Evolution of pathogens reconstructed from ancient human samples

Dissertation

in fulfillment of the requirements for the degree “Dr. rer. nat.”

of the Faculty of Mathematics and Natural Sciences at Kiel University

submitted by

Julian Susat

October 2020

First examiner: Prof. Dr. Tal Dagan

Second examiner: Prof. Dr. Eva Stukenbrock

Date of oral examination: 19.01.2021

Approved for print on: 19.01.2021

Make up your mind to act decidedly and take the consequences. No good is ever done in this world by hesitation.

Thomas Henry Huxley, British biologist 1825-1895

Summary

Infectious diseases caused by viral and bacterial pathogens have affected humans throughout history as exemplified by well-known epidemics, such as the Black Death and the Spanish Flu. Still today, many viruses and multi-drug resistant bacterial strains pose a serious threat to public health and the global economy. To overcome this threat, it is not only necessary to analyze the genomes of recent pathogenic strains but to look back in time. Present-day strains can yield only snapshots of their current genomic diversity and allow only indirect inferences about their long-term evolution. However, with ancient DNA research, it is possible to directly detect and trace changes in the genomic make-up of pathogen progenitors through time, thus providing robust historical and evolutionary evidence e.g. for changes in virulence and accurate estimation of long-term mutation rate.

In this thesis, I conducted a large-scale bioinformatic screening of shotgun sequencing libraries for the presence of pathogens. The libraries were constructed from DNA extracts of ancient human remains almost entirely from Europe (5000 BCE - 20th century CE). Because infection-induced lesions on human remains are rare and often unspecific I performed this screening without any underlying particular hypothesis and did not limit it to a single pathogen. In total, I analyzed 2255 datasets from 1881 individuals with several pathogen-specific pipelines that were individually adapted from publicly available software. The datasets were investigated for their metagenomic content. Any hits indicating the presence of specific pathogens led to the analysis of the respective genomes. Using this approach, I detected five different pathogens. This thesis particularly looks at three of them, *Yersinia pestis (Y. pestis)*, Hepatitis B virus (HBV) and parvovirus B19 (B19).

I identified and reconstructed the oldest *Y. pestis* genome (5000 BP), which helped to reliably estimate the split between the *Y*. *pestis* clade and its progenitor *Y. pseudotuberculosis*. Most of the virulence factors in this genome carried the ancestral states, suggesting a milder disease course in prehistory than in the Middle Ages when infections were caused by more virulent *Y. pestis* strains. In a *Y. pestis* genome from the 17th century CE, I detected a depletion of the *pla* virulence factor compared to previous Black Death strains. This finding prompted me to investigate the *pla* copy numbers in published *Y. pestis* genomes. Overall, 15 genomes representing post-Black Death strains (after the 14th century CE) showed this depletion. The decrease of *pla* might, together with other factors, have contributed to the disappearance of

the plague from Europe. Furthermore, I assembled 12 HBV genomes ranging from the 5th millennium BCE to the Middle Ages. Of these, 2 presented as hybrids between present-day human and non-human primate HBV genomes. However, it is unclear if these strains were the progenitors of nowadays HBV diversity or a result of recombination from an unknown ancient HBV diversity. These hybrids were only found in remains older than 3800 years and no longer exist today. For decades, it has been hypothesized that HBV transmissions between humans and non-human primates took place. The hybrids were thus the first authentic evidence for such transmissions *in vivo*. The other 13 analyzed genomes dated to the Middle Ages reflect the current European HBV diversity. In a fourth study, I screened the remains of 32 children from multiple/mass burials in Berlin (1250-1500 CE). Altogether, I diagnosed 1 *Y. pestis,* 6 HBV and 16 parvovirus B19 infections, including 2 cases with double infections. This high disease burden probably contributed to the considerable child mortality at that time.

The research conducted helps to better understand the history, distribution, long-term evolution and genetic properties of these pathogens. It expanded the temporal boundaries of ancient DNA pathogen research by reconstructing viral genomes up to 7000 years old and shows that even minute traces can be worthwhile chasing.

Zusammenfassung

Durch virale und bakterielle Erreger verursachte Infektionskrankheiten haben den Menschen im Laufe der Geschichte immer wieder heimgesucht - dies zeigen nicht zuletzt prominente Epidemien wie der Schwarze Tod und die Spanische Grippe. Auch heute noch stellen viele Viren und multiresistente Bakterienstämme eine ernsthafte Bedrohung für die öffentliche Gesundheit und die Weltwirtschaft dar. Um diese Bedrohung zu überwinden, ist es nicht nur notwendig, die Genome rezenter pathogener Stämme zu analysieren, sondern auch einen Blick in die Vergangenheit zu werfen. Analysen heutiger Viren- und Bakterienstämme können zwar ihre aktuelle genomische Vielfalt widerspiegeln, erlauben jedoch nur indirekte Rückschlüsse auf ihre langfristige evolutionäre Entwicklung. Forschung an alter DNA ermöglicht es hingegen, Veränderungen in den Genomen von Vorläufern moderner Erreger im Laufe der Zeit direkt nachzuweisen und nachzuvollziehen und damit belastbare historische und evolutionäre Belege für z.B. Veränderungen der Virulenz oder akkurate Schätzungen zur Langzeit-Mutationsrate zu liefern.

In dieser Arbeit führte ich ein groß angelegtes bioinformatisches Screening von Shotgun-Sequenzierlibraries auf das Vorhandensein von Krankheitserregern durch. Die Libraries wurden aus DNA-Extrakten alter menschlicher Überreste, welche fast ausschließlich aus Europa kamen, (5000 v. Chr. - 20. Jahrhundert n. Chr.) erstellt. Da infektionsbedingte Läsionen an menschlichen Überresten selten und oft unspezifisch sind, führte ich dieses Screening explorativ und ohne eine spezielle zugrundeliegende Hypothese, welche die Arbeit auf einen Erreger beschränkte, durch. Insgesamt analysierte ich 2255 Datensätze von 1881 Personen mit mehreren erregerspezifischen Pipelines, die ich aus öffentlich zugänglichen Softwares individuell auf meine Fragestellung anpasste. Die Datensätze wurden auf ihren metagenomischen Inhalt untersucht. Wiesen die Daten auf das Vorhandensein spezifischer Erreger hin, so wurden im Folgenden die jeweiligen Genome analysiert. Mit diesem Ansatz konnte ich fünf verschiedene Pathogene nachweisen. Diese Arbeit befasst sich insbesondere mit *Yersinia pestis* (*Y. pestis*), Hepatitis-B-Virus (HBV) und Parvovirus B19.

Ich identifizierte und rekonstruierte das älteste Genom von *Y. pestis* (3000 v. Chr.). Dadurch konnte die Spaltung zwischen der *Y. pestis*-Klade und ihrem Vorläufer *Y. pseudotuberculosis* nun zuverlässig abgeschätzt werden. Die meisten Virulenzfaktoren in diesem Genom trugen die anzestralen Varianten was auf einen eher milden Krankheitsverlauf in der prähistorischen

Zeit hindeutete. Im Mittelalter hingegen wurden Infektionen eher durch virulentere *Y. pestis*-Stämme verursacht. In einem *Y.-pestis*-Genom aus dem 17. Jahrhundert n. Chr. stellte ich im Vergleich zu früheren Pest-Stämmen (Stämme welche direkt am Schwarzen Tod beteiligt waren) eine geringere Kopiezahl des *pla*-Virulenzfaktors fest. Dieser Befund veranlasste mich, die *pla*-Kopienzahlen in bereits publizierten *Y. pestis*-Genomen zu untersuchen. In 15 Genomen (alle nach dem Schwarzen Tod, nach dem 14. Jahrhundert n. Chr.) konnte ich ebenfalls eine *pla*-Verminderung nachweisen. Die geringere pla-Kopienzahl könnte demnach, neben anderen Faktoren, zum Verschwinden der Pest aus Europa beigetragen haben.

Darüber hinaus habe ich im Rahmen dieser Arbeit 12 HBV-Genome assembliert, die vom 5. Jahrtausend v. Chr. bis zum Mittelalter datieren. Davon stellten sich zwei als Hybride zwischen menschlichen und nicht-menschlichen HBV-Genomen von Primaten heraus. Es ist jedoch unklar, ob diese Stämme die Vorfahren der heutigen HBV-Vielfalt waren oder ein Ergebnis der Rekombination aus einer unbekannten alten HBV-Vielfalt. Diese Hybriden wurden nur in Überresten gefunden, die älter als 3800 Jahre waren. Heutzutage existieren diese nicht mehr. Jahrzehntelang galt die Hypothese, dass HBV-Übertragungen zwischen Menschen und nichtmenschlichen Primaten stattgefunden haben. Die von mir identifizierten Hybriden sind nun der erste authentische Beweis für solche Übertragungen *in vivo*. Die anderen 10 analysierten Genome stammen aus dem Mittelalter und spiegelten die aktuelle europäische HBV-Vielfalt wider.

In einer vierten Studie untersuchte ich die Überreste von 32 Kindern aus Mehrfach-/Massenbestattungen in Berlin (1250-1500 n.Chr.). Insgesamt diagnostizierte ich eine *Y. pestis*, sechs HBV- und 16 Parvovirus B19-Infektionen, darunter zwei Fälle mit Doppelinfektionen. Diese hohe Krankheitslast trug wahrscheinlich zu der damals beträchtlichen Kindersterblichkeit bei.

Die in dieser Arbeit durchgeführte Forschung hilft, die Geschichte, Verbreitung, langfristige Evolution und genetischen Eigenschaften dieser Pathogene besser zu verstehen. Sie erweiterte die zeitlichen Grenzen der alten DNA-Erregerforschung durch die Rekonstruktion von bis zu 7000 Jahre alten viralen Genomen und zeigt, dass es sich lohnt, selbst winzigen Spuren nachzugehen.

Table of content

1.

General introduction

Pathogens now and then

Despite the technical and medical advances in the last 100 years, infectious agents are still a serious threat to global healthcare and cause massive multi-sectoral damage (1). Different socio-economic factors like urbanization, colonization, modification of food production and increased environmental pollution have led to the emergence and spread of several new or already known pathogenic bacteria and viruses deemed under control (2,3). Notorious are, for example, the human immunodeficiency virus (HIV), the Ebola virus, the novel coronavirus 2019-nCoV, the Zika virus and multi-drug-resistant-bacterial strains (MDR) which currently cause major problems in hospitals (4). A constant and never-ending evolutionary arms race between the host and the pathogen is ongoing (5). This model refers to the fixation of new variants in a pathogen population which leads to selective pressure on the host and vice versa (6). However, human pathogens are not just a challenge for modern societies. Research on present-day traditional forager-farmer societies with limited or no access to healthcare identified infectious diseases as the primary cause of death (72%) (7). This observation indicates that also ancient human populations, without e.g. antibiotics and vaccination, might have been heavily affected by pathogens and their corresponding diseases. Screening these ancient populations for the presence of pathogens can help answering evolutionary, biological, epidemiological and historical questions like:

- How is the ancestral genomic makeup composed?
- Can we find traces of horizontal gene transfer and recombination?
- How has the genome changed over time?
- Is synteny conserved over time?
- Can we find resistance progenitor genes against e.g. Antibiotics?
- Did the pathogenicity change over time?
- How did modern medicine influence the genome?
- What is the true long-term mutation rate?
- What was the rate of infection among ancient populations?
- Are certain historical events the result of a pandemic?
- Can specific burial circumstances like mass graves be explained?

With the help of ancient DNA (aDNA) answers to these questions can be obtained. These answers can provide us with a better understanding of pathogens and their evolution and help to successfully fight them in a modern clinical environment.

Ancient DNA research

The field of ancient DNA deals with the isolation and subsequent analysis of DNA from organic specimens which are older than 100 years. The DNA can be retrieved from different sources such as coprolites, herbaria, dental calculus, soft tissue and many others, but the most common sources are bones and teeth (8). The petrous bone is the best material to answer human related questions such as migration patterns, admixture events, immunogenetic status and the discovery of new hominin species (9–12). It is part of the temporal bone and the hardest and most dense bone in the human body (13). This density and very limited natural access to the bone results in a low-level of contamination and a high yield of endogenous DNA (14,15). A widely used material to answer pathogen related questions with the help of aDNA are teeth as they are directly connected to the bloodstream and the pulp chamber is well protected from exogenous contamination (15,16). Their direct connection to the bloodstream allows for blood-borne pathogens to be present during the infection and thus leave a trace in form of obtainable genetic material (17–19). Oral and lung pathogens that are not spread via the bloodstream can be detected by sampling and analyzing ancient dental calculus (20). Apart from that mummified tissue from crypts is an excellent material as it naturally has a reduced environmental signal and is stored in a dry surrounding which aids DNA preservation. However, this kind of sample material is very scarce (21).

aDNA research was established in 1984 when DNA was recovered from remains of the extinct South African quagga (22). This scientific achievement depended upon the invention of the Polymerase Chain Reaction (PCR) in 1983, which allowed amplification of single-copy DNA molecules to an extent that was not achievable before (23). Prior to PCR it was impossible to study DNA from ancient samples since the amount of authentic endogenous DNA is very low compared to that in a modern sample. One of the early PCR-based studies related to pathogens was the detection of *Mycobacterium tuberculosis* (TB) in human remains (24).

However, the primers used for detection of the pathogen were shown to be unspecific as the PCR would also result in a fragment based on soil contamination by closely related mycobacterial species (25). Apart from that several PCR-based aDNA studies in the 90's resulted in spectacular findings, such as dinosaur DNA that was later identified as modern contaminant and brought up the notion of aDNA research being unreliable (26–28). Despite implementing cautious guidelines regarding the work with aDNA, false-positive findings were still a problem resulting in vivid discussions regarding circumstances and conditions of the sample and the applied methods (29–34). While PCR enabled the study of aDNA at the beginning, it was the emergence of Next Generation Sequencing (NGS) in 2005 that allowed technological and scientific breakthroughs in this field. High-throughput sequencing allowed for millions of sequences to be generated per sample. Detection of infectious agents was no longer limited to PCR and a limited number of specific single loci (35). Combined with capture-based arrays to enrich the DNA of single or multiple organisms this technology enabled a completely new level of analysis and helped to overcome the limitations of targeting single loci (36). Capture-based techniques also helped to overcome the complexity of metagenomic by enriching e.g. the desired pathogen (18,37). This kind of research can provide us with information modern studies never could. For example, emergence of infectious diseases is evaluated using the modern diversity of pathogens (38,39). These evaluations, when backed up by paleopathological studies and historical records, can provide an idea about what happened. Nevertheless, they certainly suffer from limitations like taking only nowadays genetic diversity into account or trying to estimate long-term mutation rates which might be influenced by short-term effects (40–42). This time dependency of substitution rates is especially known from viruses (43,44). Information obtained by ancient pathogens can be used to develop models for the reconstruction of past epidemics or the prediction of current and future epidemics (45–47).

aDNA can be a powerful tool if the analysis is carried out carefully and pitfalls are tackled properly. Whilst an organism is alive, damage in the DNA is repaired by enzymes. However, no DNA repair mechanisms occur after death. Due to environmental factors like salt level, pH, radiation and temperature but also the presence of microorganisms and autolysis several kinds of damage can alter the nucleic acids and degrade the DNA into small fragments. These damages make analysis of the genetic material thoroughly challenging (Table 1).

Table 1. Types of DNA damage, underlying mechanisms, effects on DNA recovery and suggested solutions. Modified from (48)

Damage type	Mechanism	Effect	Solution
strand breaks	nuclease activity, microorganism degradation, desiccation, heat, chemicals, direct cleavage (hydrolysis), depurination (hydrolysis)	low quantity of surviving DNA, short molecule length	amplification of short DNA fragments (<50 - 150 bp)
miscoding lesions via hydrolyses	deamination causes miscoding lesions, adenine to hypoxanthine, cytosin to uracil, 5-methylcytosine to thymine, guanine to xanthine Terminal substitutions	base misincorporations C -> T G -> A	multiple extractions and amplifications, use of uracil-DNA glycosylase (UDG) to remove uracil and proofreading polymerase Terminal base damage can be an important authentication criterion for aDNA
blocking and miscoding lesions via oxidation	base modification 5-OH-5-methylhydantoin (blocking) 5-OH-hydantoin (blocking) 8-oxoguanosine (miscoding G -> T)	no amplification, jumping PCR, base misincorporation	special library preparation and blunt end repair

crosslinks	DNA-DNA crosslinks via alkylation DNA-protein crosslinks	no amplification	PTB (N-phenacylthaziolum bromide) to cleave crosslinks (not a standardized procedure)

Fortunately, many of the obstacles can be solved by modern lab techniques (Table 1). In contrast to modern DNA, an aDNA extract consists of a mix of endogenous organismic and exogenous environmental DNA, in which the aDNA only constitutes a small part of the whole DNA in the sample (49). However, some of the criteria in Table 1 like terminal substitution patterns and short read lengths can be used to identify and authenticate aDNA fragments in the metagenomic sample and, thus, differentiate authentic sequences from the noise.

Detection of pathogens in ancient samples

Detection of pathogens in ancient samples is not a trivial task. As mentioned above, the aDNA extract consists of a mixture of organism and for the most part, authentic pathogen DNA make up less than 0.1% of the whole DNA content (50,51). However, this number strongly depends on the pathogen, the stage of infection and the sampled material. In the easiest but also rarest cases the causative agent of a disease is already suspected from e.g. lesions on the bones or historical records. However, there are only few diseases, such as leprosy, brucellosis, treponemal disease and tuberculosis, which can lead to visible skeletal pathologies (21,52–55). These lesions can be unspecific as signs of inflammation on the bones can also result from non-pathogen related diseases.

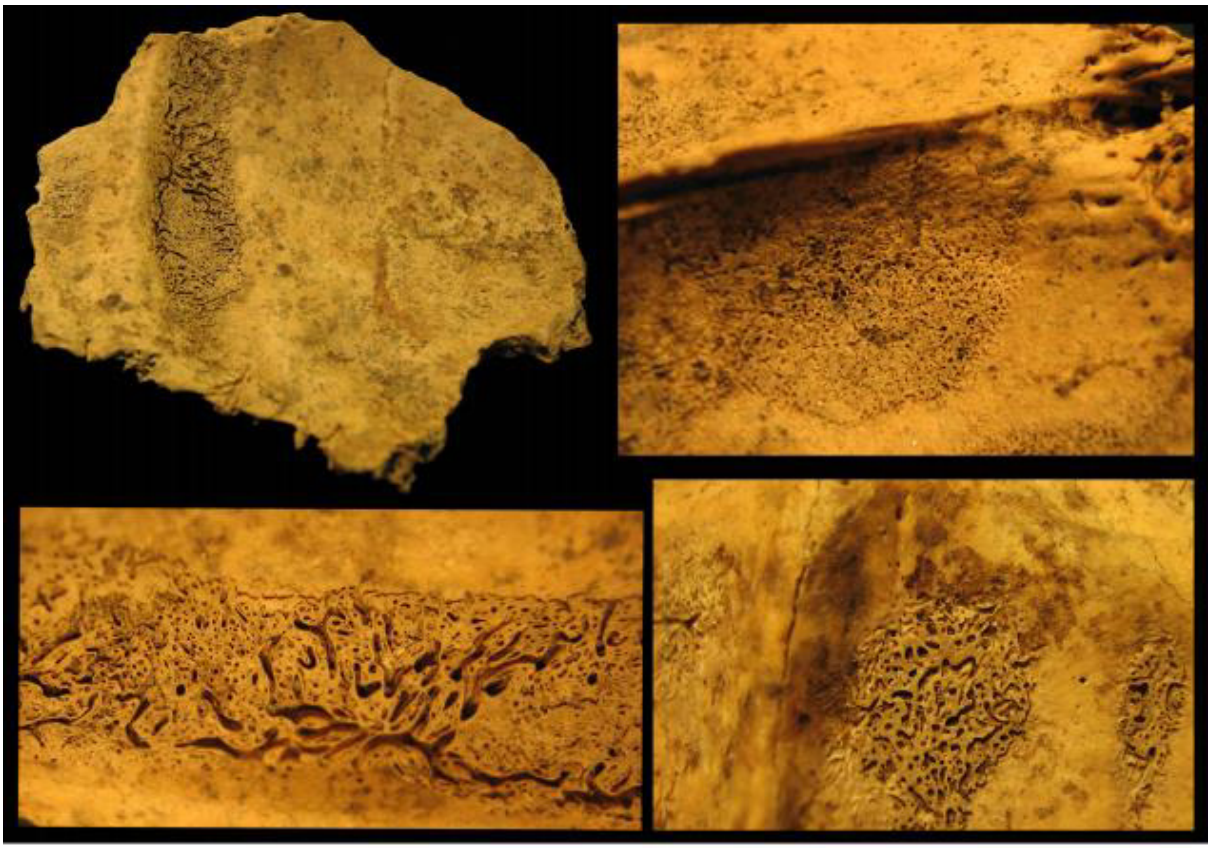


Figure 1. Endocranial patterns on the frontal and occipital bones. These lesions can be the result of inflammation, anemia, iron deficiency or the infection with a pathogen. Due to the unspecific nature further genetic investigation is needed to identify the origin. Taken from: (56)

Therefore, for most diseases a reliable and specific differential diagnosis based on ancient bone material often unfeasible and can provide only very limited insights into the pathogen´s spread and evolution (57). Furthermore, historical records are sparse, subjective and rarely date back further than to the Middle Ages. A good example of an older historical source is a stele dated to 1500BCE that depicts an Egyptian priest exhibiting signs of polio infection (58). But one must keep in mind that the absence of bone lesions or historical evidence is no evidence for the absence of a pathogen. In addition, it must be taken into consideration that the genetic distance between a modern-day pathogen and its ancient relative might be larger than expected or that there is even an unknown pathogen in the sample. Furthermore, modern contamination with closely related nonpathogenic organisms can be a problem and disturb subsequent analysis. Modern contamination might lead to an overabundance of certain sequences and therefore disguise the presence of ancient sequences or generate a false positive result. All these complications together with the general small amount of target DNA have led to different bioinformatic strategies and ways of processing the data.

Bioinformatics in aDNA

If a particular pathogen is suspected, then a simple mapping of the short generated aDNA reads against the reference is carried out using for example BWA (59). Given different parameters like the number of mismatches that are allowed, all the sequences in the sample are compared against a single reference genome. However, as mentioned above, this rarely is the case. Thus, an undirected metagenomic screening is often performed. A generally accepted approach is using BLAST or MALT which both rely on heuristic approaches when comparing samples in the dataset to genomic databases retrieved from e.g. the National Center for Biotechnology Information (NCBI) (60,61). Alternatives are for example Metaphlan that utilizes thousands of marker genes or Kraken which is based on k-mer matching (62,63). Additionally, discovery of extinct pathogens that have no recent descendant should in theory be possible. Modern studies have shown that Hidden Markov Models can detect unknown viruses in cross-validation experiments (64). Furthermore, metagenomic assembly followed by subsequent binning and reassembly has shown to be effective (65). However, given the small amount of authentic ancient DNA no one has successfully achieved this. If the initial screening results in the detection of pathogen sequences, there are several options for the generation of more pathogen sequences in the wet lab:

I) Using a targeted capture to enrich the desired sequences
II) Generating more reads from the already existing library without a capture
III) Creating a completely new library from the sample

The capture approach uses probes which are generated from one or several reference genomes (66,67). These probes are single stranded and allow for the binding of complementary DNA sequences. If the probes are generated from a pathogen genome only sequences of this pathogen should bind. All DNA that did not bind will be washed away and the bound fragments can be eluted and sequenced. Generating more reads from the existing library is only effective if there was already a large number of pathogen sequences in the initial sequencing present. Creating a new library from the sample can be an option if no capture for the desired organism is available and the initial library contains only a few sequences of interest. If eventually enough reads are generated to provide a good coverage of the reference genome, a phylogenetic analysis is carried out to put the finding in context and relation to

already known genomes of the same organism. Whole genome alignments using for example MAFFT or alignments based on Single Nucleotide Polymorphisms (SNPs) using the MultiVCFAnalyzer are generated as input for several phylogenetic programs, such as MEGA6, RAxML, MrBayes, PhyML or PAUP (68–74). Apart from this core set of tools, several more specific and in-depth analyses can be carried out to examine functionality of SNP's and evolution through time in detail. SNP can be annotated using SNPEff, molecular dating can be done using BEAST2 or a de novo assembly can be performed using for example SPAdes (75–77). As mentioned above, characteristics of aDNA sequences can be used to authenticate the ancient origin of generated reads. The software mapDamage2.0 generates damage estimation plots based on terminal substitutions of the reads and provides a measurement for the integrity of the data (78). Except mapDamage, all the programs mentioned above are also used in modern research as they were not specifically designed for aDNA studies. Nonetheless, careful modification of parameters allows to meet the specifications for working with short and damaged sequences. Moreover, it should be kept in mind that the discussed software represents only a subset of all the possible tools that can be used to analyze aDNA.

Bioinformatic problems

One of the major bottlenecks in aDNA analysis is the reliable classification of metagenomic sequences. This is also a problem with modern diagnostics where reliable classification of all sequences is still an unsolved issue (79–82). In aDNA, this classification problem is multiplied by the sheer complexity of the sample. Within millions of sequences there might only be a small fraction of sequences that belong to an organism of interest. Software like Heuristic Operations for Pathogen Screening (HOPS) can help to identify and extract these positives with respect to aDNA criteria, distribution of the sequences with respect to the reference and calculation of the edit distance (genetic distance) (83). This method, however, is still prone to reference bias as it relies on a database of genomes.

Sensitivity vs specificity and linked to that the choice of parameters presents another problem in aDNA research. In the optimal case, the screening is specific but also very sensitive. In reality, however, there is a tradeoff one would have to make sacrificing either one or the other along the way. Different parameters can lead to drastically different results when comparing datasets as there is no unified standard for workflows among the aDNA community (20). And

after all one must keep in mind that a negative screening is only showing that we are unable to find a pathogen. There might have been one we can no longer detect, or we are just unable to detect that there is one.

aDNA Patho-genomics

Until now, high-throughput sequencing combined with new bioinformatic methods has not only allowed for the detection of several ancient pathogens, but also the retrieval of complete genomes. The etiological agent of the Spanish flu was the first complete ancient pathogen genome that was published (84–90). However, the sampled soft tissue which was preserved in paraffin was from individuals that clearly died from the infection and just about 100 years old (86). These circumstances and the small genome size, which was sequenced by cycle sequencing, led to the successful reconstruction of the virus genome. The first complete genome of an ancient pathogen based on a combination of NGS and capture was published in 2011; it belongs to the bacterium *Yersinia pestis* which is the causative agent of plague (18). This bacterium is widely known as being responsible for the Black Death which was ravaging in the 14th century in Europe as one of the most disastrous pandemics in human history killing approximately half the population in its wake (91). So far, 38 genomes of this pathogen have been recovered and have helped to understand its origin as well as its spread around the world. Over time the field has expanded into several directions with in-depth studies of specific pathogens and their evolution (92). The diseases caused by *Mycobacterium leprae* (Leprosy) and *Mycobacterium tuberculosis* (Tuberculosis) can both lead to indicative traces on the bones which made them prime targets for researchers. Infections with *Yersinia pestis* leave no traces on the bones but due to the high prevalence in the middle ages and dedicated burial sites it is also a prime target. In contrast pathogens like the Hepatitis B virus, parvovirus B19, *Vibreo Cholerae*, *Helicobacter pylori* or *Salmonella enterica* are just chance findings neither backed up by traces on the bones nor historical records.

Aim of this thesis

To overcome technical limitations and detect different ancient pathogens, a systematic, large scale molecular and bioinformatic pathogen screening of dated human remains was pursued in this thesis. A positive screening should result in genome reconstruction, phylogenetic analysis, functional analysis and the evaluation of the historical, biomedical and evolutionary context. The screened samples consisted of petrous bones and teeth that rarely exhibited paleopathological evidence. Timeframe of the screened samples was restricted and ranges from 6000 BC up to the 20th century CE. During the course of this thesis approximately 2255 samples from 1881 individuals were screened for the presence of pathogens (Fig. 2, Table S1).

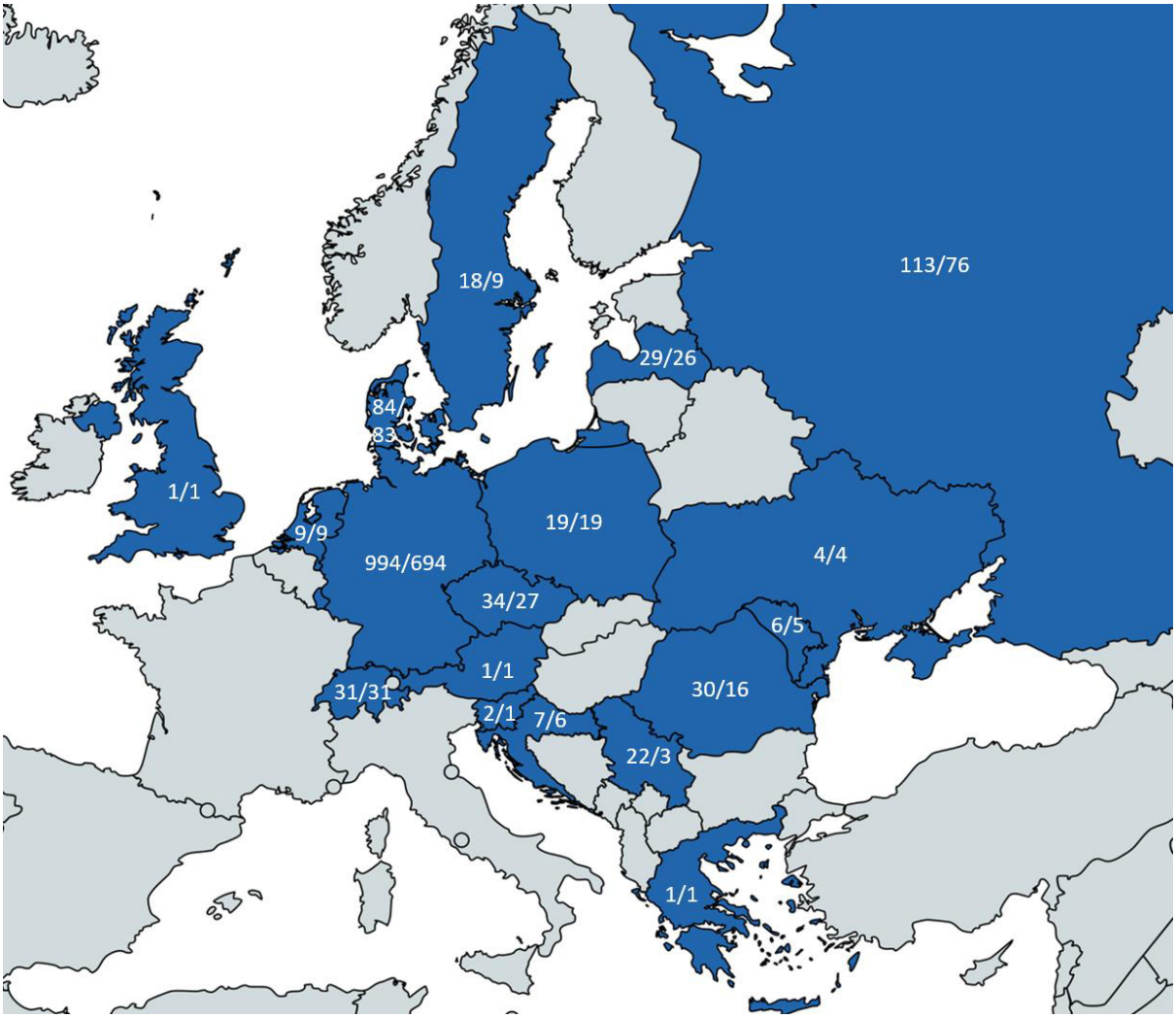


Figure 2. Sample sites in Europe. The first number is indicating the number of samples generated. The second number is indicating the number of sampled individuals. Not shown in this map: Mexico 21/21, India 13/13 and the samples sequenced in other laboratories 786/786. The map was generated using PowerPoint and mapchart.net.

This large number should make chance findings possible and in addition help addressing the following bioinformatic, evolutionary and epidemiological points:

I) Methodology and optimization of the pathogen screening
II) Origin, spread and diversity of the pathogen
III) Several individuals from the same burial site should provide an insight into infection rate and microevolution
IV) The time frame should make evolutionary changes visible and provide information on changes in virulence

1. Smith KM, Machalaba CC, Seifman R, Feferholtz Y, Karesh WB. Infectious disease and economics: The case for considering multi-sectoral impacts. Vol. 7, One Health. 2019.

2. Patz JA, Daszak P, Tabor GM, Aguirre AA, Pearl M, Epstein J, et al. Unhealthy landscapes: Policy recommendations on land use change and infectious disease emergence. Environ Health Perspect. 2004;112(10):1092–8.

3. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. Vol. 430, Nature. 2004. p. 242–9.

4. Tanwar J, Das S, Fatima Z, Hameed S. Multidrug resistance: An emerging crisis. Vol. 2014, Interdisciplinary Perspectives on Infectious Diseases. 2014.

5. Nielsen R. Molecular Signatures of Natural Selection. Annu Rev Genet. 2005;39(1):197–218.

6. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Vol. 32, Nature Genetics. 2002. p. 569–77.

7. Finch CE. Evolution of the human lifespan and diseases of aging: Roles of infection, inflammation, and nutrition. Proc Natl Acad Sci U S A. 2010;107(SUPPL. 1):1718–24.

8. Green EJ, Speller CF. Novel substrates as sources of ancient DNA: Prospects and hurdles. Vol. 8, Genes. 2017.

9. Meyer M, Kircher M, Gansauge MT, Li H, Racimo F, Mallick S, et al. A high-coverage genome sequence from an archaic Denisovan individual. Science (80-). 2012;338(6104):222–6.

10. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature. 2015;522(7555):207–11.

11. Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, et al. The Beaker phenomenon and the genomic transformation of northwest Europe. Nature. 2018;555(7695):190–6.

12. Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, et al. The genomic history of southeastern Europe. Nature. 2018;555(7695):197–203.

13. Frisch T, Sørensen MS, Overgaard S, Lind M, Bretlau P. Volume-referent bone turnover estimated from the interlabel area fraction after sequential labeling. Bone. 1998;22(6):677–82.

14. Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, et al. Optimal ancient DNA yields from the inner ear part of the human petrous bone. PLoS One. 2015;10(6).

15. Hansen HB, Damgaard PB, Margaryan A, Stenderup J, Lynnerup N, Willerslev E, et al. Comparing ancient DNA preservation in petrous bone and tooth cementum. PLoS One. 2017;12(1).

16. Margaryan A, Hansen HB, Rasmussen S, Sikora M, Moiseyev V, Khoklov A, et al. Ancient pathogen DNA in human teeth and petrous bones. Ecol Evol. 2018;8(6):3534–42.

17. Drancourt M, Aboudharam G, Signoli M, Dutour O, Raoult D. Detection of 400-year-old Yersinia pestis DNA in human dental pulp: An approach to the diagnosis of ancient septicemia. Proc Natl Acad Sci U S A. 1998;95(21):12637–40.

18. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, et al. A draft genome of Yersinia pestis from victims of the Black Death. Nature. 2011;478(7370):506–10.

19. Mühlemann B, Jones TC, De Barros Damgaard P, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature. 2018;557(7705):418–23.

20. Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, Burbano HA, et al. A Robust Framework for Microbial Archaeology. Annu Rev Genomics Hum Genet. 2017;18(1):321–56.

21. Kay GL, Sergeant MJ, Giuffra V, Bandiera P, Milanese M, Bramanti B, et al. Recovery of a medieval Brucella melitensis genome using shotgun metagenomics. MBio. 2014;5(4).

22. Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC. DNA sequences from the quagga, an extinct member of the horse family. Nature. 1984;312(5991):282–4.

23. Paabo S. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proc Natl Acad Sci U S A. 1989;86(6):1939–43.

24. Spigelman M, Lemma E. The use of the polymerase chain reaction (PCR) to detect Mycobacterium tuberculosis in ancient skeletons. Int J Osteoarchaeol. 1993;3(2):137–43.

25. Müller R, Roberts CA, Brown TA. Complications in the study of ancient tuberculosis: Presence of environmental bacteria in human archaeological remains. J Archaeol Sci. 2016;68:5–11.

26. Golenberg EM, Giannasi DE, Clegg MT, Smiley CJ, Durbin M, Henderson D, et al. Chloroplast DNA sequence from a Miocene Magnolia species. Nature. 1990;344(6267):656–8.

27. Cano RJ, Poinar HN, Pieniazek NJ, Acra A, Poinar GO. Amplification and sequencing of DNA from a 120-135-million-year-old weevil. Nature. 1993;363(6429):536–8.

