each cemetery raises some serious problems. However, most cemeteries have at least some documentary sources broadly framing them in time. The most intensely studied skeletal samples, Tirup and Westerhus, are really the only exceptions in being dated solely on archaeological evidence (Kieffer-Olsen et al. 1986, Sivěn 2005)."

"In medieval graves the position of the arms in relation to the rest of the skeleton in the grave is the only feature that systematically indicates dating of the burial within the temporal frame provided by the period of usage of the cemetery. Arm position dating is primarily based on work by Redin (1976) and Kieffer-Olsen (1993). The successive stages of arm position from A (the arms besides the body) over B (hand joint over the lower part of the abdomen and usually found in the pelvis) and C (the forearms over the upper part of the abdomen and the elbows flexed in an approximately right angle) to D (the hands placed on the shoulders, forearms often crossed over the chest) have primarily been described by Kieffer-Olsen (1993) but Jantzen et al. (1994) have slightly modified the transition dates between the various stages.".

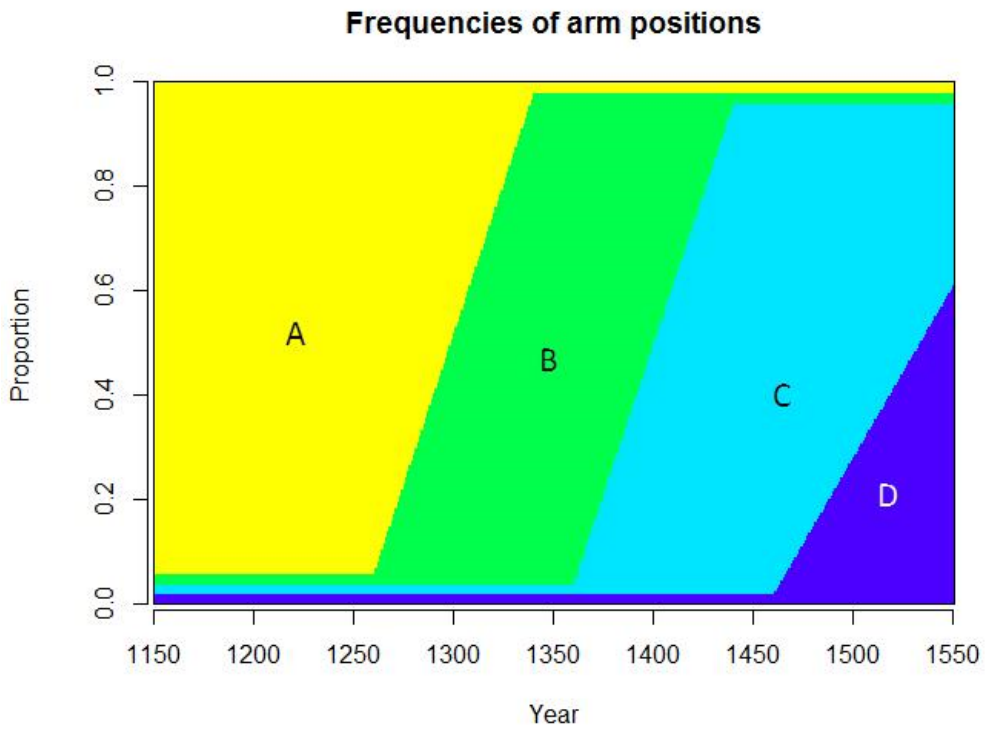

Figure 2: Arm position dating.

> **Question for Julia Gamble and Jesper Boldsen.** Can I use this diagram to broadly assign time periods to arm positions? If so, I can add greater resolution to Figure 3.
> - Position A: 1150 - 1325
> - Position B: 1250 - 1425
> - Position C: 1350 - 1550
> - Position D: 1450 - 1550

The skeletal and molecular dates of the 9 high coverage *Y. pestis* genomes are presented in Table 4 and Figure 3. Almo st all molecular dates had overlap with the archaeological dates as determined by the site occupation period and burial patterns. The exception to this pattern was individual G371 from the site of Tirup.

To investigate the dating disparity for G371, we performed 14C radiocarbon dating. The radiocarbon estimate largely agrees with the archaeological dates, with an estimated mean date of 1260 CE (+/- 75 yrs with 1 sigma). Thus there is robust evidence that the individual lived and died sometime between the late 12th and early 14th century

> **Question for Vaughan Grimes.** Is the interpretation and visualization of the radiocarbon date appropriate?

However, the associated *Y. pestis* genome is dated to the 15th century. While the 2 sigma distribution of the radiocarbon date does extends into the 15th century, there remains a substantial conflict between the date of the host and the associated pathogen.

Table 4: Summary of the *Y. pestis* molecular dates. The estimated tip date reflects the 95% highest posterior density.

ID	Region	Site	Site Occupation	Arm Position	Skeletal Date	Tip Date
G16	Ribe	Ribe Gräbrødre	1200 - 1560	C	1350 - 1550	1310 - 1388
G861 x1035	Ribe	Ribe Gräbrødre	1200 - 1560	C	1350 - 1550	1489 - 1567
G25B x98	Ribe	Ribe Gräbrødre	1200 - 1560	C	1350 - 1550	1327 - 1414
G25A	Ribe	Ribe Gräbrødre	1200 - 1560	C	1350 - 1550	1295 - 1375
G207	Ribe	Ribe Gräbrødre	1200 - 1560	C	1350 - 1550	1477 - 1551
A146 x3011	Horsens	Sejet	1150 - 1574	B	1250 - 1425	1397 - 1470
A1480 x1480	Horsens	Ole Wormsgade	1100 - 1500	?	?	1384 - 1473
G371	Horsens	Tirup	1150 - 1350	B	1250 - 1425	1419 - 1490
Gr GC 15	Viborg	Faldborg	1100 - 1600	C	1350 - 1550	1539 - 1655

