PNAS– Article

Assessing the origins of the European Plagues following the Black Death: a synthesis of genomic, historical and ecological information

Barbara Bramanti[1,2* $], Yarong Wu[3 $], Ruifu Yang[3], Yujun Cui[3 *] & Nils Chr. Stenseth[1,4 *]

[1] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Blindern, N-0316, Oslo, Norway

[2] Department of Neuroscience and Rehabilitation, Faculty of Medicine, Pharmacy and Prevention, University of Ferrara, 44121 Ferrara, Italy

[3] State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, 100071, China

[4]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, 100084, China

[*] - corresponding authors

[$] - equal contribution

Abstract

The Second Plague Pandemic started in Europe with the Black Death in 1346 and lasted until the 19th century. Based on ancient DNA studies, there is a scientific disagreement over whether the bacterium, *Yersinia pestis*, came into Europe once (Hypothesis 1), or repeatedly over the following four centuries (Hypothesis 2). Here

we synthesize the most updated phylogeny together with historical, archeological, evolutionary and ecological information. On the basis of this holistic view, we conclude that Hypothesis 2 is the most plausible. We also suggest that *Y. pestis* lineages might have developed attenuated virulence during transmission, which can explain the convergent evolutionary signals, including *pla*-decay, that appeared at the end of the pandemics.

Keywords: *Yersinia pestis* | ancient DNA | Black Death | European Plague | pathogen evolution

Significance Statement

Over the last few years there has been a great deal of scientific debate regarding whether the plague bacterium, *Yersinia pestis*, spread from a Western European reservoir during the Second Plague Pandemic, or if it repeatedly came to Europe from Asia. Here we make a synthesis of the available evidence, including genomes of ancient DNA, historical, archeological and ecological information. We conclude that the bacterium most likely came to Europe from Asia several times during the Second Plague Pandemic.

Main Text

Researchers agree that the Second Plague Pandemic was caused by *Yersinia pestis* (1-9), which arrived in Europe from Caffa transported by Genoese galleys on the Black Sea at the beginning of the Black Death (10). However, there is no consensus among researchers as to the origins of plague epidemics in Europe following the

Black Death and ravaging Europe until the 19th century, as attested by historical documents (11).

The two main theories are that one or more plague reservoirs remained in Western Europe during the entire Second Plague Pandemic (referred to in the following as Hypothesis 1) (3, 4, 8, 12), or the bacteria repeatedly invaded Europe from non-Western European reservoir(s) during the same period (referred to in the following as Hypothesis 2) (6, 7, 9, 11, 13). Here, we assess these two hypotheses using a broad spectrum of evidence including historical and archeological, genetic and evolutionary, as well as ecological information.

Results and Discussion

Assessment of the two hypotheses.

For the purpose of understanding the evolution of the plague bacteria, more than 100 ancient *Y. pestis* genomes have been published to date. The last 17 were recently reported during a short period by four distinct research groups (7-9, 12). Using most of the ancient genomes (criteria for exclusion are described in methods) along with 499 modern ones, we present here the most updated phylogeny (Fig. 1).

The updated phylogeny confirms the almost clonal nature of the Black Death strains, in comparison to all other lineages of the Second Plague Pandemic, including the strains from the *Pestis secunda* (Ber37 and Ber 45, The Netherlands (6), and BolgarCity2370, Russia (3)), which are placed on Branch 1 (see also London-Ind6330, UK (3)), as well as to all other strains, which are placed on the post-Black Death branch. There is general agreement that the post-Black Death branch was hosted in a novel wild rodent reservoir – either in Europe or outside Europe. The original hypothesis (Hypothesis 1) claims that such a plague reservoir existed in

Western Europe (14), perhaps in the Alps (15). However, a newer hypothesis (Hypothesis 2) claims that the plague reservoir was in Asia, possibly close to Eastern Europe (6, 7, 9, 11, 13).

In order to more easily view the phylogeny from the Second Plague Pandemic and to better contrast the evidence for the two hypotheses, we generated two schematic figures (Fig. 2) and a table (Table 1).

Table 1. Main differences between the two competing hypotheses proposed to explain the phylogeny of *Y. pestis* of the Second Plague Pandemic. Genomic and evolutionary, historical and archaeological, as well as ecological arguments are considered.

	Hypothesis 1	**Hypothesis 2**
Main differences		
Origin of the outbreaks	Plague established in Western European reservoirs (for example, marmots in the Alps) (3, 4, 8)	Plague was repeatedly imported from Eastern European or Asian reservoirs (6, 7, 9, 11, 13)
Transmission	Mediated by rats infected by wild rodents, as in China during the Third Pandemic	Imported by rats, humans, and goods, and subsequently spread by chains of human transmission, as in Europe during the Third Pandemic (16, 17)
Vector	*Xenopsylla cheopis* and other rodent fleas	Any ectoparasite, including *Pulex irritans*, and body lice (18, 19)
Supporting information		
Population genomics	Western European strains (the Alpine clade) are basal in the sub-branch (although of different eras). A model proposes Western European strains as ancestral sources for the transmissions (Fig. S1)	The oldest (LAI009) and the most recent strains, in addition to Bolgar, are all from Eastern Europe (Russia), as well as all the strains of the 18th century; the majority of the Western European strains in the phylogeny come from ports
History and archaeology	A hypothesis (15) suggests that the plagues from the 16th century in the Alps, were not imported by major trade centers	Multiple historical records assert that plague was imported for outbreaks associated with the Black Death strains (partially reviewed in (6, 7, 11, 13)), for BRA (4), for SPN (in the Alps), PEB10 (7) and OBS (5). Multiple records of importation are historically attested (20), particularly in harbors

Climate	No climatic signal in support, although 4 datasets of climatic proxies were from the Alps (13)	Strong signals of climate-driven introductions of plague from Asia (13)
Evolution	*Y. pestis* developed in Western European reservoir strains *pla*+/*pla*- and Delta49kb possibly as a form of adaptation to the local host (rodents).	*Y. pestis* developed *pla*+/*pla*- and Delta49kb strains possibly as a form of adaptation to the new host (humans) and/or new vectors (fleas or body lice).