28. Woodward SR, Weyand NJ, Bunnell M. DNA sequence from cretaceous period bone fragments. Science (80-). 1994;266(5188):1229–32.

29. Cooper A. Ancient DNA: Do It Right or Not at All. Science (80-). 2000;289(5482):1139b – 1139.

30. Gilbert MTP, Bandelt HJ, Hofreiter M, Barnes I. Assessing ancient DNA studies. Trends Ecol Evol. 2005;20(10):541–4.

31. Papagrigorakis MJ, Yapijakis C, Synodinos PN, Baziotopoulou-Valavani E. Insufficient phylogenetic analysis may not exclude candidacy of typhoid fever as a probable cause of the Plague of Athens (reply to Shapiro et al.). Vol. 10, International Journal of Infectious Diseases. 2006. p. 335–6.

32. Shapiro B, Rambaut A, Gilbert MTP. No proof that typhoid caused the Plague of Athens (a reply to Papagrigorakis et al.). Vol. 10, International Journal of Infectious Diseases. 2006. p. 334–5.

33. Roberts C, Ingham S. Using ancient DNA analysis in palaeopathology: A critical analysis of published papers, with recommendations for future work. Vol. 18, International Journal of Osteoarchaeology. 2008. p. 600–13.

34. Wilbur AK, Bouwman AS, Stone AC, Roberts CA, Pfister LA, Buikstra JE, et al. Deficiencies and challenges in the study of ancient tuberculosis DNA. J Archaeol Sci. 2009;36(9):1990–7.

35. Hänsch S, Cilli E, Catalano G, Gruppioni G, Bianucci R, Stenseth NC, et al. The pla gene, encoding plasminogen activator, is not specific to Yersinia pestis. BMC Res Notes. 2015;8(1).

36. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol. 2009;27(2):182–9.

37. Bos KI, Jäger G, Schuenemann VJ, Vågene ÅJ, Spyrou MA, Herbig A, et al. Parallel detection of ancient pathogens via array-based DNA capture. Philos Trans R Soc B Biol Sci. 2015;370(1660).

38. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, et al. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nat Genet. 2010;42(12):1140–3.

39. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet. 2013;45(10):1176–82.

40. Robert CA, Buikstra JE. The Bioarchaeology of Tuberculosis: A Global View on a Reemerging Disease. University Press of Florida; 2008.

41. Peck KM, Lauring AS. Complexities of Viral Mutation Rates. Sullivan CS, editor. J Virol [Internet]. 2018 May 2;92(14). Available from: https://jvi.asm.org/content/92/14/e01031-17

42. HO SYW, LANFEAR R, BROMHAM L, PHILLIPS MJ, SOUBRIER J, RODRIGO AG, et al. Time-dependent rates of molecular evolution. Mol Ecol [Internet]. 2011 Aug;20(15):3087–101. Available from: http://doi.wiley.com/10.1111/j.1365-294X.2011.05178.x

43. Duchêne S, Holmes EC, Ho SYW. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc R Soc B Biol Sci. 2014;281(1786).

44. Aiewsakun P, Katzourakis A. Time-Dependent Rate Phenomenon in Viruses. J Virol. 2016;90(16):7184–95.

45. Achtman M. Insights from genomic comparisons of genetically monomorphic bacterial pathogens. Vol. 367, Philosophical Transactions of the Royal Society B: Biological Sciences. 2012. p. 860–7.

46. Mahan MJ, Kubicek-Sutherland JZ, Heithoff DM. Rise of the microbes. Vol. 4, Virulence. 2013. p. 213–22.

47. Mathema B, Kurepina N, Fallows D, Kreiswirth BN. Lessons from molecular epidemiology and comparative genomics. Vol. 29, Seminars in Respiratory and Critical Care Medicine. 2008. p. 467–80.

48. Shapiro B, Hofreiter M. Ancient DNA Methods and Protocols. Humana Press; 2012.

49. Shapiro B, Hofreiter M. A paleogenomic perspective on evolution and gene function: New insights from ancient DNA. Vol. 343, Science. 2014.

50. Devault AM, McLoughlin K, Jaing C, Gardner S, Porter TM, Enk JM, et al. Ancient pathogen DNA in archaeological samples detected with a Microbial Detection Array. Sci Rep. 2014;4.

51. Guellil M, Kersten O, Namouchi A, Bauer EL, Derrick M, Jensen AO, et al. Genomic blueprint of a relapsing fever pathogen in 15th century Scandinavia. Proc Natl Acad Sci U S A.

2018;115(41):10422–7.

52. Andersen JG, Manchester K, Roberts C. Septic bone changes in leprosy: A clinical, radiological and palaeopathological review. Int J Osteoarchaeol. 1994;4(1):21–30.

53. Pigrau-Serrallach C, Rodríguez-Pardo D. Bone and joint tuberculosis. Eur Spine J. 2013;22(SUPPL.4).

54. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, et al. Genome-wide comparison of medieval and modern Mycobacterium leprae. Science (80-). 2013;341(6142):179–83.

55. Schuenemann VJ, Kumar Lankapalli A, Barquera R, Nelson EA, Iraíz Hernández D, Acuña Alonzo V, et al. Historic Treponema pallidum genomes from Colonial Mexico retrieved from archaeological remains. PLoS Negl Trop Dis. 2018;12(6).

56. Köhler K, Marcsik A, Zádori P, Biro G, Szeniczey T, Fábián S, et al. Possible cases of leprosy from the Late Copper Age (3780-3650 cal BC) in Hungary. Spigelman M, editor. PLoS One [Internet]. 2017 Oct 12;12(10):e0185966. Available from: https://dx.plos.org/10.1371/journal.pone.0185966

57. Ortner DJ. Differential Diagnosis of Skeletal Lesions in Infectious Disease. In: Advances in Human Palaeopathology. 2007. p. 189–214.

58. Galassi FM, Habicht ME, Rühli FJ. Poliomyelitis in Ancient Egypt? Vol. 38, Neurological Sciences. 2017. p. 375.

59. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.

60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.

61. Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. Nat Ecol Evol. 2018;2(3):520–8.

62. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012;9(8):811–4.

63. Wood DE, Salzberg SL. Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3).

64. Skewes-Cox P, Sharpton TJ, Pollard KS, DeRisi JL. Profile hidden Markov models for the detection of viruses within metagenomic sequence data. PLoS One. 2014;9(8).

65. Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GGZ, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. Nat Commun. 2014;5.

66. Mertes F, ElSharawy A, Sauer S, van Helvoort JMLM, van der Zaag PJ, Franke A, et al. Targeted enrichment of genomic DNA regions for next-generation sequencing. Vol. 10, Briefings in Functional Genomics. 2011. p. 374–86.

67. Altmüller J, Budde BS, Nürnberg P. Enrichment of target sequences for nextgeneration sequencing applications in research and diagnostics. Vol. 395, Biological Chemistry. 2014. p. 231–7.

68. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.

69. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014;514(7253):494–7.

70. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.

71. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

72. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17(8):754–5.

73. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307–21.

74. Swofford D. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts; 2003.

75. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92.

76. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.

77. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4).

78. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. In: Bioinformatics. 2013. p. 1682–4.

79. Mande SS, Mohammed MH, Ghosh TS. Classification of metagenomic sequences: Methods and challenges. Brief Bioinform. 2012;13(6):669–81.

80. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7.

81. Girotto S, Comin M, Pizzi C. Higher recall in metagenomic sequence classification exploiting overlapping reads. BMC Genomics. 2017;18:917.

82. Breitwieser FP, Lu J, Salzberg SL. A review of methods and databases for metagenomic classification and assembly. Brief Bioinform. 2018;20(4):1125–39.

83. Hübler R, Key FM, Warinner C, Bos KI, Krause J, Herbig A. HOPS: automated detection and authentication of pathogen DNA in archaeological remains. Genome Biol. 2019;20(1).

84. Basler CF, Reid AH, Dybing JK, Janczewski TA, Fanning TG, Zheng H, et al. Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. Proc Natl Acad Sci U S A. 2001;98(5):2746–51.

85. Reid AH, Fanning TG, Janczewski TA, Taubenberger JK. Characterization of the 1918 “Spanish” influenza virus neuraminidase gene. Proc Natl Acad Sci U S A. 2000;97(12):6785–90.

86. Reid AH, Fanning TG, Hultin J V., Taubenberger JK. Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. Proc Natl Acad Sci U S A. 1999;96(4):1651–6.

87. Reid AH, Fanning TG, Janczewski TA, McCall S, Taubenberger JK. Characterization of the 1918 “Spanish” Influenza Virus Matrix Gene Segment. J Virol. 2002;76(21):10717–23.

88. Reid AH, Taubenberger JK, Fanning TG. Evidence of an absence: The genetic origins of the 1918 pandemic influenza virus. Nat Rev Microbiol. 2004;2(11):909–14.

89. DeWitte SN. Mortality risk and survival in the aftermath of the medieval Black Death. PLoS One. 2014;9(5).

90. Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. Characterization of the 1918 influenza virus polymerase genes. Nature. 2005;437(7060):889–93.

91. Benedictow OJ. The Black Death: Greatest Catastrophe Ever. Hist Today. 2005;

92. Spyrou MA, Bos KI, Herbig A, Krause J. Ancient pathogen genomics as an emerging tool for infectious disease research. Vol. 20, Nature Reviews Genetics. 2019. p. 323–40.

2.

Chapter I: *Yersinia pestis* strains from Latvia show depletion of the *pla* virulence gene at the end of the second plague pandemic

Scientific Reports 2020

Doi: 10.1038/s41598-020-71530-9

Supplementary material can be found at the end of this thesis

SCIENTIFIC REPORTS
nature research

OPEN

Yersinia pestis strains from Latvia show depletion of the *pla* virulence gene at the end of the second plague pandemic

Julian Susat[1], Joanna H. Bonczarowska[1], Elīna Pētersone-Gordina[2], Alexander Immel[1], Almut Nebel[1], Guntis Gerhards[2] & Ben Krause-Kyora[1]✉

Ancient genomic studies have identified *Yersinia pestis* (*Y. pestis*) as the causative agent of the second plague pandemic (fourteenth–eighteenth century) that started with the Black Death (1,347–1,353). Most of the *Y. pestis* strains investigated from this pandemic have been isolated from western Europe, and not much is known about the diversity and microevolution of this bacterium in eastern European countries. In this study, we investigated human remains excavated from two cemeteries in Riga (Latvia). Historical evidence suggests that the burials were a consequence of plague outbreaks during the seventeenth century. DNA was extracted from teeth of 16 individuals and subjected to shotgun sequencing. Analysis of the metagenomic data revealed the presence of *Y. pestis* sequences in four remains, confirming that the buried individuals were victims of plague. In two samples, *Y. pestis* DNA coverage was sufficient for genome reconstruction. Subsequent phylogenetic analysis showed that the Riga strains fell within the diversity of the already known post-Black Death genomes. Interestingly, the two Latvian isolates did not cluster together. Moreover, we detected a drop in coverage of the pPCP1 plasmid region containing the *pla* gene. Further analysis indicated the presence of two pPCP1 plasmids, one with and one without the *pla* gene region, and only one bacterial chromosome, indicating that the same bacterium carried two distinct pPCP1 plasmids. In addition, we found the same pattern in the majority of previously published post-Black Death strains, but not in the Black Death strains. The *pla* gene is an important virulence factor for the infection of and transmission in humans. Thus, the spread of *pla*-depleted strains may, among other causes, have contributed to the disappearance of the second plague pandemic in eighteenth century Europe.

Yersinia pestis (*Y. pestis*) is the causative agent of plague that evolved from a relatively benign pathogen—*Yersinia pseudotuberculosis*—thousands of years ago[1–4]. Whilst plague is a zoonotic disease with rodents being the primary hosts of the pathogen[5–7], it has also afflicted humans for at least 5,000 years[1]. In the last two millennia, *Y. pestis* was responsible for three major epidemics, of which the second was the most infamous. This pandemic began with the Black Death in the fourteenth century (1,346–1,353). It is thought to have originated in Asia from where it quickly spread to and across the European continent killing approximately 30–50% of the population within a few years[8,9]. The second pandemic persisted for over 400 years[10] causing, among others, the Great Plague of London (1,665–1,666) and Marseille (1,720–1,722)[11]. In contrast to western Europe, the first major outbreaks of plague in the north-eastern part of the continent took place mainly in the post-Black Death period (after 1,353), for instance in Riga (1,601–3, 1,621–23 and 1,657) or during the Russian Plague (1,770–1,772)[12–14].

Ancient DNA (aDNA) studies have shown that all second-plague pandemic strains cluster together in branch 1 of the *Y. pestis* phylogeny, close to present-day representatives[15–17]. Interestingly, pathogens involved in the post-Black Death outbreaks derived from two Black Death strains from London and Barcelona[15,18]. This post-Black Death subset includes strains from Germany (Starnberg 1,433–1,523, Ellwangen 1,486–1,627),

[1]Institute of Clinical Molecular Biology, Kiel University, Rosalind-Franklin-Straße 12, 24105 Kiel, Germany. [2]Institute of Latvian History, University of Latvia, Kalpaka bulvāris 4, Riga 1050, Latvia. ✉email: b.krause-kyora@ikmb.uni-kiel.de

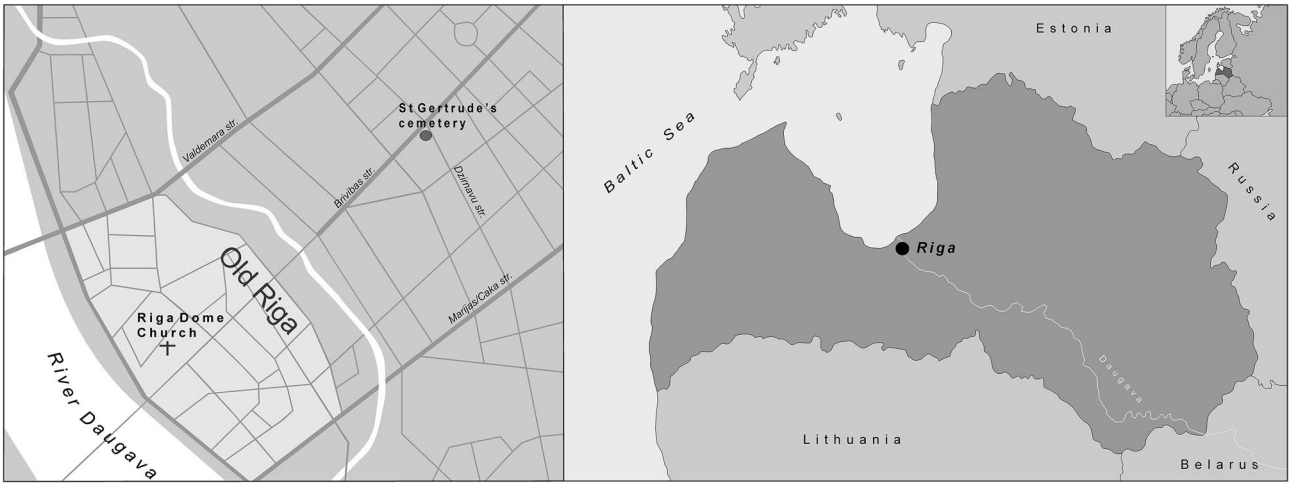


Figure 1. Cemeteries in Riga from which the samples in this study were obtained (left picture). Map of Latvia and Europe (right picture, upper right corner) with the location of Riga. (Figure was created using CorelDRAW Home & Student X8, URL: https://www.coreldraw.com/en/).

Cemetery	Context	Burial no.	Osteological sex	Genetic sex	Age (years)	*Yersinia pestis*
St. Gertrude's Church cemetery	Mass grave 1	G83	ND	No DNA	11–13	No DNA
		G157	M	No DNA	30–35	No DNA
		G488	M	M	30–35	+
		G645	F	F	20–25	+
	Mass grave 2	G103	ND	M	10–11	+
		G177	ND	M	10–12	–
		G687	M	M	45–50	–
	Burial pit	G691	ND	F	9–11	–
		G701	F	F	35–40	+
	General cemetery	G41	ND	M	14–15	–
		G92	M	M	55–60	–
		G100	M	M	20–25	–
Riga Dome Church cemetery	Mass burial pit	G143b	F	F	25–30	–
		G161c	F	M	50–60	–
		G171c	F	ND	40–50	–
	Collective burial pit	G184b	M	M	40–50	–

Table 1. List of samples from St. Gertrude's Church cemetery and Riga Dome Church cemetery. *M* male, *F* female, *ND* not determinable; +/– *Y. pestis*-positive or negative.

England (Cambridge 1,433–1,523, London 1,560–1,635), Sweden (Stans 1,485–1,635) and France (Marseille 1,720–1,722)[19]. Since no modern descendants have yet been identified, this lineage probably went extinct[10].

There are ongoing debates about the reasons as to why the disease disappeared from Europe in the eighteenth century. Improvements in living conditions, hygiene, and quarantine facilities as well as a replacement of the rodent host from black to brown rat have been suggested[20]. Other explanations revolved around the idea that evolutionary processes may have favored less virulent pathogen strains[21] or may have resulted in increased immunity of the human population[20].

Most of the second-pandemic strains analyzed to date have been isolated from human remains excavated in western Europe, with exception of two strains found in Russia (Bolgar 1,362–1,400[18], Laishevo 1,300–1,400[19]). Therefore, not much is known about the diversity and microevolution of the strains responsible for disease outbreaks in eastern Europe, which could have served as a gateway for the spread of the pathogen towards the west.

At least three plague outbreaks in Riga (Latvia) dating to the seventeenth century have been previously reported[12,13]. To confirm the presence of *Y. pestis* in Riga, we analysed samples from 16 human skeletons excavated from two cemeteries (fifteenth–seventeenth century[13]) (Fig. 1 and Table 1). The individuals had been interred in multiple or mass burials and single graves (Table 1). Four of the samples yielded positive screening results for the presence of *Y. pestis*, leading to the successful reconstruction of two complete bacterial genomes. This new data allowed us to expand the geographical range of known *Y. pestis* strains to north-eastern Europe.

Sample	Aligned reads	Coverage ≥ 1x (%)	Coverage ≥ 2x (%)	Coverage ≥ 3x (%)	Mean coverage whole genome
Chromosome					
G701	237,557	92.74	75.66	53.77	3.09x
G488	597,757	98.85	95.83	89.76	6.79x
pMT1					
G701	6,540	97.74	89.33	75.03	4.16x
G488	17,450	99.65	99.04	97.52	9.68x
pPCP1					
G701	4,544	98.75	94.42	88.24	30.07x
G488	11,893	99.80	99.14	98.72	68.28x
pCD1					
G701	6,659	99.35	96.30	89.55	5.9x
G488	26,728	99.99	99.14	98.72	20.99x

Table 2. Basic mapping statistics for the strongly positive *Y. pestis* samples after duplicate removal and quality filtering. Mapping against the CO92 chromosome (NC_003143.1), pMT1 (NC_003134.1), pPCP1 (NC_003132.1) and pCD1 (NC_003131.1).

Results

Sixteen individuals excavated from two cemeteries located in Riga, Latvia were submitted for aDNA analysis. For 14 samples, authentic ancient shotgun sequences were generated. Of these, we identified four *Y. pestis*-positive individuals using MALT in BlastN mode (Table 1 and S1). Three of them (G103, G645, G488) were buried in two mass graves and one (G701) in the burial pit of the St. Gertrude's Church cemetery. The presence of *Y. pestis*-specific reads was confirmed by mapping against a multi fasta reference containing several representatives of the genus *Yersinia* (Table S2). Furthermore, we calculated an endogenous based score (i.e. the number of *Y. pestis*-specific reads with respect to the number of total reads in the sample)[2] (Table S3). Two of the samples (G488, G701) were identified as strong positives (endogenous based score > 0.005, Table S3). For these candidates, we subsequently generated 330 and 383 million reads, respectively, to reconstruct the complete bacterial genomes. The obtained reads were mapped against the *Y. pestis* reference (CO92, NC_003143.1) containing the chromosome and all three plasmids (pCD1, NC_003131.1; pMT1, NC_003134.1; pPCP1, NC_003132.1). Both samples exhibited a genome coverage above 92% for the chromosome as well as for the three plasmids (Table 2 and Fig. S1). Terminal deamination patterns displayed damage profiles typical of authentic aDNA (Fig. S2).

We computed maximum likelihood and Bayesian trees (Figs. 2, S3 and S4) with the two genomes from Riga (G488, G701) and previously published *Y. pestis* strains (228 modern (Table S4), 36 ancient (Table S5)) and a Y. *pseudotuberculosis* genome (NZ_CP008943.1). The strains from Riga clustered with other second-plague pandemic strains from Starnberg (STA001) and Cambridge (NMS002)[19] with a strong bootstrap support of > 97% and a posterior probability of 1 (Figs. 2, S3 and S4).

Both Riga isolates had SNP alleles (p1 and p2[14]) diagnostic for branch 1 (Table S6). Furthermore, they showed differences in the SNPs p3 and p4 relative to the CO92 reference sequence. These substitutions also characterize other previously reported second-plague pandemic strains from London (England), Barcelona (Spain), Ellwangen (Germany), and l'Observance (France)[10,15,18].

The two reconstructed genomes from Riga exhibited a drastic drop in coverage of sequences mapping to the *pla* gene region (NC_003132.1:6,428–8,530), which consists of the *pla* gene, a putative transcriptional regulator and a hypothetical protein on the pPCP1 plasmid (Fig. 3). We confirmed the presence of two variants of the plasmid, i.e. one with (*pla*+) and one without (*pla*–) the *pla* region, by identifying gap-spanning reads in each of the strains (Figs. 3C,D, S5). In contrast, we noted only one bacterial chromosome per strain, indicating that the two distinct pPCP1 plasmids existed in the same bacterium. To rule out that the *pla* gene originated from contamination, a competitive mapping against the *pla* gene from environmental bacteria (e.g. *E. coli* and *Citrobacter koseri*) and *Y. pestis* showed that the reads in this region were indeed specific for *Y. pestis* and exhibited the expected damage patterns (Table S7). This observation was verified by a competitive mapping against a *pla*+ and a *pla*– reference (Fig. 3 and Table S8). The observed ratio between the coverage of the region in *pla*+ *and pla*– plasmids was about 1:10 (Table S9). To further explore this finding, we re-analyzed 28 previously published *Y. pestis* genomes from the second pandemic[10,15,18,19]. Interestingly, in 14 strains (marked in blue, Fig. 2) we detected gap-spanning reads and a depletion of the *pla* region, also indicating the presence of both *pla*+ and *pla*– plasmids (Fig. S6 and Table S10). Four more strains showed no signs of *pla*– plasmids (marked in red, Figs. 2 and S6), whereas ten others, for which the overall pPCP1 plasmid coverage was poor, were excluded as their *pla* region could not be examined reliably (marked in black, Fig. 2). The possibility that the different plasmids could be the result of a co-infection with two different strains is highly unlikely as this pattern was found in several *Y. pestis* genomes from various sites and times, and there was no evidence for such an infection on the chromosomal level.

With regard to (potentially) functional SNPs, we identified two variants of interest. First, both Riga genomes carried the T259 allele in the *pla* gene that is important for massive bacterial dissemination in the bubonic form of plague[22]. Second, in sample G488, SnpEff[23] detected a variant in the s*srA-smpB* region.

For five of the samples (G92, G143b, G161c, G488 and G701), the human endogenous DNA content in the metagenomic data was sufficient to investigate the population genetic ancestry using PCA, f3 and admixture

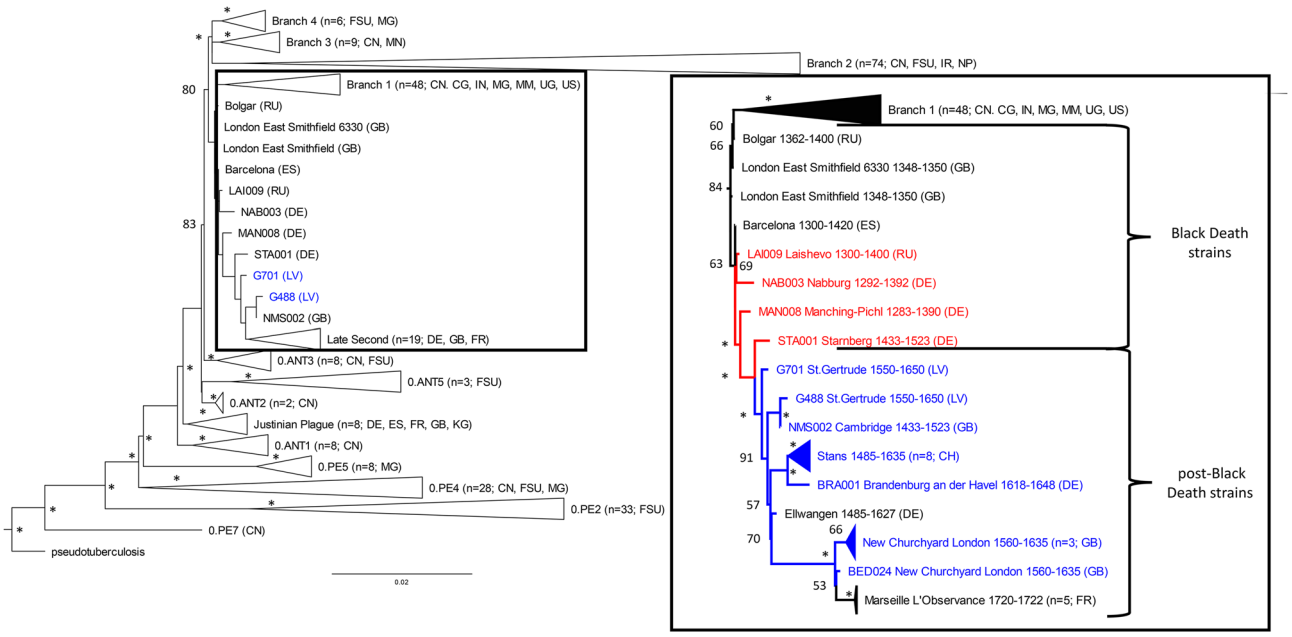


Figure 2. Maximum-likelihood tree. The tree is based on the SNP alignment (15,976 positions) of 228 modern *Y. pestis* genomes, 36 published ancient *Y. pestis* strains, one *Y. pseudotuberculosis* genome, and the two genomes G488 and G701 from St. Gertrude's Church cemetery (left panel, blue). Country abbreviation is given in brackets (DE = Germany, ES = Spain, FR = France, GB = Great Britain, US = United States, RU = Russia, LV = Latvia, CN = China, CG = Congo, FSU = Former Soviet Union, IN = India, IR = Iran, MG = Madagascar, MM = Myanmar, MN = Mongolia, NP = Nepal, UG = Uganda, KG = Kyrgyzstan, CH = Switzerland). Bootstrap values are shown on the nodes for 500 replicates. An asterisk (*) indicates a bootstrap support above 94. Strains in red exhibit no signs of the *pla*-negative plasmid. Strains in blue exhibit clear signs of the *pla*-negative plasmid. For the strains in black (within the Back Death and post-Black Death subclades), the *pla* status was inconclusive or could not be determined due to a general lack of coverage.

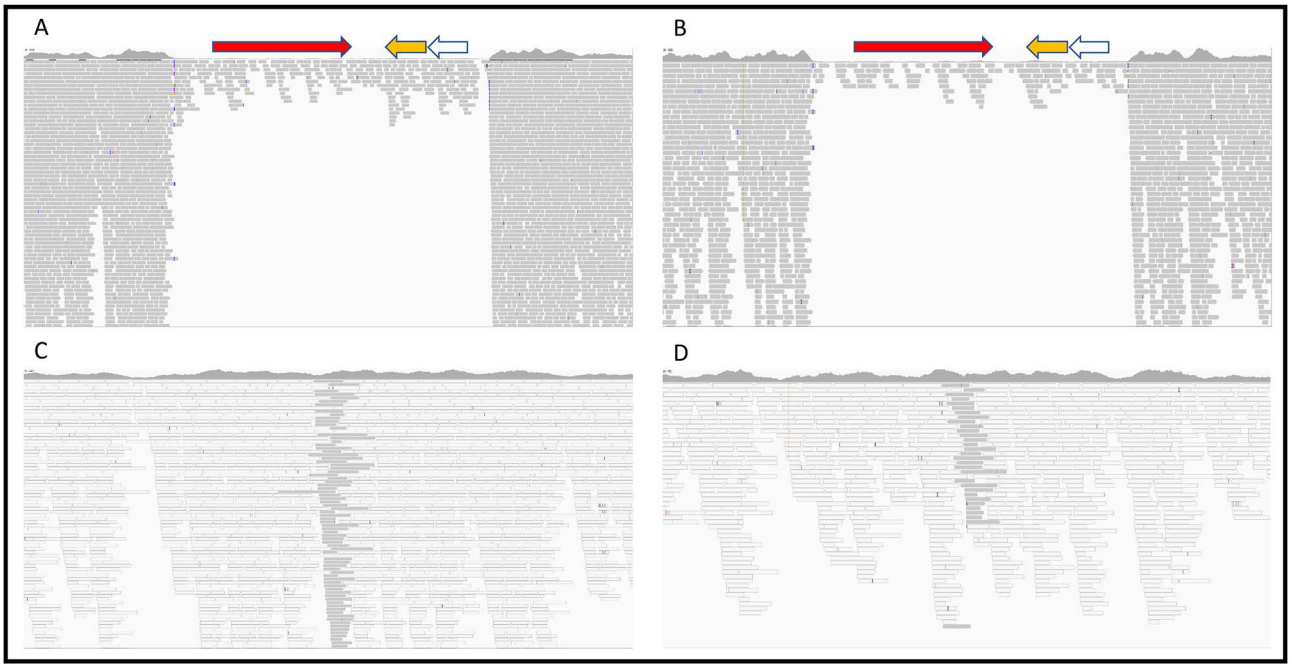


Figure 3. Coverage plots of pPCP1. (**A**, **B**) Coverage plot of the *pla* region in G701 (**A**) and G488 (**B**) depicting the *pla* gene in red, the putative transcriptional regulator in yellow and the hypothetical protein in white. G488 average coverage (ac) before the *pla* region 107x, ac of the *pla* region 7x; G701 ac before the *pla* region 47x, ac of the *pla* region 3x. Each grey line represents an obtained aDNA sequence. (**C**) Coverage plot showing reads of G701 (sequences in dark grey) specifically mapping to the gap-spanning region. A white color indicates sequences that could be mapped to both the *pla*+ and *pla*− references as they are not specific to the newly created region. (**D**) Coverage plot showing reads of G488 (sequences in dark grey) specifically mapping to the gap-spanning region.

analyses. The results indicate a northern European origin of the individuals who were genetically fairly homogeneous (Figs. S7, S8, S9 and S10).

Discussion

Here we analyzed the remains of 16 individuals who had been interred in two cemeteries (Dome Church cemetery, St. Gertrude's Church cemetery) in Riga, Latvia (fifteenth–seventeenth century) (Fig. 1 and Table 1). Four of the individuals buried in two mass graves and a burial pit of St. Gertrude's Church cemetery were infected with the *Y. pestis* bacterium. Previous analyses led to the hypothesis that people inhumed in the mass graves were rural immigrants. Our human population genetic data confirmed their northern European descent. Most likely they came to Riga during the famine of 1601–2 and stayed in shelters in the vicinity of St. Gertrude's Church[13]. Inadequate nutrition and crowded, unhygienic conditions may have increased their susceptibility to infectious diseases, including plague[24,25]. Our finding showed that the individuals buried in the mass graves as well as the small burial pit were indeed victims of plague[26,27]. Thus, we corroborate historical reports of local plague outbreaks in Riga at the beginning of the seventeenth century (1,601–3, 1,621–23 and 1,657)[12].

From two (G488 and G701) of the four infected individuals, high-quality whole genomes of *Y. pestis* were generated. They formed a subclade in branch 1 with other strains dating to the post-Black Death pandemic (Fig. 2). They all originated from Black Death genomes from London[15] and Barcelona[18]. Although the two Riga strains were found in the same cemetery (in one mass grave and the burial pit), they showed a surprisingly large genetic distance. Based on this observation and the close proximity of G488 to NMS002 (Cambridge[19]) (Fig. 2), we assume that G488 had not evolved from G701 during a single outbreak. This may indicate that the two different strains were responsible for two subsequent plague events[12,13].

As no modern strains fall into the post-Black Death subclade, which includes the two genomes from Riga, it could represent a *Y. pestis* lineage that went extinct[10]. Although the strains differ widely in terms of time and geography, their positions within the subclade—like pearls on a string—suggest continuous in situ evolution. This observation agrees with the hypothesis that the second plague pandemic reached Europe from Asia rather in one than in multiple waves. If the pandemic had multiple sources, one would expect a higher diversity. Thus, the post-Black Death outbreaks were likely due to established plague reservoirs within Europe. Little is known about natural plague foci and hosts during the second pandemic. Several rodent species, e.g. marmots and squirrels, from the Alps and the steppe have been mentioned as possible reservoirs[20].

The majority (ratio 10:1) of the pPCP1 plasmid copies carried by the two Riga strains G488 and G701 did not possess the *pla* region. Based on the chromosome data we did not see any signs of coinfection with different strains that could explain the presence of two different pPCP1 plasmids. In a re-evaluation of already published second-plague pandemic genomes, we identified 14 strains carrying both plasmid types (*pla*− and *pla*+) (Table S10, Fig. S6). All 14 strains were dated to the post-Black Death period whereas four strains which only had copies of the *pla*+ plasmid were present during the Black Death (Fig. 2). The *pla* region consists of the *pla* gene, a putative transcriptional regulator and a hypothetical protein. Up to now, the role of the putative transcriptional regulator and the hypothetical protein with respect to *pla* expression and function has remained unclear. Several studies have shown that *pla* is an important factor for the infection of and transmission in humans. The pla protease activates host fibrinolysis and allows the spread of the bacteria to the lymphatic nodes[28] as well as their extensive growth in the lower airways[29,30]. *Pla*− mutants of *Y. pestis* have been associated with reduced dissemination and virulence[22,30,31]. In mice, *pla*− strains transmitted by fleas were reported to cause primary septicemic plague while no bubonic forms were detected[31]. Therefore, *pla* seems to influence the infection routes leading to bubonic and pneumonic plague, while the septicemic route remains unaffected[32]. Therefore, the extremely high mortality of the Black Death could possibly be explained by the functional features of *pla*. The pandemic was mainly associated with a rapid spread of the severe bubonic form via the flea vector. In contrast, subsequent outbreaks could have been influenced by a slower transmission rate and a different disease course (primarily septicemic rather than bubonic) due to the *pla*-depleted strains (Fig. 2).

Horizontal gene transfer, selective gene loss, and changes in regulatory pathways have been described in the evolution and adaptation of *Y. pestis* to new host environments[33,34]. These mechanisms together with the genomic structure (e.g. IS-like elements or transposases) of the plasmids could have led to a loss of *pla* or a *pla*-depletion. *Pla*− *Y. pestis* strains have been isolated worldwide[35,36] and *pla*− plasmids are known from *Pestoides* strains[37]—enzoonotic pathogens that infect rodents[38], but rarely humans[39]. The *pla*-depleted post-Black Death strains could therefore be an adaptation to a specific host environment, perhaps in rodents.

The strain from individual G488 possessed a mutation in the *ssrA* gene, which has not yet been described in other strains. Together with the smpB protein, ssrA creates a complex responsible for maintaining cellular homeostasis[40–42]. The important role of this system in bacterial pathogenicity was shown in previous studies on several bacteria such as *Salmonella*, *Y. pseudotuberculosis* and *Y. pestis*[43–45]. Interestingly, in a mouse model, *Y. pestis* strains with a deletion of *ssrA* and *smpB* had limited ability to disseminate and colonize target organs. Moreover, mice which were intranasally immunized with this mutated strain were protected against pulmonary infection with a fully virulent strain[46]. In view of these findings, a SNP in the *ssrA* gene is intriguing. However, the functional implications of this single point mutation remain unclear.

Y. pestis strains responsible for the Post-Black Death plague showed a *pla*-depletion in the pPCP1 plasmid. This change in the pathogen genome can probably be attributed to an adaptation of the bacterium to a new host environment (e.g. rodents). In humans, the depletion may have favorably influenced the disease course, the probability of disease manifestation after a flea bite and/or the transmission rate. These positive effects fit with the evolutionary hypothesis that less virulent strains may, at least in part, explain the easing and ultimate disappearance of plague from Europe[21]. Nevertheless, it is imperative to obtain more *Y. pestis* genomes and conduct

additional functional studies to create a more comprehensive picture of the bacteria responsible for the second plague pandemic.