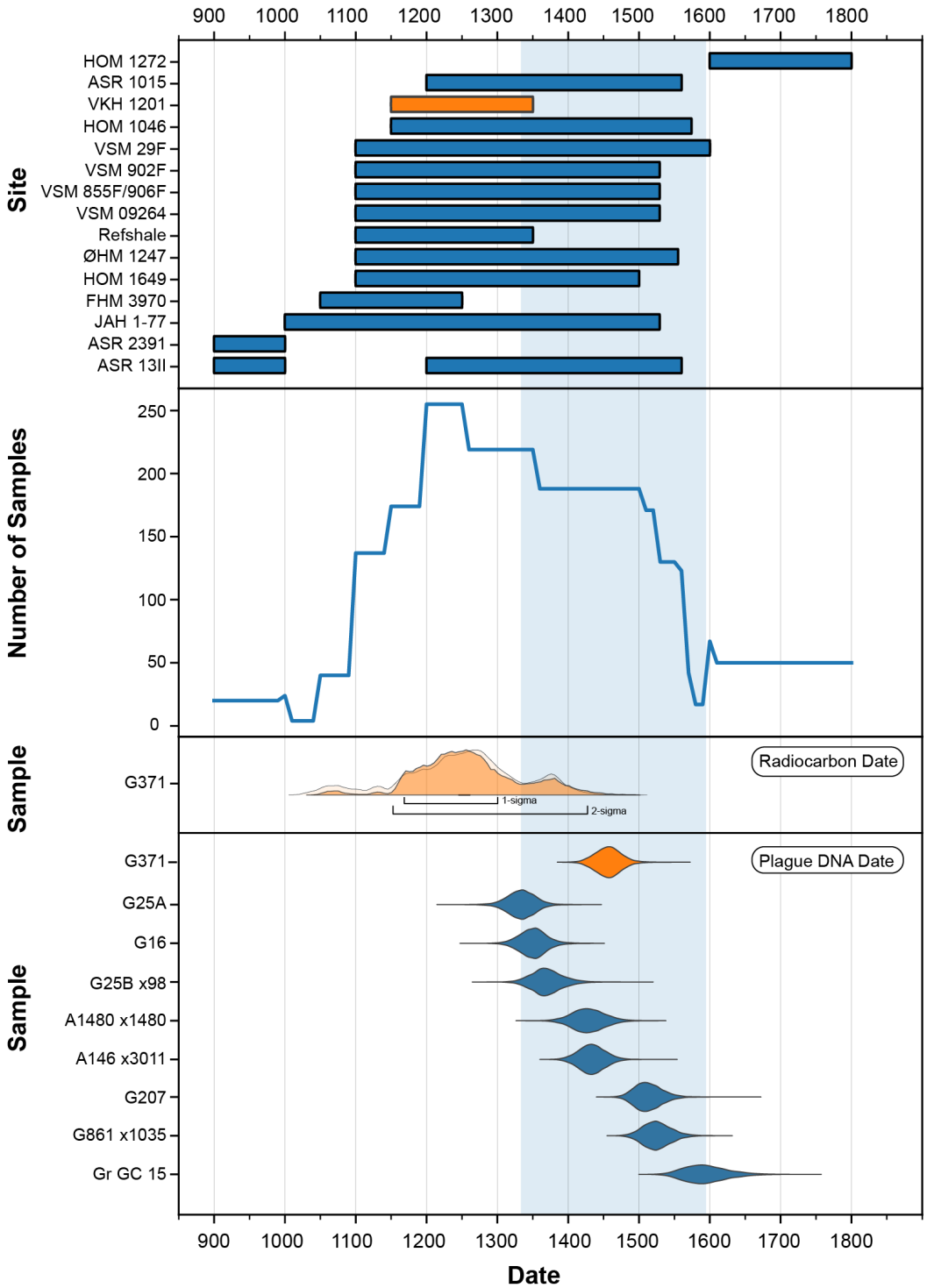

Figure 3: Timeline of archaeological sites and plague-positive individuals. The blue shaded range spans the highest probability period from the oldest to the youngest sample. The color orange indicates the sample and site with disparate dates.