Hypothesis 1 is supported by a phylogenetic analysis based on the currently available ancient genomes, which infers high posterior probability for a Western European source of the transmissions on the post Black Death branch (see Fig. S1). However, as the dataset includes 41 ancient genomes from Western Europe against only 8 strains from Eastern Europe (including Gdansk and Riga), the proposed origins from Western Europe are likely to be biased toward a European reservoir due to a size-effect bias. Notably, the most basal genome LAI009 (4) (the Black Death's lineage), Bolgar (at the root of Branch 1), and the most recent genome (CHE1 (7)), all originated from Western Russia, implying that they might have been closer to a putative Asian or Eastern European reservoir. This continuity does represent strong evidence in support of Hypothesis 2.

Using only genomic data, Hypothesis 1 might be seen as the most parsimonious hypothesis, since it proposes an internal source for all western Eurasian outbreaks. However, for two locations (Pestbacken, Sweden 1710 (PEB10), and Marseille, France 1722 (OBS)), an origin from the Ottoman Empire is historically and archaeologically well supported (7). Thus, Hypothesis 1 needs to account for a back-and-forth spread, which reintroduced plague on two occasions to the Ottoman Empire and back again to Western Europe. Notably, none of the strains from the 18th century appear to have originated in Western Europe according to historical sources (7, 9).

Hypothesis 1 assumes the existence of a wild rodent plague reservoir in the Alps, which is not supported by ecological evidence (13). Instead, a study of more than 7,000 historical plague outbreaks and 15 tree-ring datasets (4 of which from the Alps), found climatic signals in support of frequent re-importations of plague from Asia into Eastern and Western European harbors (13).

Intriguingly, only a few genotyped strains are nodes on the backbone of the post-Black Death branch: the strains of the Black Death itself, the strain from Gdansk 1425-1469 and the strains from London (BED, 16th-17th century). While the strains of the Black Death were notoriously imported into Western Europe from the Mongol Empire via Caffa in Crimea (10), both Gdansk and London were very active harbors also in historical times and were very often hit by plague. Interestingly, *Yersinia pestis* was also recovered from a rat found in Gdansk. Although the genome is partial, due to the different SNPs profile, it is clear that the strain from the rat could not have infected the victim (Gdansk8) (9). Being a port, Gdansk may indeed have hosted diverse importations of infected rats in the period 1425-1469, as it happened in European harbors during the Third Pandemic (16).

Hypothesis 2 is consistent with the ecological as well as with the historical evidence (Fig. 2 and Table 1). The only Western European sub-cluster, the 'Alpine cluster' formed by LBG, STN, BRA, LAR and SPN, may naturally be explained by the circulation of soldiers and troops in Europe during the Thirty Years War (1618-1648), which made up human chains of transmission with historically documented epidemic events (20, 21). For three strains, SPN from the Italian Alps, LAR from the French Alps and BRA, from Northern Germany, the relationship with the time of the Thirty

Years War is historically and archaeologically documented (4, 7, 12). Human chains of transmission, which do not require the presence of rats to start and sustain an epidemic, might explain the circulation of the plague within Europe over long periods of time. They might be due to interpersonal contacts, crowding, infected parasites in clothes or goods, or contact with infected pets. Several chains of human transmission within Europe could be reconstructed for cases of the last century (16, 17).

Two convergent mutations.

To better understand the evolution of *Y. pestis*, we examined two more mutations, which were recently discovered in ancient strains. In the most recent sub-clade of the Second Pandemic, starting with BED, there is a 49kbp deletion with unknown function. This deletion was also present in the last lineage of the First Pandemic, and, in both cases, might have accounted for the decline of the pandemic (7). We found the same mutation in the Rostov 2033 strain in the 18th century clade (Fig.1 and 2). By contrast, a second strain found in the same cemetery in Rostov (Rostov 2039) has a different SNPs-pattern and lacks the chromosomal deletion.

Another mutation, the depletion of the *pla*-gene on the plasmid pPCP1, has recently been proposed as the cause of the disappearance of the Second Plague Pandemic in the 18th century (8), given that the *pla*-gene is an important virulence factor of *Y. pestis*. We checked for the presence of the *pla*+/*pla*- plasmids in all published ancient strains. The ratio in coverage depth between *pla* and the whole pPCP1 plasmid indicates the status of *pla*-loss in an organism (Fig. 3). If the depth of *pla* is significantly lower than that of pPCP1, it might properly be concluded that the *pla*-gene was lost in some pPCP1 plasmids. Our analyses show that the ratio of *pla* in the Black Death and post-Black Death genomes appears to be different when

compared with the pre-historic and the First pandemic lineages (Fig. 1). We have also checked randomly selected modern *Y. pestis* genomes in different lineages, and their depth of *pla* and pPCP1 are quite consistent, indicating no other *pla*-loss in modern plagues. By contrast, the generalized depletion of *pla* extensively observed during the post-Black Death era and at the end of the First Pandemic (Fig. 1) seems to be consistent. Given that the sequencing data were generated by several different research groups, a systemic error during sequencing is unlikely.