Material and methods

Historical background and sample selection. Individuals excavated from the cemeteries Riga Dome Church and St. Gertrude's Church cemetery, Riga, were selected for a bacterial pathogen screening (Fig. 1). At Riga Dome Church cemetery, three mass burial pits were found at the northern wall of the church. A total of 87 individuals were identified and archaeological finds suggest that they date to the seventeenth century[47]. Unlike people buried elsewhere in the cemetery, the individuals in the mass burial pits had been buried in coffins, and none of them showed evidence of peri-mortem trauma[48]. Therefore, it is likely that they had succumbed to a mass mortality event that was not related to violence. Indeed, they could be victims of one of the seventeenth century plague epidemics. Historical sources suggest that during the plague epidemic of 1,657, the churches of Riga ran out of intramural burial space, which was normally reserved for people of a higher social status[48]. It is possible that during the epidemic, the dead were buried in larger pits outside the church, albeit in a dedicated area and in coffins to signify their higher social status. In addition, smaller collective burials close to the aforementioned mass burial pits were found. In these, four to nine people were buried in several layers, without coffins. For DNA analysis, three samples were selected from the mass burial pits with coffins, and one from the smaller collective burial pits (Table 1).

The other site selected for aDNA analysis was St. Gertrude's Church cemetery. The graveyard was in use between the fifteenth and seventeenth century, and back then located outside the city wall (Fig. 1). The cemetery was divided into four distinct burial areas: a general cemetery, two mass graves, and a small mass burial pit. In total, 709 individuals were excavated from this cemetery, and 283 of these were buried in the two mass graves (south-eastern mass grave, or mass grave 1, and north-western mass grave, or mass grave 2)[13]. The bodies in the mass graves were placed on top of each other, reaching seven to eight layers in total. Both archaeological (based on artefacts) and radiocarbon analysis dated the two mass graves to the end of the sixteenth or the beginning of the seventeenth century, although they are not necessarily contemporaneous[49]. The chronology of the small mass burial pit with 15 buried individuals is not entirely clear. The demographic profile of the mass graves with comparatively few children < 7 years of age and, again, the lack of evidence of peri-mortem trauma, suggested that these people had died in a non-violent mass mortality event[13,48]. Moreover, the date of both mass graves coincides with two historically documented plague epidemics (1,601–3 and 1,621–23) and a famine (1,601–2)[12]. For DNA analysis, four samples were selected from mass grave 1, three from mass grave 2, two from the grave pit and three from the general cemetery (Table 1).

DNA isolation, sequencing and postprocessing. All lab work was carried out in a dedicated ancient DNA facility. DNA extraction of 16 tooth samples from Riga, sequencing (Illumina HiSeq4000 (2 × 75 bp)) and subsequent clip and merge was performed as described in Krause-Kyora et al.[50]. Reads that were shorter than 25 bp were discarded after merging. Two samples (G83 and G157) did not yield sufficient DNA and were not processed further (Table 1).

Pathogen screening. Screening of the 14 samples was carried out with the software MALT using a custom database consisting of complete bacterial genomes downloaded from NCBI (24.01.2019)[51]. A sequence identity threshold of 90% was set and the alignment mode was changed to SemiGlobal. The analysis was done using the following command:

malt-run -mode BlastN -e 0.001 -id 85 -alignmentType SemiGlobal -index $index -inFile $FASTQCM -output $OUT

where $index is the index file, $FASTQCM is the clipped and merged file and $OUT is the output file.

The resulting alignments were visually inspected using MEGAN 6[52].

***Y. pestis* alignment.** All four positive samples from the screening (G103, G488, G645, G701) were mapped against the *Y. pestis* reference genome including the three plasmids pCD1, pMT1 and pPCP1 (NC_0031431.1; NC_003131.1; NC_003134.1; NC_003132.1). The mapping was conducted using BWA with the following command line[53]:

bwa aln -n 0.01 -l 300 $INDEX $FASTQCM $OUT

where $INDEX is the reference, $FASTQCM is the input file and $OUT is the output file. Minimum mapping quality was set to 0. Subsequent duplicate removal was performed using DeDup version 0.11.3, part of the EAGER pipeline[54], with default parameters. Subsequently, the Binary Alignment Map (BAM) files were filtered for a mapping quality of 30 using the following command in samtools[55]:

samtools view -bq $INT $IN > $OUT

where $INT is the mapping quality, $IN is the input BAM file and $OUT is the output BAM file.

Sample authentication. To further authenticate the positive samples, an endogenous score was used to calculate the number of *Y. pestis*-specific reads with respect to the number of total reads in the sample[2]. Positive samples were mapped against a multi fasta reference containing several representatives of the genus *Yersinia* (Table S2) and the three *Y. pestis*-specific plasmids (pCD1, pMT1, pPCP1). The mapping was done as described

above followed by a duplicate removal step. Mapped reads were filtered for a mapping quality greater than 30 using samtools as described above.

Applying samtools idxstats, an endogenous based score was used to evaluate the potential of the sample being positive for *Y. pestis*. The score was calculated with the following formula:

$$\frac{(YPS - (YS))}{M} \times 1000$$

where YPS is the number of reads specifically mapping to *Y. pestis*; YS is the maximum number of reads mapping specifically to any *Yersinia* species with the exception of *Y. pestis* and M is the total number of merged reads in the sample. By using the maximum number of reads mapping to another species of the genus *Yersinia*, the score takes into account different sources of contamination other than *Y. pseudotuberculosis*. Two of the four positive samples exhibited a score higher than 0.005, a value which indicates a strong positive candidate (Table S3).

VCF-generation and phylogenetic analysis. VFC files were generated for the two positive samples using GATK and the *Y. pestis* reference with the following command line[56]:

GenomeAnalysisTK.jar-T UnifiedGenotyper -R $REF -I $IN -out_mode EMIT_ALL_SITES -o $OUT

where $REF is the reference, $IN is the input BAM file and $OUT is the VCF output. These two VCFs together with a VCF dataset of 264 strains (Table S4 and S5, mapped (see *Y. pestis* alignment) and generated as described above) and one *Y. pseudotuberculosis* (NZ_CP008943.1) genome were used as an input for the MultiVCFAnalyzer to generate a SNP-based multiple alignment[57]. The analysis was carried out with the following command:

Java -jar MultiVCFAnalyzer_0-85-1.jar NA $REF NA $OUT F 30 3 0.9 0.9 NA

where $REF is the *Y. pestis* reference and $OUT is the output folder. *Y. pseudotuberculosis*-specific variants were removed from the SNP table to improve visual resolution of the tree. The resulting SNP table was also filtered for SNPs in repetitive regions and homoplasies[16]. The SNP table was reformatted to multi fasta and nexus format that were used as input for RAxML and MrBayes[58,59]. RAxML was executed with the following command line:

RaxmlHPC -f a -x 12345 -p 12345 -#500 -m GTRGAMMA -s $IN -o $OG -n $NAME

where $IN is the input file, $OG is the name of the outgroup and $NAME is the outfile name.

MrBayes was executed applying the GTR model with the default number of generations.

Resulting trees were visualized using FigTree[60].

SNP effect analysis. VCFs from the Y. *pestis*-positive samples were filtered for SNPs with a coverage of at least 3x, a calling quality of 30 and a 90% majority call. This set of filtered SNPS was provided to SNPEff to analyze the effects of SNPs in the ancient genomes[23]. SNPEff was executed with the following command line:

java -jar snpEff.jar $REF $IN > $OUT

where $REF is the SNPEff reference, $IN is the input VCF file and $OUT is the annotated SNPEff output file. The annotated VCF file was filtered for high-impact SNPs and evaluated.

pPCP1 and *pla* analysis. To ascertain that the coverage in the low-coverage region was not introduced artificially by environmental contamination, we carried out a competitive mapping as described above. The fasta reference that was used contained the *Y. pestis* reference sequence of the pPCP1 plasmid and also the *pla* genes of *Citrobacter* and *E. coli* (Table S7). Gap-bridging reads were analyzed by removing the low-coverage region from the pPCP1 plasmid. Seqkit was used with the following command lines[61]:

seqkit subseq -r 1:6428 $IN > $OUT1

seqkit subseq -r 8530:9612 $IN > $OUT2

where $IN is the pPCP1 reference (NC_003132.1) and $OUT1/$OUT2 are parts of the new plasmid without the *pla* region. Both output files were concatenated to form a complete plasmid. An additional mapping with the pPCP1 plasmid from which the *pla* region was removed (*pla-)* and the regular pPCP1 plasmid (*pla*+) as references was performed (Table S8). These references were used to identify reads which were truly unique for the new region in the *pla*− plasmid. All the non-specific reads received a mapping quality of zero as the exact same region was present in both references. The reads mapping exclusively to the new region were extracted and compared with the reads from the initial mapping against *Y. pestis* and gap-bridging reads were identified. Based on these exclusive mapping reads, the coverage for the *pla*− plasmid was calculated. The coverage of the *pla* region (NC_003132.1, 6,428–8,529) in the regular mapping was used to estimate the number of *pla*+ plasmids. Based on these two values a ratio was calculated to estimate the proportions.

Human genetic analyses. *Mapping and aDNA damage patterns.* Shotgun sequence data was mapped to the human genome build hg19 using BWA with a reduced mapping stringency parameter "-n 0.01" to account for mismatches in aDNA[54,62]. Duplicates were removed. Authenticity of the aDNA fragments was assessed using mapDamage 2.0[63,64]. After validation of the terminal damage patterns, the first two positions from the 5′ end of the fastq-reads were trimmed.

Genotyping. Alleles were drawn at random from a pool of 1,233,013 SNPs (commonly used for population genetic analyses) in a pseudo-haploid manner using a custom script as described previously[65–68]. Datasets had to have at least 10,000 SNPs to be considered for further analysis[68]. Five individuals from Riga (G92, G143b, G161c, G488 and G701) passed this filtering step.

Population genetic analyses. The five genotype datasets from Riga were merged with a dataset of previously published 59 West Eurasian populations genotyped on the aforementioned 1,233,013 SNPs using the program mergeit from the EIGENSOFT package[69]. Principal component analysis (PCA) was performed using the software smartpca and the 'lsqproject' option[69]. Admixture modelling was done with the software ADMIXTURE on the same populations as used in the PCA. The number of ancestral components k ranged from 4 to 12[70]. Cross-validation was performed for every admixture model and the model with the highest accuracy was determined by the lowest cross-validation error.

F3 outgroup statistics. f3 outgroup statistics were run as a part of the Admixtools package in the form of *f3* (Riga; test; Mbuti) using for test the same populations as used in the PCA and ADMIXTURE analysis[70].

Material transfer agreement. The authors confirm that the sampling and the ancient DNA analysis of human bones from the fifteenth–seventeenth century archaeological features unearthed in Riga, Latvia were performed with permission from the institution which headed the excavations and the anthropological analysis (Institute of Latvian History, University of Latvia).

Data availability

The aligned sequences are available through the European Nucleotide Archive under Accession Number PRJEB36413.

Received: 20 March 2020; Accepted: 4 August 2020
Published online: 03 September 2020

References

1. Rasmussen, S. *et al.* Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* **163**, 571–582 (2015).
2. Valtuena, A. A. *et al.* The Stone Age plague and its persistence in Europe. *Curr. Biol.* **27**, 3683–3691 (2017).
3. Spyrou, M. A. *et al.* Analysis of 3800-year old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nat. Commun.* **9**, 2234. https://doi.org/10.1038/s41467-018-04550-9 (2018).
4. Rascovan, N. *et al.* Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic Decline. *Cell* **176**, 295–305 (2019).
5. Anisimov, A. P., Lindler, L. E. & Pier, G. B. Intraspecific diversity of *Yersinia pestis*. *Clin. Microbiol. Rev.* **17**, 434–464 (2004).
6. Gage, K. L. & Kosoy, M. Y. Natural history of plague: perspectives from more than a century of research. *Annu. Rev. Entomol.* **50**, 505–528 (2005).
7. Roelle, J. E., Miller, B. J., Godbey, J. L. & Biggins, D. E. Recovery of the black-footed ferret progress and continuing challenges. *US Geol. Surv. Sci. Invest. Rep.* **5293**, 213–231 (2006).
8. Morelli, G. *et al. Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat. Genet.* **42**, 1140–1143 (2010).
9. Ziegler, P. *The Black Death* (Harper Collins, New York, 1969).
10. Bos, K. I. *et al.* Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *Elife* **5**, e12994. https://doi.org/10.7554/eLife.12994 (2016).
11. Byrne, J. P. *Encyclopedia of the Black Death* (ABC-CLIO, Santa Barbara, 2012).
12. Napiersky, J. G. L. *Bodeckers Chronik livländischer und rigascher Ereignisse [Bodecker's chronicle of Livonian and Riga events] (in German)* (Kymmel's Buchhandlung, Riga, 1890).
13. Petersone-Gordina, E., Roberts, C., Millard, A. R., Montgomery, J. & Gerhards, G. Dental disease and dietary isotopes of individuals from St Gertrude Church cemetery, Riga, Lativia. *PLoS ONE* **13**, e0191757. https://doi.org/10.1371/journal.pone.0191757 (2018).
14. John, T. A. *Bubonic Plague in Early Modern Russia: Public Health and Urban Disaster* (Oxford University Press, Oxford, 2003).
15. Bos, K. I. *et al.* A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506–510 (2011).
16. Cui, Y. *et al.* Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* **110**, 577–582 (2013).
17. Namouchi, A. *et al.* Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proc. Natl. Acad. Sci. USA* **115**, E11790–E11797. https://doi.org/10.1073/pnas.1812865115 (2018).
18. Spyrou, M. A. *et al.* Historical *Y pestis* genomes reveal the European Black Death as the source of ancient and modern plague pandemics. *Cell Host Microbe* **19**, 874–881 (2016).
19. Spyrou, M. A. *et al.* Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nat. Commun.* **10**, 4470. https://doi.org/10.1038/s41467-019-12154-0 (2019).
20. Appleby, A. B. The disappearance of plague: a continuing puzzle. *Econ. Hist. Rev.* **33**, 161–173 (1980).
21. Lenski, R. E. Evolution of plague virulence. *Nature* **334**, 473–474 (1988).
22. Zimbler, D. L., Schroeder, J. A., Eddy, J. L. & Lathem, W. W. Early emergence of *Yersinia pestis* as a severe respiratory pathogen. *Nat. Commun.* **6**, 7487. https://doi.org/10.1038/ncomms8487 (2015).
23. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
24. Dirks, R. *et al.* Social responses during severe food shortages and famine. *Cur. Anthropol.* **21**, 21–44 (1980).
25. Wisner, B. Adams, J. Control of communicable diseases and prevention of epidemics. Rīgas pilsētas valde *WHO.* www.who.int/entity/water_sanitation_health/hygiene/emergencies/em2002chap11.pdf (2019)
26. Pīrangs, H. *Rīgas kapsētas [Graveyards of Riga] (in Latvian)* (Rīgas pilsētas valde, Riga, 1932).
27. Rusovs, B. *Livonijas kronika [Chronicle of Livonia] (in Latvian)* (Valters un Rapa, Riga, 1926).
28. Degen, J. L., Bugge, T. H. & Goguen, J. D. Fibrin and fibrinolysis in infection and host defense. *J. Thromb. Haemost.* **5**, 24–31 (2007).
29. Lathem, W. W. *et al.* Post-transcriptional regulation of the *Yersinia pestis* cAMP receptor protein Crp and impact on virulence. *MBio* **5**, e01038-13. https://doi.org/10.1128/mBio.01038-13 (2014).
30. Perry, R. D. & Fetherston, J. D. *Yersinia pestis*—etiologic agent of plague. *Clin. Microbiol. Rev.* **10**, 35–66 (1997).

31. Sebbane, F., Jarrett, C. O., Gardner, D., Long, D. & Hinnebusch, B. J. Role of the *Yersinia pestis* plasminogen activator in the incidence of distinct septicemic and bubonic forms of flea-borne plague. *Proc. Natl. Acad. Sci. USA* **103**, 5526–5530 (2006).
32. Sebbane, F. *et al.* Adaptive response of *Yersinia pestis* to extracellular effectors of innate immunity during bubonic plague. *Proc. Natl. Acad. Sci. USA* **103**, 11766–11771 (2006).
33. Zhou, D. & Yang, R. Molecular Darwinian evolution of virulence in *Yersinia pestis*. *Infect. Immun.* **77**, 2242–2250 (2009).
34. Sun, Y. C., Jarret, C. O., Bosio, C. F. & Hinnebusch, B. J. Retracing the evolutionary path that led to flea-borne transmission of *Yersinia pestis*. *Cell Host Microbe* **15**, 578–584 (2014).
35. Cavalcanti, Y. V. N., Leal, N. C. & De Almeida, A. M. P. Typing of *Yersinia pestis* isolates from the state of Ceará, Brazil. *Lett. Appl. Microbiol.* **35**, 543–547 (2002).
36. Filippov, A. A., Solodovnikov, N. S., Kookleva, L. M. & Protsenko, O. A. Plasmid content in *Yersinia pestis* strains of different origin. *FEMS. Microbiol. Lett.* **67**, 45–48 (1990).
37. Garcia, E. *et al.* Pestoides F, an atypical *Yersinia pestis* strain from the former Soviet Union. *Adv. Exp. Med. Biol.* **603**, 17–22 (2007).
38. Achtman, M. *et al.* Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc. Natl. Acad. Sci. USA* **101**, 17837–17842 (2004).
39. Rajanna, C. *et al.* Characterization of pPCP1 plasmids in *Yersinia pestis* strains isolated from the former Soviet Union. *Int. J. Microbiol.* https://doi.org/10.1155/2010/760819 (2010).
40. Gueneau de Novoa, P. & Williams, K. P. The tmRNA website: reductive evolution of tmRNA in plastids and other endosymbionts. *Nucl. Acids. Res.* **32**, D104–D108. https://doi.org/10.1093/nar/gkh102 (2004).
41. Keiler, K. C. Biology of trans-translation. *Annu. Rev. Microbiol.* **62**, 133–151 (2008).
42. Dulebohn, D., Choy, J., Sundermeier, T., Okan, N. & Karzai, A. W. Trans-translation: the tmRNA-mediated surveillance mechanism for ribosome rescue, directed protein degradation, and nonstop mRNA decay. *Biochemistry* **46**, 4681–4693 (2007).
43. Okan, N. A., Bliska, J. B. & Karzai, A. W. A role for the SmpB-SsrA system in Yersinia pseudotuberculosis pathogenesis. *PLoS Pathog.* **2**, e6. https://doi.org/10.1371/journal.ppat.0020006 (2006).
44. Baumler, A. J., Kusters, J. G., Stojiljkovic, I. & Heffron, F. Salmonella typhimurium loci involved in survival within macrophages. *Infect. Immun.* **62**, 1623–1630 (1994).
45. Julio, S. M., Heithoff, D. M. & Mahan, M. J. ssrA (tmRNA) plays a role in Salmonella enterica serovar *Typhimurium pathogenesis*. *J. Bacteriol.* **182**, 1558–1563 (2000).
46. Okan, N. A., Mena, P., Benach, J. L., Bliska, J. B. & Karzai, A. W. The smpB-ssrA mutant of *Yersinia pestis* functions as a live attenuated vaccine to protect mice against pulmonary plague infection. *Infect. Immun.* **78**, 1284–1293 (2010).
47. Tilko, S. Apbedīšanas tradīcijas Doma viduslaiku kapsētā [Mortuary Practices in the Dom Medieval Cemetery] (in Latvian). Senā Rīga: Pētījumi pilsētas arheoloģijā un vēsturē [Ancient Riga: Studies in the archaeology and history of the city]. *Rīga: Latvijas vēstures institūta apgāds.* 273- 289 (1998).
48. Gerhards, G. Traumas un ievainojumi Rīgas 13.-18. gadsimta iedzīvotājie [Trauma and Injuries of Riga 13th–18th Century Population] (in Latvian). Senā Rīga: pētījumi pilsētas arheoloģijā un vēsturē [Ancient Riga: Studies in the archaeology and history of the city] (in Latvian). *Rīga: Latvijas vēstures institūta apgāds*. 128–148 (2012).
49. Petersone-Gordina, E., *et al.* Investigating dietary life histories and mobility in children buried in St Gertrude Church cemetery, Riga, Latvia (15th–17th centuries AD). *Archaeometry* **62**, 3–18 (2020).
50. Krause-Kyora, B. *et al.* Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. *Nat. Commun.* **9**, 1569. https://doi.org/10.1038/s41467-018-03857-x (2018).
51. Vågene, Å. J. *et al.* Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* **2**, 520–528 (2018).
52. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
53. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. Peltzer, A. *et al.* EAGER: efficient ancient genome reconstruction. *Genome Biol.* **17**, 60. https://doi.org/10.1186/s13059-016-0918-z (2016).
55. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
56. McKenna, A. *et al.* The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
57. Bos, K. I. *et al.* Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
58. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
59. Ronquist, F. *et al.* Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
60. Rambaut, A. FigTree version 1.4.3. https://tree.bio.ed.ac.uk/software/figtree/.
61. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962. https://doi.org/10.1371/journal.pone.0163962 (2016).
62. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
63. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. MapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* **29**, 1682–1684 (2013).
64. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. USA* **104**, 14616–14621 (2007).
65. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
66. Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211 (2015).
67. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
68. Lamnidis, T. C. *et al.* Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nat. Commun.* **9**, 5018. https://doi.org/10.1038/s41467-018-07483-5 (2018).
69. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190. https://doi.org/10.1371/journal.pgen.0020190 (2006).
70. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

Acknowledgements

J.B. was funded by the International Max Planck Research School for Evolutionary Biology. This study was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the following projects: numbers 2901391021 (CRC 1266) and 390870439 (EXC 2150 - ROOTS) to A.N. and B.K.-K.

Author contributions

B.K.-K., G.G. and E.P. conceived and designed the research. J.S., A.I. and B.K.-K. generated and analyzed the ancient DNA data. J.S., J.B., E.P., G.G., A.N. and B.K.-K. interpreted the findings. J.S., J.B., E.P., G.G., A.N. and B.K.-K. wrote the manuscript.

Funding

Open Access funding provided by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-71530-9.

Correspondence and requests for materials should be addressed to B.K.-K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

3.

Chapter II: Late infection diagnosis: Rudolf Virchow´s skull collection indicates 5000-year-old hunter-gatherer already plagued by *Yersinia pestis*

Currently submitted to Cell

Status: With the Editor

Supplementary material can be found at the end of this thesis

Title:

Late infection diagnosis: Rudolf Virchow's skull collection indicates 5000-year-old hunter-gatherer already plagued by *Yersinia pestis*

Authors:

Julian Susat[1,§], Harald Lübke[2,§], Alexander Immel[1], Ute Brinker[2], Aija Macane[3], John Meadows[2,4], Britta Steer[5], Andreas Tholey[5], Ilga Zagorska[6], Guntis Gerhards[6], Ulrich Schmölcke[2], Mārcis Kalniņš[6], Andre Franke[1], Elīna Pētersone-Gordina[6], Barbara Tessman[7], Mari Tõrv[8], Stefan Schreiber[1,9], Christian Andree[10], Valdis Bērziņš[6], Almut Nebel[1], Ben Krause-Kyora[1,*]

*** corresponding authors**: b.krause-kyora@ikmb.uni-kiel.de
§ equally contributing authors

Affiliations:
[1]Institute of Clinical Molecular Biology, Kiel University, Rosalind-Franklin-Str. 12, 24105 Kiel, Germany
[2]Centre for Baltic and Scandinavian Archaeology (ZBSA), Schleswig-Holstein State Museums Foundation Schloss Gottorf, Schlossinsel 1, 24837 Schleswig, Germany
[3]Department of Historical Studies, University of Gothenburg, PO Box 200, SE405 30 Göteborg, Sweden
[4]Leibniz Laboratory for AMS Dating and Isotope Research, Kiel University, Max-Eyth-Str. 11-13, 24118 Kiel, Germany
[5]Systematic Proteomics & Bioanalytics, Institute for Experimental Medicine, Kiel University, Niemannsweg 11, 24105 Kiel, Germany.
[6]Institute of Latvian History, University of Latvia, Kalpaka bulv. 4, LV 1050, Riga, Latvia
[7]Berlin Society of Anthropology, Ethnology and Prehistory, c/o Museum of Pre- and Protohistory, Geschwister-Scholl-Str. 6, 10117 Berlin, Germany
[8]Department of Archaeology, Institute of History and Archaeology, Tartu University, Tartu, Estonia, Jakobi 2, 51014, Tartu, Estonia
[9]Department of General Internal Medicine, University Hospital Schleswig-Holstein, Kiel University, Rosalind-Franklin-Str. 12, 24105 Kiel, Germany
[10]Research Center of Medical History, Kiel University, Breiter Weg 10, 24105 Kiel, Germany

Summary

Background

The plague bacterium *Yersinia pestis* (*Y. pestis*), one of the most dreaded pathogens of all times, has infected humans since prehistory. As not much is known about its emergence and early spread, we sequenced and analysed a 5000-year-old *Y. pestis* genome from northern Europe.

Methods

From the site of Riņņukalns, Latvia, skeletal material of four individuals belonging to a group of complex hunter-fisher-gatherers (5350-5150 cal BP) was subjected to metagenomic sequencing. The generated data was used for pathogen screening and genomic analysis.

Results

We reconstructed a *Y. pestis* genome (RV 2039) from Riņņukalns remains that had been stored in the Anthropological Rudolf Virchow Collection in Berlin. RV 2039 was the first in a series of ancient strains that evolved shortly after the split of *Y. pestis* from its antecessor *Y. pseudotuberculosis* about 7000 years ago. The genomic composition showed a pattern which is associated with a decreased transmission efficiency and virulence.

Conclusions

The genomic characteristics of RV 2039 are consistent with the hypothesis that this very early *Y. pestis* form was likely less transmissible and maybe even less virulent than later evolved strains with crowd disease potential. In addition, an infection with RV 2039 may have manifested as a chronic disease. These characteristics support the scenario of a prehistoric plague pandemic, albeit caused by a chronic rather than an acute form of the disease. The finding of *Y. pestis* in a Riņņukalns individual opens the possibility that hunter-gatherers might have contributed to the zoonotic emergence of *Y. pestis*.

Introduction

The archaeological shell midden site of Riņņukalns is located in Latvia, next to the River Salaca that flows into the Baltic Sea (figure 1). The shell midden itself consists of alternating layers of unburnt freshwater mussel shells, burnt mussel shells and fish bones that were deposited by human activity in a short time period of 100-200 years in the early 6th millennium BP.[1,2]

Figure 1. Map with the site Riņņukalns from where the individuals presented in this study were recovered.

In 1875 the local amateur archaeologist Carl Georg Count Sievers (1814-1879) conducted the first systematic excavation of Riņņukalns. Sievers detected in two single graves the skeletal remains of a 12- to 18-year old woman (RV 1852) and a 20- to 30-year-old man (RV 2039, figure 2, appendix table 1), covered by intact midden layers.

Figure 2. a) Cranium and **b)** mandible of individual RV 2039 rediscovered in the Anthropological Rudolf Virchow Collection of the Berlin Society of Anthropology, Ethnology and Prehistory (Berliner Gesellschaft für Anthropologie, Ethnologie und Urgeschichte - BGAEU), where it had apparently been kept for the last 150 years.

Due to the stratigraphic position, he assumed a prehistoric context of these burials.[3] This interpretation was considered wrong at the time and therefore met with fierce resistance.[4,5] For confirmation of his hypothesis and osteological investigations, Sievers sent the crania of the specimens RV 1852 and RV 2039, among other human remains from Riņņukalns, to his future mentor and friend, the already world-famous German physician Rudolf Virchow (1821-1902) in Berlin.[6] Virchow is still today highly renowned for establishing the fields of cellular pathology - the basis of modern medicine - and social medicine. In 1869, because of his avid interest in pre- and protohistory, he became the founder of the Berlin Society of Anthropology (today known as Berliner Gesellschaft für Anthropologie, Ethnologie und Urgeschichte -

BGAEU).[7] Despite his worldwide reputation, Virchow lent amateur researchers such as Heinrich Schliemann or Sievers his expertise for their archaeological projects without being deterred by opposing views of his contemporaries. After being convinced by Sievers of the importance of the finds, Virchow examined the complete Riņņukalns material and published a first report.[5,6] After World War II, the Riņņukalns crania seemed to have disappeared and their whereabouts were deemed unknown - until recently. It was only in 2011 that the skulls were rediscovered in the re-inventoried Anthropological Rudolf Virchow Collection of the BGAEU, where they had apparently been kept for the last 150 years.[8] Concurrently with the search for the crania, new field work on the Riņņukalns site led to the detection of two additional burials, those of an adult male (2017/1) and a neonate (2018/1) (appendix table 1).[2] Also for these skeletal remains, the archaeological stratigraphy indicated a prehistoric origin. Indeed, all four specimens, including RV 1852 and RV 2039 from the Virchow Collection, were subsequently radiocarbon dated to 5350-5150 cal BP (appendix table 1), thus confirming Sievers' original assessment.[2] The individuals buried in the midden belonged to a group of complex hunter-fisher-gatherers.[10] They most probably lived in semi-permanent or permanent settlements on the banks of the River Salaca. The valuable aquatic food resources certainly provided enough food for year-round habitation.[2]

Since not much is known about the genomic composition of hunter-gatherers who lived in northeastern Europe 5000 years ago or about their infectious disease burden, we subjected the four Riņņukalns individuals to an ancient DNA (aDNA) analysis that also included a pathogen screening. Surprisingly, in one male, we identified the genome of *Yersinia pestis*, the infectious agent responsible for at least three historical plague epidemics.[11] Our finding presents the hitherto first evidence of this bacterium in a hunter-gatherer and indicates the

Baltic region as an important locale in the very early phases of *Y. pestis* evolution and diversification.

Methods

Anthropological Rudolf Virchow Collection

This collection of prehistorical and historical human skulls, which Rudolf Virchow (RV) started to assemble in 1870, is preserved today by the BGAEU. BGEAU-RV 1852 and BGEAU-RV 2039 are the official collection labels of the samples investigated in this study, hereafter referred to as RV 1852 and RV 2039.

Anthropological analysis

The human remains were examined according to standard osteological methods (see appendix page one for additional Information).[5] For the genetic analyses, both petrous bones and teeth were collected for the individuals RV 1852 and 2017/1; for 2018/1 a petrous bone and for RV 2039 a tooth were sampled.

aDNA extraction and sequencing

All lab work was carried out in a dedicated aDNA facility. DNA extraction from bone/tooth powder, library preparation with unique index combinations for each sample, sequencing (Illumina HiSeq4000 (2x75bp)) and subsequent pre-processing of the sequence data was performed following established protocols.[12]

Yersinia pestis genetic analyses

All samples were screened for pathogens using parameters as described in Krause-Kyora et al. 2018 and Susat et al. 2020.[12,13] The *Y. pestis*-positive sample RV 2039 was further processed as described in Susat et al. 2020 with the following exceptions: 1) MapDamage 2.0 was used to rescale the terminal damage that remained after the half-UDG library preparation (14). 2)

The table with single-nucleotide polymorphisms (SNPs) was not filtered for repetitive regions and homoplasies.[13] 3) In total, 277 strains were used in the phylogenetic analysis (appendix table 2). 4) RAxML was executed with 1000 bootstraps. 5) MrBayes was executed with 5,000,000 generations.

Molecular dating

A subset of 42 representative *Y. pestis* genomes was used to generate a SNP-based alignment with the MultiVCFAnalyzer (appendix table 3).[13,15] Based on this alignment RAxML calculated a rooted ML tree. The resulting tree was utilized as a starting tree for BEAST2 version 2.6.[16] Dating analysis followed previous work using an uncorrelated relaxed clock with lognormal distribution and the GTR+G4 substitution model.[15,17] Furthermore, a coalescent constant population size was assumed. The dates published for previous historical *Y. pestis* genomes were transformed to BP by setting the year 1950 as age 0 (with the mean as the age and the boundaries as an interval for a prior uniform distribution). Two different runs were executed in BEAST2. One run in which the RAxML starting tree was provided and BEAST2 was allowed to alter the tree, and another in which no starting tree was provided. Multiple versions from the same run were combined and post burn-in data was visualized using Tracer version 1.7.1.[18] All effective sample size (ESS) values were at least greater than 250 (appendix table 4). LogCombiner and TreeAnnotator from BEAST2 were used to combine the result files and to generate Maximum Clade Credibility trees.

SNP effect and analysis of virulence factors

The vcf file was used by SnpEFF for the annotation of SNPs. RV 2039 had 106 unique SNPs with a coverage > 3X and a support of 80% of the reads (appendix table 5). Coverage of virulence genes was calculated and plotted using Gnuplot version 5.2.

Human population genetic analyses

Mapping of reads, SNP genotyping, genetic sex determination, contamination estimation, principal component analysis, ADMIXTURE, f3 outgroup statistics and *qpADM* were conducted as described previously.[19]

Proteomic analysis

See appendix pages two to four for Information.

Results

We generated genome-wide shotgun sequences from various skeletal elements (teeth, petrous bones) of the four Riņņukalns individuals (appendix table 6). First, we screened the datasets for the presence of known bacterial and viral pathogens, using an established in-house pipeline.[12,13] In sample RV 2039 (figure 2; dated to 5350-4980 cal BP), we observed *Y. pestis*-specific reads that showed typical aDNA damage patterns and length distributions, thus supporting the ancient origin of the sequences (appendix figure 1). The other extracts did not yield signs of *Y. pestis*. For RV 2039, we reconstructed the *Y. pestis* genome (i.e. chromosome, plasmids pCD1 and pPCP1) at high coverage (appendix table 7). The plasmid pMT1 lacked a 20-kb region containing the virulence factor *ymt;* this characteristic was reported previously for other ancient *Y. pestis* strains.[17] Phylogenetic analysis with 277 previously published ancient and modern *Y. pestis* genomes and a *Y. pseudotuberculosis* genome (appendix table 2) placed RV 2039 basal to all known *Y. pestis* bacteria (figure 3, appendix figures 2 and 3). The Riņņukalns strain had its own clade and marked the first branching event after the split from the bacterium *Y. pseudotuberculosis*. RV 2039 was clearly separated from all later dating Neolithic and Bronze Age strains (figure 3). To estimate divergence times, we performed molecular clock analyses using BEAST2 with two independent runs. The first run is based on a rooted starting tree which BEAST2 is allowed to alter, in the second run no starting tree was provided (figure 4, appendix figures 4 and 5). The results of both analyses were congruent and showed that RV 2039 diverged from all other *Y. pestis* bacteria around 7100 cal BP (with starting tree 8693-5669 cal BP; without starting tree: 8895-5733 cal BP). The split between *Y. pestis* and *Y. pseudotuberculosis* was dated to about 7400 cal BP (with starting tree 9069-5782 cal BP; without starting tree 9223-5833 cal BP).

Investigation of the genomic structure of RV 2039 led to the discovery of 106 exclusive SNPs that are not known to have an effect on virulence factors (appendix table 5). Apart from the *ymt* gene, which abscence is specific for the Neolithic/Bronze Age clade, no other virulence factor was missing (appendix figure 6). The virulence associated genes *ureD, flhD, pde2* and the plasminogen activator *pla* exhibited the ancestral states as already shown for the Neolithic and Bronze Age lineages.[17,20] The results of the genomic and phylogenetic analyses were supported by the archaeological age of RV 2039, rendering it one of the earliest lineages in *Y. pestis* evolution.

Second, the human endogenous DNA content in the four datasets was sufficient to perform kinship and population genomic analyses in the four Riņņukalns individuals (appendix table 1). Kinship analysis did not provide evidence of relatedness among the four individuals buried in the shell midden. All four carried a large genomic ancestry component that is specific for hunter-gatherers. In particular, they had a high affinity to hunter-gatherers of Eastern Europe (appendix figures 7, 8 and 9) who inhabited a large area stretching from the Baltic Sea to the Pontic-Caspian steppe.[21]

In order to confirm the presence of *Y. pestis* at the protein level in RV 2039, we performed an LC-MS based proteomics analysis. We identified three different peptides from three proteins that are specific for *Y. pestis* (appendix table 8 and figure 10), together with about 118 human proteins.