Phylodynamics

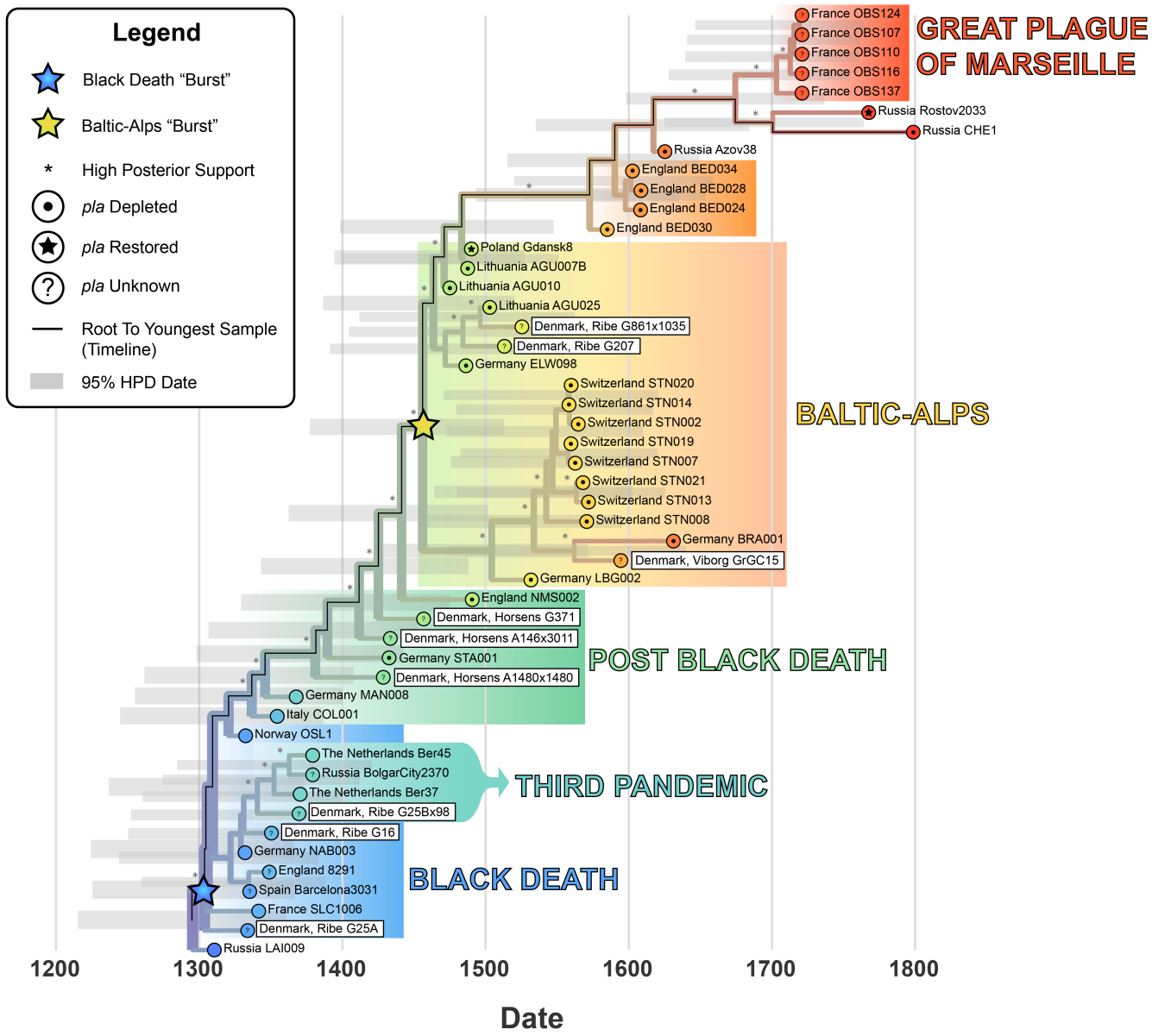

Figure 4: A time-scaled phylogeny of the Second Plague Pandemic.

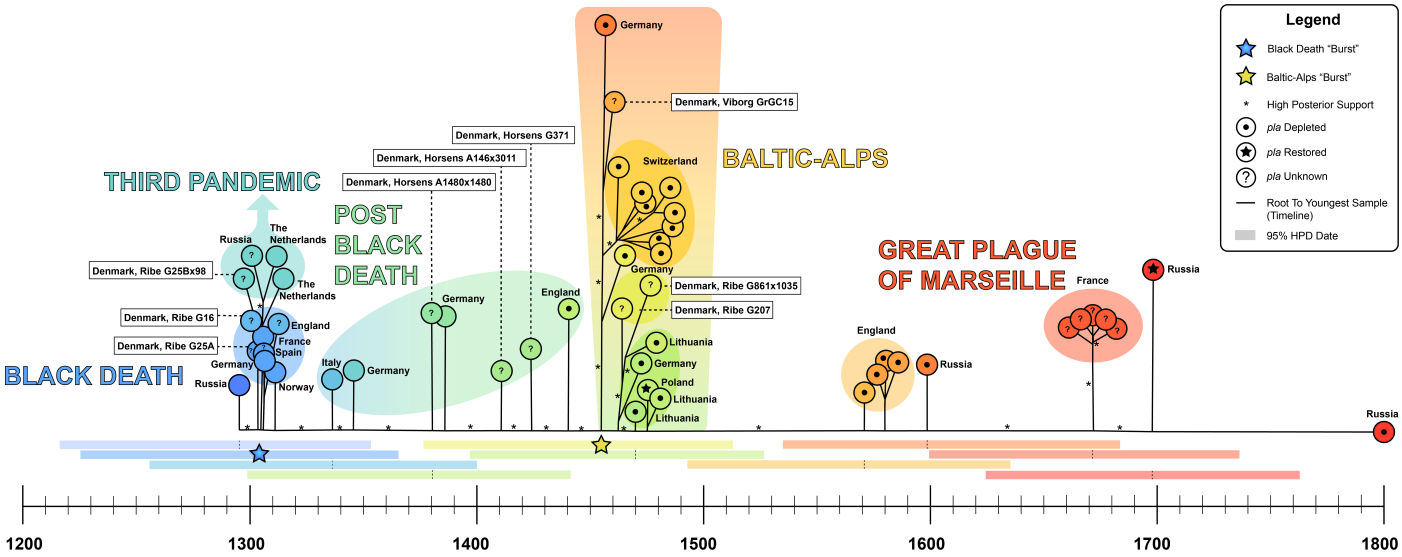

Figure 5: A timeline phylogeny of the Second Plague Pandemic.