It seems that full *pla*-strains were slightly depleted at the end of the Second Pandemic (8), with the same phenomenon at the end of the First Pandemic. Notably, however, Rostov2033, one of the most recent genomes of the Second Pandemic, shows full reads pPCP1 plasmids, whereas CHE, the most recent historical strain, shows very slight *pla*-decay (Fig. 3). This observation is not fully in agreement with the proposed hypothesis that *pla*-depletion contributed to the end of the pandemic. An alternative explanation for this phenomenon (8) is that the differences observed in the full *pla*-plasmids might be due to different forms of plague. In particular, bubonic plague and pneumonic plague need the *pla*-gene to develop, whereas primary septicemic plague does not (8). It seems that plague existed in all three forms, at least from the time of the First Pandemic, however, this does not add any specific evolutionary information to the observed variability.

The evolution of the *pla*-gene.

We propose an evolutionary hypothesis for the presence of lineages with *pla* decay. One of the optimized survival strategies for an emerging pathogen is to balance its virulence to the main host with its transmission strategy. This trade-off hypothesis was previously demonstrated for *Y. pestis* (22, 23). This mechanism would allow the

bacterium to reduce virulence and enhance the time of survival of the host and, consequently, of the pathogen (24). After experiencing the Black Death and successive waves, the *pla* decay strains might have attempted to acquire a fitness advantage reducing their virulence by increasing the time-to-death. Indeed, we observe among the victims only *pla*+/*pla*- mixed strains, whereas *pla*- lineages might have survived longer in the host population, providing a milder form of the illness. The Eastern European/Asia clade of the 18th century (including CHE1) further lost the 49kb region, which can be the result of an extension of a virulence attenuated pattern. Such events of attenuated virulence might have occurred multiple times in the *Y. pestis* evolutionary history, and left out host-adapted lineages, such as for 0.PE2 and 0.PE4 (25). Therefore, the possible virulence reduction caused by *pla* decay and loss of the 49k region is not necessarily the reason for the extinction of plague at the end of the First and Second Pandemics, but might be the result of a form of adaptation to a new host, which may be the wild rodent in the putative Western European reservoir (Hypothesis 1), a new host in the Asian reservoir or the human host (Hypothesis 2), as well as their vectors. We observed that the newly published strains from Lariey (French Alps, (12)) do not show *pla*-decay, in contrast to other Alpine lineages (SPN). This evidence might exclude the hypothesis of an adaptation to a host in a Western European reservoir. Thus, we tentatively propose that this mechanism of *pla*-decay would support the presence of human-to-human transmission chains mediated by human ectoparasites (fleas and body lice) during plague pandemics in Europe, the plausibility of which has previously been demonstrated (16-18).

Conclusion

Altogether, the most consistent interpretation of the current information is in support of Hypothesis 2. This implies that there must have existed a reservoir outside of Western Europe (with the Ottoman Empire, Persia and Central Asia as possible candidates (7, 9)). Such a reservoir could then fed, with multiple introductions, the Second Plague Pandemic outbreaks in Western Europe along Northern and Southern trade routes (6, 26). In addition there might have been continuous recirculation of plague in (Western and Eastern) Europe with the movement of goods and troops. The recirculation of plague within Western Europe, mediated by trade and travel, might mimic the presence of a local reservoir, assuring the long-lasting re-emergence of epidemics on the continent. Such recirculation of plague could explain the collateral branches on the phylogenetic tree – an example of which may be the Thirty Years War Clade (1618-1648) evidenced in Fig. 2. Additional ancient strains from Asia and Eastern Europe, and more accurate dating and historical contextualization, will provide further evidence to clarify the phylogeny and evolution of the *Y. pestis* pathogen.

Methods

SNP calling and evaluation

All raw reads of 111 published ancient genomes were downloaded from NCBI SRA and EBI ENA databases. For the clones from each sampling site, only one genome with the highest quality (sequencing depth) as reported in corresponding publications was chosen, and genomes covering less than 20% of the chromosome length of the CO92 assembly (GCF_000009065.1) were excluded from further analysis, which resulted in a dataset with a total of 75 ancient genomes (Dataset S1). We trimmed and quality filtered raw reads using Trimmomatic v0.38 (27), and reads shorter than

30 bp and below a quality score of 20 were discarded. Subsequently, the filtered reads were mapped against the CO92 assembly with BWA mem model (v0.7.17) (28) and the aligned reads were extracted from bam files using SAMtools (v1.9) (29) view command (-bF 4) and then different runs of the same sample were merged using SAMtools merge command. Sequences with more than 10 soft and hard clipped alignments were filtered out by samclip, and duplicates were removed using Picard's MarkDuplicates module. SNP calling was performed using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK v3.8) (30) under the "EMIT_ALL_CONFIDENT_SITES" option with a minimum confidence threshold 10; a vcf file for every ancient genome was produced and SNPs that were close to each other by less than 20 bp were excluded.

A total of 499 modern genomic assemblies of *Y. pestis* available in NCBI Genbank database on 19th October, 2020 were downloaded (Dataset S2) and then aligned to the CO92 assembly using NASP's convert_external_genome command (31), which was based on MUMmer's nucmer and delta-filter modules (v3.23) (32). A fasta file 1-to-1 position aligned with a reference fasta for every modern genome was created.