Discussion

We reconstructed a 5000-year-old *Y. pestis* genome (RV 2039) from the remains of an adult man (dated to 5350-4980 cal BP) who was buried in the shell midden of Riņņukalns, Latvia. This diagnosis was confirmed independently by detection of three *Y. pestis*-specific proteins in the sample. RV 2039 is basal to all known ancient or modern *Y. pestis* bacteria. It represents a very early, independent lineage that emerged about 7000 years ago, only a few hundred years after the split of the *Y. pestis* clade from its antecessor *Y. pseudotuberculosis*. Our date of 7400 cal BP for this split is around 1000 years earlier than previously reported estimates.[15,17,20,22,23]

RV 2039 marks the beginning of *Y. pestis* evolution and is the first in a series of ancient strains that eventually became extinct. RV 2039 and its slightly younger relative, the Gok2 genome from a Neolithic farmer in Sweden (dated to 5040-4867 cal BP), are on separate branches and both are distinct from later dating Neolithic and Bronze Age strains.[15] Apart from RV 2039 and Gok2, four more very early *Y. pestis* genomes have been identified in human remains from Estonia, Lithuania and Sweden indicating northeastern Europe, especially the Baltic region, as a geographical hotspot in the diversification of the bacterium.[15,17,22,23]

The estimated emergence date of the RV 2039 lineage suggests that the large-scale branching and geographic radiation of *Y. pestis* occurred between 7000 and 5000 years ago, coinciding with the beginning of the Neolithic period in Europe (~7500-5000 cal BP). This major epidemiological and demographic transition was characterized by the introduction of agriculture, animal domestication and sedentism as well as a general increase in population size.[24] These developments are thought to have laid the foundation for the long-term

evolutionary changes in host-pathogen ecology that ultimately led to the emergence of so-called crowd diseases, i.e. acute and efficiently transmitted infections that probably first appeared in large and densely packed urban communities.[25,26] Therefore, our finding of *Y. pestis* in a complex hunter-gatherer group is rather surprising. Indeed, compared with large-scale settlements, small and sparse populations such as hunter-gatherers offer only a very limited and usually locally restricted pool of susceptible hosts for a pathogen with crowd disease potential.[25]

Archaeological and isotope evidence clearly showed that the infected individual from Riņņukalns followed a life-style and diet (based on freshwater resources) that are typically associated with hunter-gatherers of the Baltic region at that time.[27] The human genomic data support this cultural affiliation and indicate a strong link with Eastern hunter-gatherers that roamed the forest steppe zone between the Black and Baltic Sea during the 6th millennium cal BP.[21] Riņņukalns' 'gateway' location and archaeological artefacts show that the site was an integral part of a long-distance exchange via the river networks of the East European Plain.[28,29,30] Such an exchange system was evidently maintained through intensive and regular contacts between human groups. Thus, the presence of *Y. pestis* in the man from Riņņukalns indicates that also hunter-gatherers may have contributed to the diversification of the early *Y. pestis* lineages across Eurasia.

Modern *Y. pestis* can be transmitted from animals to humans.[20] It is possible that hunter-gatherers, who frequently killed rodents for food or personal decoration, contracted *Y. pestis* or its antecessor *Y. pseudotuberculosis* directly from these animals. Interestingly, at the Riņņukalns site, beaver (*Castor fiber*) was the most frequently recorded species among the

archaeozoological finds excavated by Sievers.[31] Beavers are a common carrier of *Y. pseudotuberculosis*, which directly precedes our early *Y. pestis* strain.[32] Therefore, hunter-gatherers may have played an important role in the zoonotic emergence and early evolution of *Y. pestis.*

Although the hunter-gatherer from Riņņukalns was infected with *Y. pestis*, it is not clear if or to what extent he was actually affected by the plague. RV 2039 did not yet have the genetic components for flea adaptation needed to cause the bubonic form of the disease. However, the genomic data suggest that it was theoretically capable of causing pneumonic plague. The absence of the I259T mutation in the *pla* gene might have compromised its dissemination capabilities, allowing only localized outbreaks of pneumonic plague.[33] Also, with regard to other virulence factors, RV 2039 consistently showed the ancestral alleles typical of *Y. pseudotuberculosis*, an enteropathogenic bacterium that is known to cause relatively mild disease.[34] RV 2039 shared many of its ancestral characteristics with the other Neolithic / Bronze Age strains. It was only millennia later (ca. 3800 cal BP) that *Y. pestis* had acquired all mutations for flea-based transmission.[20] This step is considered crucial in *Y. pestis* evolution since it required the death of the host to assure perpetuation of the bacterium via the flea vector.[35] Thus, subsequent selective pressures facilitated mechanisms promoting the invasion of human tissues, bacteremia and lethality that turned the bubonic plague into an acute and highly deadly disease.[35]

Interestingly, in the RV 2039 dataset, the *Y. pestis*-specific reads were very frequent though they had been generated by shotgun sequencing without any prior enrichment. Many of the other prehistoric *Y. pestis* genomes were also assembled using only shotgun data.[15,17,22,23] The

abundance of *Y. pestis* reads in the aDNA extracts suggests a high bacterial load in the bloodstream of the diseased at the time of death. Remarkably, infection experiments with different *Yersinia* species in mice have shown a negative correlation between the number of bacteria during the terminal stage and their virulence.[35] It is tempting to hypothesize that this correlation also holds for the different ancient *Y. pestis* strains. If true, prehistoric strains with a high bacterial load may have had a lower virulence.

Admittedly, so far there is no experimental data about the pathogenicity of ancient *Y. pestis* strains. However, it cannot be excluded that RV 2039 and the other early forms were less transmissible or perhaps even less virulent than the more evolved strains, leading to a type of 'proto-plague'. In addition, an infection with early *Y. pestis* may have manifested as a chronic disease. This hypothesis is underscored by the observation that almost all infected prehistoric individuals (apart from Gok2) represented sporadic cases and were regularly interred in single or multiple burials, but not in mass graves such as those known from the historical plague epidemics.[17,22,23] Also the man from Riņņukalns was carefully buried, just as the other three contemporaneous individuals in whom we did not find any evidence of *Y. pestis*. These findings argue against an infestation of the whole group with a bacterium that kills its hosts within a few days. The geographical and temporal distribution of all the prehistoric cases reported so far – scattered over the huge Eurasian continent and a period of about 2 millennia – is consistent with a long-standing plague pandemic, albeit one that is associated with a chronic rather than an acute disease course. Intriguingly, the distribution of *Y. pestis* resembles the pattern observed for another common disease at that time, hepatitis B.[36]

Our study highlights how important the Anthropological Rudolf Virchow Collection is for research on human populations in the past. In addition, it shows the power of modern aDNA sequencing technology to detect ancient pathogens, especially bacteria such as *Y. pestis* that do not leave telltale lesions on bone. With the scientific toolkit available at the time, Virchow was in no position to diagnose plague on the Riņņukalns cranium. However, thanks to his progressive scientific approach the remains excavated by Sievers were stored in his collection, where they survived the vicissitudes of time unscathed, so that a later diagnosis was still possible, even after 145 years.

Acknowledgments

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 290391021 – SFB 1266 and under Germany`s Excellence Strategy – EXC 2167 – 390884018 and EXC 2150 – 390870439. The anthropological and archaeological research on Riņņukalns is part of the project 'Riņņukalns, ein neolithischer Süsswassermuschelhaufen im Norden Lettlands und seine Bedeutung für die steinzeitliche Kulturentwicklung im östlichen Baltikum' funded by the DFG (project number 335674082), and of the base funding project of the University of Latvia 'Letonica, diaspora and intercultural communication' (ZD2015/AZ85). We thank the Berlin Society of Anthropology, Ethnology and Prehistory for the opportunity to re-examine the skulls and take the necessary samples for the analyses. We acknowledge financial support by Land Schleswig-Holstein within the funding programme Open Access Publikationsfonds.

Declaration of Interests

The authors declare no conflict of interests.

References

1. Bērziņš V, Brinker U, Klein C, et al. New research at Riņņukalns, a Neolithic freshwater shell midden in northern Latvia. Antiquity 2014;88: 715–732.
2. Brinker U, Bērziņš V, Ceriņa A, et al. (2020) Two burials in a unique freshwater shell midden: insights into transformations of Stone Age hunter-fisher daily life in Latvia. Archaeol. Anthropol. Sci. 2020;12: 97 [DOI: 10.1007/s12520-020-01049-7].
3. Sievers CG. Ein normannisches Schiffsgrab bei Ronneburg und die Ausgrabung des Rinnehügels. Sitzung vom 18. Oktober 1875. Verhandlungen der Berliner Gesellschaft für Anthropologie, Ethnologie und Urgeschichte 1875. Zeitschrift für Ethnologie 1875;7: 214 – 223.
4. Grewingk, C. Der Kauler- und der Rinne-Kaln am Burtnecksee in Livland. Neunzigste Sitzung der Dorpater Naturforscher-Gesellschaft am 28. Januar 1876. Sitzungsberichte der Dorpater Naturforscher-Gesellschaft in den Jahren 1875 bis 1877. 1878;4: 206–225.
5. Brinker U, Meinel D, Teßmann B, Lübke H. Die menschlichen Skelettreste der zur anthropologischen Rudolf-Virchow-Sammlung gehörenden Kollektion des Fundplatzes Riņņukalns im Norden Lettlands – Resultat eines Forschungsstreites des 19. Jahrhunderts im damaligen Livland. Mitteilungen der Berliner Gesellschaft für Anthropologie, Ethnologie und Urgeschichte 2018;39: 35–54.
6. Virchow R Bericht über eine archäologische Reise nach Livland. Sitzung vom 20. October 1877. Verhandlungen der Berliner Gesellschaft für Anthropologie, Ethnologie und Urgeschichte 1877. Zeitschrift für Ethnologie 1877;9:365–437.
7. Andree C. Rudolf Virchow als Prähistoriker. 3 Bde. Böhlau, Köln, Wien; 1976-1986.
8. Lübke H, Brinker U, Meadows J, Bērziņš V, Zagorska New research on the human burials of Riņņukalns, Latvia. In: Grünberg J, GrmascB, Larsson L, Orschiedt J, Meller H (ed) Mesolithic Burials–Rites, Symbols and Social Organization of Early Postglacial Communities, Proceedings of the International Conference, Halle (Saale), 18th–21st September 2013. 2016: 241–256.
9. Ames K.M. Complex Hunter-Gatherers. In: Smith C. (eds) Encyclopedia of Global Archaeology. Springer, New York, NY: 2014:1613-1621. https://doi.org/10.1007/978-1-4419-0465-2.

10. Brooks S, Suchey JM. Skeletal age determination based on the os pubis: a comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. Hum Evol 1990;5: 227–238.
11. Wagner DM, Klunk J, Harbeck M, et al. Yersinia pestis and the Plague of Justinian 541-543 AD: A genomic analysis. Lancet Infect Dis 2014;14: 319–326.
12. Krause-Kyora B, Nutsua M, Beohme L, et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. Nat Commun 2018;9: 1569.
13. Susat J, Bonczarowska JH, Pētersone-Gordina, et al. Yersinia pestis strains from Latvia show depletion of the pla virulence gene at the end of the second plague pandemic. Currently under review at Scientific reports.
14. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 2013;29: 1682-1684.
15. Rascovan N, Sjögren KG, Kristiansen K, et al. Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. Cell 2019;176: 295-305.
16. Bouckaert R, Vaughan TG, Barido-Sottani J, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 2019. https://doi.org/10.1371/journal.pcbi.1006650.
17. Andrades Valtueña A, Mittnik A, Key FM, et al. The Stone Age Plague and Its Persistence in Eurasia. Curr Biol 2017;27: 3683-3691.
18. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 2018;67: 901-904.
19. Immel A, Ţerna S, Simalcsik A, et al. Gene-flow from steppe individuals into Cucuteni-Trypillia associated populations indicates long-standing contacts and gradual admixture. Sci Rep 2020;10.
20. Demeure CE, Dussurget O, Mas Fiol G, Le Guern AS, Savin C, Pizarro-Cerdá J. Yersinia pestis and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. Genes and Immunity 2019;20: 357-370.
21. Mathieson I, Alpaslan-Roodenberg S, Posth C, et al. The genomic history of southeastern Europe. Nature 2018;555: 197-203.
22. Rasmussen S, Allentoft ME, Nielsen K, et al. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. Cell 2015;163: 571-582.

23. Spyrou MA, Tukhbatova RI, Wang CC, et al. Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. Nat Commun 2018;9.
24. Müller J, Diachenko A (2019) Tracing long-term demographic changes: The issue of spatial scales. PLoS ONE 14(1): e0208739. https://doi.org/10.1371/journal.pone.0208739.
25. Wolfe ND, Dunavan CP, Diamond J. Origins of major human infectious diseases. Nature 2007;447: 279-283
26. Düx A, Lequime S, Patrono LV, et al. Measles virus and rinderpest virus divergence dated to the sixth century BCE. Science 2020;368: 1367-1370.
27. Meadows J, Bērziņš V, Legzdiņa D, et al. Stone-age subsistence strategies at Lake Burtnieks, Latvia Journal of Archaeological Science: Reports 2018;17: 992–1006.
28. Loze I. Some aspects of research on Middle Neolithic amber in the Lake Lubāns depression. In: Butrimas A (ed), *Baltic Amber: Proceedings of the International Interdisciplinary Conference: Baltic Amber in Natural Sciences, Archaeology and Applied Arts, 13–18 September 2001, Vilnius, Palanga, Nida.* Acta Academiae Artium Vilnensis, 2001;22: 125–133.
29. Núñez M, Franzén P. Implications of Baltic amber finds in northern Finland 4000–2000 BCE. *Archaeologia Lituana* 2011;12: 10–24.
30. Kriiska A. Foreign materials and artefacts in the 4th and 3rd millennia BCE Estonian Comb Ware Complex. In: Espak P, Läänemets M, Sazonov V (eds), *When Gods Spoke. Researches and Reflections on Religious Phenomena and Artefacts. Studia in honorem Tarmo Kulmar*, 2015:107−124. Tartu: Tartu Ülikooli ajaloo ja arheoloogia Instituudi.
31. Rütimeyer L. Ueber die Thierreste des Rinnekalns. Sitzungsberichte Dorpater Naturforscher-Gesellschaft 1877;4: 534–539.
32. Gaydos JK, Zabek E, Raverty S. Yersinia pseudotuberculosis septicemia in a beaver from Washington State. J Wildl Dis 2009;45: 1182-1186.
33. Zimbler DL, Schroeder JA, Eddy JL, Lathem WW. Early emergence of Yersinia pestis as a severe respiratory pathogen. Nat Commun 2015;6.
34. Bertelli L, Masetti R, Bardasi G, et al. Two cases of abdominal pain in children with mesenteric lymphadenitis due to yersinia pseudotuberculosis infection. J Pediatr 2014;165.

35. Brubaker RR. Factors promoting acute and chronic diseases caused by yersiniae. Clin Microbiol Rev 1991;4: 309-324.
36. Krause-Kyora B, Susat J, Key FM, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. Elife. 2018;7.

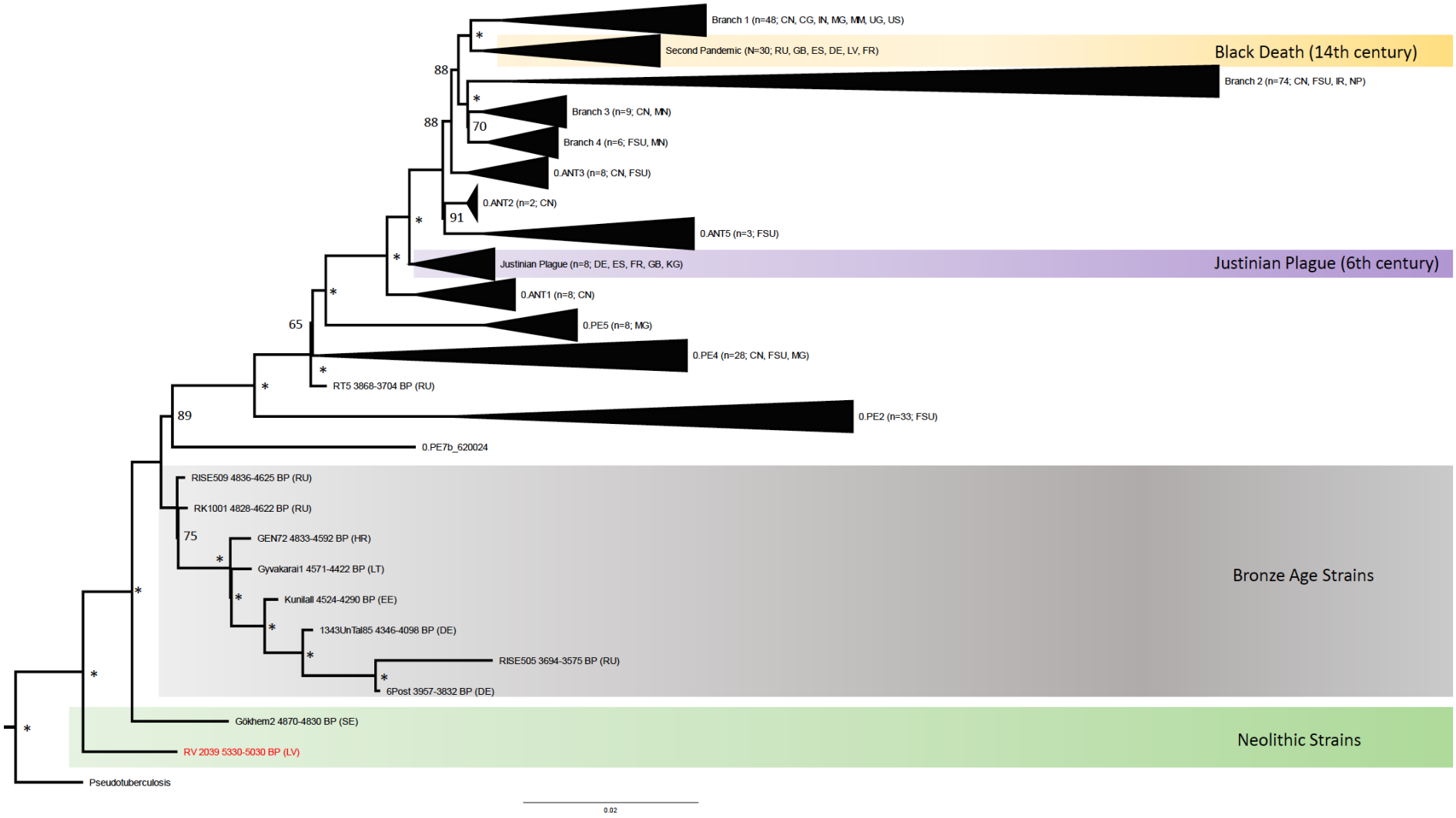

Branch 1 (n=48; CN, CG, IN, MG, MM, UG, US)
Second Pandemic (N=30; RU, GB, ES, DE, LV, FR)
Black Death (14th century)
Branch 2 (n=74; CN, FSU, IR, NP)
Branch 3 (n=9; CN, MN)
Branch 4 (n=6; FSU, MN)
0.ANT3 (n=8; CN, FSU)
0.ANT2 (n=2; CN)
0.ANT5 (n=3; FSU)
Justinian Plague (n=8; DE, ES, FR, GB, KG)
Justinian Plague (6th century)
0.ANT1 (n=8; CN)
0.PE5 (n=8; MG)
0.PE4 (n=28; CN, FSU, MG)
RT5 3868-3704 BP (RU)
0.PE2 (n=33; FSU)
0.PE7b_620024
RISE509 4836-4625 BP (RU)
RK1001 4828-4622 BP (RU)
GEN72 4833-4592 BP (HR)
Gyvakarai1 4571-4422 BP (LT)
Kunila11 4524-4290 BP (EE)
1343UnTal85 4346-4098 BP (DE)
RISE505 3694-3575 BP (RU)
6Post 3957-3832 BP (DE)
Bronze Age Strains
Gökhem2 4870-4830 BP (SE)
RV 2039 5330-5030 BP (LV)
Neolithic Strains
Pseudotuberculosis
88
88
70
91
65
89
75
0.02

Figure 3. Maximum-likelihood tree. The tree is based on the SNP alignment (18169 positions) of 228 modern *Y. pestis* genomes, 48 published ancient *Y. pestis* strains, one *Y. pseudotuberculosis* genome and the strain from Riņņukalns (RV 2039 in red). Country abbreviation is given in brackets (DE = Germany, ES = Spain, FR = France, GB = Great Britain, US = United States, RU = Russia, LV = Latvia, CN = China, CG = Congo, FSU = Former Soviet Union, IN = India, IR = Iran, MG = Madagascar, MM = Myanmar, MN = Mongolia, NP = Nepal, UG = Uganda, KG = Kyrgyzstan, CH = Switzerland, HR = Croatia, EE = Estonia, LT = Lithuania, SE = Sweden). Neolithic strains are highlighted in green, Bronze Age strains in grey, strains from the Justinian Plague in purple and strains from the Black Death in yellow. Bootstrap values are shown on the nodes for 500 replicates and an asterisk (*) represents a bootstrap support above 95.

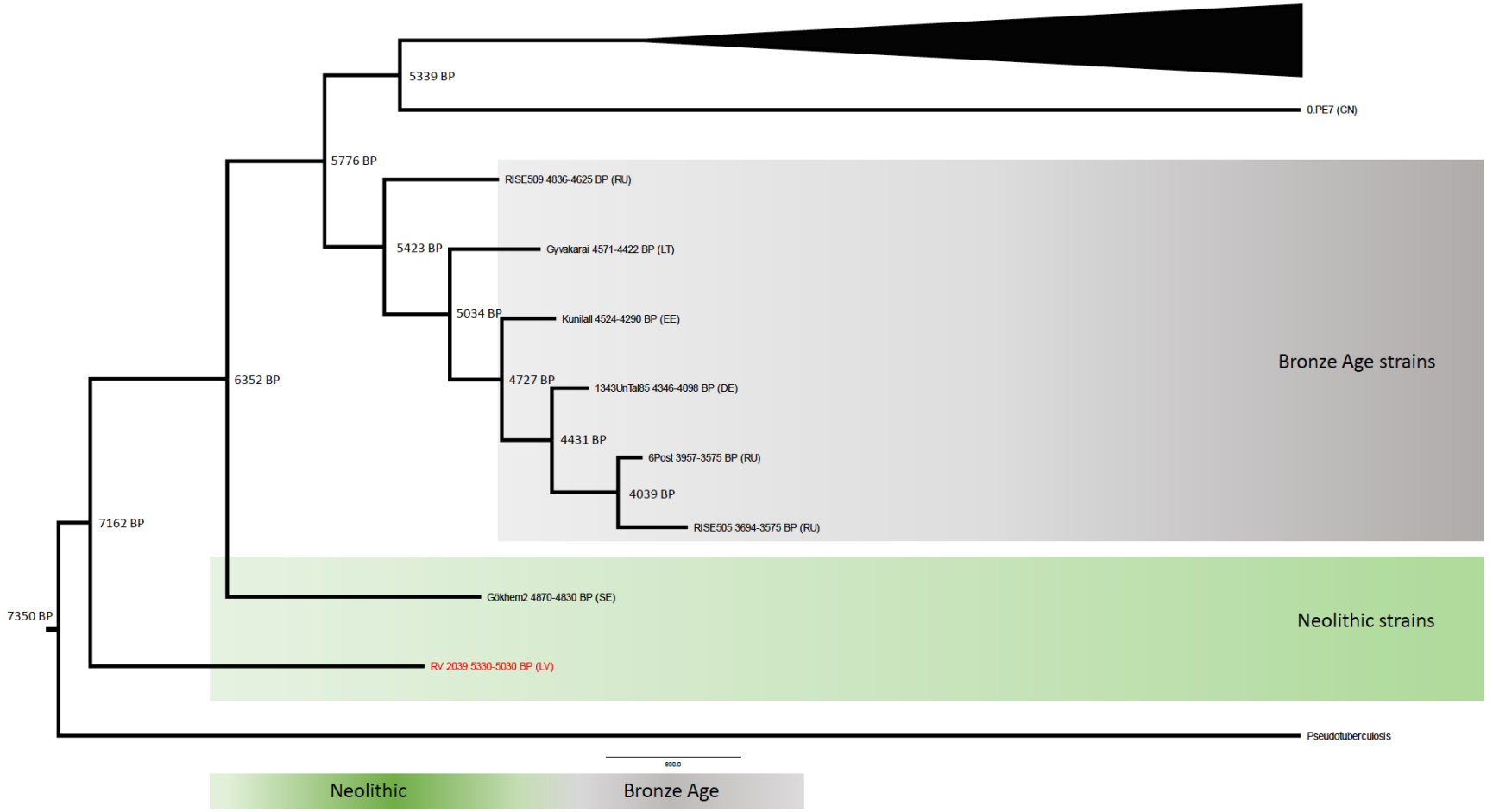

5339 BP
0.PE7 (CN)
5776 BP
RISE509 4836-4625 BP (RU)
5423 BP
Gyvakarai 4571-4422 BP (LT)
5034 BP
Kunila II 4524-4290 BP (EE)
6352 BP
4727 BP
1343UnTal85 4346-4098 BP (DE)
4431 BP
6Post 3957-3575 BP (RU)
4039 BP
RISE505 3694-3575 BP (RU)
Bronze Age strains
7162 BP
Gökhem2 4870-4830 BP (SE)
7350 BP
RV 2039 5330-5030 BP (LV)
Neolithic strains
Pseudotuberculosis
800.0
Neolithic
Bronze Age

Figure 4. Molecular clock estimation (without starting tree) of modern and ancient Y. *pestis* strains. BEAST2 maximum clade credibility tree based on 41 modern and ancient Y. *pestis* genomes and the strain from Riņņukalns (RV 2039 in red). The final tree was summarized from 900 million states and 810252 trees showing divergence dates as tree nodes. All dates are given as years before the present (BP). Country abbreviation is given in brackets (DE = Germany, RU = Russia, LV = Latvia, CN = China, EE = Estonia, LT = Lithuania, SE = Sweden). Neolithic strains are highlighted in green and Bronze Age strains in grey.

4.

Chapter III: Neolithic and medieval virus genomes reveal complex evolution of hepatitis B

eLife 2018

Doi: 10.7554/eLife.36666

Supplementary material can be found at the end of this thesis

eLIFE elifesciences.org

SHORT REPORT

Neolithic and medieval virus genomes reveal complex evolution of hepatitis B

Ben Krause-Kyora[1,2†*], Julian Susat[1†], Felix M Key[2], Denise Kühnert[2,3], Esther Bosse[1,4], Alexander Immel[1,2], Christoph Rinne[5], Sabin-Christin Kornell[1], Diego Yepes[4], Sören Franzenburg[1], Henrike O Heyne[6,7,8], Thomas Meier[9,10], Sandra Lösch[11], Harald Meller[12], Susanne Friederich[12], Nicole Nicklisch[12,13], Kurt W Alt[12,13,14,15], Stefan Schreiber[1,16], Andreas Tholey[4], Alexander Herbig[2], Almut Nebel[1], Johannes Krause[2*]

[1]Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany; [2]Max Planck Institute for the Science of Human History, Jena, Germany; [3]Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland; [4]Systematic Proteomics & Bioanalytics, Institute for Experimental Medicine, Kiel University, Kiel, Germany; [5]Institute of Pre- and Protohistoric Archaeology, Kiel University, Kiel, Germany; [6]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, United States; [7]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, United States; [8]Program in Medical and Population Genetics, Broad Institute of MIT & Harvard, Cambridge, United States; [9]Institute for Pre- and Protohistory and Near Eastern Archaeology, Heidelberg University, Heidelberg, Germany; [10]Heidelberg Center for the Environment, Heidelberg University, Heidelberg, Germany; [11]Department of Physical Anthropology, Institute of Forensic Medicine, University of Bern, Bern, Switzerland; [12]State Office for Heritage Management and Archaeology Saxony-Anhalt, State Museum of Prehistory, Halle, Germany; [13]Danube Private University, Krems, Austria; [14]Department of Biomedical Engineering, University Hospital Basel, University of Basel, Basel, Switzerland; [15]Integrative Prehistory and Archaeological Science, University of Basel, Basel, Switzerland; [16]Clinic for Internal Medicine, University Hospital Schleswig-Holstein, Kiel, Germany

***For correspondence:** b.krause-kyora@ikmb.uni-kiel.de (BK-K); krause@shh.mpg.de (JK)

[†]These authors contributed equally to this work

Competing interests: The authors declare that no competing interests exist.

Funding: See page 12

Received: 14 March 2018
Accepted: 09 May 2018
Published: 10 May 2018

Reviewing editor: Stephen Locarnini, Doherty Institute, Australia

Abstract The hepatitis B virus (HBV) is one of the most widespread human pathogens known today, yet its origin and evolutionary history are still unclear and controversial. Here, we report the analysis of three ancient HBV genomes recovered from human skeletons found at three different archaeological sites in Germany. We reconstructed two Neolithic and one medieval HBV genome by *de novo* assembly from shotgun DNA sequencing data. Additionally, we observed HBV-specific peptides using paleo-proteomics. Our results demonstrated that HBV has circulated in the European population for at least 7000 years. The Neolithic HBV genomes show a high genomic similarity to each other. In a phylogenetic network, they do not group with any human-associated HBV genome and are most closely related to those infecting African non-human primates. The ancient viruses appear to represent distinct lineages that have no close relatives today and possibly went extinct. Our results reveal the great potential of ancient DNA from human skeletons in order to study the long-time evolution of blood borne viruses.
DOI: https://doi.org/10.7554/eLife.36666.001

Introduction

The hepatitis B virus (HBV) is one of the most widespread human pathogens, with worldwide over 250 million people being infected, and an annual death toll of about 1 million globally (**WHO, 2017**). Infection of liver cells with HBV leads to acute hepatitis B, which is self-limiting in about 90–95% of cases. In about 5–10% of infected individuals virus clearance fails and patients develop chronic infection of hepatitis B, which puts them at lifelong elevated risk for liver cirrhosis and liver cancer (hepatocellular carcinoma). HBV is usually transmitted by contact with infectious blood, in highly endemic countries often during birth (**WHO, 2017**).

HBV has a circular, partially double-stranded DNA genome of about 3.2kbp that encodes four overlapping open reading frames (P, pre-S/S, pre-C/C, and X). Based on the genomic sequence diversity, HBVs are currently classified into eight genotypes (A-H) and numerous subgenotypes that show distinct geographic distributions (**Castelhano et al., 2017**). All genotypes are hypothesised to be primarily the result of recombination events (**Littlejohn et al., 2016**; **Simmonds and Midgley, 2005**). To a lesser extent, HBV evolution is also driven by the accumulation of point mutations (**Schaefer, 2007**; **Araujo, 2015**).

Despite being widespread and well-studied, the origin and evolutionary history of HBV are still unclear and controversial (**Littlejohn et al., 2016**; **Souza et al., 2014**). HBVs in non-human primates (NHP), for instance in chimpanzees and gorillas, are phylogenetically closely related to, and yet distinct from, human HBV isolates, supporting the notion of an Africa origin of the virus (**Souza et al., 2014**). Molecular-clock-based analyses dating the origin of HBV have resulted in conflicting estimates with some as recent as about 400 years ago (**Zhou and Holmes, 2007**; **Souza et al., 2014**). These observations have raised doubts about the suitability of molecular dating approaches for reconstructing the evolution of HBV (**Bouckaert et al., 2013** , **Souza et al., 2014**). Moreover, ancient DNA (aDNA) research on HBV-infected mummies from the 16th century AD revealed a very close relationship between the ancient and modern HBV genomes (**Kahila Bar-Gal et al., 2012**; **Patterson Ross et al., 2018**), indicating a surprising lack of temporal genetic changes in the virus during the last 500 years (**Patterson Ross et al., 2018**). Therefore, diachronic aDNA HBV studies are necessary, in which both the changes in the viral genome over time as well as the provenance and age of the archaeological samples are investigated, to better understand the origin and evolutionary history of the virus.

Here, we report the analysis of three complete HBV genomes recovered from human skeletal remains from the prehistoric Neolithic and Medieval Periods in Central Europe. Our results show that HBV already circulated in the European population more than 7000 years ago. Although the ancient forms show a relationship to modern isolates they appear to represent distinct lineages that have no close modern relatives and are possibly extinct today.

Results and discussion

We detected evidence for presence of ancient HBV in three human tooth samples as part of a metagenomic screening for viral pathogens that was performed on shotgun sequencing data from 53 skeletons using the metagenomic alignment software MALT (**Vågene et al., 2018**). The remains of the individuals were excavated from the Neolithic sites of Karsdorf (Linearbandkeramik [LBK], 5056–4959 cal BC) and Sorsum (Tiefstichkeramik group of the Funnel Beaker culture, 3335–3107 cal BC) as well as from the medieval cemetery of Petersberg/Kleiner Madron (1020–1116 cal AD), all located in Germany (**Figure 1**, **Figure 1—figure supplements 1–3**). After the three aDNA extracts had appeared HBV-positive in the initial virus screening, they were subjected to deep-sequencing without any prior enrichment resulting in 367 to 419 million reads per sample (**Table 1**). A principal component analysis (PCA) of the human DNA recovered from Karsdorf (3-fold genomic coverage) revealed that the sample clusters tightly with other contemporary early Neolithic individuals from the LBK (**Figure 1—figure supplement 4**). The genetic makeup of the early LBK agriculturalists was previously found quite distinct from the preceding western hunter-gatherers of Europe. The genetic shift between both populations was interpreted as a result of early farmers migrating from Western Anatolia into Central Europe introducing agriculture (**Lazaridis et al., 2014**; **Haak et al., 2015**). The almost 2000 years younger Sorsum individual (1.2-fold genomic coverage) clusters in the PCA most closely with individuals from the contemporary Funnel Beaker culture that inhabited

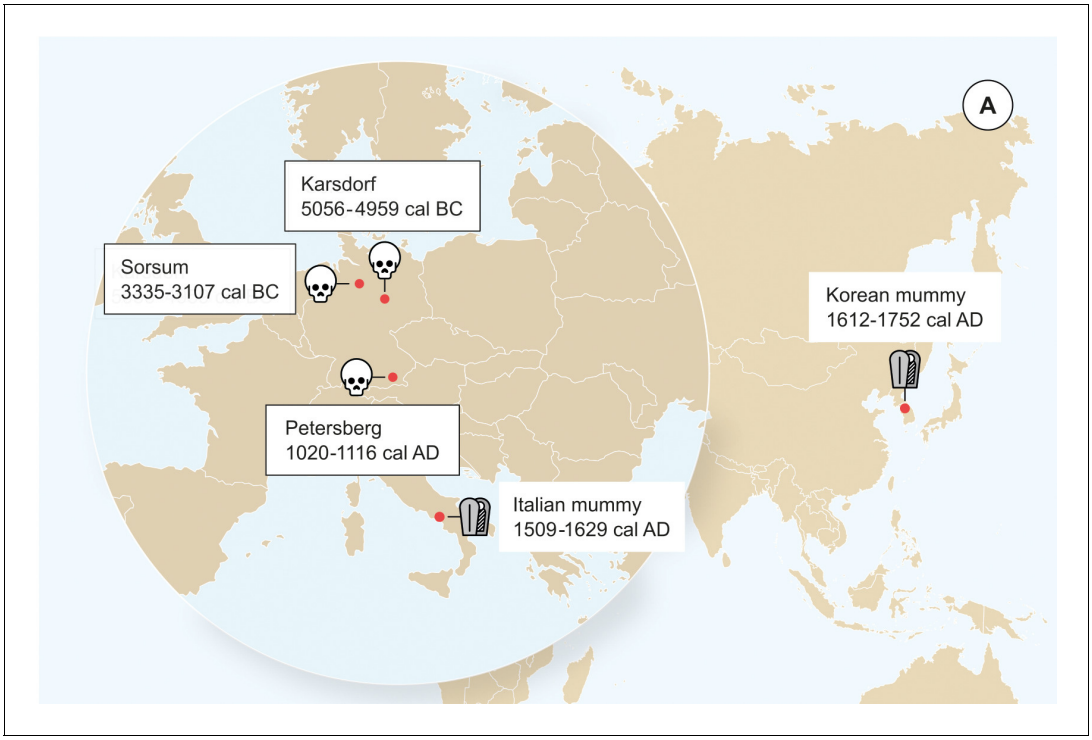


Figure 1. Origin of samples. Geographic location of the samples from which ancient HBV genomes were isolated. Radiocarbon dates of the specimens is given in two sigma range. Icons indicate the sample material (tooth or mummy). HBV genomes obtained in this study are indicated by black frame.
DOI: https://doi.org/10.7554/eLife.36666.002
The following figure supplements are available for figure 1:

Figure supplement 1. Skull of the investigated Karsdorf individual 537 is from a male with an age at death of around 25–30 years.
DOI: https://doi.org/10.7554/eLife.36666.003
Figure supplement 2. Mandible fragment of the Sorsum individual XLVII 11 analyzed in this study is from a male.
DOI: https://doi.org/10.7554/eLife.36666.004
Figure supplement 3. Skull of the analyzed Petersberg individual from grave 820 is from a male with an age at death of around 65–70 years.
DOI: https://doi.org/10.7554/eLife.36666.005
Figure supplement 4. Principal Component Analysis (PCA) of the human Karsdorf and Sorsum samples together with previously published ancient populations projected on 27 modern day West Eurasian populations (not shown) based on a set of 1.23 million SNPs (***Mathieson et al., 2015***).
DOI: https://doi.org/10.7554/eLife.36666.006
Figure supplement 5. Damage plots showing deamination patterns of hg19-specific reads for the HalfUDG-treated libraries of (**a**) Karsdorf, (**b**) Sorsum, (**c**) Petersberg.
DOI: https://doi.org/10.7554/eLife.36666.007
Figure supplement 6. Damage plots showing deamination patterns of HBV-specific reads for the HalfUDG-treated libraries of (**a**) Karsdorf, (**b**) Sorsum, (**c**) Petersberg.
DOI: https://doi.org/10.7554/eLife.36666.008
Figure supplement 7. MS/MS spectrum of the proteotypic HBV-peptide DLLDTASALYR from the HBV-protein external core antigen (residues 58–68).
DOI: https://doi.org/10.7554/eLife.36666.009
Figure supplement 8. Principal Component Analysis (PCA) of the human Karsdorf and Sorsum samples together with previously published ancient populations projected on 27 modern day West Eurasian populations (shown in gray) based on a set of 1.23 million SNPs (***Mathieson et al., 2015***).
DOI: https://doi.org/10.7554/eLife.36666.010
Figure supplement 9. Principal Component Analysis (PCA) of the human Petersberg sample projected on 27 modern day West Eurasian populations based on a set of 1.23 million SNPs (***Mathieson et al., 2015***).
DOI: https://doi.org/10.7554/eLife.36666.011

Northern Germany at the end of the fourth millennium BCE (***Figure 1—figure supplement 4***). This population was previously shown to be quite admixed, as a result of a spatial and temporal overlap of early Neolithic farmers and remaining western hunter-gatherers for almost 2000 years (***Bollongino et al., 2013***; ***Haak et al., 2015***). The Petersberg individual (2.9-fold genomic coverage),

Table 1. Results of the genome reconstruction

	*Merged reads	Length of HBV consensus sequence	Mean HBV coverage	Gaps in the consensus sequence at nt position	*Mapped reads HBV	*Mapped reads human	Mean human coverage	Human genomes/ HBVgenomes
Karsdorf	386,780,892	3183	104X	2157–2175; 3107–3128; 3133–3183	10,718	122,568,310	2.96X	1: 35.1
Sorsum	367,574,767	3182	47X	-	3249	9,856,001	1.17X	1: 40.2
Petersberg	419,413,082	3161	46X	880–1000; 1232–1329; 1331–1415; 1420–1581; 1585–1598	2125	105,476,677	2.88X	1: 16

*number.

nt, nucleotide.