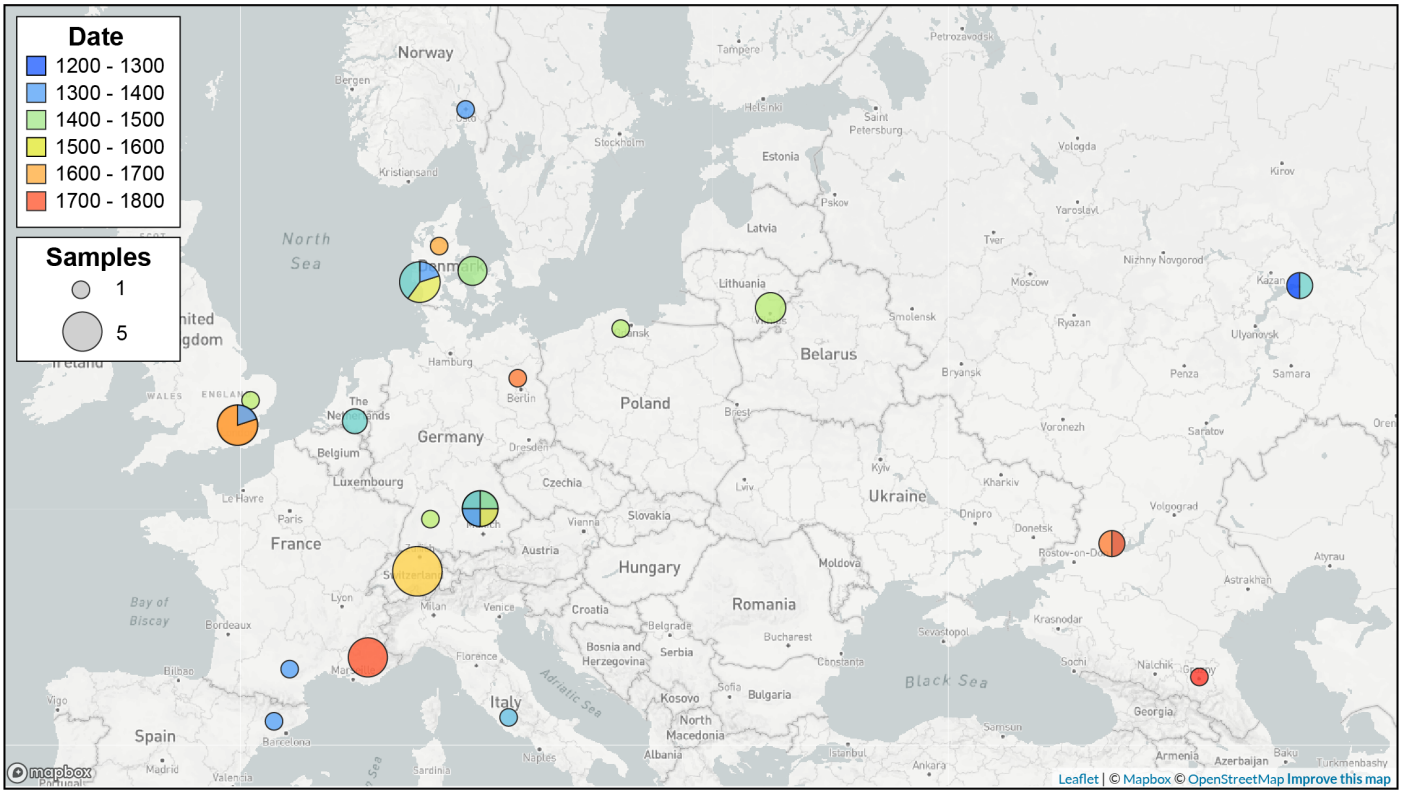

Figure 6: Geographic distribution of Second Pandemic samples used in the *Yersinia pestis* phylogeny.

The Black Death

Two individuals from Ribe, G25A and G16, cluster with *Y. pestis* strains dated to the 14th century. This cluster is thought to be associated with the Black Death (1346-1353), as the genomes are nearly identical and have been isolated from all across Europe, including France, Spain, England, Germany, and Norway. These genomes mark the first observation of *Y. pestis* in Denmark and is congruent with historical records that document the arrival of the Black Death in Ribe in 1349 [2]. However, the genome associated with G25A has the lowest sequencing depth, with a mean chromosomal depth of 3.8 molecules per nucleotide.

> **Kat's Note**:
> - G25A has the lowest sequencing depth, with a mean chromosomal depth of 3.8 molecules per nucleotide. This might mean that diagnostic positions for the post-Black Death clade are missing as "false" negatives. This would make it seem earlier/less-derived if those mutations are 'missing'. I will investigate!

Similar to the rest of Western Europe, *Y. pestis* appears suddenly in Denmark in the 14th century. No evidence of *Y. pestis* was found in Denmark during the preceding centuries, suggesting plague was a relatively new disease for medieval Danish populations. If true, this would imply an immunologically naive population and may have important implications for the study of health and immunity in both past and present Denmark.

Unfortunately, the high degree of genetic similarity means that the branching patterns, and thus dispersal of *Y. pestis*, cannot be resolved during this period. Our understanding of the Black Death clade is that of a "Burst" radiation, with limited genetic diversity spreading rapidly over a vast geographic area. This pattern is typical of epidemic spread, and has also been observed at the advent of the Third Plague Pandemic.

Ancestors of the Third Pandemic

Interestingly, the *Y. pestis* genome associated with individual G25Bx98 (who was found in the same grave as G25A?) is genetically distinct from the earlier strains from Ribe. This isolate falls within a clade of high epidemiological significance, which is the ancestral group giving rise to the Third Pandemic of plague and the rest of Branch 1.

The phylogenetic position of samples within this clade has been hypothesized to reflect a "backward" migration of plague from Northern Europe into Asia. The *Y. pestis* genome retrieved from G25Bx98 tentatively supports this hypothesis, as it falls basal to the more derived strains from The Netherlands and Russia. However, directionality cannot be robustly inferred from four samples alone, particularly given the strong Western European sampling bias of Second Pandemic samples. It will be an important avenue of future research to further develop the relationship between Northern European plague and the only Second Pandemic lineage that is known to persist until the present.

Post Black Death

In contrast to the genetic homogeneity observed across Europe during the Black Death period, isolates of plague in the post-Black Death period are easily distinguished. Three genomes collected from three different sites near Horsens derive from independent emergences, despite having temporal overlap. This has also been observed in Germany during this period, as samples collected in relatively close proximity are genetically distinct. A product of this increased genetic diversity means

that branching patterns are well resolved in the post-Black Death period. Unfortunately, the geographic origins and dispersal of plague are still challenging to reconstruct, as the number of genomes (N=7) is sparsely sampled relative to the minimum number of countries (N=4) that are implicated.

The pattern of independent emergence is the defining dynamic of plague during this period as all *Y. pestis* collected after the Black Death, but prior to the Early Modern Period, are unique lineages. The transition captured here, from little genetic diversity spread across a continent to significant diversity accumulating within a country, may indicate “boom-bust” dynamics [3].

> The epidemiological interpretation of this transition requires more thought.

Another defining characteristic of the post-Black Death period is depletion of a key virulence factor, the *plasminogen activator* (*pla*). Figure 7 compares the sequencing depth of the *pla* gene to its corresponding plasmid pPCP1 across samples from the Second Pandemic. Two linear trends are observed separating samples with ‘normal’ gene-to-plasmid ratios from samples that have a ‘depleted’ gene-to-plasmid ratio. This event was previously observed [Susat 2020 Yersinia Pestis Strains|[4]]], and found in samples from as early as the 15th century.

The three Danish genomes from Horsens have temporal overlap with the 15th century and may potentially capture the transition from to a *pla*-depleted state. Unfortunately, the sequencing depth of the pPCP1 plasmid in these samples is insufficient for statistical analysis. The results of a targeted enrichment for the pPCP1 plasmid will be the subject of a forthcoming publication.

> Ravneet has completed this experiment and has exciting results, stay tuned!