All vcf files and aligned fasta files were used to aggregate sample calls into matrices with NASP's matrix module. For ancient genomes, a SNP would be called when supported by >=3 reads and >= 90% allele frequency. After manually validating the calls for ancient genomes with notably longer branch lengths (SPN strains, BSS31, and SLC1006) according to their published SNP lists, we got a final dataset of 12,608 polymorphic loci (Dataset S3).

Phylogenetic analyses and geographic extent of sampled sites

A fasta file, concatenated of all SNP sites, was used to generate a maximum-likelihood tree with 1,000 fast bootstrap replicates using IQ-TREE (v1.6.5) (33)with the option -m MFP+ASC to infer the best substitution model and account for ascertainment bias correction. Then FigTree (v1.4.3) was used to visualize the generated tree. The packages of ggplot2, maptoos and maps in R (v3.6.1) were used to mark the archaeological site locations of samples from the Second Plague Pandemic. The longitude and latitude for each site was taken from the website mapcoordinates (https://www.mapcoordinates.net/en).

Phylognetic analyses with calculation of MCMC posterior probability

A maximum-likelihood tree for 47 ancient genomes during the Second Pandemic was rebuilt using RAxML(v8.2.11) (34) with 100 replicates and GTRGAMMA model. It was rerooted to strain LAI009 and transformed into newic format using FigTree. We used the ReorderData function in evobiR (v1.1, R package) to match the source records (country names) to the order of tips on the phylogenetic tree. Then the make.simmap function in phytools (v0.7-70, R package) (35) was used to perform stochastic source mapping based on ARD model and the tip states on the tree, with 10000 generations of MCMC sampling every 100 generations.

pPCP1 and *pla* analysis

Samtools depth command was used to count the depth of whole pPCP1 plasmid and *pla*-gene for each sample from bam files. The packages of ggplot2 and ggsignif were used to obtain the boxplots and group-wise comparisons (Wilcoxon-test) of the ratio between the depth of *pla* and that of whole pPCP1 plasmid among three subclades of the Second Pandemic. The coverage plots of pPCP1 in CHE1 (Fig. 3) was visualized using Integrative Genomics Viewer(IGV, v2.8.11) (36).

Data availability

Publicly available genomes are listed in Dataset S1 and Dataset S2.

Funding

This work was supported by National Key Program for Infectious Diseases of China (No. 2018ZX10101003 and 2018ZX10714-002) and has received funding from the University of Ferrara under the Bando per il finanziamento della ricerca scientifica "Fondo per l'Incentivazione alla Ricerca" (FIR)-2020. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

B.B. conceived the work; N.C.S. established the author team; B.B., Y.C. and N.C.S. designed research; Y.W. analyzed the data, designed and generated the phylogeny; B.B. wrote the paper with contributions from Y.W., R.Y., Y.C. and N.C.S.; R.Y. and N.C.S. supervised the work.

Competing interests

The authors declare no competing interests.

Acknowledgement

Katharine Rose Dean is thanked for constructive comments on the manuscript, including improving the language. Two anonymous reviewers on an earlier version of this paper are greatly thanked for comments and suggestions which helped us to improve the paper.

References

1. Haensch S, *et al.* (2010) Distinct clones of *Yersinia pestis* caused the black death. *PLoS Pathog* 6(10):e1001134.
2. Bos KI, *et al.* (2011) A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478(7370):506-510.
3. Spyrou MA, *et al.* (2016) Historical *Y. pestis* Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host Microbe* 19(6):874-881.
4. Spyrou MA, *et al.* (2019) Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nat Commun* 10(1):4470.
5. Bos KI, *et al.* (2016) Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife* 5:e12994.
6. Namouchi A, *et al.* (2018) Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proc Natl Acad Sci U S A* 115(50):E11790-E11797.
7. Guellil M, *et al.* (2020) A genomic and historical synthesis of plague in 18th century Eurasia. *Proc Natl Acad Sci U S A* 117(45):28328-28335.
8. Susat J, *et al.* (2020) *Yersinia pestis* strains from Latvia show depletion of the *pla* virulence gene at the end of the second plague pandemic. *Sci Rep* 10(1):14628.
9. Morozova I, *et al.* (2020) New ancient Eastern European *Yersinia pestis* genomes illuminate the dispersal of plague in Europe. *Philos Trans R Soc Lond B Biol Sci* 375(1812):20190569.

10. Piazza MD (1980) *Cronaca* (Palermo : Renzo Mazzone).

11. Bramanti B, Stenseth NC, Walløe L, & Lei X (2016) Plague: A Disease Which Changed the Path of Human Civilization. *Advances in experimental medicine and biology* 918:1-26.

12. Seguin-Orlando A*, et al.* (2021) No particular genomic features underpin the dramatic economic consequences of 17th century plague epidemics in Italy. *iScience*:102383.

13. Schmid BV*, et al.* (2015) Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci U S A* 112(10):3020-3025.

14. Ell SR (1984) Immunity as a Factor in the Epidemiology of Medieval Plague. *Reviews of Infectious Diseases* 6(6):866-879.

15. Carmichael AG (2014) Plague Persistence in Western Europe: A Hypothesis. *The Medieval Globe* Vol. 1 : No. 1 , Article 8.

16. Bramanti B, Dean KR, Walloe L, & Chr Stenseth N (2019) The Third Plague Pandemic in Europe. *Proc Biol Sci* 286(1901):20182429.

17. Dean KR, Krauer F, & Schmid BV (2019) Epidemiology of a bubonic plague outbreak in Glasgow, Scotland in 1900. *R Soc Open Sci* 6(1):181695.