DOI: https://doi.org/10.7554/eLife.36666.012

however, showed genetic affinities in the PCA with modern day central European populations. All three ancient human individuals are therefore in agreement with the archeological evidence and radiocarbon dates for their respective time of origin. Together with typical aDNA damage patterns (***Figure 1—figure supplements 5–6***), the human population genetic investigation supports the ancient origin of the obtained datasets.

For successful HBV genome reconstruction, we mapped all metagenomic sequences to 16 HBV reference genomes (eight human genotypes (A-H) and 8 NHPs from Africa and Asia) that are representative of the current HBV strain diversity (***Supplementary file 6***). The mapped reads were used for a *de novo* assembly, resulting in contigs from which one ancient HBV consensus sequence per sample was constructed. The consensus genomes are 3161 (46-fold coverage), 3182 (47-fold coverage), and 3183 (104-fold coverage) nucleotides in length, which falls in the length range of modern HBV genomes and suggests that we successfully reconstructed the entire ancient HBV genomes (***Table 1***, ***Figure 2—figure supplements 1–3***). Further, when we conducted liquid chromatography-mass spectrometry (LC-MS) based bottom-up proteomics on tooth material from the three individuals, we identified in the Karsdorf and Petersberg samples a peptide that is part of the very stable HBV core protein, supporting the presence and active replication of HBV in the individuals' blood (***Figure 1—figure supplement 7***).

Phylogenetic network analysis was carried out with a dataset comprised of 493 modern HBV strains representing the full genetic diversity. Strikingly, the Neolithic HBV genomes did not group with any human strain in the phylogeny. Instead, they branched off in two lineages and were most closely related to the African NHP genomes (***Figure 2***, 93% similarity). Although the two Neolithic strains were recovered from humans who had lived about 2000 years apart, they showed a higher genomic similarity to each other than to any other human or NHP genotype. Still, their genomes differed by 6% from each other and may therefore be considered representatives of two separate lineages. They did, however, differ less than 8% from the African NHP strains and should therefore not be called a separate genotype (***Figure 2—figure supplement 4***). The genome from the 1000-year-old Petersberg individual clustered with modern D4 genotypes.

Owing to continuous recombination over time, different gene segments or modules of the ancestral genomes can show up in various subsequent virus generations. Such precursors have been postulated (***Simmonds and Midgley, 2005***) and their existence is supported by the results of our recombination analysis (***Figure 2—figure supplements 5–8***, ***Figure 2—source data 1***). Some fragments of the Karsdorf sequences appeared to be very similar to modern human (G, E) and African NHP genotypes, and the Sorsum genome partially showed a high similarity to the human genotypes G, E and B. (***Figure 2—figure supplements 5–8***, ***Figure 2—source data 1***). Given the close relationship between the two Neolithic virus genomes, it is also conceivable that the older HBV from Karsdorf could have been a distant source for the younger Sorsum virus (***Figure 2—figure supplements 5–8***, ***Figure 2—source data 1***). The closer relationship between the Neolithic and the NHP strains compared to other human strains is noteworthy and may have involved reciprocal cross-species transmission at one or possibly several times in the past (***Simmonds and Midgley, 2005***; ***Souza et al., 2014***; ***Rasche et al., 2016***).

Taken together, our results demonstrate that HBV already existed in Europeans 7000 years ago and that its genomic structure closely resembled that of modern hepatitis B viruses. Both Neolithic

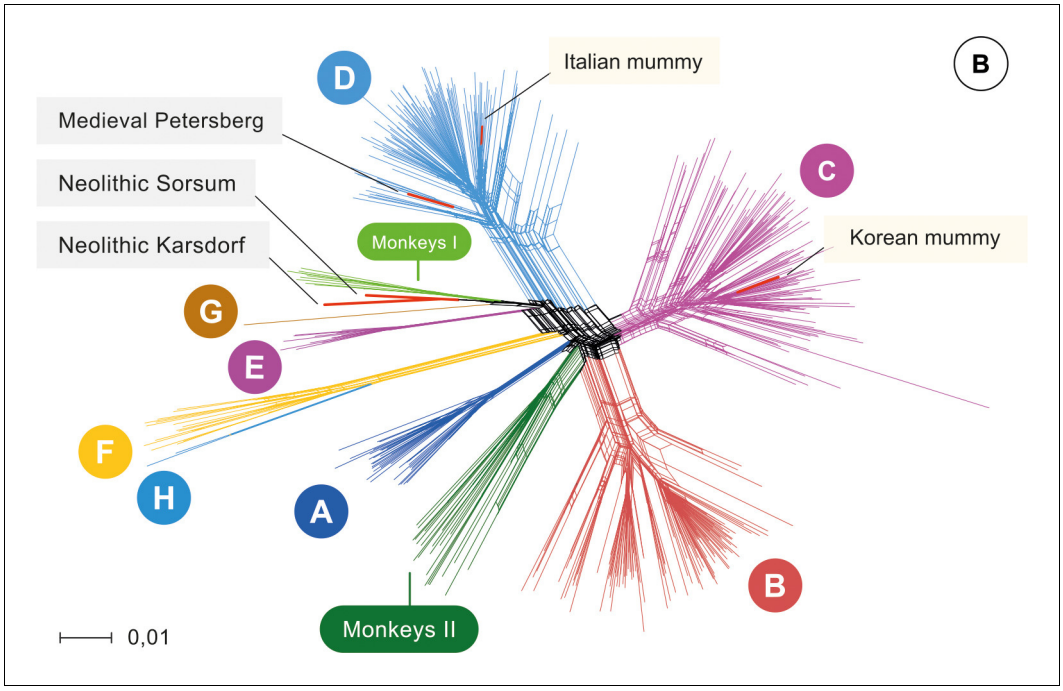


Figure 2. Network. Network of 493 modern, two published ancient genomes (light yellow box), and three ancient hepatitis B virus (HBV) obtained in this study (grey box). Colors indicate the eight human HBV genotypes (**A–H**), two monkey genotypes (Monkeys I, African apes and Monkeys II, Asian monkeys) and ancient genomes (red).
DOI: https://doi.org/10.7554/eLife.36666.013

The following source data and figure supplements are available for figure 2:

Source data 1. Results of the recombination analysis using the methods RDP, GENECOV, Chimera, MaxChi, BootScan, SiScan, 3Seq within the RDP v4 software package with all modern full reference genomes (n = 493) and five ancient genomes.
DOI: https://doi.org/10.7554/eLife.36666.023

Source data 2. Multiple sequence alignment of the 493 representative and five ancient HBV genomes.
DOI: https://doi.org/10.7554/eLife.36666.024

Source data 3. Maximum-likelihood tree based on the multiple sequence alignment of the 493 representative and five ancient HBV genomes with 2000 replicates.
DOI: https://doi.org/10.7554/eLife.36666.025

Source data 4. Neighbour-Joining tree based on the multiple sequence alignment of the 493 representative modern and five ancient HBV genomes with 10000 replicates.
DOI: https://doi.org/10.7554/eLife.36666.026

Figure supplement 1. Consensus sequence of the Karsdorf HBV genome.
DOI: https://doi.org/10.7554/eLife.36666.014

Figure supplement 2. Consensus sequence of the Sorsum HBV genome.
DOI: https://doi.org/10.7554/eLife.36666.015

Figure supplement 3. Consensus sequence of the Petersberg HBV genome.
DOI: https://doi.org/10.7554/eLife.36666.016

Figure supplement 4. Genetic (hamming) distance of our three ancient HBV genomes compared to all 493 reference genomes.
DOI: https://doi.org/10.7554/eLife.36666.017

Figure supplement 5. BootScan analysis of the sequence Karsdorf.
DOI: https://doi.org/10.7554/eLife.36666.018

Figure supplement 6. BootScan analysis of the sequence Sorsum.
DOI: https://doi.org/10.7554/eLife.36666.019

Figure supplement 7. BootScan analysis of the sequence Petersberg.
DOI: https://doi.org/10.7554/eLife.36666.020

Figure supplement 8. SimPlot analysis of (**a**) Karsdorf, (**b**) Sorsum and (**c**) Petersberg.
DOI: https://doi.org/10.7554/eLife.36666.021

Figure 2 continued on next page

Figure 2 continued

Figure supplement 9. Plot of phylogenetic root-to-tip distance relative to sampling time (TempEst).
DOI: https://doi.org/10.7554/eLife.36666.022

viruses fall between the present-day modern human and the known NHP diversity. Therefore, it can be hypothesized that although the two Neolithic HBV strains are no longer observed today and thus may reflect two distinct clades that went extinct, they could still be closely related to the remote ancestors of the present-day genotypes, which is supported by signs of ancient recombination events. More ancient precursors, intermediates and modern strains of both humans and NHPs need to be sequenced to disentangle the complex evolution of HBV. As this evolution is characterized by recombination and point mutations and may further be complicated by human-ape host barrier crossing (***Simmonds and Midgley, 2005***; ***Souza et al., 2014***; ***Rasche et al., 2016***), genetic dating is not expected to yield meaningful results. This is additionally supported by a TempEst analysis (***Rambaut et al., 2016***) that shows very little temporal signal (***Figure 2—figure supplement 9***). It should, however, be noted that the oldest genome (Karsdorf) was found in an individual that belonged to a population of early farmers that had migrated in the previous few hundred years from the Near East into central Europe. One might speculate that the close proximity to recently domesticated animals, changes in subsistence strategy as well as the adopted sedentary lifestyle might have contributed to the spread of HBV within Neolithic human populations.

Based on our analysis, HBV DNA can reliably be detected in tooth samples that are up to 7000 years old. Ancient HBV has so far only been identified in soft tissue from two 16th-century mummies (***Kahila Bar-Gal et al., 2012***; ***Patterson Ross et al., 2018***). The aDNA analysis of HBV from prehistoric skeletons, which facilitates evolutionary studies on a far-reaching temporal scale, has not been described up to now. One explanation for the difficulty of a molecular HBV diagnosis in bones is that the virus infection does not leave lesions on skeletal remains that would allow researchers to select affected individuals *a priori*, as it is the case for instance for leprosy (***Schuenemann et al., 2013***). The diagnosis of an HBV infection in skeletal populations is purely a chance finding and is thus more probable in a large-scale screening.

Overall, HBV biomolecules seem to be well preserved in teeth: Avoiding biases from DNA capture and reference-based mapping we could reconstruct three HBV genomes by *de novo* assembly from shotgun data and even observed HBV-specific peptides. The ratio of HBV genomes to the human genome in our samples was rather high and similar in all three samples (Karsdorf 35:1, Sorsum 40.2:1 and Petersberg 16:1). As there is no evidence that HBV DNA is more resistant to post-mortem degradation than human DNA, the high rate of HBV compared to human DNA may reflect the disease state in the infected individuals at the time of death. High copy numbers of viral DNA in the blood of infected individuals are associated with acute HBV infection, or reactivation of chronic HBV. Thus, it seems likely that the death of the ancient individuals is related to the HBV infection, but might not be the direct cause of death as fulminant liver failure is rather rare in modern day patients. The HBV infection might have instead contributed to other forms of lethal liver failure such as cirrhosis or liver cancer.

In view of the unexpected complexity of our findings, we envisage future diachronic HBV studies that go beyond the temporal and geographic scope of our current work.

Materials and methods

Human remains

The LBK settlement of Karsdorf, Saxony-Anhalt, Germany, is located in the valley of the river Unstrut. Between 1996 and 2010 systematic excavations were conducted at Karsdorf that led to the discovery of settlements and graves from the Neolithic to the Iron Age (***Behnke, 2007***; ***2011***; ***2012***). The LBK is represented by 24 longhouses in north-west to south-east orientation that were associated with settlement burials (***Veit, 1996***). The investigated individual 537 is a male with an age at death of around 25–30 years (***Figure 1—figure supplement 1***), dated to 5056–4959 cal BC (KIA 40357–6116 ± 32 BP) (***Brandt et al., 2014***; ***Nicklisch, 2017***).

The gallery grave of Sorsum, Lower-Saxony, Germany, is typologically dated to the Tiefstichkeramik (group of the Funnelbeaker culture). Sorsum is exceptional as it was built into the bedrock. During the excavations (1956–1960) of the grave chamber around 105 individuals were recovered (***Claus, 1983***; ***Czarnetzki, 1966***). Individual XLVII 11 analyzed in this study is a male (***Figure 1—figure supplement 2***) and dates to 3335–3107 cal BC (MAMS 33641–4501 ± 19 BP).

The medieval cemetery of Petersberg/Kleiner Madron, Bavaria, Germany, lies on a hill top at 850 meters asl and 400 meters above the floor of the Inn Valley. On the eastern part of the cemetery members of a priory were buried that was most likely established in the late 10th century. Written sources document its existence from 1132 onwards (***Meier, 1998***). During systematic excavations (1997–2004) in the southeastern part of the churchyard the remains of individuals buried in 99 graves were uncovered. The examined individual in grave 820 is a male with an age at death of around 65–70 years (***Lösch, 2009*** - ***Figure 1—figure supplement 3***) dating to 1020–1116 cal AD (MAMS 33642–982 ± 17 BP).

DNA extraction and sequencing

The DNA extractions and pre-PCR steps were carried out in clean room facilities dedicated to aDNA research. Teeth were used for the analyses. The samples from Petersberg and Sorsum were processed in the Ancient DNA Laboratory at Kiel University and the sample from Karsdorf in the Ancient DNA Laboratory of the Max Planck Institute for the Science of Human History (MPI SHH) in Jena. All procedures followed the guidelines on contamination control in aDNA studies (***Warinner et al., 2017***; ***Key et al., 2017***). The teeth were cleaned in pure bleach solution to remove potential contaminations prior to powdering. Fifty milligrams of powder were used for extraction following a silica-based protocol (***Dabney et al., 2013***). Negative controls were included in all steps.

From each sample, double-stranded DNA sequencing libraries (UDGhalf) were prepared according to an established protocol for multiplex high-throughput sequencing (***Meyer and Kircher, 2010***). Sample-specific indices were added to both library adapters via amplification with two index primers. Extraction and library blanks were treated in the same manner. For the initial screening, the library of the individual from Karsdorf was sequenced on 1/50 of a lane on the HiSeq 3000 (2 × 75 bp) at the MPI SHH in Jena and the libraries from Petersberg and Sorsum were sequenced on the Illumina HiSeq 4000 (2 × 75 bp) platform at the Institute of Clinical Molecular Biology, Kiel University, using the HiSeq v4 chemistry and the manufacturer's protocol for multiplex sequencing. Deep-sequencing for each of the three samples was carried out on two lanes on the Illumina HiSeq 4000 platform at the Institute of Clinical Molecular Biology, Kiel University.

Clip and merge

The datasets produced for all ancient samples contained paired-end reads with varying numbers of overlapping nucleotides as well as artificial adapter sequences. We used ClipAndMerge version 1.7.3, a module of the EAGER pipeline (***Peltzer et al., 2016***), to clip adapter sequences, merge corresponding paired-end reads in overlapping regions and to trim the resulting reads. We used the default options with the following command:

```
java -jar ClipAndMerge.jar -in1 $FASTQ1 -in2 $FASTQ2 \
-f AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC \
-r AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTA \
-l 25 -qt -q 20 -o $output_file
```

where $FASTQ1 and $FASTQ2 are the two gzipped FASTQ input files

Adapter clipping

ClipAndMerge uses an overlap alignment of the respective forward or reverse adapter with the 3' end of each read in order to remove sequencing adapter sequences. Regions at the 3' end of each read that were contained in the alignment were clipped. Reads that were shorter than 25 nucleotides after adapter clipping or contained only adapter sequences (adapter dimers) were removed. All remaining reads were then used in the merging step.

Merging of overlapping paired reads

Merging was performed for all remaining paired reads with a minimum overlap of 10 nucleotides and at most 5% mismatches in the overlap region. The algorithm selected the maximal overlap fulfilling these criteria. The consensus sequence was generated using the nucleotides in the overlap regions from the read with the higher PHRED quality score, maximizing the quality of the resulting read.

Quality trimming

In a final step, ClipAndMerge performed quality trimming of the reads and all nucleotides with PHRED scores smaller than 20 were trimmed from the 3' end of each read. Finally, all reads with fewer than 25 nucleotides after quality trimming were removed. The resulting high-quality reads were used for the alignment.

Virus screening

Screening of the datasets was carried out with the software MALT using the ncbi-viral database as a reference. A sequence identity threshold of 85% was set and the alignment mode was changed to SemiGlobal. The analysis was carried out using the following command:

malt-run –mode BlastN -e 0.001 -id 85 –alignmentType SemiGlobal –index $index –inFile $FASTQCM –output $OUT

where $index is the index file, $FASTQCM is the clipped and merged file and $OUT is the output the file.

The resulting alignments were visually inspected using MEGAN 6. Reads mapping to the hepatitis B reference in the database (NC_003977.2) were extracted and verified using a discontiguous megablast against the virus taxa (taxid: 10239) with default parameters.

HBV alignment

For identification of the genotypes, samples were aligned against one reference for each of the eight hepatitis B genotypes available in the NCBI hepatitis B genotyping project (https://www.ncbi.nlm.nih.gov/projects/genotyping/view.cgi?db=2) (***Supplementary file 1***). Additionally, eight NHP strains were used. All references were combined in one FASTA file and a competitive mapping was performed using BWA. The mapping was carried out using the following command:

bwa aln -n 0.01 l 300 $INDEX $FASTQCM $OUT

where $INDEX is the reference, $FASTQCM is the input file and $OUT is the output file.

Minimum mapping quality was set to 0.

Duplicate removal

We used DeDup version 0.11.3, part of the EAGER pipeline,***Peltzer et al., 2016*** to identify and remove all duplicate reads in the sample specific BAM files (***Supplementary file 2***) with the default options and the following command:

java -jar DeDup.jar -i $IN -o $OUT

where $IN is the input BAM file and $OUT is the output BAM file.

Extracting mapped reads

After duplicate removal the resulting BAM files were converted to SAM files using SAMtools version 0.1.19-96b5f2294a with default parameters and the following command:

samtools view -h -o $OUT $IN

where $OUT is the SAM output file and $IN is the BAM input file. Reads from the SAM file where converted to FASTQ using the following awk script:

awk '/[FMR]/{print '@"$1'\n'$10'\n+\n'$11}' $IN > $OUT where $IN is a SAM file and $OUT is the resulting FASTQ file containing all the mapped reads.

De novo assembly

The de novo assembly was performed using the SPAdes genome assembler version v3.9.0 (***Bankevich et al., 2012***) with the following settings:

spades.py -t 20 m 500 k

11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,59,61,63,65,67,69,71,73,75,77,79,81,83,85,87,89,91,93,95,97,99,101,103,105,107,109,111,113,115,117,119,121,123,125,127 s $IN -o $OUT

where $IN is a FASTQ file containing the mapped reads and $OUT is the output folder for SPAdes.

Resulting contigs for each K-value were checked and the one which spawned the longest contig was selected for further processing (***Supplementary file 3***).

Mapping of contigs

Contigs were mapped against the multi FASTA file containing all 16 references. The following command was used:

bwa mem $INDEX $IN $OUT

where $INDEX is the reference, $IN the file containing the contig/contigs and $OUT is the resulting BAM file.

Consensus generation

For genomic reconstruction of the ancient HBV strains, the results of the alignments were inspected visually with IGV version 2.3.92 (***Thorvaldsdóttir et al., 2013***). Information about contig order and direction were used for the construction of a consensus sequence. Bases that were soft clipped in the alignment were cut off using SeqKit software version 0.7.0 and realigned to the 16 references as described above. This was done because of the circular genome structure of HBV. Big contigs needed to be split to preserve genomic order with respect to the reference sequences (***Supplementary file 4***).

Remapping raw reads against the consensus sequence

Raw reads of each sample were mapped to their corresponding consensus sequence using the software CircularMapper version 1.93.4 and the following command line:

java -jar CircularGenerator.jar -e $E -i $IN -s '$N'

where $E is the length of elongation, $IN is the input file and $N is the name of the target sequence.

bwa aln -t 8 $IN $R -n 0.01-l 300-f $OUT

where $IN is the elongated consensus sequence, $R is the file containing the clipped and merged reads and $OUT is the output file.

bwa samse $RE $IN $R -f $OUT

where $RE is the elongated reference, $IN is the bwa aln output, $ R is the file containing the clipped and merged reads and $OUT is the output file.

java -jar realign-1.93.4.jar -e $E -i $IN -r $OR

where $E is the length of elongation, $IN is the output of bwa samse and $OR is the unmodified consensus.

Phylogenetic analysis

Hepatitis B reference strains for apes were collected using edirect with the following command:

esearch -db pubmed -query 'hepatitis B AND Orangutan OR hepatitis B AND Gibbon OR hepatitis B AND Gorilla OR hepatitis B AND Chimpanzee OR hepatitis B AND Orang-utan' | elink -target nuccore | efetch -format fasta > $OUT

where $OUT is the output file in fasta format containing all sequences from the papers containing the search keys.

To control the received sequences a multiple sequence alignment using the linsi algorithm contained in MAFFT version 7.310 was carried out. The following command was used:

linsi $IN > $OUT

where $IN is the input file containing the retrieved sequences and $OUT is the multiple sequence alignment.

The alignment was visually inspected in AliView (v. 1.18.1) and sequences that differed from the majority were removed. This step was necessary due to the unrestricted esearch command which, by

chance, could also return non-primate sequences. After filtering the set contained 74 ape infecting HBV strains.

Using the 74 ape strains and 5497 non-recombinant genomes available at hpvdb (https://hbvdb.ibcp.fr/HBVdb/HBVdbDataset?seqtype=0) clustering was carried out with UClust v 1.1.579 (**Edgar, 2010**). The clustering with an identity threshold of 97% yielded 493 representative HBV genomes. Combining them with the five ancient strains a multiple sequence alignment was carried out using Geneious version 10.1.2 (**Kearse et al., 2012**) with a 65% similarity cost matrix, a gap open penalty of 12 and a gap extension penalty of 3. The multiple sequence alignment was stripped of any sites (columns) that had gaps in more than 95% of sequences. The complete alignment including all modern and ancient genomes as multi-fasta is available in **Figure 2—source data 2**. The alignment was used to construct a network with the software SplitsTree v4 (**Huson and Bryant, 2006**), creating a NeighborNet (**Bryant and Moulton, 2004**) with uncorrected P distances.

The same multiple sequence alignment was used for the generation of Maximum-Likelihood (ML) and Neighbour-Joining (NJ) Trees. MEGA7 version 7170509-x86_64 with the following command line was used:

Megacc -a $MAO -d $IN -o $OUT

where $MAO is the megacc configuration file, $IN is the multiple alignment and $OUT is the output directory. For both trees 1408 informative sites and Jukes-Cantor substitution model were used. Bootstrap replicates are 2000 for ML and 10000 for NJ. The trees are provided in **Figure 2—source data 3** and **4**.

Molecular clock analysis

The evolution of hepatitis B virus over time is unclear with regard to its evolutionary rate and the role of recombination. Previous studies have attempted to detect a molecular clock-like signature without success. We investigate if the ancient genomes presented here allow a molecular clock analysis using TempEst v1.5.1 (**Rambaut et al., 2016**). The data set shows little positive correlation between genetic divergence and sampling time (correlation coefficient 0.075) and there is very little temporal signal (TempEst R2 = 0.006, see **Figure 2—figure supplement 9**). Therefore, we refrain from further dating analysis.

Recombination analysis

We performed recombination analysis using all modern full reference genomes (n = 493) and five ancient genomes used for the network analysis (see above in Phylogenetic analysis). The methods RDP, GENECOV, Chimera, MaxChi, BootScan, SiScan, 3Seq within RDP v4 (**Martin et al., 2015**) with a window size of 100 nt and the parameter set to circular genome with and without outgroup reference (results are provided in **Figure 2—source data 1**) and SimPlot v 3.5.1 (**Lole et al., 1999**, **Figure 2—figure supplements 5–8**) were applied to the data set.

LC-MS/MS analysis and database searches

Proteins were extracted from powdered tooth samples (50 mg) using a modified filter-aided sample preperation (FASP) protocol as previously described (**Warinner et al., 2014**; **Cappellini et al., 2014**). In-filter trypsin digested samples were analyzed on a Dionex Ultimate 3000 nano-HPLC coupled to a Q Exactive mass spectrometer (Thermo Scientific, Bremen). The samples were washed on a trap column (Acclaim Pepmap 100 C18, 10 mm ×300 µm, 3 µm, 100 Å, Dionex) for 5 min with 3% acetonitrile (ACN)/0.1% TFA at a flow rate of 30 µL/min prior to peptide separation on an Acclaim PepMap 100 C18 analytical column (15 cm ×75 µm, 3 µm, 100 Å, Dionex). A flow rate of 300 nL/min using eluent A (0.05% formic acid (FA)) and eluent B (80% ACN/0.04% FA) was used for gradient separation as follows: linear gradient 5 ± 50% B in 60 min, 50 ± 95% B in 5 min, 95% B for 10 min, 95 ± 5% B in 1 min, and equilibration at 5% B for 12 min. Spray voltage applied on a metal-coated PicoTip emitter (30 µm tip size, New Objective, Woburn, Massachusetts, US) was 1.25 kV, with a source temperature of 250°C. Full scan MS spectra were acquired from 5 to 145 min between 300 and 2,000 m/z at a resolution of 60,000 at m/z 400 (automatic gain control [AGC] target of 1E6; maximum ion injection time [IIT] of 500 ms). The five most intense precursors with charge states 2 + used were selected with an isolation window of 1.6 m/z and fragmented by HCD with normalized collision

energies of 25. The precursor mass tolerance was set to 10 ppm, and dynamic exclusion (30 s) was enabled.

Acquired spectra were analyzed by database searches using Proteome Discoverer (PD) 2.2.0.388 with the search engines SequestHT (Thermo Scientific). Searches were performed against a combined database built by the combination of the full Swiss protein database (468,716 entries, downloaded from Uniprot, December, 21st, 2017), hepatitis B data base (seven entries, downloaded from Uniprot, December, 7th, 2017) and common laboratory contaminants (115 entries, downloaded from Uniprot, August, 15th, 2014). The following settings were used for the search: semi-tryptic specificity; two missed cleavage sites; mass tolerances of 10 ppm for precursors and for fragment masses 0.02 Da (HCD) and 0.5 Da (CID); static modifications: carbamidomethylation on Cys; dynamic modifications: oxidation of Met, Lys and Pro. An additional search was performed using 12 FASTA files from in silico translated DNA sequences. The DNA sequences were obtained from previous DNA sequencing of the samples.

A nearly complete y-ion series and two b-ion fragments allow for an assignment of the full peptide sequence. The peptide was identified in the biological sample from Petersberg with four peptide spectral matches, showing that the detection of this peptide is not a random event. Moreover, the same peptide could also be identified in the second biological sample from Karsdorf (not shown); blank runs between the LC-MS/MS runs of the two samples rule out potential artifacts due to sample carryover.

Note that the MS/MS method applied here does not allow us to distinguish leucine (L) or isoleucine (I) residues. Manual permutation of the leucine residues in the above stated sequence followed by a BLAST search (default search parameters) led to the identification of the HBV-protein external core antigen in all cases with the exception of the combinations D**II**DTASALYR and DLLDTASA**I**YR; these two variants were reported by BLAST search as the proteins hypothetical protein CR988_04570 [*Treponema sp.*] and anti-GFP antibody [synthetic construct] with the HBV-protein external core antigen listed at rank 3. However, these proteins were not found in the genomic data. Hence, despite the uncertainty of the I/L assignment, the MS/MS data support the genomic finding of an HBV infection.

Human population genetic analyses

Mapping of the adapter-clipped and merged FASTQ files to the human reference genome hg19 was done using BWA (***Li and Durbin, 2010***) with the following command line:

bwa aln -n 0.01 l 300 $INDEX $FASTQCM $OUT

where $INDEX is the reference, $FASTQCM is the input file and $OUT is the output file. The duplicate removal after mapping was executed as described above.

The mapped sequencing data were transformed into the *Eigenstrat* format (***Price et al., 2006***) and merged with a dataset of 1,233,013 SNPs (***Haak et al., 2015***; ***Mathieson et al., 2015***). Using the software Smartpca (***Patterson et al., 2006***) the three samples and previously published ancient populations were projected onto a base map of genetic variation calculated from 32 West Eurasian populations (***Figure 1—figure supplements 4***, ***8*** and ***9***).

Sex determination

Sex was assessed based on the ratio of sequences aligning to the X and Y chromosomes compared to the autosomes (***Skoglund et al., 2013***).

Acknowledgements

We are grateful to the following people and institutions for providing samples, support, and advice: Bodo Krause-Kyora, Hildegard Nelson (Referat A1 Archäologische Dokumentation, Niedersächsisches Landesamt für Denkmalpflege), Ulrike Weller (Sammlungsverwaltung Archäologie Landesmuseum Hannover) and Britta Steer for technical assistance with proteomics sample preparation. This work was supported by the Collaborative Research Centre 1266 *Scales of Transformation*, the Excellence Cluster 306 *Inflammation at Interfaces*, the Medical Faculty of Kiel University, the Max Planck Society and the European Research Council (ERC) starting grant APGREID (to JK). Excavations and analysis of the archaeological site of Karsdorf were supported by the German Research Foundation (DFG) Grant of Kurt W Alt (Al 287-7-1) and Harald Meller (Me 3245/1–1). Analysis of the

archaeological site of Petersberg/Kleiner Madron was supported by a grant of the VolkswagenStiftung to Thomas Meier and by a PhD-Fellowship of the Ludwig-Maximilians-University, Munich to Sandra Lösch.

Additional information

Funding

Funder	Grant reference number	Author
Collaborative Research Center	1266	Ben Krause-Kyora Almut Nebel
Swiss National Science Foundation	PMPDP3_171320/1	Denise Kühnert
Deutsche Forschungsgemeinschaft	HE7987/1-1	Henrike O Heyne
Deutsche Forschungsgemeinschaft	Me 3245/1-1	Harald Meller
Deutsche Forschungsgemeinschaft	Al 287-7-1	Kurt W Alt
European Research Council	APGREID	Johannes Krause

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Ben Krause-Kyora, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Methodology, Writing—original draft, Project administration, Writing—review and editing; Julian Susat, Felix M Key, Alexander Herbig, Data curation, Formal analysis, Investigation, Writing—original draft, Writing—review and editing; Denise Kühnert, Esther Bosse, Formal analysis, Investigation; Alexander Immel, Formal analysis, Investigation, Writing—original draft; Christoph Rinne, Harald Meller, Susanne Friederich, Stefan Schreiber, Resources; Sabin-Christin Kornell, Investigation; Diego Yepes, Sören Franzenburg, Formal analysis; Henrike O Heyne, Formal analysis, Writing—original draft; Thomas Meier, Sandra Lösch, Resources, Investigation; Nicole Nicklisch, Resources, Formal analysis, Investigation; Kurt W Alt, Resources, Formal analysis; Andreas Tholey, Resources, Supervision, Methodology, Writing—original draft; Almut Nebel, Resources, Formal analysis, Funding acquisition, Writing—original draft, Writing—review and editing; Johannes Krause, Conceptualization, Resources, Data curation, Supervision, Funding acquisition, Methodology, Writing—original draft, Project administration, Writing—review and editing

Author ORCIDs

Ben Krause-Kyora http://orcid.org/0000-0001-9435-2872
Felix M Key http://orcid.org/0000-0003-2812-6636
Johannes Krause http://orcid.org/0000-0001-9144-3920

Ethics

Human subjects: The human remains are prehistoric European specimens, so consent was not required. No decedent groups claim responsibility or ancestry to those people.

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.36666.037
Author response https://doi.org/10.7554/eLife.36666.038

Additional files

Supplementary files

- Supplementary file 1. Accession numbers for the reference genomes used in the first alignment step to catch HBV diversity in the sample. Since monkey HBV strains are not classified into genotypes the column is left blank.

DOI: https://doi.org/10.7554/eLife.36666.027

- Supplementary file 2. Number of reads mapping against the references shown in ***Supplementary file 1*** before and after duplicate removal.

DOI: https://doi.org/10.7554/eLife.36666.028

- Supplementary file 3. Number of contigs and combined contig length of the de novo assembly for chosen K-values.

DOI: https://doi.org/10.7554/eLife.36666.029

- Supplementary file 4. Final consensus length after retrieving gap information from the multiple sequence alignment with Geneious.

DOI: https://doi.org/10.7554/eLife.36666.030

- Supplementary file 5. Number of reads mapping against hg19 before and after duplicate removal and percentage of the genome where coverage is at least one.

DOI: https://doi.org/10.7554/eLife.36666.031

- Supplementary file 6. Basic statistics for the mapping against the references shown in table S1. Shown are mean coverage, mean coverage for the covered region, genome length, number of missing bases and covered bases

DOI: https://doi.org/10.7554/eLife.36666.032

- Transparent reporting form

DOI: https://doi.org/10.7554/eLife.36666.033

Data availability

Raw sequence read files have been deposited at the European Nucleotide Archive under accession no. PRJEB24921

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Krause J	2018	High-throughput sequence data for Neolithic and Medieval virus genomes reveal complex evolution of Hepatitis B	https://www.ebi.ac.uk/ena/data/view/PRJEB24921	Publicly available at the European Nucleotide Archive (accession no. PRJEB24921)

References

Araujo NM. 2015. Hepatitis B virus intergenotypic recombinants worldwide: an overview. *Infection, Genetics and Evolution* **36**:500–510. DOI: https://doi.org/10.1016/j.meegid.2015.08.024

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**:455–477. DOI: https://doi.org/10.1089/cmb.2012.0021, PMID: 22506599

Behnke HJ. 2007. Süßes aus der Bronzezeit – Siedlungsgunst im Unstruttal bei Karsdorf, Burgenlandkreis. Die Ausgrabungen 2004. In: *Archäologie in Sachsen-Anhalt N.F.* **4** p. 63–86.

Behnke HJ. 2011. Erste Siedler der Linienbandkeramik in der Karsdorfer Feldflur. Ergebnisse der Ausgrabungen im Jahr 2005. In: *Archäologie in Sachsen-Anhalt N.F.* **5** p. 184–199.

Behnke HJ. 2012. Siedlungsgunst im Unstruttal bei Karsdorf, Burgenlandkreis: Ergebnisse der Grabungen 2006 und 2007. In: *Archäologie in Sachsen-Anhalt N. F.* **6** p. 35–70.

Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, Sell C, Fajkosová Z, Powell A, Burger J. 2013. 2000 years of parallel societies in stone age central Europe. *Science* **342**:479–481. DOI: https://doi.org/10.1126/science.1245049, PMID: 24114781

Bouckaert R, Alvarado-Mora MV, Pinho JR. 2013. Evolutionary rates and HBV: issues of rate estimation with Bayesian molecular methods. *Antiviral Therapy* **18**:497–503. DOI: https://doi.org/10.3851/IMP2656, PMID: 23792904

Brandt G, Knipper C, Nicklisch N, Ganslmeier R, Klamm M, Alt KW. 2014. Settlement burials at the Karsdorf LBK site, Saxony-Anhalt, Germany - biological ties and residential mobility. In: Whittle A, Bickle Penny (Eds). *Early Farmers. The View from Archaeology and Science*. Oxford University Press, British Academy. p. 95–114.

Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* **21**:255–265. DOI: https://doi.org/10.1093/molbev/msh018, PMID: 14660700

Cappellini E, Gentry A, Palkopoulou E, Ishida Y, Cram D, Roos A-M, Watson M, Johansson US, Fernholm B, Agnelli P, Barbagli F, Littlewood DTJ, Kelstrup CD, Olsen JV, Lister AM, Roca AL, Dalén L, Gilbert MTP. 2014. Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae). *Zoological Journal of the Linnean Society* **170**:222–232. DOI: https://doi.org/10.1111/zoj.12084

Castelhano N, Araujo NM, Arenas M. 2017. Heterogeneous recombination among Hepatitis B virus genotypes. *Infection, Genetics and Evolution* **54**:486–490. DOI: https://doi.org/10.1016/j.meegid.2017.08.015, PMID: 28827173

Claus M. 1983. Das neolithische Felsenkammergrab auf dem Halsberg bei Sorsum, Stadt Hildesheim. *Die Kunde N. F.* :91–122.

Czarnetzki A. 1966. *Die menschlichen Skelettreste aus vier neolithischen Steinkisten Hessens und Niedersachsens*. Tübingen: Ungedr. Diss. Univ. Tübingen.

Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, Meyer M. 2013. Complete mitochondrial genome sequence of a middle pleistocene cave bear reconstructed from ultrashort DNA fragments. *PNAS* **110**:15758–15763. DOI: https://doi.org/10.1073/pnas.1314445110, PMID: 24019490

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:2460–2461. DOI: https://doi.org/10.1093/bioinformatics/btq461, PMID: 20709691

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, et al. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**:207–211. DOI: https://doi.org/10.1038/nature14317, PMID: 25731166

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**:254–267. DOI: https://doi.org/10.1093/molbev/msj030, PMID: 16221896

Kahila Bar-Gal G, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, Kim TH, Kim SB, Grant PR, Pappo O, Spigelman M, Shouval D. 2012. Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* **56**:1671–1680. DOI: https://doi.org/10.1002/hep.25852, PMID: 22610996

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649. DOI: https://doi.org/10.1093/bioinformatics/bts199, PMID: 22543367

Key FM, Posth C, Krause J, Herbig A, Bos KI. 2017. Mining metagenomic data sets for ancient DNA: recommended protocols for authentication. *Trends in Genetics* **33**:508–520. DOI: https://doi.org/10.1016/j.tig.2017.05.005, PMID: 28688671

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**:409–413. DOI: https://doi.org/10.1038/nature13673, PMID: 25230663

Li H, Durbin R. 2010. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* **26**:589–595. DOI: https://doi.org/10.1093/bioinformatics/btp698, PMID: 20080505

Littlejohn M, Locarnini S, Yuen L. 2016. Origins and evolution of hepatitis B virus and hepatitis D virus. *Cold Spring Harbor Perspectives in Medicine* **6**:a021360. DOI: https://doi.org/10.1101/cshperspect.a021360, PMID: 26729756

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology* **73**:152–160. PMID: 9847317

Lösch S. 2009. Paläopathologisch-anthropologische und molekulare Untersuchungen an mittelalterlichen und frühneuzeitlichen Bevölkerungsgruppen. Dissertation zum Erwerb des Doktorgrades der Humanbiologie an der Medizinischen Fakultät der Ludwig-Maximilians-Universität zu München. Medizinischen Fakultät der Universität München.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**:vev003. DOI: https://doi.org/10.1093/ve/vev003, PMID: 27774277

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JM, Carbonell E, Gerritsen F, et al. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**:499–503. DOI: https://doi.org/10.1038/nature16152, PMID: 26595274

Meier T. 1998. Ein mittelalterliches Kloster mit Friedhof auf dem Kleinen Madron/Petersberg bei Flintsbach a. *Inn. Landkreis Rosenheim, Oberbayern*. In: *Das Archäologische Jahr in Bayern*. p. 127–129.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**:pdb.prot5448. DOI: https://doi.org/10.1101/pdb.prot5448, PMID: 20516186

Nicklisch N. 2017. Spurensuche an Skeletten. Paläodemografische und epidemiologische Untersuchungen an neolithischen und frühbronzezeitlichen Bestattungen aus dem Mittelelbe-Saale-Gebiet im Kontext populationsdynamischer Prozesse. In: *Forschungsberichte des Landesmuseums für Vorgeschichte Halle*. **11**

Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genetics* **2**:e190. DOI: https://doi.org/10.1371/journal.pgen.0020190, PMID: 17194218

Patterson Ross Z, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, Poinar D, Douglas MW, Eden JS, Holmes EC, Poinar HN. 2018. The paradox of HBV evolution as revealed from a 16th century mummy. *PLoS Pathogens* **14**:e1006750. DOI: https://doi.org/10.1371/journal.ppat.1006750, PMID: 29300782

Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. 2016. EAGER: efficient ancient genome reconstruction. *Genome Biology* **17**:60. DOI: https://doi.org/10.1186/s13059-016-0918-z, PMID: 27036623

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**:904–909. DOI: https://doi.org/10.1038/ng1847, PMID: 16862161

Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2**:vew007. DOI: https://doi.org/10.1093/ve/vew007

Rasche A, Souza B, Drexler JF. 2016. Bat hepadnaviruses and the origins of primate hepatitis B viruses. *Current Opinion in Virology* **16**:86–94. DOI: https://doi.org/10.1016/j.coviro.2016.01.015, PMID: 26897577

Schaefer S. 2007. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World Journal of Gastroenterology* **13**:14–21. DOI: https://doi.org/10.3748/wjg.v13.i1.14, PMID: 17206751

Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, Herbig A, Economou C, Benjak A, Busso P, Nebel A, Boldsen JL, Kjellström A, Wu H, Stewart GR, Taylor GM, Bauer P, Lee OY, Wu HH, Minnikin DE, et al. 2013. Genome-wide comparison of medieval and modern Mycobacterium leprae. *Science* **341**:179–183. DOI: https://doi.org/10.1126/science.1238286, PMID: 23765279

Simmonds P, Midgley S. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. *Journal of Virology* **79**:15467–15476. DOI: https://doi.org/10.1128/JVI.79.24.15467-15476.2005, PMID: 16306618

Skoglund P, Storå J, Götherström A, Jakobsson M. 2013. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science* **40**:4477–4482. DOI: https://doi.org/10.1016/j.jas.2013.07.004

Souza BF, Drexler JF, Lima RS, Rosário MO, Netto EM. 2014. Theories about evolutionary origins of human hepatitis B virus in primates and humans. *The Brazilian Journal of Infectious Diseases* **18**:535–543.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**:178–192. DOI: https://doi.org/10.1093/bib/bbs017, PMID: 22517427

Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, Spyrou MA, Andrades Valtueña A, Huson D, Tuross N, Bos KI, Krause J. 2018. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology & Evolution* **2**:520–528. DOI: https://doi.org/10.1038/s41559-017-0446-6, PMID: 29335577

Veit U. 1996. *Studien zum Problem der Siedlungsbestattungen im europäischen Neolithikum, Münster*. New York: Waxmann.

Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, Burbano HA, Orlando L, Krause J. 2017. A robust framework for microbial archaeology. *Annual Review of Genomics and Human Genetics* **18**:321–356. DOI: https://doi.org/10.1146/annurev-genom-091416-035526, PMID: 28460196

Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S, Speller C, Hendy J, Charlton S, Luder HU, Salazar-García DC, Eppler E, Seiler R, Hansen LH, Castruita JA, Barkow-Oesterreicher S, et al. 2014. Pathogens and host immunity in the ancient human oral cavity. *Nature Genetics* **46**:336–344. DOI: https://doi.org/10.1038/ng.2906, PMID: 24562188

WHO. 2017. *Global Hepatitis Report*.

Zhou Y, Holmes EC. 2007. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *Journal of Molecular Evolution* **65**:197–205. DOI: https://doi.org/10.1007/s00239-007-0054-1, PMID: 17684696

5.

Chapter IV: Mass child burials in early Berlin. Victims of pathogens or general harsh circumstances?

This chapter was prepared as part of a bigger project investigating the burial site at the Petriplatz in Berlin

Supplementary material can be found at the end of this thesis

Title:

Mass child burials in early Berlin. Victims of pathogens or general harsh circumstances?

Author:

Julian Susat

Introduction

Despite being one of the youngest capitals in Europe, Berlin has a vivid architectural and cultural history. In some cases, real estate development and constructional work led to discovery and excavation of former parts of the city. In early 2007, restructuring of the area around Scharrenstraße/Am Petriplatz in Berlin-Mitte revealed vestiges of a church and its adjoining graveyard. A large-scale excavation of the site enabled identification of the church remnants to be those of the St.-Petri-Church, which in its first form, was mentioned in 1285 (1). During the course of its existence the church was rebuilt five times due to repeated fires and bad constructional work. In its last iteration, the church was deconstructed after the Second World War as a result of severe weapon damage. The associated graveyard was closed in 1717 which was a common practice for burial grounds in the city at that time (2). Skeletal remains of 3718 individuals buried in 3123 graves were uncovered during the excavation (3). Ninety percent of the human remains were excavated from single burials, which points to ordinary circumstances of death. However, several graves with multiple individuals were also discovered (2-12 individuals). There was no evidence of uncareful and chaotic burial practices that are often the case during pandemics, war and natural disasters (4). Nevertheless, a relatively big proportion of juvenile individuals was observed among the buried, which could be due to general harsh conditions in the Middle Ages that fostered the spread of infectious diseases among the population (5). Combined with malnutrition and physical labor, these circumstances led to the general high mortality rate (50%) of children in the Middle Ages (6).

To further investigate the mass burials, skeletal remains of 32 juvenile individuals from 8 burials were screened for the presence of bacterial and viral pathogens. As there was no initial expectation of a particular pathogen based on historical and osteological evidence, the screening was not restricted to a specific organism. In the examined samples traces of the Hepatitis B virus (HBV), Parvovirus B19 (B19) and *Yersinia pestis* (*Y. pestis*) were detected.

Results

Archaeological dating

Contextual dating revealed that the graves were dug during the early times of the city Berlin (Range 1250-1549, Table 1 and supplementary table 1).

Table 1. Graves and dating

Grave	Number of individuals	Number of infected individuals	Detected pathogens	Dating
1	3	1	B19	1550-1599
2	7	4	B19, HBV and *Y. pestis*	1500-1549
3	5	4	B19 and HBV	1400-1449
4	2	1	B19	1250-1299
5	3	3	B19 and HBV	1250-1299
6	5	3	B19 and HBV	1500-1549
7	4	3	B19	1350-1399
8	3	1	B19	1400-1499

Pathogen screening

Thirty-two juvenile individuals from eight graves where screened for the presence of viral and bacterial pathogens. The screening revealed the presence of HBV (n=6) in four graves, Parvovirus B19 (n=16) in eight graves and *Y. pestis* (n=1) in one grave of the sampled population (Table 1 and supplementary table 1).

Hepatitis B virus

Samples that were HBV positive in the screening were subjected to a virus capture. Subsequently the combined reads (initial sequencing and captured data) were mapped against a multi fasta reference containing modern, ancient and non-human primate HBV strains (Supplementary table S2). The mapping reads were extracted, and a de novo assembly led to the reconstruction of seven genomes (Table 2 and supplementary table S3). Signs of an infection with two different HBV strains were observed for one individual (3129B). The reconstructed HBV genomes were genotyped using two different methods which showed congruent results for all typed strains and the genotypes G, D and A were noted. Double infection of the 3129B individual consisted of genotypes G and D. Completeness of

the reconstructed genomes ranged from 41% in 3009G (excluding N's in the sequence) up to 99% in 2993B (Table 2).

Table 2. Ancient HBV and B19 genome properties

HBV							
Sample	Genotype of the closest modern reference	Sequence ID to the closest modern reference	Number of N's in Sequence	Genome length	Possible serotype	Insertions/deletions	Predicted HBeAG Status
2993B	A2	98.5%	4	3221	adw2	6-nt in C; 33nt in PreS1/-	positive
3129B	G	55.7%	870	3215	adw2	36-nt in PreC; 30nt in PreS1/-	negative
3129B	D1	89.8%	153	3182	ayw2	-/33-nt	positive
3009E	D3	82.7%	291	3182	ayw2/ayw3	-/33-nt	positive
3009B	D1	68.1%	548	3091	ayw2	-/33-nt	positive
2884A	D3	97%	65	3182	ayw2/ayw3	-/33-nt	positive
3009G	D3	28.4%	1437	ambigous	ayw2/ayw3	-/33-nt	positive
B19					NS1 and VP1/2 status		
3129C	2	68.7%	1383	5005	Incomplete		
3060C	2	94.6%	128	4762	complete		

Each strain depicted the expected pattern of insertions and deletions with respect to its genotype (7). HBeAg (HBe antigen) status was positive for all strains except for the G strain from the double infection. Traces of all the expected HBV ORF's could be found in nearly all the strains. Both strains from the double infection exhibited no signs of the Polymerase, while the strain 3009G lacked evidence for the presence of the X and the Core genes. In a phylogenetic Network, the strains from Berlin clustered together with the predicted genotypes (Figure 1).

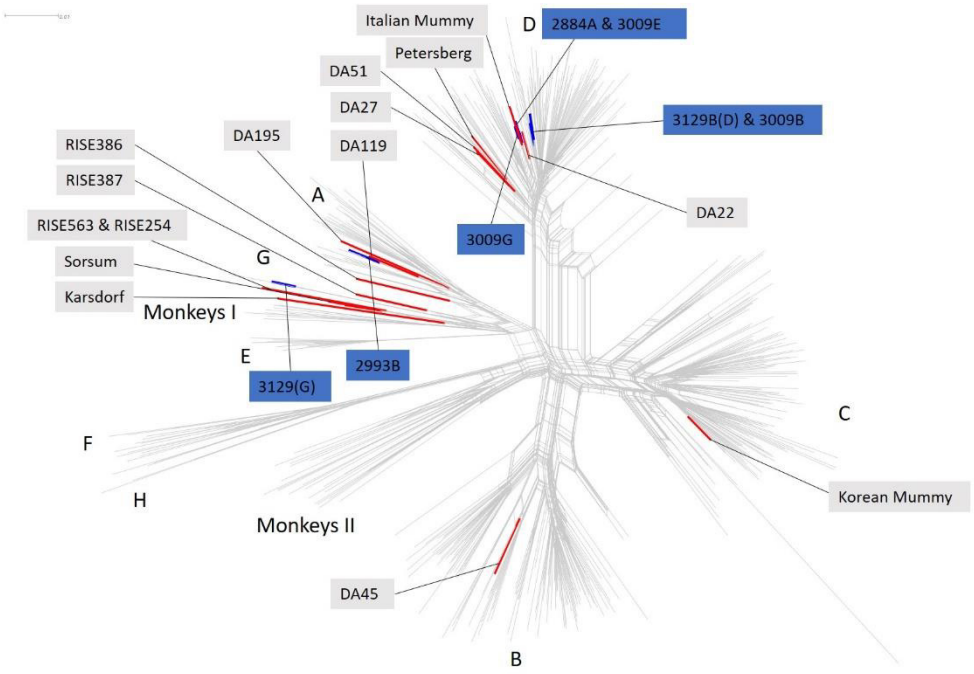


Figure 1. Phylogenetic network of 493 representative modern and 23 ancient HBV genomes. Depicted distance is uncorrected P. Published ancient genomes are highlighted in red. Strains from the Petriplatz are highlighted in blue and the label is also blue. The different clades of HBV are indicated by the capital letters. Monkeys I include strains from gibbons and orangutans, Monkeys II strains from gorillas and chimpanzees.

Parvovirus B19

Samples that were B19 positive were also subjected to a virus capture. After capture the samples were mapped against a multi fasta file representing the modern and known ancient diversity of B19 (Supplementary table 4). Due to insufficient coverage recovery of the B19 genomes was only possible for two out of 16 positive samples (3129C and 3060C, supplementary table 5). Genotyping based on dataset 1 revealed the closest relationship to genotype 2 (See additional supplementary information, dataset 1). Whilst the genome of the 3060C strain was fully recovered with complete NS1 and VP1/2 ORF's, the 3129C strain genome did not reach that level of quality exhibiting only short fragments of NS1 and VP1/2 (Table 2). A Maximum likelihood phylogeny based on 77 modern, 10 published (dataset 2) ancient sequences and the two strains from Berlin was constructed (See additional supplementary information, dataset 2). Both strains clustered together in their own branch which is basal to the VK strains from Greenland and Sweden, the RISE569 from Czech Republic and the modern strains belonging to genotype 2.

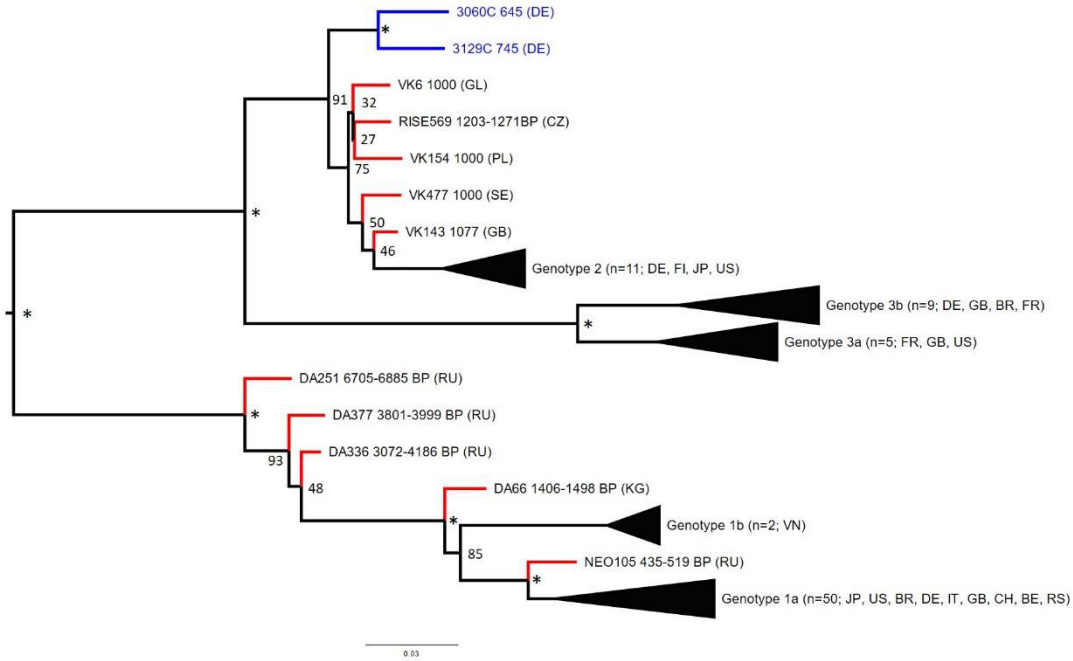


Figure 2. Maximum-likelihood tree of 77 modern and 12 ancient B19 genomes. Country abbreviation is given in brackets (DE = Germany, GL = Greenland, CZ = Czechia, PL = Poland, SE = Sweden, GB = Great Britain, FL = Finland, JP = Japan, US = United States, BR = Brazil, FR = France, RU = Russia, KG = Kyrgyzstan, VN = Vietnam, IT = Italy, BE = Belgium, RS = Serbia). Bootstrap values are shown for 500 replicates. An asterisk (*) indicates bootstrap support above 94%. Branches of published ancient genomes are highlighted in red. Strains from the Petriplatz are highlighted in blue. For strains with C14-dating the date is given in BP. Strains without BP just depict the age of the sample based on e. g. contextual dating.

Yersinia pestis

Reconstruction of the *Y. pestis* genome revealed insufficient coverage of the reference genome (>=1x 43.22%; >=2x 39.15%; >=3x 13.05%). The SNP based phylogeny depicted a low bootstrap support of 34 and severe lengthening of the branch. Based on these results no further analyses were conducted.

Discussion

Several mass burials with a large proportion of juvenile individuals were uncovered from the medieval St.-Petri-Church cemetery in Berlin. Osteological analysis did not provide clear evidence of any infectious disease. Thus, a subset of 32 subadult individuals from 8 mass graves was screened for the presence of bacterial and viral pathogens. Six of the individuals showed molecular evidence of an infection with HBV. Evidence for B19 virus infection was found in 16 samples, one of which was also positive for *Y.pestis*.

HBV is a partially double stranded DNA virus and the causing agent of hepatitis B. This is one of the most widespread viral infectious diseases persisting in about 260M people each year (8). Mild symptoms of the HBV infection include general ill-health, loss of appetite, nausea, vomiting, body aches and fever. Nevertheless, a chronic state of hepatitis B is regarded as a strong risk factor for the development of the hepatocellular carcinoma and the liver cirrhosis (9). There is evidence that the HBV has been infecting humans for as long as 7000 years and its genome shows a high degree of conservation through time (10,11). The main modes of transmission include perinatal infection, contact with contaminated body fluids and sexual contact (12). The six infected individuals represent nearly 19% of the tested population. Nowadays, an infection rate of above 8% of the population is considered to be “high prevalence” (13). With respect to the general living conditions in the Middle Ages, this number might be realistic. However, it must be treated with caution as the analyzed samples represent only a fraction of the whole population. As the HBV was detected in juvenile individuals, a vertical mode of infection seems to be the most likely. Infants born by a HBV infected mother that is positive for HBeAg, the Hepatitis B e antigen, have a 70 to 90% chance of being infected (14–16). Furthermore, it was shown that infection during childbirth leads to a chronic state in approximately 90% of cases (14–17).

HBV exhibits a clear phylogeographic distribution pattern, which was also observed for the HBV genomes (genotypes A, G and D) recovered in this study (Figure 1)(18). Nowadays, genotypes D and A are common in Europe, whereas genotype G is prevalent in France, Germany and the United States and is only responsible for a small number of infections (19). It is possible that in the modern distribution pattern differs from the pattern present in the Middle Ages. However, the presence of recovered genotypes in medieval Berlin is not unlikely.

The double infection of individual 3129B with the genotypes D and G is an interesting finding. Due to a 12 amino acid insertion in the Core protein the replication of genotype G is limited. Thus, mono-infections with genotype G are uncommon and a so called “helper virus” is needed (20–23). Double infections including this genotype are mostly seen in combination with genotype A, C, D, H and F (24–29). Interestingly infections with genotype G are strongly associated with certain risk groups like intravenous drug users and immunocompromised patients (24). Infections with the HBV in general are more frequent in immune deficient populations that show lower rates of spontaneous recovery (30). This might support idea of a general bad health status among the buried children. All the strains examined in this study displayed high conservation with the expected genomic order of the HBV. Several reconstructed strains lacked particular genes and gene regions. It is, however, most likely due to the nature of ancient DNA and the relatively bad preservation of the genetic material.

Parvovirus B19 is a linear single stranded DNA virus which is responsible for several diseases, such as arthropathy, transient aplastic crisis, hydrops fetalis/fetal loss and erythema infectiosum (Fifth disease) (31). B19 possesses a high level of infectivity with at least 40% of the worldwide adult population exhibiting evidence of a past infection (32,33). Recent findings of ancient B19 sequences, as old as 6000 years, revealed its long history with humans (34). Initial symptoms of an infection (fever, malaise and headache) are unspecific (35). Transmission of B19 occurs via the respiratory route, through contact with blood-derived products and from mother to fetus (36). If the mother is infected there is a 33% chance that the child might get infected too (37). Under normal circumstances, the infection is cleared by the immune response within two weeks, providing a lifelong protection against the virus. However, in some cases (for instance in an immunocompromised host), the infection can lead to a transient aplastic crisis which might enable a life-threatening condition (38). The 16 children represent 50% of the population, which is comparable to the seroprevalence of 38.1%-72.7% observed in modern German individuals of six to ten years of age (39). In comparison to HBV, B19 was more abundant with at least one infected individual for each of the screened graves. The placement of the two strains in the phylogeny forming their own distinct clade is noteworthy. However, as 3129C is of questionable quality and 3060C has a 94.6% sequence identity to a genotype 2 strain it is safe to say that at least 3060C is belonging to genotype 2.

In contrast to HBV, the distribution of B19 genotypes is not spatially and temporally uniform. Nowadays genotype 2 is mainly found in elderly adults in northern Europe (40). This phenomenon was attributed to the idea that genotype 2 was replaced by genotype 1 which originated in the second half of the 20th century (41). However, the detection of ancient B19 sequences clearly belonging to genotype 1 rendered this hypothesis obsolete (34). The discovery of ancient genotype 2 sequences as old as 1300 years is supporting the plausibility of the B19 genomes from Berlin (34). Unfortunately, only two B19 genomes could be recovered and the diversity of B19 genotypes in the mass graves might be bigger.

Traces *of Y. pestis* was found in one of the examined samples. *Yersinia pestis* is the causative agent of plague best known for its role in the Black Death which ravaged Europe in the 14th century killing approximately 50% of its population (42). Unfortunately, the sample did not provide enough material for a sufficient in-depth analysis of the strain and a reliable phylogeny. The remains of the infected individual were dated to 1500-1549 - nearly 200 years after the Black Death. However, *Y. pestis* was still a common pathogen in central Europe at that time (43–46). There is historical evidence that the inhabitants of Berlin suffered from plague outbreaks in 1502 and 1516, which is in concordance with the dating of the graves and would explain the presence of *Y. pestis* in Berlin in that time period (47–49).

Unfortunately, the limited number of samples did not allow for reliable statistics. However, based on the observed numbers, some speculations can be made. Skeletal remains of seven B19-positive and four HBV-positive individuals were dated to the 16th century. This could suggest a higher rate of infection in the 16th century compared to the timeframes before, which might be a result of a bigger population density in the 16th century. Population of Berlin increased from approximately 2000 to 11.000 residents between 1250 and 1576, which depicts an increase by nearly 500% (50,51). Moreover, the number of residential buildings increased from 220 to 1070 between 1250 and 1450 (52). Whilst these numbers are just estimations, they provide a general idea about the growing population of Berlin, which might have been a fertile ground for the propagation and spread of pathogens like HBV, B19 and *Y. pestis.*

We discovered the presence of several pathogens in the remains of 20 out of 32 juvenile individuals buried in the mass graves dated to the 13th -16th centuries. The general circumstances during early days of Berlin might have fostered the spread of disease and

therefore the high rate of infected individuals is no surprise. It is within the range of plausible to assume that these infections have contributed to the early death of these individuals and the presence of the mass burials. As it is not clear if the infected suffered from the disease other explanations, such as malnutrition, must be considered. More individuals from the St.-Petri-Church graveyard should be examined in order to better estimate viral and bacterial loads in medieval Berlin and medieval populations in general.

Material and Methods

DNA isolation, sequencing and postprocessing

All lab work was carried out in a dedicated ancient DNA facility. DNA extraction of the samples from 32 individuals from Berlin, sequencing (Illumina HiSeq4000 (2 × 75 bp)) and subsequent clip and merge was performed as described in Krause-Kyora et al. (53). Reads that were shorter than 25 bp were discarded after merging.

Bacteria and Virus screening

Bacterial screening was carried out with the software MALT 0.4.1 using a custom database only consisting of compete bacterial genomes downloaded from NCBI (24.01.2019)(54). A sequence identity threshold of 90% was set and the alignment mode was changed to SemiGlobal. The analysis was carried out using the following command:

malt-run --mode BlastN -e 0.001 -id 90 --alignmentType SemiGlobal --index $index --inFile $FASTQCM --output $OUT

where $index is the index file, $FASTQCM is the clipped and merged file and $OUT is the output the file. Virus screening was also performed using MALT ant the ncbi-viral database as a reference. The sequence identity threshold was changed to 85%. The resulting alignments were visually inspected using MEGAN 6 (55).

Yersinia pestis alignment

The Y.pestis positive sample was mapped against the Yersinia pestis reference genome including the three plasmids (NC_0031431.1; NC_003131.1; NC_003134.1; NC_003132.1). The mapping was carried out using BWA and Bowtie2 with the following command lines (56,57):

bwa aln -n 0.01 -l 300 $INDEX $FASTQCM $OUT

bowtie2 --mp 1,1 --ignore-quals --score-min L,0,-0.05 -x $INDEX –r $FASTQCM

where $INDEX is the reference, $FASTQCM is the input file and $OUT is the output file. Minimum mapping quality was set to 0.

HBV alignment

HBV positive samples were aligned against a dataset representing the modern and known ancient diversity of HBV. The dataset comprised of one modern reference for each of the eight hepatitis B genotypes available in the NCBI hepatitis B genotyping project (https://www.ncbi.nlm.nih.gov/projects/genotyping/view.cgi?db=2), 16 ancient human strains, and eight modern monkey strains (Supplementary table 2). All references were combined in one FASTA file and a mapping was performed as described above.

B19 alignment

Parvovirus B19 positive samples were aligned against a dataset representing the modern and known ancient diversity of B19 diversity. The dataset comprised of: Three modern Genotype 1 strains, three modern Genotype 2 strains, two modern Genotype 3 strains, one Bovine Parvovirus Strain, five ancient Genotype 1 strains and five ancient Genotype 2 strains (Supplementary table 4). All references were combined in one FASTA file and mapping was performed as described above.

Sample concatenation

BAM files from different samples and mapping approaches but belonging to one individual were merged using SAMtools merge (58).

Duplicate removal

Duplicates were identified and removed using the software DeDup version 0.11.3 with the default options and the following command (59):

```
java -jar DeDup.jar -i $IN -o $OUT
```

where $IN is the input BAM file and $OUT is the output BAM file.

Extracting mapped reads

After duplicate removal the resulting BAM files were converted to SAM files using SAMtools with default parameters and the following command:

samtools view -h -o $OUT $IN

where $OUT is the SAM output file and $IN is the BAM input file. Reads from the SAM file where converted to FASTQ using the following awk script:

awk '/^[FMR]/{ print “@”$1”\n”$10”\n+\n”$11 }' $IN > $OUT

where $IN is a SAM file and $OUT is the resulting FASTQ file containing all the mapped reads.

<u>De novo assembly</u>

The de novo assembly for B19 and HBP positive samples was performed using the SPAdes genome assembler version v3.9.0 with the following settings (60):

spades.py -t 20 -m 500 -k 11,13,15,17,19,21,23,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,59,61,63,65,67,69,71,73,75,77,79,81,83,85,87,89,91,93,95,97,99,101,103,105,107,109,111,113,115,117,119,121,123,125,127 -s $IN -o $OUT

where $IN is a FASTQ file containing the mapped reads and $OUT is the output folder for SPAdes.

Resulting contigs for each K-value were checked and the one which spawned the longest contig was selected for further processing (Supplementary table 6).

<u>Mapping of contigs HBV/B19</u>

Contigs were mapped against the respective multi FASTA file containing all the references mentioned above. The following command was used:

bwa mem $INDEX $IN $OUT

where $INDEX is the reference, $IN the file containing the contig/contigs and $OUT is the resulting BAM file.

Consensus generation

For genomic reconstruction the results of the remapping were visually inspected in IGV (61). A custom awk script taking into account mapping information for each contig was used to generate consensus sequences for B19 and HBV genomes. Subsequent curation of the resulting genomes was done manually.

Remapping raw reads against the consensus sequence

Raw reads of each sample were mapped to their corresponding consensus sequence using the software CircularMapper and the following command line (59):

```
java -jar CircularGenerator.jar -e $E -i $IN -s "$N"
```

where $E is the length of elongation, $IN is the input file and $N is the name of the target sequence.

```
bwa aln -t 8 $IN $R -n 0.01 -l 300 -f $OUT
```

where $IN is the elongated consensus sequence, $R is the file containing the clipped and merged reads and $OUT is the output file.

```
bwa samse $RE $IN $R -f $OUT
```

where $RE is the elongated reference, $IN is the bwa aln output, $ R is the file containing the clipped and merged reads and $OUT is the output file.

```
java -jar realign-1.93.4.jar -e $E -i $IN -r $OR
```

where $E is the length of elongation, $IN is the output of bwa samse and $OR is the unmodified consensus.

Phylogenetic analysis of *Y.pestis*

A VFC file was for the positive samples using GATK version and the Y. pestis reference with the following command line (62):

GenomeAnalysisTK.jar-T UnifiedGenotyper -R $REF -I $IN -out_mode EMIT_ALL_SITES -o $OUT

where $REF is the reference, $IN is the input BAM file and $OUT is the VCF output. These to VCF's together with a VCF dataset of 249 strains (Table S6, 7., generated as described above) was used as an input for the MultiVCFAnalyzer to generate a SNP based multiple alignment. The analysis was carried out with the following command (63):

Java –jar MultiVCFAnalyzer_0-85-1.jar NA $REF NA $OUT F 30 3 0.9 0.9 NA

Where $REF is the Y. pestis reference and $OUT is the output folder. Y. pseudotuberculosis specific variants were removed from the SNP table to improve visual resolution of the tree. The SNP table was reformatted to multi fasta and nexus format which as used as Input for RAxML (64). RAxML was executed as described in Susat et al. 2020 (65).

Phylogenetic analysis and assignment of HBV

Phylogenetic assignment of the newly found ancient HBV strains was done using three different approaches.

- Online assignment
  The seven genomes were analyzed using https://hbv.geno2pheno.org/ with default parameters

- Needleman-Wunsch Algorithm
  5497 modern HBV reference genomes containing Genotypes (A-H) were downloaded from https://hbvdb.ibcp.fr/HBVdb/HBVdbDataset?seqtype=0. Redundant sequences were removed using usearch version with the following command (66):

usearch –cluster_fast $INPUT –id 0.99 - centroids $OUTPUT

Where $INPUT is the in file containing the 5497 reference genomes and $OUTPUT is the output file. The clustering resulted in 2339 remaining sequences (this dataset can be found on the attached CD. Filename: HBV_needle.fasta). Using the needleall command from the EMBOSS package each ancient genome was aligned to each modern genome in a pairwise manner (67). The alignment was carried out with the following command and default parameters:

Needleall –asequence $GENOME –bsequnce $REFERENCES

Where $GENOME is the ancient HBV genome and $REFERENCES is the file containing the centroid sequences. Similarity to the closest relative was identified using awk.

- Neighbor Net

Hepatitis B reference strains for monkeys were collected using edirect with the following command (68):

esearch -db pubmed -query "hepatitis B AND Orangutan OR hepatitis B AND Gibbon OR hepatitis B AND Gorilla OR hepatitis B AND Chimpanzee OR hepatitis B AND Orang-utan" | elink -target nuccore | efetch -format fasta > $OUT

where $OUT is the output file in fasta format containing all sequences from the papers containing the search keys.

To control the received sequences a multiple sequence alignment using the linsi algorithm contained in MAFFT was carried out (69). The following command was used:

linsi $IN > $OUT

where $IN is the input file containing the retrieved sequences and $OUT is the multiple sequence alignment.

The alignment was visually inspected in AliView and sequences that differed from the majority were removed (70). This step was necessary due to the unrestricted research command which, by chance, could also return non-primate sequences. After filtering the set contained 74 ape infecting HBV strains. Using the 74 ape strains and 5497 clustering was carried out using usearch with an identity threshold of 97% percent. This clustering yielded 493 representative HBV genomes. 23 ancient published HBV genomes were added to this representative dataset (this dataset excluding the HBV sequences from Berlin can be found on the attached CD. Filename: HBV_net.fasta). A multiple sequence alignment using MAFFT was carried out as described above and phylogentic informative blocks were extracted using GBLOCKS with default parameters (71). The resulting alignment was used to construct a network with the software SplitsTree, creating a NeighborNet with uncorrected P distances (72).

Phylogenetic analysis and assignment of Parvovirus B19

Phylogenetic assignment of the newly found ancient HBV strains was done using two different approaches.

- Needleman-Wunsch Algorithm

Dataset 1 was used and needleall was executed as described above (Supplementary additional information).

- Phylogenetic trees

Phylogenetic Analysis of B19 was performed using dataset 2 (Supplementary additional information) and the ancient strains published in Mühlemann et al. 2018 (34). A MAFFT alignment followed by GBLOCKS as described above was performed. A maximum likelihood phylogeny was generated using PhyML with the TN93 substitution model, estimated proportion of invariable sites, SPR tree topology search and 1000 bootstrap replicates (73).