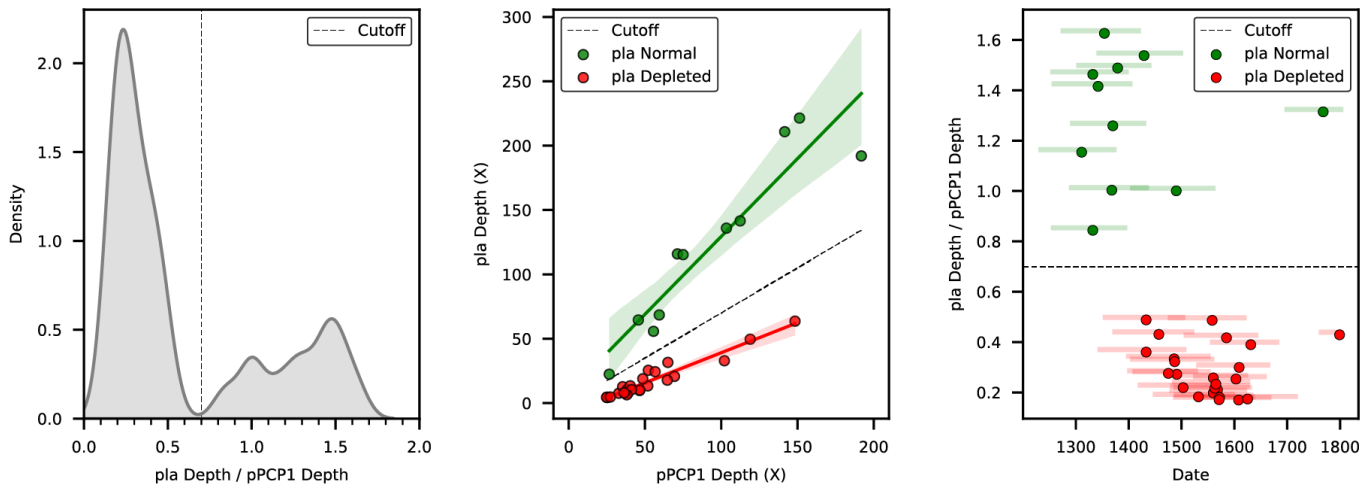

Figure 7: Depletion of the *pla* virulence factor. Left: The observed multimodal distribution of *pla*/pPCP1 ratios across the Second Pandemic. Middle: Linear regressions of *pla* depth on pPCP1 depth. Right: *pla* depletion events over time.

Early Modern Plague

Following the Medieval Period, the evolutionary dynamic of *Y. pestis* changes yet again and a second “Burst” event occurs. Similar to the Black Death “Burst”, this event begins with a series of divergences in quick succession, producing highly similar genomes in contemporaneous individuals from Germany, Lithuania, and Poland. Following these divergences, two major lineages emerge.

The first lineage, termed the Baltic-Alps, has been previously observed in the Baltic countries (Lithuania, Poland, Germany) and the Alps (Switzerland). Three Danish genomes fall within the Baltic-

Alps clade, including two samples from coastal Ribe and one sample from inland Viborg. The addition of these samples strengthens the geographic ties of this lineage to the Baltic Region. The epidemiological significance of this localization has been speculated to indicate the formation of a novel plague reservoir within Europe.

The second lineage that emerges in this period gives rise to later epidemics of plague in Russia, England, and France, including the Great Plague of Marseille (1720-1721). No Danish plague in this study is associated with this lineage. As sampling efforts intensify, it will be interesting to see if these two lineages remain geographically distinct, and to investigate what might be obstructing the flow of plague between them.

Conclusion

To Be Done post co-author discussion of results.

Methods

DNA Extraction

Ancient DNA laboratory work was performed in dedicated clean-room facilities at the McMaster Ancient DNA Centre (Hamilton, ON). A single root of each individual's molar was sectioned into two subsamples (50-100 mg) using a circular Dremel. Demineralisation and digestion were performed as previously described [5] and DNA extraction was conducted using a specialized protocol designed for ancient DNA [6]. Reagent blanks were introduced as negative controls to monitor DNA contamination in subsequent steps.

Yersinia pestis PCR screening

An initial plague-screening PCR was performed in duplicate on 1:10 extract dilutions using a *pla* assay [7]. In brief, the PCR primers used in this study target the 3'UTR of the *pla* gene which has reduced sequence similarity in non-*Yersinia* species. The forward primer used is thus far known to be identical to *Yersinia pestis* only. A second round of pla PCR was performed for all extracts that amplified in the initial PCR, using the 1:10 dilution in duplicate and the original concentration in duplicate. In total, 6 PCR replicates were performed for each plague-positive sample.

Shotgun Sequencing

The extracted DNA of plague-positive individuals was converted into Illumina sequencing libraries using a modified protocol [8] and quantified using an Illumina library qPCR assay. Sample libraries were then pooled at equimolar concentrations while negative controls were sequenced at maximum volume input to maximize detection of contaminant organisms. Paired-end sequencing was performed on an Illumina HiSeq 1500 platform (Farncombe Family Digestive Health Research Institute, Hamilton, ON).

Targeted Sequencing

In-solution enrichment for the pan-genome of *Yersinia pestis* was performed with a previously designed bait-set [7] and using the myBaits v4 protocol. The following modifications were incorporated to improve recovery of degraded and divergent DNA sequences: 5uL library input, 100ng bait concentration, hybridisation at 60°C, 16–24h hybridisation capture, and two rounds of enrichment. The enriched libraries were quantified using an Illumina library quantification qPCR assay and pooled at maximum input volume (13uL) due to low concentration. Following pooling, libraries were size-selected on anagarose gel to retain 150–500bp fragments which corresponds to molecule lengths of approximately 15–365bp without the adapter sequences. Paired-end sequencing was performed on an Illumina HiSeq 1500 platform at the Farncombe Metagenomics Facility (Hamilton, ON).

Comparative Genomes

Ancient unassembled genomes from the Second Plague Pandemic were identified using NCBImeta 10 and downloaded from the SRA database in FASTQ format using the SRA Toolkit [11] (Table 7).

Y. pestis strain CO92 was used as the reference genome for sequence alignment and annotation (Table 8). All assembled genomes belonging to phylogenetic branch 0.ANT3 were downloaded and used as an outgroup to root the maximum likelihood phylogeny (Table 9).

Genomic Alignment

Pre-processing and alignment to the reference genome was performed using the nf-core/eager pipeline, a reproducible workflow for ancient genome reconstruction [12]. A multiple sequence alignment was constructed using the Snippy Core module of the Snippy pipeline [13]. The output alignment was filtered to only include chromosomal variants and to exclude sites that had no more than 30% missing data (ie. no more than 30% of samples having an ambiguous nucleotide) . The value of 30% was selected as the most permissive threshold where the number of shared, parsimony-informative sites (240) was still larger than the number of singleton sites (234) which are observed in a single genome (Figure 8). Furthermore, a 30% threshold was selected as ambiguous nucleotides made up no more than 10% of the alignment (9).