18. Dean KR*, et al.* (2018) Human ectoparasites and the spread of plague in Europe during the Second Pandemic. *Proc Natl Acad Sci U S A* 115(6):1304-1309.

19. Barbieri R, Drancourt M, & Raoult D (2020) The role of louse-transmitted diseases in historical plague pandemics. *Lancet Infect Dis*.

20. Alfani G (2013) Plague in seventeenth-century Europe and the decline of Italy: an epidemiological hypothesis. *European Review of Economic History* 17(4):408-430.

21. Lammert G (1890) Geschichte der Seuchen, Hungers-und Kriegsnoth zur Zeit des dreissigjährigen Krieges.

22. Hinnebusch BJ, Jarrett CO, & Bland DM (2017) "Fleaing" the Plague: Adaptations of *Yersinia pestis* to Its Insect Vector That Lead to Transmission. *Annu Rev Microbiol* 71:215-232.

23. Cui Y*, et al.* (2020) Evolutionary selection of biofilm-mediated extended phenotypes in *Yersinia pestis* in response to a fluctuating environment. *Nat Commun* 11(1):281.

24. Lathem WW, Price PA, Miller VL, & Goldman WE (2007) A plasminogen-activating protease specifically controls the development of primary pneumonic plague. *Science* 315(5811):509-513.

25. Sebbane F, Uversky VN, & Anisimov AP (2020) *Yersinia pestis* Plasminogen Activator. *Biomolecules* 10(11).

26. Bramanti B, Namouchi A, Schmid BV, Dean KR, & Stenseth NC (2019) Reply to Barbieri et al.: Out of the Land of Darkness: Plague on the fur trade routes. *Proc Natl Acad Sci U S A* 116(16):7622-7623.

27. Bolger AM, Lohse M, & Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114-2120.

28. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.arXiv:1303.3997.

29. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.

30. Van der Auwera GA*, et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11 10 11-11 10 33.

31. Sahl JW, *et al.* (2016) NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. *Microb Genom* 2(8):e000074.

32. Kurtz S, *et al.* (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5(2):R12.

33. Nguyen LT, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32(1):268-274.

34. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312-1313.

35. Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3(2):217-223.

36. Robinson JT, *et al.* (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24-26.

Figure legends

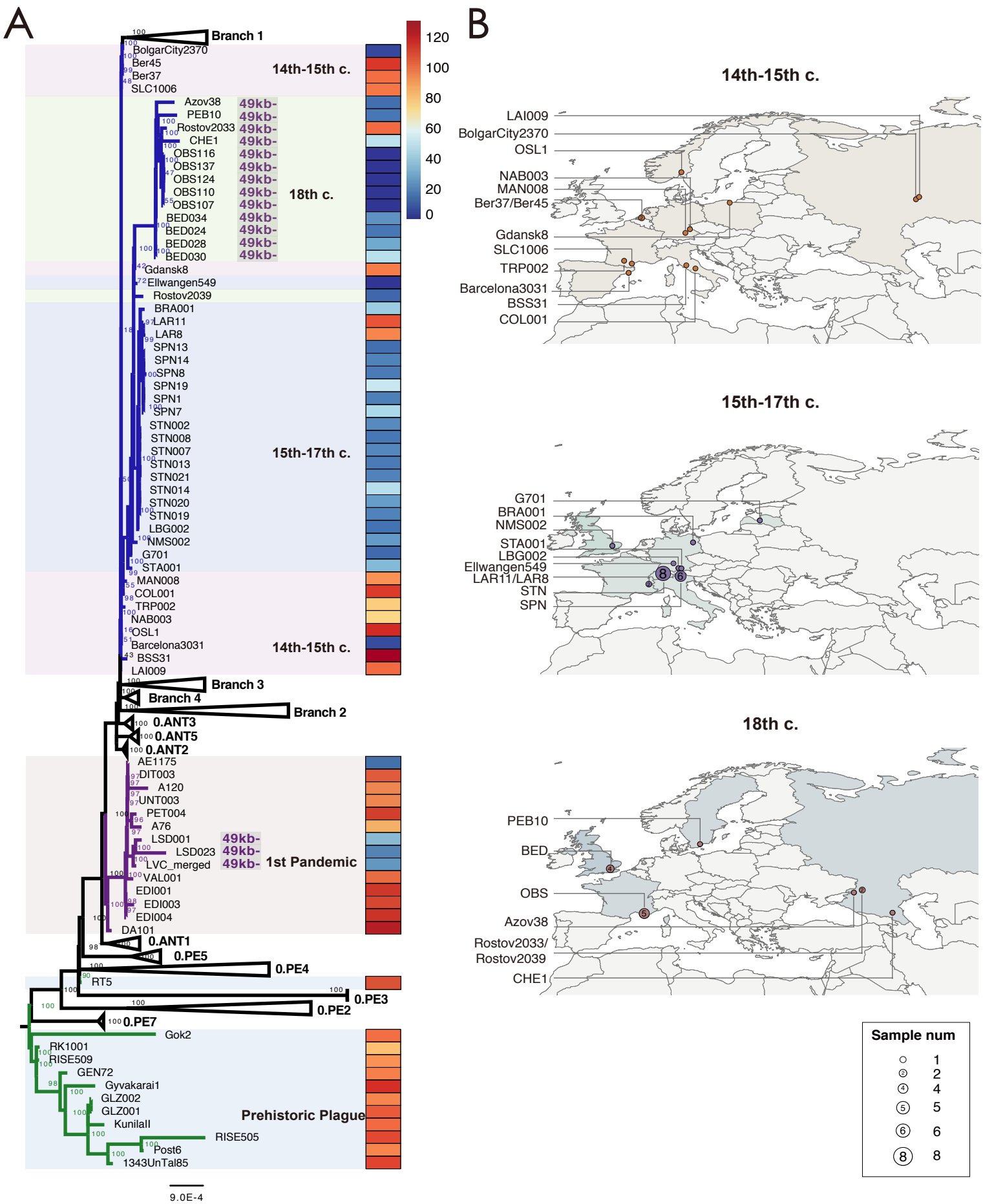
Fig 1. Phylogeny and archaeological site locations of ancient genomes. (A) A maximum likelihood phylogeny was obtained with 574 genomes of *Y. pestis* (including 75 ancient genomes) involved, based on 12,608 SNPs. The numbers at each node indicate the bootstrap values of 1,000 replicates. Branches highlighted in blue correspond to the Second Pandemic, which is subdivided in three groups: the 14th-15th century group, which also includes the Black Death and the *Pestis secunda* (1357-1366) strains, the 15th-17th century group, and the 18th century group (which includes also the BED genomes for homogeneity). Branches in purple correspond to the First Pandemic and branches in green correspond to the prehistoric plague. The ratio between the depth of *pla* and that of the entire pPCP1 plasmid for all ancient