1. Die Vielgefprüfte Petrikirche. Vossische Zeitung [Internet]. 1928; Available from: https://archive.vn/20150222232222/http://zefys.staatsbibliothek-berlin.de/dfg-viewer/?set[image]=5&set[zoom]=default&set[debug]=0&set[double]=0&set[mets]=http://zefys.staatsbibliothek-berlin.de/oai/?tx_zefysoai_pi1%255Bidentifier%255D%3D6515195c-283e-48d6
2. Berliner geschriebene Zeitungen aus den jahren 1713 bis 1717 und 1735. Ein Beitrag zur Preußischen Geschichte unter König Friedrich Wilhelm I. Schriften des Vereins für die Geschichte Berlins. 1902;(38).
3. Melisch CM. Die Bedeutung des Petriplatzes aus archäologischer Sicht. Bericht zu den Ausgrabungen 2007-2009. In: Der Bär von Berlin. Berlin: Verein für die Geschichte Berlins; 2010. p. 9–24.
4. Lange H. Ausgrabungen am ehemaligen Heiliggeist-Hospital an der Spandauer Straße, Berlin-Mitte. In: Großstadtdenkmalpflege Erfahrungen und Perspektiven, Jahrbuch 1996. Berlin; 1997. p. 33–6.
5. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, et al. Genome-wide comparison of medieval and modern Mycobacterium leprae. Science (80-). 2013;341(6142):179–83.
6. Volk AA, Atkinson JA. Infant and child death in the human environment of evolutionary adaptation. Evol Hum Behav [Internet]. 2013 May;34(3):182–92. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1090513812001237
7. Kramvis A. Genotypes and Genetic Variability of Hepatitis B Virus. Intervirology [Internet]. 2014;57(3–4):141–50. Available from: https://www.karger.com/Article/FullText/360947
8. Hepatitis B [Internet]. 2020 [cited 2020 Sep 10]. Available from: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b
9. Lai C-L, Yuen M-F. Prevention of hepatitis B virus-related hepatocellular carcinoma with antiviral therapy. Hepatology [Internet]. 2013 Jan;57(1):399–408. Available from: http://doi.wiley.com/10.1002/hep.25937
10. Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. Elife. 2018;7.
11. Mühlemann B, Jones TC, De Barros Damgaard P, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature.

2018;557(7705):418–23.

12. Kwon SY, Lee CH. Epidemiology and prevention of hepatitis B virus infection. Korean J Hepatol [Internet]. 2011;17(2):87. Available from: http://e-cmh.org/journal/view.php?doi=10.3350/kjhep.2011.17.2.87
13. MacLachlan JH, Cowie BC. Hepatitis B Virus Epidemiology. Cold Spring Harb Perspect Med [Internet]. 2015 May 1;5(5):a021410–a021410. Available from: http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a021410
14. Okada K, Kamiyama I, Inomata M, Imai M, Miyakawa Y, Mayumi M. E Antigen and Anti-E in the Serum of Asymptomatic Carrier Mothers as Indicators of Positive and Negative Transmission of Hepatitis B Virus to Their Infants. N Engl J Med [Internet]. 1976 Apr;294(14):746–9. Available from: http://www.nejm.org/doi/abs/10.1056/NEJM197604012941402
15. Beasley RP, Trepo C, Stevens CE, Szmuness W. The e antigen and vertical transmission of hepatitis B surface antigen. Am J Epidemiol. 1977;105(2):94–8.
16. Borgia G, Gentile I. Vertical transmission of hepatitis B virus: challenges and solutions. Int J Womens Health [Internet]. 2014 Jun;605. Available from: http://www.dovepress.com/vertical-transmission-of-hepatitis-b-virus-challenges-and-solutions-peer-reviewed-article-IJWH
17. The influence of age on the development of the hepatitis B carrier state. Proc R Soc London Ser B Biol Sci [Internet]. 1993 Aug 23;253(1337):197–201. Available from: https://royalsocietypublishing.org/doi/10.1098/rspb.1993.0102
18. Velkov S, Ott J, Protzer U, Michler T. The Global Hepatitis B Virus Genotype Distribution Approximated from Available Genotyping Data. Genes (Basel) [Internet]. 2018 Oct 15;9(10):495. Available from: http://www.mdpi.com/2073-4425/9/10/495
19. Lin C-L, Kao J-H. Hepatitis B Virus Genotypes and Variants. Cold Spring Harb Perspect Med [Internet]. 2015 May 1;5(5):a021436–a021436. Available from: http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a021436
20. Zaaijer HL, Boot HJ, van Swieten P, Koppelman MHGM, Cuypers HTM. HBsAg-negative mono-infection with hepatitis B virus genotype G. J Viral Hepat [Internet]. 2011 Nov;18(11):815–9. Available from: http://doi.wiley.com/10.1111/j.1365-2893.2010.01397.x
21. Chudy M, Schmidt M, Czudai V, Scheiblauer H, Nick S, Mosebach M, et al. Hepatitis B

virus genotype G monoinfection and its transmission by blood components. Hepatology [Internet]. 2006 Jul;44(1):99–107. Available from: http://doi.wiley.com/10.1002/hep.21220

22. Sayan M, Dogan C. Hepatitis B Virus Genotype G Infection in a Turkish Patient Undergoing Hemodialysis Therapy. Hepat Mon [Internet]. 2012 Feb 1;12(2):118–21. Available from: https://sites.kowsarpub.com/hepatmon/articles/70336.html
23. Li K, Zoulim F, Pichoud C, Kwei K, Villet S, Wands J, et al. Critical Role of the 36-Nucleotide Insertion in Hepatitis B Virus Genotype G in Core Protein Expression, Genome Replication, and Virion Secretion. J Virol [Internet]. 2007 Sep 1;81(17):9202–15. Available from: https://jvi.asm.org/content/81/17/9202
24. Cornelissen M, Zorgdrager F, Bruisten SM, Bakker M, Berkhout B, van der Kuyl AC. Widespread hepatitis B virus genotype G (HBV-G) infection during the early years of the HIV epidemic in the Netherlands among men who have sex with men. BMC Infect Dis [Internet]. 2016 Dec 10;16(1):268. Available from: http://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-016-1599-7
25. Kato H. Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G. Hepatology [Internet]. 2002 Apr;35(4):922–9. Available from: http://doi.wiley.com/10.1053/jhep.2002.32096
26. Erhardt A, Göbel T, Ludwig A, Lau GKK, Marcellin P, van Bömmel F, et al. Response to antiviral treatment in patients infected with hepatitis B virus genotypes E-H. J Med Virol [Internet]. 2009 Oct;81(10):1716–20. Available from: http://doi.wiley.com/10.1002/jmv.21588
27. Beggel B, Neumann-Fraune M, Döring M, Lawyer G, Kaiser R, Verheyen J, et al. Genotyping hepatitis B virus dual infections using population-based sequence data. J Gen Virol [Internet]. 2012 Sep 1;93(9):1899–907. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.043042-0
28. Sánchez LV, Tanaka Y, Maldonado M, Mizokami M, Panduro A. Difference of Hepatitis B Virus Genotype Distribution in Two Groups of Mexican Patients with Different Risk Factors. Intervirology [Internet]. 2007;50(1):9–15. Available from: https://www.karger.com/Article/FullText/96307
29. Araujo NM, Araujo OC, Silva EM, Villela-Nogueira CA, Nabuco LC, Parana R, et al. Identification of novel recombinants of hepatitis B virus genotypes F and G in human

immunodeficiency virus-positive patients from Argentina and Brazil. J Gen Virol [Internet]. 2013 Jan 1;94(1):150–8. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.047324-0

30. Pol S. Management of HBV in immunocompromised patients. Liver Int [Internet]. 2013 Feb;33:182–7. Available from: http://doi.wiley.com/10.1111/liv.12055

31. Qiu J, Söderlund-Venermo M, Young NS. Human Parvoviruses. Clin Microbiol Rev [Internet]. 2017 Jan 2;30(1):43–113. Available from: https://cmr.asm.org/content/30/1/43

32. Ihara T, Furusyo N, Hayashi T, Toyoda K, Murata M, Hayashi J. A population-based epidemiological survey of human parvovirus B19 infection: a project of the Kyushu and Okinawa Population Study (KOPS). Arch Virol [Internet]. 2013 Dec 18;158(12):2465–72. Available from: http://link.springer.com/10.1007/s00705-013-1746-z

33. Zhang L, Cai C, Pan F, Hong L, Luo X, Hu S, et al. Epidemiologic study of human parvovirus B19 infection in East China. J Med Virol [Internet]. 2016 Jul;88(7):1113–9. Available from: http://doi.wiley.com/10.1002/jmv.24459

34. Mühlemann B, Margaryan A, De Barros Damgaard P, Allentoft ME, Vinner L, Hansen AJ, et al. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. Proc Natl Acad Sci U S A. 2018;115(29):7557–62.

35. Anderson MJ, Higgins PG, Davis LR, Willman JS, Jones SE, Kidd IM, et al. Experimental Parvoviral Infection in Humans. J Infect Dis [Internet]. 1985 Aug 1;152(2):257–65. Available from: https://academic.oup.com/jid/article-lookup/doi/10.1093/infdis/152.2.257

36. Heegaard ED, Brown KE. Human Parvovirus B19. Clin Microbiol Rev [Internet]. 2002 Jul;15(3):485–505. Available from: https://cmr.asm.org/content/15/3/485

37. Prospective study of human parvovirus (B19) infection in pregnancy. Public Health Laboratory Service Working Party on Fifth Disease. BMJ [Internet]. 1990 May 5;300(6733):1166–70. Available from: http://www.bmj.com/cgi/doi/10.1136/bmj.300.6733.1166

38. Rajput R, Sehgal A, Jain D, Sen R, Gupta A. Acute Parvovirus B19 Infection Leading to Severe Aplastic Anemia in a Previously Healthy Adult Female. Indian J Hematol Blood Transfus [Internet]. 2012 Jun 10;28(2):123–6. Available from:

http://link.springer.com/10.1007/s12288-011-0112-0

39. RÖHRER C, GÄRTNER B, SAUERBREI A, BÖHM S, HOTTENTRÄGER B, RAAB U, et al. Seroprevalence of parvovirus B19 in the German population. Epidemiol Infect [Internet]. 2008 Nov 16;136(11):1564–75. Available from: https://www.cambridge.org/core/product/identifier/S0950268807009958/type/journal_article

40. Norja P, Hokynar K, Aaltonen L-M, Chen R, Ranki A, Partio EK, et al. Bioportfolio: Lifelong persistence of variant and prototypic erythrovirus DNA genomes in human tissue. Proc Natl Acad Sci [Internet]. 2006 May 9;103(19):7450–3. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0602259103

41. Norja P, Eis-Hübinger AM, Söderlund-Venermo M, Hedman K, Simmonds P. Rapid Sequence Change and Geographical Spread of Human Parvovirus B19: Comparison of B19 Virus Evolution in Acute and Persistent Infections. J Virol [Internet]. 2008 Jul 1;82(13):6427–33. Available from: https://jvi.asm.org/content/82/13/6427

42. Benedictow OJ. The Black Death: Greatest Catastrophe Ever. Hist Today. 2005;

43. Bos KI, Herbig A, Sahl J, Waglechner N, Fourment M, Forrest SA, et al. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. Elife. 2016;5(JANUARY2016).

44. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, et al. A draft genome of Yersinia pestis from victims of the Black Death. Nature. 2011;478(7370):506–10.

45. Spyrou MA, Tukhbatova RI, Feldman M, Drath J, Kacki S, Beltrán De Heredia J, et al. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. Cell Host Microbe. 2016;19(6):874–81.

46. Spyrou MA, Keller M, Tukhbatova RI, Scheib CL, Nelson EA, Andrades Valtueña A, et al. Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. Nat Commun. 2019;10(1).

47. Albertz H, Bodo H. Die Chronik Berlins. 2nd Editio. Chronik; 1986.

48. Bauer R. Berlin: Illustrierte Chronik, Bd. 1. Bis 1870. Helvetica Chimica Acta; 1988. 421 p.

49. Schwebel O. Geschichte der Stadt Berlin 2 Bde. Brachvogel & Rauft; 1888.

50. Schneider W, Gottschalk W. Berlin. Eine Kulturgeschichte in Bildern und Dokumenten.

Berlin: Gustav Kiepenheuer; 1983.

51. Peters G. Kleine Berliner Baugeschichte: Von der Stadtgründung bis zur Bundeshauptstadt. Berlin: Stapp Verlag; 1995.
52. Schwenk H. Berliner Stadtentwicklung von A bis Z. Edition Luisenstadt; 1998.
53. Krause-Kyora B, Nutsua M, Boehme L, Pierini F, Pedersen DD, Kornell SC, et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. Nat Commun. 2018;9(1).
54. Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. Nat Ecol Evol. 2018;2(3):520–8.
55. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17(3):377–86.
56. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
57. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
58. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
59. Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome Biol. 2016;17:60.
60. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
61. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol [Internet]. 2011 Jan 1;29(1):24–6. Available from: http://www.nature.com/articles/nbt.1754
62. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
63. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014;514(7253):494–7.

64. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
65. Susat J, Bonczarowska JH, Pētersone-Gordina E, Immel A, Nebel A, Gerhards G, et al. Yersinia pestis strains from Latvia show depletion of the pla virulence gene at the end of the second plague pandemic. Sci Rep [Internet]. 2020 Dec 3;10(1):14628. Available from: http://www.nature.com/articles/s41598-020-71530-9
66. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics [Internet]. 2010 Oct 1;26(19):2460–1. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq461
67. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet [Internet]. 2000 Jun;16(6):276–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168952500020242
68. National Center for Biotechnology Information. Entrez Direct: E-utilities on the Unix Command Line [Internet]. Entrez Programming Utilities Help. Available from: https://www.ncbi.nlm.nih.gov/books/NBK179288/
69. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.
70. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics [Internet]. 2014 Nov 15;30(22):3276–8. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btu531
71. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol [Internet]. 2000 Apr 1;17(4):540–52. Available from: http://academic.oup.com/mbe/article/17/4/540/1127654
72. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Mol Biol Evol [Internet]. 2006 Feb 1;23(2):254–67. Available from: http://academic.oup.com/mbe/article/23/2/254/1118872/Application-of-Phylogenetic-Networks-in
73. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the

performance of PhyML 3.0. Syst Biol. 2010;59(3):307–21.

6.

Conclusion and outlook

Conclusion and outlook

In this thesis, an undirected large-scale biomolecular and bioinformatic screening was performed resulting in data from 2255 samples (1881 individuals) that was investigated for the presence of human pathogens (Table S1). The broad timeframe of the samples (5000 BCE to 1950 CE) makes the number of analyzed per year samples insufficient according to modern epidemiological standards. Nevertheless, in ancient DNA research this sample size is unrivaled. Due to the undirected screening, this huge dataset should also enable chance findings as prior evidence of a pathogens has been rare. The samples were screened for bacterial and viral pathogens using a BLASTN like approach to identify potential candidates based on a reference database of complete viral and bacterial genomes. Despite spanning a timeframe of 7000 years a lion's share of the samples dates the Middle Ages and the Neolithic (Figure 1).

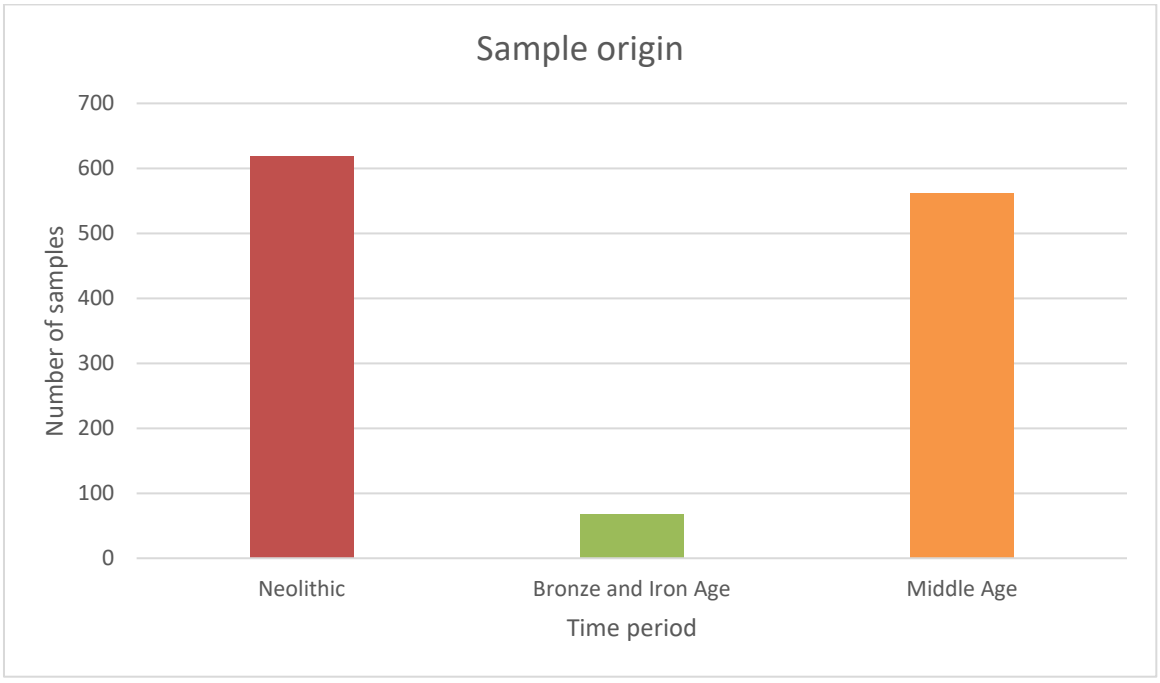


Figure 1. Temporal sample origin of 1247 representative samples.

Interestingly, the Bronze, Iron and Roman Age are periods were a lot of human remains have been discovered but in-depth molecular analysis are sparse. This is the result of cremation as a common burial practice throughout these periods (1). Cremation leads to destruction of the DNA rendering sampling of the burnt remains pointless (2).

Currently the most used material for detecting ancient pathogens is the human tooth (2,3). However, if available also petrous bones, ribs and fragments of the skull and spine were screened for the presence of pathogens. The screening with the undirected BLASTN approach was carried out with relaxed parameters to account for evolutionary distance and the ambition to catch the whole variety of the metagenomic mix. In the end, this procedure resulted in the detection of five different pathogens (Table 1). Due to the promising initial screening results the Hepatitis B virus (HBV), *Yersinia pestis* (*Y. pestis*) and parvovirus B19 (B19) were investigated further. As not all samples contained enough reads for assembly and in-depth analysis a total of 12 HBV (not all are part of this thesis), three *Y. pestis* and two B19 genomes were reconstructed.

Table 1. Results of the pathogen screening for five different pathogens. For example, in 175 samples the presence of Hepatitis B virus (HBV) was at least supported by one read. In 61 samples the presence was supported by at least ten reads and in 21 samples by at least 100 reads.

Pathogen	At least 1 read	At least 10 reads	At least 100 reads
HBV	175	61	21
B19	260	117	35
Y. pestis	257	47	13
Mycobacterium leprae (*M. leprae)*	339	49	8
Mycobacterium tuberculosis (*M. tuberculosis*)	820	369	61

Despite the screening being undirected, prominent pathogens like *M. tuberculosis, M. leprae and Y. pestis* were detected. These pathogens are very are well known among the aDNA community and several genomes have been published (Table 2).

Table 2. List of ancient pathogen genomes recovered from archeological or historical specimens. Source: (4), changed

Pathogen	Infectious disease	Method of retrieval	Number of genomes	References
M. leprae	Leprosy	NGS Microarray-based capture followed by NGS	27	(5–8)
M. tuberculosis	Tuberculosis	NGS Microarray-based capture followed by NGS	19	(9–11)
Y. pestis	Bubonic, pneumonic and septicaemic plague	NGS Microarray-based capture followed by NGS	58	(12–19)
HBV	Hepatitis B	NGS Microarray-based capture followed by NGS	14	(20–22)
B19	Erythema infectiosum (fifth disease in children)	In-solution capture followed by NGS	10	(23)

M. Leprae, for instance, is a prime target for aDNA researchers as leprosy is a disease which can lead to specific lesions on the bones. It led to visible disfigurement and was deemed incurable which resulted in great fear within the affected populations. This resulted in a long history of isolating infected people to contain the spread of the disease. Leper colonies, also known as leprosaria, became widespread in the Middle ages and some leprosaria were operating until the middle of the 20th century CE. This in turn led to dedicated leprosy graveyards which have a been a great target for aDNA researchers

Comparing the number of samples with more than 100 reads for a certain pathogen in Table 1. to the numbers of published genomes for the respective pathogen in Table 2. leads to the assumption that the pathogen screening was very successful, but these numbers are just telling half the story. Determination of true and false positives can be a complex task. For *Yersinia pestis* a scoring system with defined criteria was implemented (12). No such scores, however, exist for other pathogens. Each of these results was individually analyzed. A great number of the *Mycobacterium leprae* and virtually all *Mycobacterium tuberculosis* hits turned out to be false positives due to mycobacterial cross contamination. Mycobacteria are ubiquitous in soil environments and aquatic environments (24). The best studied members of this genus which are also a part of aDNA research are *M. leprae* and *M. tuberculosis* but apart from these there are more than 175 additional members in the genus (25).

With *M. marinum* and *M. kansaii* being genetically very similar to *M. tuberculosis* sequences originating from these two can lead to misassignments and false positives using a blast like approach (26,27). When looking at the assigned sequences, it became obvious that all the TB hits were based on cross contamination as there were lots of mismatches or piles of sequences in repetitive regions. Paleopathological evidence for *M. tuberculosis* was only present for a handful of samples (Figure 2) and even these did not yield positive sequences.

Figure 2. Spine from the Tuberculosis Archive in Heidelberg (Germany) exhibiting severe deformation. The osteomyelitis that caused this deformation can be the result of an infection with several different bacteria and is therefore unreliable to pinpoint the responsible agent. Picture was taken by: Dr. Oswinde Bock-Hensley

A reason for not getting positive *M. tuberculosis* results might be the nature of the disease itself. Progression can be very slow and under normal circumstances the pathogen is cleared by the host´s immune system (28). Only in later stages of the disease damage of lung tissue can lead to bacteremia and systemic infection which is mandatory for later detection in the bone material (29). After a closer look at the generated data, reads belonging to *Acinetobacter johnsonii* could be recovered (Not part of this thesis). It is known that Acinetobacter infection can also lead to osteomyelitis and subsequent bone deformation (30,31).

M. lepromatosis and *M. haemophilum* are closely related to *M. leprae* leading to the same misassignment problem as with *M. tuberculosis*. After inspection and curation of the positive samples the 339 initial reports of *M. leprae* boiled down to eight positive and seven usable samples, dated to the Middle Ages from German burial sites. A usable sample was defined as one with enough authentic pathogen sequences to justify enrichment or deeper sequencing as just a few reads were not providing sufficient information. One of the samples yielded a complete *M. leprae* genome which was phylogenetically analyzed (Not part of this thesis). Some of these bones depicted typical evidence (Figure 3) for a severe infection with *M. leprae* (32,33). After all, initial assumptions based on paleopathology and historical records can be helpful but also misleading like in the tuberculotic spine and all possible reasons should be taken into consideration.

Figure 3. Leprosy skull from the Leprosarium in Odense (Denmark). The skull exhibits degradation of the bone around the nasal area and the maxilla which are typical signs of a severe infection with *M. Leprae*. Picture was taken by: Ben Krause-Kyora

Material limitations and false expectations

Despite the huge number of 2255 screened samples, in the end only a small fraction contained authentic and sufficient amount of ancient pathogen DNA to enable reliable analysis and investigation. There are several reasons as to why only relatively few positives were detected. The first and most obvious reason is the sample material. The lack of soft tissue limits the detection of possible pathogens by a great margin (34). All pathogens that are not causing a systemic infection and just reside in specific organs are missed. During the course of a disease enough molecules of the agent must be produced to enable later detection. It has been shown that intact viral particles can persist for days and even weeks on dry hydrophobic surfaces (35). However, on the long run, taphonomic and diagenetic processes might just simply destroy any

DNA or RNA evidence that was present (36). It is widely accepted that over time, RNA is less stable than DNA but in a cool and dry environment it can keep its structure and functionality for up to 30.000 years (37).

One obvious reason could be, that our expectations regarding the disease burden during human history might just be too high. One good example is the spread of infectious diseases during the first epidemiological transition also known as the Neolithic transition. About 7500 years ago this transition introduced a more sedentary lifestyle involving plant and animal domestication (38). This period was also characterized by an increase in population size and formation of dense settlements. It has been hypothesized that this change in lifestyle was accompanied by the rise of infectious diseases (39). Denser populations could foster the spread of a disease, proximity to animals would promote a stable conduit for exposure of humans to animal pathogens and vice versa and connections between villages would support the spread of diseases (40). But up to now, except for sporadic cases, no direct evidence for large scale infections have been found (41). A plausible reason might be that, despite settlements being formed and people living closer together, overall population density was too low to maintain and spread infectious diseases. Measles, rubella and pertussis require at least several hundred thousand people to sustain within the population (42,43). The same seems to be true for the Middle Age also partly known as the dark age (500-1000 CE), referring to the deterioration of Western Europe after the fall of the Roman Empire. During that time cities were severely affected by overcrowding subsequently fostering outbreaks of infectious diseases (6). *Y. pestis* and *M. leprae* both led to common and historically well documented diseases but apart from that only few other infectious agents have been sporadically recovered from medieval remains (6,13,14,32,44).

Bioinformatic approach

The verification of each detected pathogen started with mapping against the most prominent reference used in aDNA literature. If no literature was available, mapping against the most prominent reference genome on NCBI (National Center for Biotechnology Information) was carried out. This mapping provided a good idea about the initial quality of the data and gave a clear indication if deeper sequencing or capture methods were needed. However, after this step, each virus or bacteria demanded a specific bioinformatic pipeline (Table 3).

Table 3. Different bioinformatic pathways from the initial mapping to the construction of phylogeny.

	Y. *pestis*	HBV	B19
Screening	MALT (45)	MALT	MALT
Mapping	single modern reference BWA (46)	Several modern/ancient References BWA and Bowtie2 (47)	Several modern/ancient References BWA and Bowtie2
Assembly (tools and level)	GATK (48), VCF	Spades (49), custom script ,manual curation, Complete Genome	Spades, custom script ,manual curation, Complete Genome
Multiple Sequence Alignment (tools and level)	MultiFCVAnalyzer (11), based on SNPs (manual curation)	MAFFT (50) and GBLOCKS (51), based on complete genomes	MAFFT and GBLOCKS, based on complete genomes
Phylogenetics (tools)	RAxML (52), MrBayes (53)	Splitstree (54), MEGA6 (55)	PhyML, MrBayes
Additional analysis	Analysis of virulence factors, SNP effect, BEAST2 (56), detection of deletions and insertions	Detection of recombination, investigation of open reading frames, subgenotype analysis	-

For instance, the HBV-positive samples were additionally mapped against a file containing references from all HBV genotypes using BWA as well as Bowtie2. This was done for identification of the true genotype as it is known that recombination between different strains is a common phenomenon in HBV (57,58). Using two mappers in this step should ensure that the whole HBV variety in the sample is detected. Due to the short genome size of HBV and B19 the results of this mapping were inspected manually and gave clear indication of the

genotype and possible recombination of genotypes. Subsequently, all the mapped reads were used as an input for a de novo assembly. The multiple sequence alignment and phylogenetic analysis for both viruses were based on the whole genome. On the contrary, aDNA studies of *Y. pestis* suggested that mapping against one modern reference to be sufficient. Moreover, due to the length of the *Y. pestis* genome, no manual curation was performed. An SNP based multiple sequence alignment was preferred as it reduced the dataset to a fraction of size that contained only phylogenetically informative positions. This SNP dataset led to a faster analysis in subsequent steps and allowed for manual curation and correction.

Technical limitations

As mentioned in the introduction up to now, no unknown and completely new pathogen has been recovered from an ancient sample. One of the biggest problems is the database bias. Nowadays genetic databases contain a great number organisms which are of high relevance for public health or industrial purposes. However, a wide range of information is missing (59). These databases reflect only a small proportion of the worldwide diversity and subsequently bias the screening process. Furthermore, curation of records and metadata is a problem substantially affecting the reliability of databases. Up to date, there has been no way to work around these limitations and discover something completely unknown in an aDNA sample. Protocols and Methods that are able to detect new viral or bacterial species in modern metagenomic samples could work in ancient ones. However, the tiny amount of actual aDNA sequences compromises this line of work right from the start.

Sadly, new sequencing technologies like Oxford nanopore and PacBio cannot be used for the generation of longer and more informative sequences as the ancient sequences are naturally of short length (60). De novo genome assembly on aDNA data is a delicate task as the quality of the assembly depends on coverage, read accuracy and read length (61,62). Additionally, aDNA datasets are often metagenomic and only multi-layer approaches which integrate different steps, can be successful (63).

Furthermore, some bacteria might have been pathogenic for humans in the past and due to certain events lost their pathogenicity over time. This phenomenon is known from several organisms (64–66). These bacteria might show up in the results of the metagenomic screening. However, since they are not regarded as pathogens today, they go unnoticed. A general screening for the presence of certain pathogenicity motives might be a way to overcome this

obstacle but these targets should be chosen wisely. Results obtained with this line of work should be handled with caution as this method will also exhibit a high rate of false positives.

During this thesis other applications were also tested to examine if better results could be achieved in comparison to the BLASTN approach. A profile-hidden markov model (HMM) based approach (unpublished) was conducted to screen HBV positive samples trying to evaluate the sensitivity with respect to only a few reads being present and the short length of the sequences. In contrast to a normal mapping, profile-HHMs are probabilistic models based on position-specific information from a multiple sequence alignment (67). As it is a probabilistic approach, homologous sequences with low identity can be recognized using HMMs. The models were able to detect fewer HBV reads and in some cases, only one read exhibited a good e-value. Due to a higher computational demand, which resulted in a longer runtime and even weaker signal of this method, the BLASTN based approach was favored. After all, HMMs suffer from the same database flaw as the model construction relies on a set of reference sequences. However, there might still be lots of potential in profile HMM's for the detection of pathogens in aDNA.

The detection of RNA based pathogens in aDNA samples is another complex topic. Up to date, just a single RNA based pathogen has been recovered from ancient remains. Using frozen lung tissue as a source material, researchers were able to reconstruct all eight RNA segments of the virus responsible for the Spanish flu in 1918 (68–73). In theory, however, RNA molecules, should be stable for a longer time given specific circumstances (74).

Another problem is the blind trust towards software and the provided mapping results or metagenomic assignment. As mentioned above, cross contamination and misassignments are still a problem in aDNA research. Most of the modern bioinformatic software that is available is not finetuned for aDNA conditions. Even dedicated aDNA software might have problems discovering very small numbers of pathogen sequences. Detection of HBV is a good example for explaining this problem. After the discovery of HBV in several ancient samples (20–22), already generated datasets were again screened for the presence of HBV and a lot of possible candidates were noted. Initially trusting pure numbers and relying on software to report positive matches allowed a small number of positive reads to go unnoticed. In the initial screening, which generates about 5 million reads, only 30 of them might belong to HBV. Compared to bacterial pathogens which have a much bigger genome, 30 initial HBV reads

could already justify generating more data as in theory there should be enough reads to cover the whole genome. The same problem was encountered while detecting the *pla*- (*pla* negative) plasmids in *Y. pestis*. Considering the numbers alone would hide the phenomenon within statistics about average coverage. In these cases, manual inspection is necessary to confirm the presence or absence of coverage in the region. This shows that working with aDNA can be a very delicate affair and highlights the great potential in already generated and published datasets.

Yersinia pestis

I could show that *Y. pestis* has been infecting humans up for at least 7000 years and that it might have lost some of its pathogenicity after the decrease of *pla+* (*pla* positive) pPCP1 plasmids (Chapter 1 and 2). Using dated phylogeny, I confirmed previously known split dates and estimated a new split date with *Y. pseudotuberculosis* (Chapter 2). Due to its historical importance and the devastation it brought it will always have a special place within the aDNA community. These findings show that its history is still not completely written. The loss of the *pla* gene is interesting for several reasons. From modern studies it is known that *pla*- strains exist in nature and mainly reside within rodent populations (75,76). These strains exhibit highly reduced virulence and are unable to infect the lungs preventing them from causing the bubonic plague (77–79). Given already published datasets the evolution of post Black Death strains resembles pearls on a string where at one point in time a decrease of the *pla*+ plasmids has started. Based on this observation one would assume that the decline would occur in a linear way. With respect to the *pla+*/*pla*- ratios I found it might as well be an independent event that has happened in different strains at different geographical locations. Another interesting factor with direct connection to the loss of the *pla* region would be the natural rodent host. Models have shown that the primary mode of plague transmission in Europe occurred predominantly through human ectoparasites and was not driven by rodents (80). *Rattus rattus*, which was the predominant rat in the Middle Ages, was evolutionary adapted to warmer climate and the medieval rat population was only large enough to maintain a plague epidemic in major towns but not in the countryside (81–83). *Pla+* strains are also associated with an enhanced transmissibility to the arthropod vector as the bacteremia induced by them is consistently higher than the one of *pla*- strains (84). The Black Death (1346-

1353) killed nearly half of the Europeans in its wake leaving a thinned host population behind (85). Furthermore, there is possible evidence for microevolutionary changes in the human population which are most likely associated with bacterial pathogens, such as like *Y. pestis* (86). Given these factors rodent populations might have been a more attractive host to *Y. pestis* than humans. All these facts and circumstances regarding the loss of the *pla* gene might explain why subsequent outbreaks during the post Black Death era never reached a pandemic level. It would be interesting to evaluate if the *pla*- plasmid results in an overall fitness advantage or just in certain situations. Apart from these reasons, there might have been other explanations for the post Black Death history of *Y. pestis*. Another open question would be when did the descendants of the Black Death strains died out. The most recent descendants were discovered at the L'Observance plague cemetery dating to the 1722 plague of Marseille (18). There might be a chance that more descendants of the Black Death strains can be discovered in other burial sites. As it is known that *Y. pestis* evolved from *Y. pseudotuberculosis* by acquiring certain pathogenicity genes, reliable split dating for this event is very important (87–89). However, until now there is no consensus about the split as estimations range from at least 30000 to 5000 years (12,87,88,90). Recent discovery of basal *Y. pestis* lineages estimated that the split is closer to 7000 years, but more strains and data is needed to verify this split (Susat et al 2020).

Hepatitis B virus

HBV was initially detected in 175 samples (Table 1). During the course of this thesis, 12 HBV genomes from five German burial sites (Petersberg, Karsdorf, Sorsum, Berlin, Lauchheim (not published)) were reconstructed and analyzed (Chapter 3 and 4). The strain from Karsdorf is up to date the earliest direct evidence of a human pathogen dating 7000 years back (Chapter 3). That HBV is one of the few viruses detected in ancient samples is not a coincidence. Its partially double stranded circular DNA and small genome size make it a prime candidate for aDNA research. It has been shown that the proto human HBV arose within the last 30mya and is much older than assumed (Chapter 3). The current HBV diversity consists of eight genotypes (A-H) with a clear phylogeographic distribution (91). Interestingly, this modern distribution is not reflected in all the ancient HBV strains (Figure 4), which suggests a more complicated ancestry than previously assumed (20).