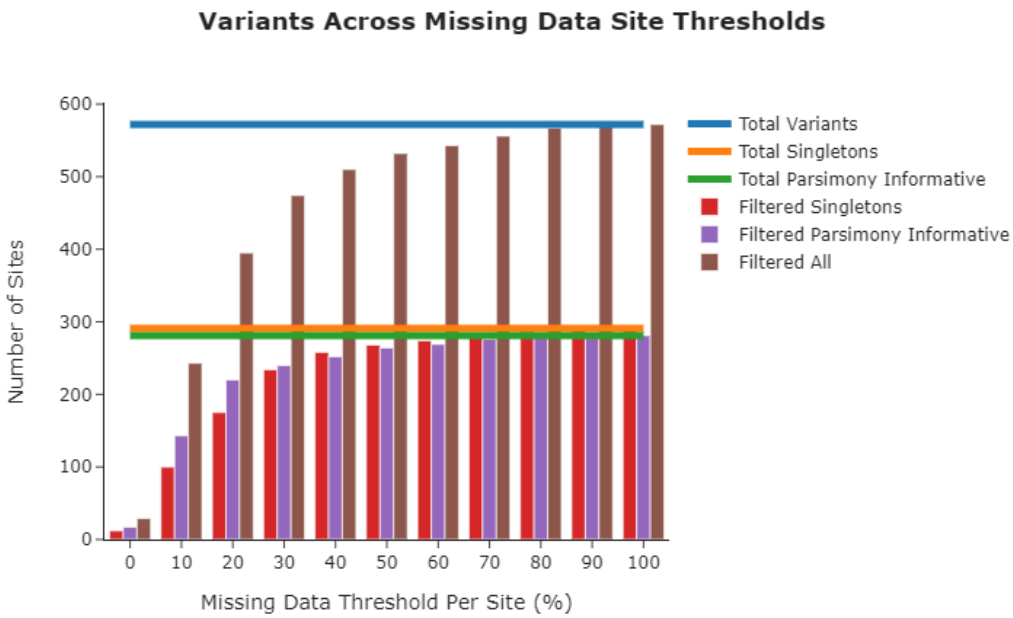

Figure 8: The number of variant positions used in the multiple alignment according to different missing data thresholds.

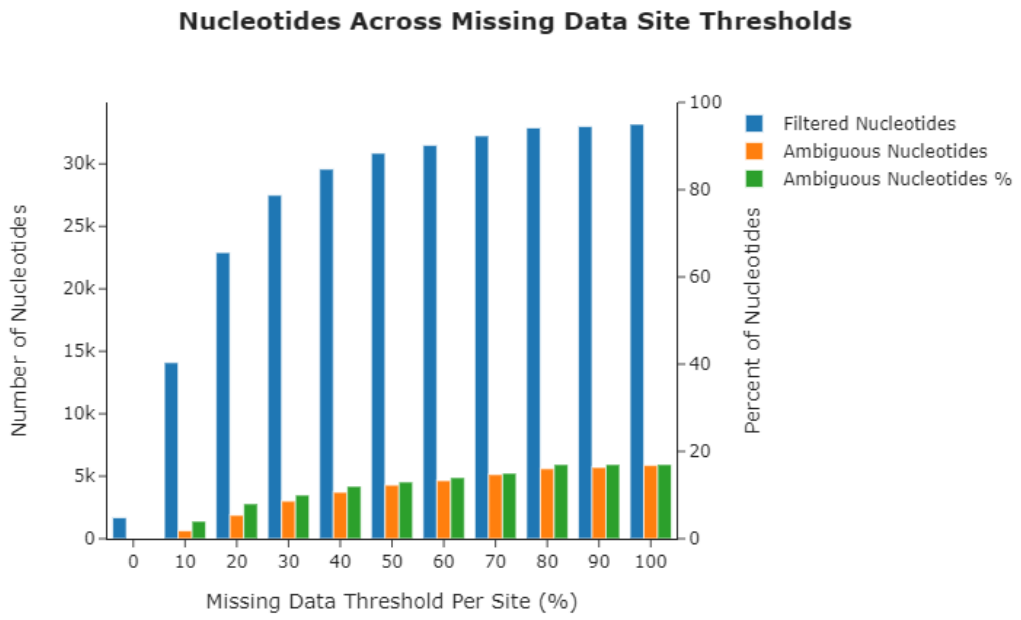

Figure 9: The number of ambiguous nucleotides used in the multiple alignment according to different missing data thresholds.

Phylogeny

Model selection was performed using Modelfinder which identified the K3Pu+F+I model as the optimal choice based on the Bayesian Information Criterion (BIC) [14]. A maximum-likelihood phylogeny was then estimated across 10 independent runs of IQTREE [15]. Branch support was evaluated using 1000 iterations of the ultrafast bootstrap approximation, with a threshold of 95% required for strong support [16].

Phylodynamics

To evaluate the degree of temporal signal present, a Bayesian Evaluation of Temporal Signal (BETS) was first performed. Briefly, each candidate model was tested using the correct collection dates of all samples and then compared to the same model with all collection dates assumed to be contemporaneous. Bayes factors (BF) were calculated by comparing the marginal likelihoods of each model, as estimated with a generalized stepping stone (GSS) computation across 100 chains each sampled over 1,000,000 generations.

The BETS analysis revealed decisive support for temporal signal (dates vs. no dates) using both the strict clock (SC) and the uncorrelated lognormal relaxed clock (UCLN) (Table 5). A comparison of the strict vs. relaxed clocks using collection date produced decisive support for the relaxed clock. Therefore, a time-scaled phylogeny with tip-dating was estimated using a relaxed clock and diffuse normal priors centered around the mean collection date.

Table 5: Bayesian Evaluation of Temporal Signal (BETS) summary.

Model	Abbrev.	Dates	Likelihood	Bayes Factor (Dates)	Bayes Factor (Model)
Strict Clock	SC	Yes	-5948088	749	–
		No	-5948837	–	–

Model	Abbrev.	Dates	Likelihood	Bayes Factor (Dates)	Bayes Factor (Model)
Relaxed Clock	UCLN	Yes	**-5947948**	715	140
		No	-5948663	–	–

References

1. **Leprosy in Medieval Denmark — Osteological and epidemiological analyses**
   Jesper L. Boldsen
   Anthropologischer Anzeiger (2009) http://www.jstor.org/stable/29543069

2. **The Black Death**
   Kristina Lenz, Nils Hybel
   Scandinavian Journal of History (2016) https://journals.scholarsportal.info/details/03468755/v41i0001/54_tbd.xml
   DOI: 10.1080/03468755.2015.1110533