419 genomes is shown in the rightmost heatmap, with a color scale ranging from 0 (dark
420 blue) to 130+ (dark red). (B) Geographic distribution of the three waves during the
421 Second Pandemic.

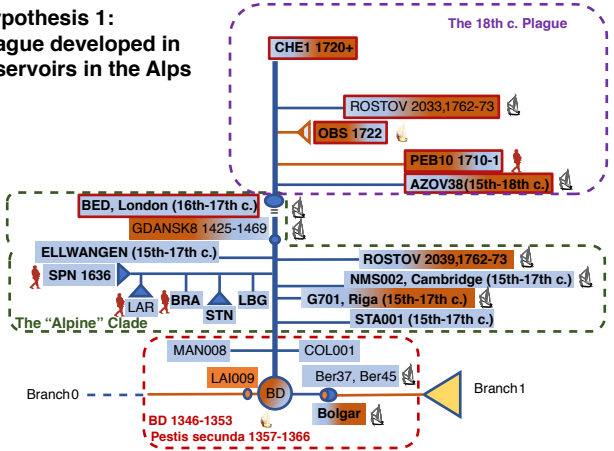
422 **Fig. 2. Schematic comparison between the two main hypotheses for the**
423 **interpretation of the *Y. pestis* phylogeny of the Second Plague Pandemic.**
424 Historic and evolutionary information is included in the schematic figures. In addition
425 to the symbols explained in the figure, we outlined in red the strains showing the
426 49kb deletion. *Pla*-decay (meaning both, full or partial absence of the *pla*-gene) is
427 indicated by the names in bold.

428 **Fig 3. The decay of the *pla*-gene.** (A) Depth plot of the pPCP1 plasmid in strain
429 CHE1 using IGV. The annotated genes of the pPCP1 plasmid are marked with blue
430 bars. The average sequencing depth of whole pPCP1 plasmid is 195.65X, while the
431 average sequencing depth of the *pla*-region is 96.04X. (B) Group-wise comparison of
432 the ratio between the depth of *pla* and that of whole pPCP1 plasmid among three
433 waves of the Second Pandemic. Boxplots depict the upper, median, and lower
434 quartiles of the ratios, individual dots indicate outliers that lie outside of 1.5 times the
435 interquartile range, and vertical lines indicate the range of all ratios except for outliers.
436 The *p*-value of group-wise comparison using the Wilcoxon-test are labeled on the top,
437 two of which are statistically significant ($p<0.05$). Data of Fig. 3b are provided in
438 Dataset S4.