3. **Boom-bust dynamics in biological invasions: towards an improved application of the concept**
   David L. Strayer, Carla M. D'Antonio, Franz Essl, Mike S. Fowler, Juergen Geist, Sabine Hilt, Ivan Jarić, Klaus Jöhnk, Clive G. Jones, Xavier Lambin, … Jonathan M. Jeschke
   Ecology Letters (2017-10)
   DOI: 10.1111/ele.12822 · PMID: 28834087

4. ***Yersinia pestis* strains from Latvia show depletion of the pla virulence gene at the end of the second plague pandemic**
   Julian Susat, Joanna H. Bonczarowska, Elīna Pētersone-Gordina, Alexander Immel, Almut Nebel, Guntis Gerhards, Ben Krause-Kyora
   Scientific Reports (2020-09-03) https://www.nature.com/articles/s41598-020-71530-9
   DOI: 10.1038/s41598-020-71530-9

5. **New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains**
   Carsten Schwarz, Regis Debruyne, Melanie Kuch, Elizabeth McNally, Henry Schwarcz, Andrew D. Aubrey, Jeffrey Bada, Hendrik Poinar
   Nucleic Acids Research (2009-06-01) https://doi.org/10.1093/nar/gkp159
   DOI: 10.1093/nar/gkp159

6. **Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments**
   J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C. Valdiosera, N. Garcia, S. Paabo, J.-L. Arsuaga, M. Meyer
   Proceedings of the National Academy of Sciences (2013-09-24) http://www.pnas.org/cgi/doi/10.1073/pnas.1314445110
   DOI: 10.1073/pnas.1314445110

7. ***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

8. **Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing**
   M. Meyer, M. Kircher

Cold Spring Harbor Protocols (2010-06-01) http://www.cshprotocols.org/cgi/doi/10.1101/pdb.prot5448
DOI: 10.1101/pdb.prot5448

9. **Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform**
Martin Kircher, Susanna Sawyer, Matthias Meyer
Nucleic Acids Research (2012-01-01) https://academic.oup.com/nar/article/40/1/e3/1287690
DOI: 10.1093/nar/gkr771

10. **NCBImeta: efficient and comprehensive metadata retrieval from NCBI databases**
Katherine Eaton
Journal of Open Source Software (2020) https://doi.org/10.21105/joss.01990
DOI: 10.21105/joss.01990

11. **ncbi/sra-tools**
NCBI - National Center for Biotechnology Information/NLM/NIH
(2021-05-18) https://github.com/ncbi/sra-tools

12. **Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager**
James A. Fellows Yates, Thiseas C. Lamnidis, Maxime Borry, Aida Andrades Valtueña, Zandra Fagernäs, Stephen Clayton, Maxime U. Garcia, Judith Neukamm, Alexander Peltzer
PeerJ (2021-03-16) https://peerj.com/articles/10947
DOI: 10.7717/peerj.10947

13. **Snippy: Rapid haploid variant calling and core genome alignment.**
Torsten Seemann
(2020-03-08) https://github.com/tseemann/snippy

14. **ModelFinder: fast model selection for accurate phylogenetic estimates**
Subha Kalyaanamoorthy, Bui Quang Minh, Thomas K. F. Wong, Arndt von Haeseler, Lars S. Jermiin
Nature Methods (2017-06) http://www.nature.com/articles/nmeth.4285
DOI: 10.1038/nmeth.4285

15. **IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era**
Bui Quang Minh, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, Robert Lanfear
Molecular Biology and Evolution (2020-05-01) https://academic.oup.com/mbe/article/37/5/1530/5721363
DOI: 10.1093/molbev/msaa015

16. **UFBoot2: Improving the Ultrafast Bootstrap Approximation**
Diep Thi Hoang, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, Le Sy Vinh
Molecular Biology and Evolution (2018-02-01) https://academic.oup.com/mbe/article/35/2/518/4565479
DOI: 10.1093/molbev/msx281

Appendix

Plague Detection

Table 6: Plague false positive summary.

Arch ID	Project ID	Site Code	PCR	Human (%)	Plague (%)	Chrom	pCD1	pMT1	pPCP1
G70 x212	R44	ASR 1015	3/6	1.48	0.00	0.1	0.1	0.1	0.0
G860	R39	ASR 1015	5/6	0.09	?	?	?	?	?
G364	R43	ASR 1015	4/6	?	?	?	?	?	?
K1167 x1167	P235	ASR 13 II	3/6	?	?	?	?	?	?
A21 x23	D25	ØHM 1247	4/6	0.01	0.00	0.05	0.1	0.0	0.0
G260 K539 x876	R27	VSM 09264	3/6	?	?	?	?	?	?

Comparative Genomes

Table 7: Second Pandemic genomes metadata.

Strain	Date	Branch	Country	Accession
STN021	1485 - 1635	1.PRE1	Switzerland	SAMEA5818830
STN020	1485 - 1635	1.PRE1	Switzerland	SAMEA5818829
STN019	1485 - 1635	1.PRE1	Switzerland	SAMEA5818828
STN014	1485 - 1635	1.PRE1	Switzerland	SAMEA5818826
STN013	1485 - 1635	1.PRE1	Switzerland	SAMEA5818825
STN008	1485 - 1635	1.PRE1	Switzerland	SAMEA5818822
STN007	1485 - 1635	1.PRE1	Switzerland	SAMEA5818821
STN002	1485 - 1635	1.PRE1	Switzerland	SAMEA5818818
STA001	1420 - 1630	1.PRE1	Germany	SAMEA5818817
NMS002	1475 - 1536	1.PRE1	England	SAMEA5818815
NAB003	1292 - 1392	1.PRE1	Germany	SAMEA5818811
MAN008	1283 - 1390	1.PRE1	Germany	SAMEA5818809
LBG002	1455 - 1632	1.PRE1	Germany	SAMEA5818808
LAI009	1300 - 1400	1.PRE0	Russia	SAMEA5818806
ELW098	1485 - 1627	1.PRE1	Germany	SAMEA5818805
BRA001	1618 - 1648	1.PRE1	Germany	SAMEA5818803
BED034	1560 - 1635	1.PRE1	England	SAMEA5818801
BED030	1560 - 1635	1.PRE1	England	SAMEA5818800
BED028	1560 - 1635	1.PRE1	England	SAMEA5818799
BED024	1560 - 1635	1.PRE1	England	SAMEA5818798

Strain	Date	Branch	Country	Accession
SLC1006	1279 - 1389	1.PRE1	France	SAMEA5054093
OSL1	1270 - 1390	1.PRE1	Norway	SAMEA5054092
Ber45	1300 - 1400	1.PRE2	The Netherlands	SAMEA5054090
Ber37	1300 - 1400	1.PRE2	The Netherlands	SAMEA5054089
BolgarCity2370	1362 - 1400	1.PRE3	Russia	SAMEA3937654
Barcelona3031	1300 - 1420	1.PRE1	Spain	SAMEA3937653
OBS137	1720 - 1720	1.PRE1	France	SAMEA3713715
OBS124	1720 - 1720	1.PRE1	France	SAMEA3713714
OBS116	1720 - 1720	1.PRE1	France	SAMEA3713713
OBS110	1720 - 1720	1.PRE1	France	SAMEA3713712
OBS107	1720 - 1720	1.PRE1	France	SAMEA3713711
8291	1348-1350	1.PRE1	England	SAMN00715800
COL001	1300 - 1400	1.PRE1	Italy	SAMEA7293136
CHE1	1500 - 1800	1.PRE1	Russia	SAMEA7293135
Rostov2033	1762 - 1773	1.PRE1	Russia	SAMEA7313236_38
Azov38	1400 - 1700	1.PRE1	Russia	SAMEA7313243_45
Gdansk8	1400 - 1700	1.PRE1	Poland	SAMEA7313246_49
AGU010	1435 - 1477	1.PRE1	Lithuania	SAMEA6651390
AGU025	1441 - 1612	1.PRE1	Lithuania	SAMEA6637004
AGU007B	1463 - 1632	1.PRE1	Lithuania	SAMEA6637002

Table 8: Reference genome metadata.

Strain	Date	Branch	Country	Accession
CO92	1992	1.ORI1	United States of America	SAMEA1705942

Table 9: Outgroup genomes metadata.

Strain	Date	Branch	Country	Accession
231	1947	0.ANT3	Kyrgyzstan	SAMN02777961
A-1486	1966	0.ANT3	Kyrgyzstan	SAMN05149973
790	1961 - 1976	0.ANT3	Kyrgyzstan	SAMN02769799
CMCC38001	1979	0.ANT3	China	SAMN02403043
CMCC21106	2001	0.ANT3	China	SAMN02403038
A1956001	1956	0.ANT3	China	SAMN02403019
42091	1999	0.ANT3	China	SAMN02403004
42082	1995	0.ANT3	China	SAMN02403003

Phylogeny

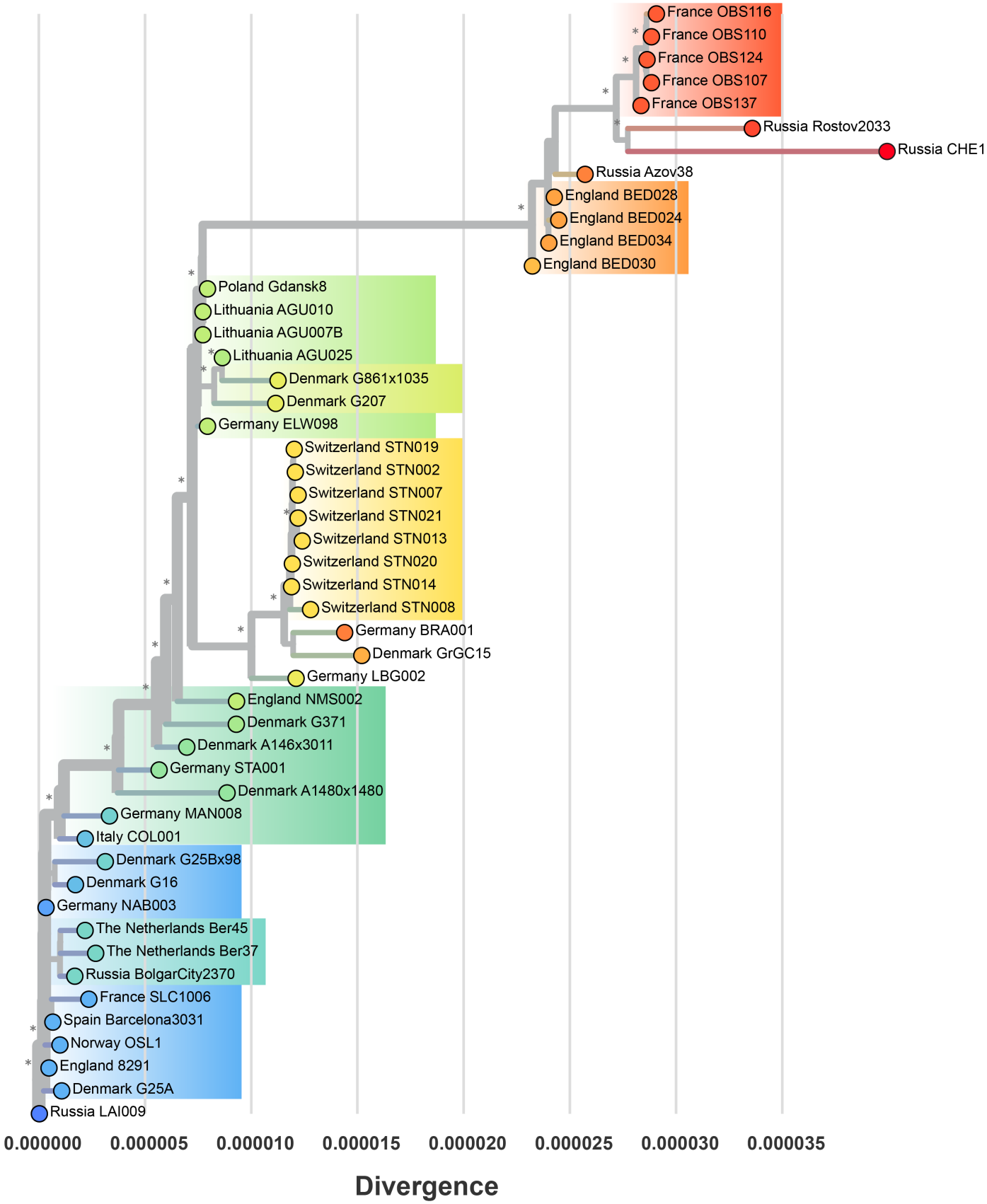

Figure 10: A divergence-scaled phylogeny of the Second Plague Pandemic. Asterisks indicate branches with strong statistical support.