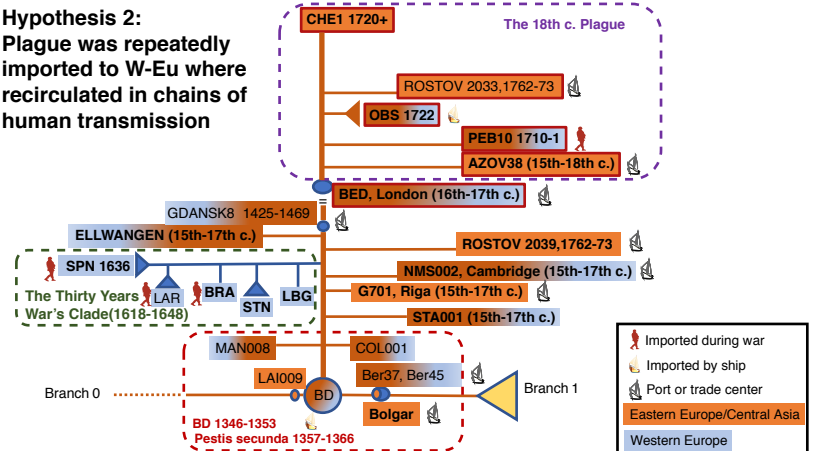
439


A
B
Branch 1
BolgarCity2370
Ber45
Ber37
SLC1006
14th-15th c.
Azov38
PEB10
Rostov2033
CHE1
OBS116
OBS137
OBS124
OBS110
OBS107
BED034
BED024
BED028
BED030
49kb-
18th c.
Gdansk8
Ellwangen549
Rostov2039
BRA001
LAR11
LAR8
SPN13
SPN14
SPN8
SPN19
SPN1
SPN7
STN002
STN008
STN007
STN013
STN021
STN014
STN020
STN019
LBG002
NMS002
G701
STA001
15th-17th c.
MAN008
COL001
TRP002
NAB003
OSL1
Barcelona3031
BSS31
LAI009
14th-15th c.
Branch 3
Branch 4
Branch 2
0.ANT3
0.ANT5
0.ANT2
AE1175
DIT003
A120
UNT003
PET004
A76
LSD001
LSD023
LVC_merged
VAL001
EDI001
EDI003
EDI004
DA101
1st Pandemic
0.ANT1
0.PE5
0.PE4
RT5
0.PE3
0.PE2
0.PE7
Gok2
RK1001
RISE509
GEN72
Gyvakarai1
GLZ002
GLZ001
KunilaII
RISE505
Post6
1343UnTal85
Prehistoric Plague
9.0E-4
120
100
80
60
40
20
0
14th-15th c.
LAI009
BolgarCity2370
OSL1
NAB003
MAN008
Ber37/Ber45
Gdansk8
SLC1006
TRP002
Barcelona3031
BSS31
COL001
15th-17th c.
G701
BRA001
NMS002
STA001
LBG002
Ellwangen549
LAR11/LAR8
STN
SPN
18th c.
PEB10
BED
OBS
Azov38
Rostov2033/
Rostov2039
CHE1
Sample num
1
2
4
5
6
8

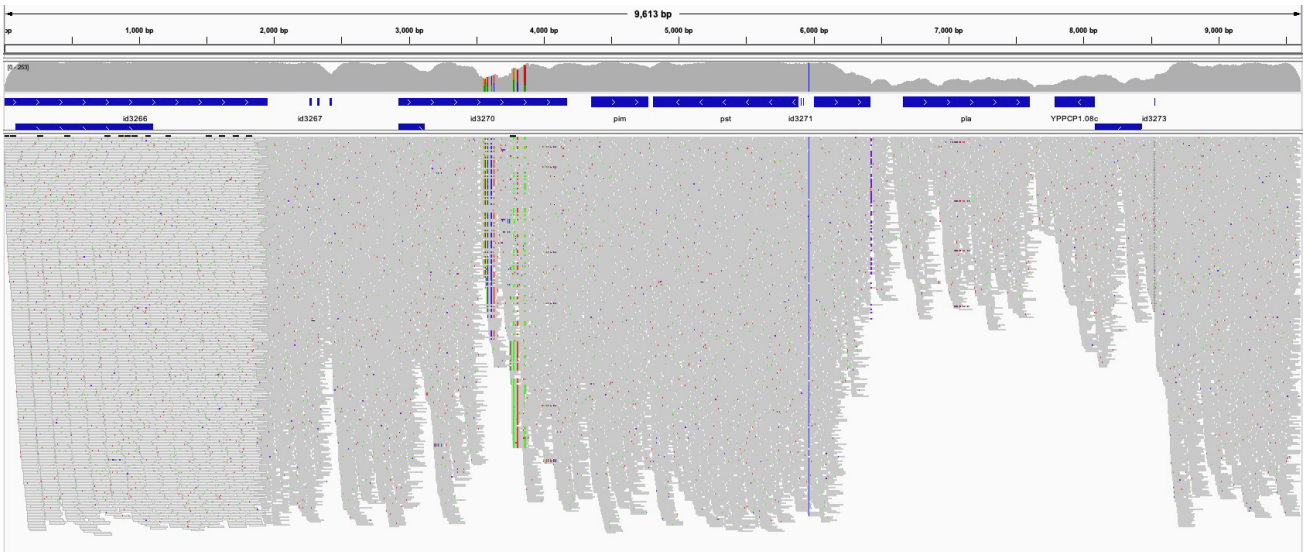
Hypothesis 1:
Plague developed in reservoirs in the Alps



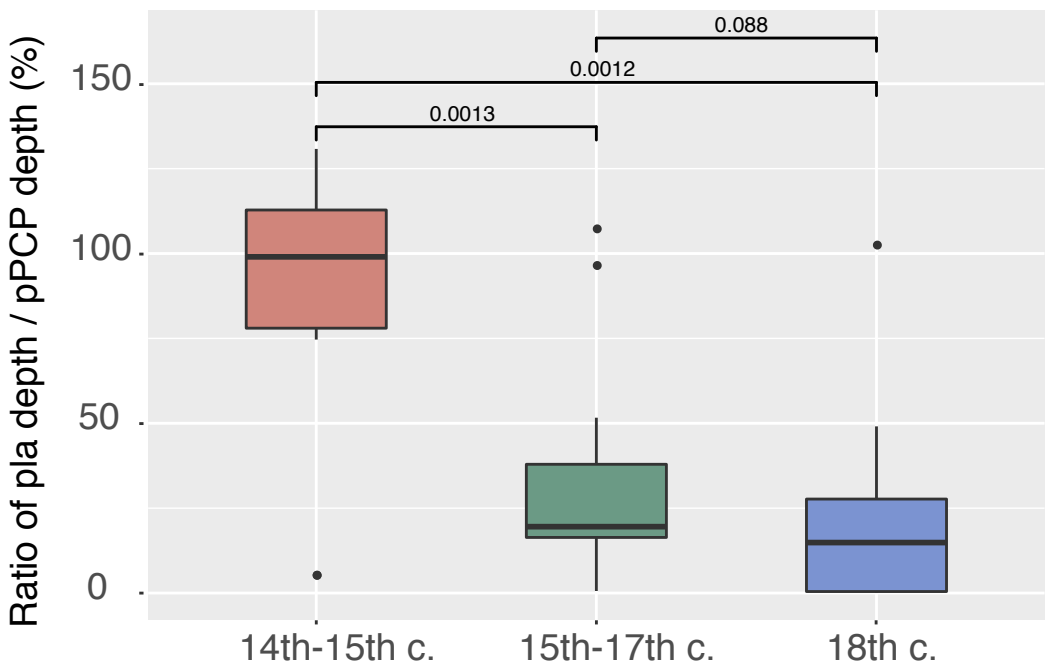
Hypothesis 2:
Plague was repeatedly imported to W-Eu where recirculated in chains of human transmission



A


9,613 bp
1,000 bp
2,000 bp
3,000 bp
4,000 bp
5,000 bp
6,000 bp
7,000 bp
8,000 bp
9,000 bp
id3266
id3267
id3270
pim
pst
id3271
pla
YPPCP1.08c
id3273

B


Ratio of pla depth / pPCP depth (%)
150
100
50
0
0.088
0.0012
0.0013
14th-15th c.
15th-17th c.
18th c.

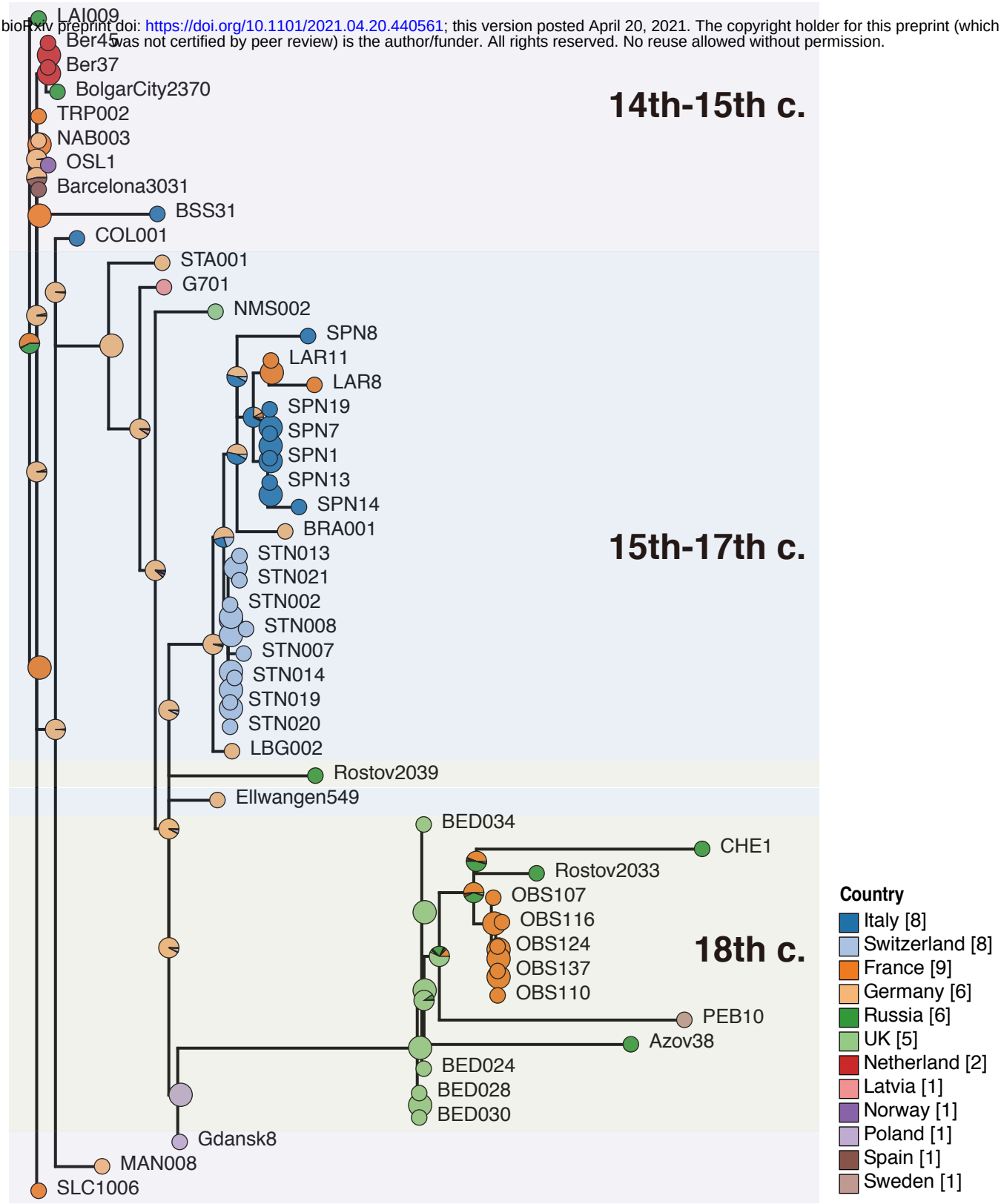


Fig. S1. Accessed transmission among different regions. The phylogenetic tree was obtained using Bayesian MCMC method implemented by phytools package for R. The pie chart at the internal node of the MCMC tree indicated the ancestor source probability of its descendent clades. Except for the Gdansk8 node, all other internal nodes of the tree were sourced from Western Europe, including the node of 14th-15th c., which appears to have originated in France or Germany with a high posterior probability. See the text for an explanation.