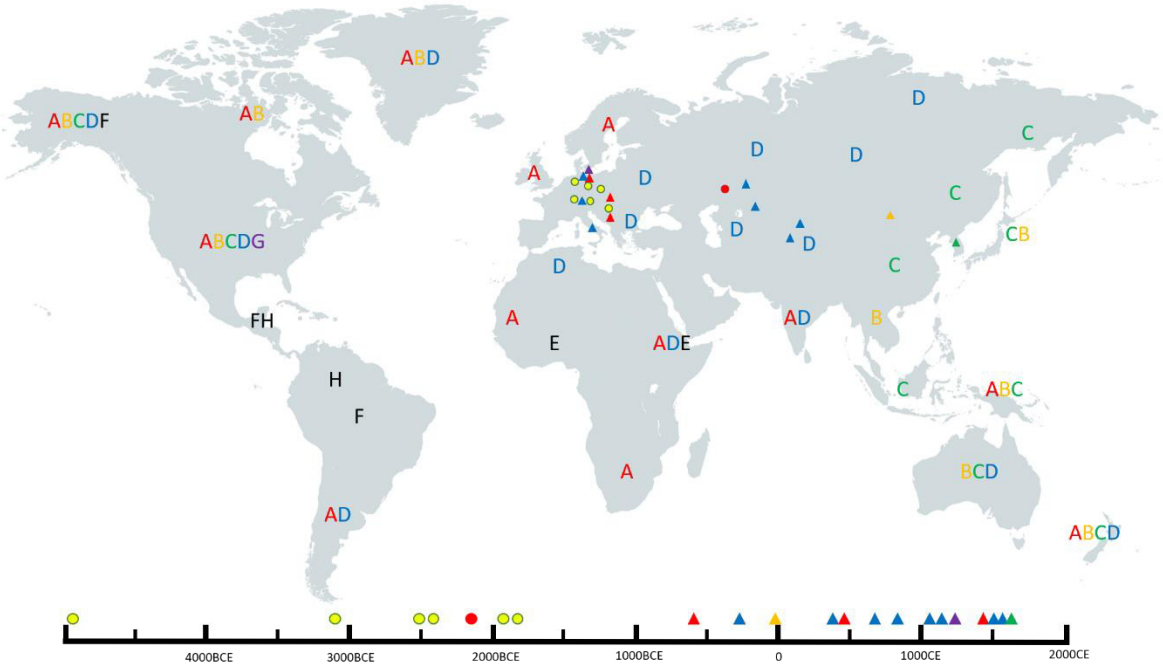


Figure 4. Geographical distribution of analyzed ancient samples with timeframe and modern distribution of genotypes. Samples older than 1000BCE are depicted as dots, younger ones as rectangles. The samples are colored according to their genotype. Hybrid strains with close relationship to non-human primate strains are depicted as yellow dots. Ancient genotype A samples were found in regions that are dominated by genotype D today. The Ancient genotype B sample was found in a region that is today dominated by genotype C. The appearance of an ancient genotype G strain in central Europe is also interesting but not unlikely as this genotype still contributes a small fraction to HBV infections in Europe (91). This map contains HBV strains from (20–22,92) as well as unpublished strains. The map was generated using PowerPoint and mapchart.net.

It has been hypothesized that HBV spread together with modern humans as they migrated out of Africa. This idea, however, is contested by the basal phylogenetic positions of the strains F and H as these strains are exclusively found in the Americas (58,93). As recombination plays an important role in the generation of new HBV genotypes and massive waves of migration might have had huge influence on the distribution and formation of genotypes, the complete picture is blurred (57,58). Detection of several strains that are genetically close to old world monkey strains supports the hypothesis of successful reciprocal cross-species transmission between humans and non-human primates (57,94–96). However, the number of transmission events remains unclear. Since up to now no recent relatives of these mosaic strains have been found, these lineages could have gone extinct. Furthermore, it is unclear if theses ancient

strains are the progenitors of present-day strains or if they are hybrids originating from strains that have not been found yet. Due to the convention that HBV sequences with a nucleotide diversity up to 8% are grouped into one genotype, the found mosaic strains are part of the African non-human primate strains (97–99). Interestingly, the genomic organization between HBV strains through time depicts a strongly conserved genome structure. From the oldest discovered genome to modern genomes the overall order and ORF´s are the same. This seems to be a result of the densely packed genome with partially overlapping genes. Nowadays an infection rate above 8% of populations is considered to be "high prevalence" (100). We showed that human populations during history can exhibited infection rates up to 19% (Chapter 4). This high rate of infection could be a result of perinatal transmission. Infants of HBV infected mothers have a high chance of being infected themselves and exhibit an increased risk of developing a chronic infection (101,102). Nowadays the infection rates are higher for males than for females as they are more prone to several risk factors like alcohol consumption and smoking which can lead to a severe outcome of the disease (103–106). Results of sex determination in ancient HBV positive samples also exhibited a higher number of infected males. The traits that lead to an unhealthier lifestyle, such as impulsivity and aggression, have been associated with the male sex and have been with us since the advent of humanity (107). After all, infections with HBV seem to be common throughout human history and with a certain sample size of ancient strains we might be able to construct complete and comprehensive picture of its evolutionary and epidemiologic history. Furthermore, it would be interesting to examine the coinfection rate with Hepatitis D Virus (HDV) as it is a sub-viral agent that cannot replicate without the HBV being present (108). However, as HDV is an RNA virus, not a single sample exhibited signs of coinfection with HBV and HDV.

Parvovirus B19

I reconstructed two Parvovirus genomes from child burials in Berlin dating to the late Middle Age (Chapter 4). Genotyping revealed that these strains are belonging to genotype 2. Compared to HBV, B19 depicts a semi stable distribution around the globe. Whilst genotype 1 can be found worldwide, genotype 2 is mainly found in northern Europe and genotype 3 is common in Sub-Saharan, West Africa, South America and France (109–111). Despite the close

genetic identity to genotype 2, phylogenetic analysis depicts an early branching event separating the Berlin strains from the other ancient and present members of genotype 2. This might hint to the rise of a new genotype/subgenotype. More data is needed to support this hypothesis. Furthermore, 50% of the buried children exhibited genetic evidence a of B19 infection. At first staggering, this rate of infection is comparable to the seroprevalence of modern German children from the age of six to ten (112). As Parvovirus is a single stranded DNA virus, it is surprising that its traces could be found in these medieval skeletons and samples up to 6900 years old (23). Depurination and deamination occurs about four times faster in ssDNA than in dsDNA (113). It might be the protective role of an intact viral capsid providing the DNA with shelter from the damaging influences. This mechanism has been observed in canine Parvovirus (114). Interestingly, It has been shown that the genetic distance between the modern and the ancient strains is as big as it is within the modern diversity (23). This level of genetic stability through time also seems to be a result of the genome structure of B19. In contrast to HBV, there are no overlapping ORF's and up to 12 different mRNA's can be generated through alternative splicing and polyadenylation (115). It has been shown that most of the B19 genome is under negative selection as functional aspects of the proteins are limiting the fixation of new mutations (116). Based on modern heterochronous datasets the inferred substitution rate for B19 is unusually high with 1.0-4.0 x 10^{-4} nucleotide substitutions per site per year (s/s/y) (117–119). This substitution rate is more similar to those of RNA viruses than single or double-stranded DNA viruses (120). Considering the information gained from ancient B19 sequences, the estimated substitution rate based on modern and ancient genomes is approximately an order of magnitude lower and more in line with other single-stranded DNA viruses (23). Further discovery of ancient B19 sequences might decrease the substitution rate even more. Investigating the spread and distribution of B19 through time could be fascinating. For this purpose, more ancient sequences are needed.

Concluding remark

Ancient pathogen genomics can help to complete the puzzle of bacterial and viral evolution. Even small details can give us more understanding, solve questions modern studies never could answer and gives us insight from times long gone. During my doctoral research I discovered several ancient pathogens and generated new information about their

evolutionary history. The *Y. pestis* strains from Riga exhibited frequency loss of the *pla* gene which is associated with the dissemination of the bacteria. This might be an adaption to a changing environment which posed new pressure upon the bacteria that favored a decrease of this pathogenicity factor. I showed that the origin of *Y. pestis,* the split from its predecessor *Y. pseudotuberculosis,* is more recent than previously assumed. The Neolithic hybrid HBV strains helped us to better understand the relationship between human HBV and non-human primate HBV strains. These strains expanded the known diversity of HBV and verified assumptions about the transmission between humans and monkeys. The mass graves from Berlin shed light on the pathogen load during the Late Middle Ages.

To further investigate the metagenomic composition of aDNA samples the development of new methodology tackling the high number of false positives and the small number of authentic reads is inevitable. Despite the initial lack of good results with profile HMMs it would still be interesting to generate and test different models. Furthermore, close collaboration with dedicated experts is necessary to fully comprehend and understand each finding completely.

Regarding future research in the field of ancient infectious diseases it would be interesting to rescreen our samples using different approaches and maybe more focused on a single pathogen. With respect to the recent publication of several ancient smallpox strains this would be a prime candidate (121). Additionally, as the number of ancient genomes is steadily increasing a more in-depth look into genomic properties of ancient pathogens could lead to the detection of variations only present in ancient genomes. These might go unnoticed since most of the literature regarding genetic variation in pathogens is based on research done with modern strains. Furthermore, phages represent an interesting topic as well, as the presence of highly specific phages could be used as a marker to verify the presence of the respective pathogen.

1. Rebay-Salisbury K. Cremations: fragmented bodies in the Bronze and Iron Ages. In: Body Parts and Bodies Whole: Changing Relations and Meanings. Oxbow Books Oxford; 2010. p. 64–71.

2. Hansen HB, Damgaard PB, Margaryan A, Stenderup J, Lynnerup N, Willerslev E, et al. Comparing ancient DNA preservation in petrous bone and tooth cementum. PLoS One. 2017;12(1).

3. Margaryan A, Hansen HB, Rasmussen S, Sikora M, Moiseyev V, Khoklov A, et al. Ancient pathogen DNA in human teeth and petrous bones. Ecol Evol. 2018;8(6):3534–42.

4. Spyrou MA, Bos KI, Herbig A, Krause J. Ancient pathogen genomics as an emerging tool for infectious disease research. Vol. 20, Nature Reviews Genetics. 2019. p. 323–40.

5. Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, et al. Mycobacterium leprae genomes from a British medieval leprosy hospital: Towards understanding an ancient epidemic. BMC Genomics. 2014;15(1).

6. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, et al. Genome-wide comparison of medieval and modern Mycobacterium leprae. Science (80-). 2013;341(6142):179–83.

7. Krause-Kyora B, Nutsua M, Boehme L, Pierini F, Pedersen DD, Kornell SC, et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. Nat Commun. 2018;9(1).

8. Schuenemann VJ, Avanzi C, Krause-Kyora B, Seitz A, Herbig A, Inskip S, et al. Ancient genomes reveal a high diversity of Mycobacterium leprae in medieval Europe. PLoS Pathog. 2018;14(5).

9. Chan JZM, Sergeant MJ, Lee OYC, Minnikin DE, Besra GS, Pap I, et al. Metagenomic analysis of tuberculosis in a mummy. Vol. 369, New England Journal of Medicine. 2013. p. 289–90.

10. Kay GL, Sergeant MJ, Zhou Z, Chan JZM, Millard A, Quick J, et al. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in

Europe. Nat Commun. 2015;6.

11. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014;514(7253):494–7.

12. Andrades Valtueña A, Mittnik A, Key FM, Haak W, Allmäe R, Belinskij A, et al. The Stone Age Plague and Its Persistence in Eurasia. Curr Biol. 2017;27(23):3683-3691.e8.

13. Spyrou MA, Tukhbatova RI, Feldman M, Drath J, Kacki S, Beltrán De Heredia J, et al. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. Cell Host Microbe. 2016;19(6):874–81.

14. Spyrou MA, Keller M, Tukhbatova RI, Scheib CL, Nelson EA, Andrades Valtueña A, et al. Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. Nat Commun. 2019;10(1).

15. Spyrou MA, Tukhbatova RI, Wang CC, Valtueña AA, Lankapalli AK, Kondrashin V V., et al. Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. Nat Commun. 2018;9(1).

16. De Barros Damgaard P, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, et al. 137 ancient human genomes from across the Eurasian steppes. Nature. 2018;557(7705):369–74.

17. Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren KG, et al. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. Cell. 2015;163(3):571–82.

18. Bos KI, Herbig A, Sahl J, Waglechner N, Fourment M, Forrest SA, et al. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. Elife. 2016;5(JANUARY2016).

19. Feldman M, Harbeck M, Keller M, Spyrou MA, Rott A, Trautmann B, et al. A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. Mol Biol Evol. 2016;33(11):2911–23.

20. Mühlemann B, Jones TC, De Barros Damgaard P, Allentoft ME, Shevnina I, Logvin A, et al. Ancient hepatitis B viruses from the Bronze Age to the Medieval period. Nature.

2018;557(7705):418–23.

21. Kahila Bar-Gal G, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, et al. Tracing hepatitis B virus to the 16th century in a Korean mummy. Hepatology. 2012;56(5):1671–80.

22. Patterson Ross Z, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, et al. The paradox of HBV evolution as revealed from a 16thcentury mummy. PLoS Pathog. 2018;14(1).

23. Mühlemann B, Margaryan A, De Barros Damgaard P, Allentoft ME, Vinner L, Hansen AJ, et al. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. Proc Natl Acad Sci U S A. 2018;115(29):7557–62.

24. Hruska K, Kaevska M. Mycobacteria in water, soil, plants and air: A review. Vol. 57, Veterinarni Medicina. 2012. p. 623–79.

25. Tortoli E. Microbiological features and clinical relevance of new species of the genus Mycobacterium. Clin Microbiol Rev. 2014;27(4):727–52.

26. Veyrier FJ, Dufort A, Behr MA. The rise and fall of the Mycobacterium tuberculosis genome. Vol. 19, Trends in Microbiology. 2011. p. 156–61.

27. Wang J, Behr MA. Building a better bacillus: The emergence of Mycobacterium tuberculosis. Front Microbiol. 2014;5(APR).

28. Chai Q, Zhang Y, Liu CH. Mycobacterium tuberculosis: An adaptable pathogen associated with multiple human diseases. Vol. 8, Frontiers in Cellular and Infection Microbiology. 2018.

29. Jilani TN, Avula A, Zafar Gondal A, Siddiqui AH. Active Tuberculosis [Internet]. StatPearls. 2020. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30020618

30. Davis KA, Moran KA, McAllister CK, Gray PJ. Multidrug-resistant Acinetobacter extremity infections in soldiers. Emerg Infect Dis. 2005;11(8):1218–24.

31. Wong D, Nielsen TB, Bonomo RA, Pantapalangkoor P, Luna B, Spellberg B. Clinical and pathophysiological overview of Acinetobacter infections: A century of challenges. Clin Microbiol Rev. 2017;30(1):409–47.

32. Boldsen JL. Epidemiological approach to the paleopathological diagnosis of leprosy.

Am J Phys Anthropol. 2001;115(4):380–7.

33. Boldsen JL. Leprosy in medieval Denmark - Osteological and epidemiological analyses. Anthropol Anzeiger. 2009;67(4):407–25.

34. Aufderheide AC. Progress in soft tissue paleopathology. Vol. 284, Journal of the American Medical Association. 2000. p. 2571–3.

35. Firquet S, Beaujard S, Lobert PE, Sané F, Caloone D, Izard D, et al. Survival of enveloped and non-enveloped viruses on inanimate surfaces. Microbes Environ. 2015;30(2):140–4.

36. Marinho ANDR, Miranda NC, Braz V, Ribeiro-dos-Santos ÂK, De Souza SMFM. Paleogenetic and taphonomic analysis of human bones from Moa, Beirada, and Zé Espinho Sambaquis, Rio de Janeiro, Brazil. In: Memorias do Instituto Oswaldo Cruz. 2006. p. 15–23.

37. Yashina S, Gubin S, Maksimovich S, Yashina A, Gakhova E, Gilichinsky D. Regeneration of whole fertile plants from 30,000-y-old fruit tissue buried in Siberian permafrost. Proc Natl Acad Sci U S A. 2012;109(10):4008–13.

38. Bellwood P, Oxenham M. The expansions of farming societies and the role of the Neolithic demographic transition. In: The Neolithic Demographic Transition and its Consequences. 2008.

39. McKeown RE. The Epidemiologic Transition: Changing Patterns of Mortality and Population Dynamics. Am J Lifestyle Med. 2009;3(1_suppl):19S-26S.

40. Fournié G, Pfeiffer DU, Bendrey R. Early animal farming and zoonotic disease dynamics: modelling brucellosis transmission in Neolithic goat populations. R Soc Open Sci [Internet]. 2017 Feb;4(2):160943. Available from: https://royalsocietypublishing.org/doi/10.1098/rsos.160943

41. Fuchs K, Rinne C, Drummer C, Immel A, Krause-Kyora B, Nebel A. Infectious diseases and Neolithic transformations: Evaluating biological and archaeological proxies in the German loess zone between 5500 and 2500 BCE. Holocene. 2019;29(10):1545–57.

42. Dobson AP, Carper ER. Infectious Diseases and Human Population History. Bioscience. 1996;46(2):115–26.

43. May RM, Anderson RM. Infectious Diseases of Humans: Dynamics and Control book. Oxford University Press; 1991.

44. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, et al. A draft genome of Yersinia pestis from victims of the Black Death. Nature. 2011;478(7370):506–10.

45. Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, et al. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. Nat Ecol Evol. 2018;2(3):520–8.

46. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.

47. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

48. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res [Internet]. 2010 Sep 1;20(9):1297–303. Available from: http://genome.cshlp.org/cgi/doi/10.1101/gr.107524.110

49. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.

50. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30(14):3059–66.

51. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol [Internet]. 2000 Apr 1;17(4):540–52. Available from: http://academic.oup.com/mbe/article/17/4/540/1127654

52. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

53. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17(8):754–5.

54. Huson DH, Bryant D. Application of Phylogenetic Networks in Evolutionary Studies. Mol Biol Evol [Internet]. 2006 Feb 1;23(2):254–67. Available from: http://academic.oup.com/mbe/article/23/2/254/1118872/Application-of-Phylogenetic-Networks-in

55. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.

56. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2019;15(4).

57. Simmonds P, Midgley S. Recombination in the Genesis and Evolution of Hepatitis B Virus Genotypes. J Virol [Internet]. 2005 Dec 15;79(24):15467–76. Available from: https://jvi.asm.org/content/79/24/15467

58. Littlejohn M, Locarnini S, Yuen L. Origins and Evolution of Hepatitis B Virus and Hepatitis D Virus. Cold Spring Harb Perspect Med [Internet]. 2016 Jan 4;6(1):a021360. Available from: http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a021360

59. Pollock J, Glendinning L, Wisedchanwet T, Watson M. The madness of microbiome: Attempting to find consensus “best practice” for 16S microbiome studies. Vol. 84, Applied and Environmental Microbiology. 2018.

60. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One. 2012;7(3).

61. Myers Jr EW. A history of DNA sequence assembly. it - Inf Technol [Internet]. 2016 Jan 28;58(3). Available from: https://www.degruyter.com/view/j/itit.2016.58.issue-3/itit-2015-0047/itit-2015-0047.xml

62. Nagarajan N, Pop M. Sequence assembly demystified. Nat Rev Genet. 2013;14:157–67.

63. Seitz A, Nieselt K. Improving ancient DNA genome assembly. PeerJ. 2017;2017(4).

64. Laanto E, Bamford JKH, Laakso J, Sundberg L-R. Phage-Driven Loss of Virulence in a

Fish Pathogenic Bacterium. Brockhurst MA, editor. PLoS One [Internet]. 2012 Dec 31;7(12):e53157. Available from: https://dx.plos.org/10.1371/journal.pone.0053157

65. Linz B, Ivanov Y V., Preston A, Brinkac L, Parkhill J, Kim M, et al. Acquisition and loss of virulence-associated factors during genome evolution and speciation in three clades of Bordetella species. BMC Genomics [Internet]. 2016 Dec 30;17(1):767. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-3112-5

66. Maury MM, Chenal-Francisque V, Bracq-Dieye H, Han L, Leclercq A, Vales G, et al. Spontaneous Loss of Virulence in Natural Populations of Listeria monocytogenes. Freitag NE, editor. Infect Immun [Internet]. 2017 Nov 21;85(11). Available from: https://iai.asm.org/lookup/doi/10.1128/IAI.00541-17

67. Krogh A, Brown M, Mian IS, Sjölander K, Haussler D. Hidden Markov Models in computational biology applications to protein modeling. J Mol Biol. 1994;235(5):1501–31.

68. Reid AH, Fanning TG, Janczewski TA, McCall S, Taubenberger JK. Characterization of the 1918 “Spanish” Influenza Virus Matrix Gene Segment. J Virol. 2002;76(21):10717–23.

69. Reid AH, Taubenberger JK, Fanning TG. Evidence of an absence: The genetic origins of the 1918 pandemic influenza virus. Nat Rev Microbiol. 2004;2(11):909–14.

70. Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. Characterization of the 1918 influenza virus polymerase genes. Nature. 2005;437(7060):889–93.

71. Reid AH, Fanning TG, Hultin J V., Taubenberger JK. Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. Proc Natl Acad Sci U S A. 1999;96(4):1651–6.

72. Reid AH, Fanning TG, Janczewski TA, Taubenberger JK. Characterization of the 1918 “Spanish” influenza virus neuraminidase gene. Proc Natl Acad Sci U S A. 2000;97(12):6785–90.

73. Basler CF, Reid AH, Dybing JK, Janczewski TA, Fanning TG, Zheng H, et al. Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. Proc Natl Acad Sci

U S A. 2001;98(5):2746–51.

74. Fordyce SL, Kampmann ML, van Doorn NL, Gilbert MTP. Long-term RNA persistence in postmortem contexts. Vol. 4, Investigative Genetics. 2013.

75. Filippov AA, Solodovnikov NS, Kookleva LM, Protsenko OA. Plasmid content in Yersinia pestis strains of different origin. FEMS Microbiol Lett. 1990;67(1–2):45–8.

76. Cavalcanti YVN, Leal NC, De Almeida AMP. Typing of Yersinia pestis isolates from the state of Ceará, Brazil. Lett Appl Microbiol. 2002;35(6):543–7.

77. Sodeinde OA, Subrahmanyam YVBK, Stark K, Quan T, Bao Y, Goguen JD. A surface protease and the invasive character of plague. Science (80-). 1992;258(5084):1004–7.

78. Welkos SL, Friedlander AM, Davis KJ. Studies on the role of plasminogen activator in systemic infection by virulentYersinia pestisstrain C092. Microb Pathog [Internet]. 1997 Oct;23(4):211–23. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0882401097901546

79. Sebbane F, Lemaître N, Sturdevant DE, Rebeil R, Virtaneva K, Porcella SF, et al. Adaptive response of Yersinia pestis to extracellular effectors of innate immunity during bubonic plague. Proc Natl Acad Sci U S A. 2006;103(31):11766–71.

80. Dean KR, Krauer F, Walløe L, Lingjærde OC, Bramanti B, Stenseth NC, et al. Human ectoparasites and the spread of plague in Europe during the Second Pandemic. Proc Natl Acad Sci U S A. 2018;115(6):1304–9.

81. Shrewsbury JFD. A History of Bubonic Plague in the British Isles. Cambridge; 1970.

82. Becker K. Rattus Rattus. In: Niethammer J, Krapp F, editors. Handbuch der Säugetiere Europas. Akademische Verlagsgesellschaft Wiesbaden; 1978. p. 382–400.

83. Gillespie H. Rattus Rattus [Internet]. 2004. Available from: https://animaldiversity.org/site/accounts/information/Rattus_rattus.html

84. Lorange EA, Race BL, Sebbane F, Hinnebusch BJ. Poor vector competence of fleas and the evolution of hypervirulence in Yersinia pestis. J Infect Dis. 2005;191(11):1907–12.

85. Benedictow OJ. The Black Death: Greatest Catastrophe Ever. Hist Today. 2005;

86. Immel A, Key FM, Szolek A, Barquera R, Robinson MK, Harrison GF, et al. Analysis of

genomic DNA from medieval plague victims suggests effect of 2 Yersinia pestis on human immunity genes. Mol Biol Evol. Currently under revision.

87. Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, et al. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. Proc Natl Acad Sci U S A. 2013;110(2):577–82.

88. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. Yersinia pestis, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proc Natl Acad Sci [Internet]. 1999 Nov 23;96(24):14043–8. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.96.24.14043

89. Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, et al. Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. Proc Natl Acad Sci [Internet]. 2004 Sep 21;101(38):13826–31. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0404012101

90. Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B, et al. Microevolution and history of the plague bacillus, Yersinia pestis. Proc Natl Acad Sci [Internet]. 2004 Dec 21;101(51):17837–42. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0408026101

91. Velkov S, Ott J, Protzer U, Michler T. The Global Hepatitis B Virus Genotype Distribution Approximated from Available Genotyping Data. Genes (Basel) [Internet]. 2018 Oct 15;9(10):495. Available from: http://www.mdpi.com/2073-4425/9/10/495

92. Krause-Kyora B, Susat J, Key FM, Kühnert D, Bosse E, Immel A, et al. Neolithic and medieval virus genomes reveal complex evolution of hepatitis B. Elife. 2018;7.

93. Locarnini S, Littlejohn M, Aziz MN, Yuen L. Possible origins and evolution of the hepatitis B virus (HBV). Vol. 23, Seminars in Cancer Biology. 2013. p. 561–75.

94. MacDonald DM, Holmes EC, Lewis JCM, Simmonds P. Detection of Hepatitis B Virus Infection in Wild-Born Chimpanzees (Pan troglodytes verus): Phylogenetic Relationships with Human and Other Primate Genotypes. J Virol. 2000;74(9):4253–7.

95. de Carvalho Dominguez Souza BF, Drexler JF, de Lima RS, de Oliveira Hughes Veiga do Rosário M, Netto EM. Theories about evolutionary origins of human hepatitis B virus in primates and humans. Brazilian J Infect Dis [Internet]. 2014 Sep;18(5):535–43.

Available from: https://linkinghub.elsevier.com/retrieve/pii/S1413867014000622

96. Rasche A, Souza BF de CD, Drexler JF. Bat hepadnaviruses and the origins of primate hepatitis B viruses. Curr Opin Virol [Internet]. 2016 Feb;16:86–94. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1879625716000183

97. Okamoto H, Tsuda F, Sakugawa H, Sastrosoewignjo RI, Imai M, Miyakawa Y, et al. Typing Hepatitis B Virus by Homology in Nucleotide Sequence: Comparison of Surface Antigen Subtypes. J Gen Virol [Internet]. 1988 Oct 1;69(10):2575–83. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-69-10-2575

98. Norder H, Couroucé A-M, Magnius LO. Complete Genomes, Phylogenetic Relatedness, and Structural Proteins of Six Strains of the Hepatitis B Virus, Four of Which Represent Two New Genotypes. Virology [Internet]. 1994 Feb;198(2):489–503. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0042682284710609

99. Sugauchi F, Mizokami M, Orito E, Ohno T, Kato H, Suzuki S, et al. A novel variant genotype C of hepatitis B virus identified in isolates from Australian Aborigines: complete genome sequence and phylogenetic relatedness The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank databases und. J Gen Virol [Internet]. 2001 Apr 1;82(4):883–92. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-82-4-883

100. MacLachlan JH, Locarnini S, Cowie BC. Estimating the global prevalence of hepatitis B. Lancet [Internet]. 2015 Oct;386(10003):1515–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0140673615611163

101. Okada K, Kamiyama I, Inomata M, Imai M, Miyakawa Y, Mayumi M. E Antigen and Anti-E in the Serum of Asymptomatic Carrier Mothers as Indicators of Positive and Negative Transmission of Hepatitis B Virus to Their Infants. N Engl J Med. 1976;294(14):746–9.

102. Beasley RP, Trepo C, Stevens CE, Szmuness W. The e antigen and vertical transmission of hepatitis B surface antigen. Am J Epidemiol. 1977;105(2):94–8.

103. Tanaka K, Hirohata T, Takeshita S, Hirohata I, Koga S, Sugimachi K, et al. Hepatitis B Virus, cigarette smoking and alcohol consumption in the development of hepatocellular carcinoma: A Case-Control study in Fukuoka, Japan. Int J Cancer. 1992;51(4):509–14.

104. Jee SH, Ohrr H, Sull JW, Samet JM. Cigarette smoking, alcohol drinking, hepatitis B, and risk for hepatocellular carcinoma in Korea. J Natl Cancer Inst. 2004;96(24):1851–6.

105. Ritchie H, Roser M. Smoking [Internet]. 2013. Available from: https://ourworldindata.org/smoking

106. Ritchie H. Alcohol Consumption [Internet]. Alcohol Consumption. 2018. Available from: https://ourworldindata.org/alcohol-consumption

107. Schulte MT, Ramo D, Brown SA. Gender differences in factors influencing alcohol use and drinking progression among adolescents. Clin Psychol Rev [Internet]. 2009 Aug;29(6):535–47. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0272735809000701

108. Negro F. Hepatitis D Virus Coinfection and Superinfection. Cold Spring Harb Perspect Med [Internet]. 2014 Nov 1;4(11):a021550–a021550. Available from: http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a021550

109. Hubschen JM, Mihneva Z, Mentis AF, Schneider F, Aboudy Y, Grossman Z, et al. Phylogenetic Analysis of Human Parvovirus B19 Sequences from Eleven Different Countries Confirms the Predominance of Genotype 1 and Suggests the Spread of Genotype 3b. J Clin Microbiol [Internet]. 2009 Nov 1;47(11):3735–8. Available from: https://jcm.asm.org/content/47/11/3735

110. Norja P, Hokynar K, Aaltonen L-M, Chen R, Ranki A, Partio EK, et al. Bioportfolio: Lifelong persistence of variant and prototypic erythrovirus DNA genomes in human tissue. Proc Natl Acad Sci [Internet]. 2006 May 9;103(19):7450–3. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0602259103

111. Parsyan A, Szmaragd C, Allain J-P, Candotti D. Identification and genetic diversity of two human parvovirus B19 genotype 3 subtypes. J Gen Virol [Internet]. 2007 Feb

1;88(2):428–31. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.82496-0

112. RÖHRER C, GÄRTNER B, SAUERBREI A, BÖHM S, HOTTENTRÄGER B, RAAB U, et al. Seroprevalence of parvovirus B19 in the German population. Epidemiol Infect [Internet]. 2008 Nov 16;136(11):1564–75. Available from: https://www.cambridge.org/core/product/identifier/S0950268807009958/type/journal_article

113. Lindahl T. Instability and decay of the primary structure of DNA. Nature [Internet]. 1993 Apr;362(6422):709–15. Available from: http://www.nature.com/articles/362709a0

114. Chapman MS, Rossmann MG. Single-stranded DNA–protein interactions in canine parvovirus. Structure [Internet]. 1995 Feb;3(2):151–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0969212601001460

115. Ganaie SS, Qiu J. Recent Advances in Replication and Infection of Human Parvovirus B19. Front Cell Infect Microbiol [Internet]. 2018 Jun 5;8. Available from: https://www.frontiersin.org/article/10.3389/fcimb.2018.00166/full

116. Stamenković GG, Ćirković VS, Šiljić MM, Blagojević J V., Knežević AM, Joksić ID, et al. Substitution rate and natural selection in parvovirus B19. Sci Rep [Internet]. 2016 Dec 24;6(1):35759. Available from: http://www.nature.com/articles/srep35759

117. Shackelton LA, Holmes EC. Phylogenetic Evidence for the Rapid Evolution of Human B19 Erythrovirus. J Virol [Internet]. 2006 Apr 1;80(7):3666–9. Available from: https://jvi.asm.org/content/80/7/3666

118. Norja P, Eis-Hübinger AM, Söderlund-Venermo M, Hedman K, Simmonds P. Rapid Sequence Change and Geographical Spread of Human Parvovirus B19: Comparison of B19 Virus Evolution in Acute and Persistent Infections. J Virol [Internet]. 2008 Jul 1;82(13):6427–33. Available from: https://jvi.asm.org/content/82/13/6427

119. Toppinen M, Perdomo MF, Palo JU, Simmonds P, Lycett SJ, Söderlund-Venermo M, et al. Bones hold the key to DNA virus history and epidemiology. Sci Rep [Internet]. 2015 Dec 27;5(1):17226. Available from: http://www.nature.com/articles/srep17226

120. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet [Internet]. 2008 Apr 4;9(4):267–76. Available from: http://www.nature.com/articles/nrg2323

121. Mühlemann B, Vinner L, Margaryan A, Wilhelmson H, Castro CDLF, Allentoft ME, et al. Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science (80-). 2020;369(6502).

7.

Declaration

Herewith, I confirm that the submitted thesis is completely the result of my own work. Apart from the advice of my supervisors, all sources and cooperation partners are listed within the thesis. This thesis has not been submitted elsewhere. It has been carried out in strict accordance with the rules of Good Scientific Practice of the *Deutsche Forschungsgesellschaft*. In addition, no academic degree has ever been withdrawn.

8.

Acknowledgements

First, I like to thank my parents for letting me choose my own way without (too much) questioning my choices and supporting me on my journey. Thanks to all the members of my family for the good time we share. Especially my sister and my brother. I am glad to have you. It is always a great pleasure.

Many thanks to Prof. Dr. Tal Dagan for being my first supervisor and showing interest for my research topic. Prof. Dr. Krause-Kyora receives my highest appreciation for guiding me through my time as a PhD student. Thanks for your support and trust in what I have done. You gave me the freedom to explore new territories and discover things on my own. But you were also there when I needed guidance and focus.

Thanks to all the members of the ancient DNA and the human longevity group. Without you it would have been a lonesome journey. All the funny things we experienced together, the nice discussions in the office, the slow internet, and the still ongoing struggle for getting a good screen and cables will always have a place in my heart. Thanks to Joanna Bonczarowska and Katharina Fuchs for corrections and suggestions on how to improve my work. Thanks to Alexander Immel and Joanna Bonczarowska for your work on the manuscripts that we published together. Thanks to Prof. Dr. Almut Nebel for your input to the manuscripts and support throughout the years. And Thanks to all the researchers I had the pleasure working with.

And last but not least I thank my friends for your constant support.

9.

General introduction & Conclusion and outlook

Supplementary material

Table S1. Samples, individuals and sample sites that were investigated in this thesis. Additional to the 1469 samples from 1095 individuals listed here 786 samples from 786 individuals were screened. These samples were not extracted and sequenced by the aDNA laboratory in Kiel. Country abbreviation is given in brackets (RS = Serbia, LU = Luxembourg, DE = Germany, MX = Mexico, DK = Denmark, MD = Moldova, SE = Sweden, CH = Switzerland, CZ = Czechia, RU = Russia, LV = Latvia, NL = Netherlands, UA = Ukraine, AT = Austria, HR = Croatia, SI = Slovenia, GR = Greece, PL = Poland, IN = India, IQ = Iraq.

Number of samples	Number of individuals	Organism	Sample site or Country
22	3	Rat	Caricin Grad (RS)
2	2	Human	Loschbour (LU)
3	3	Human	Lossow (DE)
131	101	Human	Niedertiefenbach (DE)
15	7	Human	Petersberg (DE)
21	21	Human	Puxcatan (MX)
81	81	Human	St Jorgen (DK)
21	1	Human	Strande (DE)
5	5	Human	Trebur (DE)
3	3	Hornless catle	Halle (DE)
1	1	Human	Rungholt (DE)
1	1	Auerochs	Zoological museum Kiel (DE)
47	36	Human	Trebur (DE)
1	1	Human	Göttingen (DE)
3	2	Human	Ribe (DK)
6	5	Human	Gordinesti and Pocrovca (MD)
1	1	Human	Petersberg (DE)
3	3	Rind	Halle (DE)
1	1	Auerochs	Bensheim (DE)
18	9	Human	Gotland (SE)
26	13	Human	St. Catharinenkirche (DE)
5	5	Human	Gertrudenfriedhof, Oldenburg (DE)
9	9	Soil samples	Zürich (CH)
27	18	Human	Berlin (DE)
23	23	Human	Fellbach-Oeffingen (DE)
1	1	Human	Klobuky (CZ)
10	10	Human	Kudachurt (RU)
28	14	Human	Lauchheim (DE)
36	36	Human	Lübeck (DE)
1	1	Soil	Mang de Bargen (DE)
28	21	Human	Mikulov (CZ)
5	5	Human	Podivin (CZ)
20	10	Pinniped	Zoologcal collection, Kiel (DE)
2	2	Human	Regensburg (DE)
11	11	Human	Reigoldswil (CH)
30	30	Human	Rendsburg Neuwerk (DE)

2	2	Human	Riga (LV)
26	18	Human	Castle Gottorf (DE)
64	32	Human	Sorsum (DE)
9	9	Soil	Lübeck (DE)
19	13	Bee	Germany
106	68	Human	Lauchheim (DE)
4	4	Human	Niedertiefenbach (DE)
2	2	Human	Künzing (DE)
6	6	Human	Altendorf (DE)
5	5	Human	Leeuwarden (NL)
17	17	Human	Warburg (DE)
17	17	Human	Altendorf (DE)
23	13	Human	Niederpöring (DE)
1	1	Human	Rovantsi (UA)
1	1	Human	Wöllersdorf (AT)
1	1	Squirrel	Castle Homburg (DE)
2	1	Human	Zumberak (HR, SI)
10	10	Human	Rimbeck (DE)
103	66	Human	Sakhtysh (RU)
14	11	Human	Rinnukalns (LV)
4	4	Human	Göttingen (DE)
19	19	Human	Fellbach-Oeffingen (DE)
20	10	Human	Melaten (DE)
1	1	Human	Greece
5	5	Human	Bribir and Dubrovnik (HR)
4	4	Human	Deventer (NL)
43	23	Human	Berlin (DE)
19	11	Human	Gepios (GR)
4	4	Human	Berlin (DE)
21	21	Human	Hospital of the holy ghost Lübeck (DE)
1	1	Human	Kerameikos (GR)
30	16	Human	Transylvania (RO)
13	13	Human	Riga (LV)
5	3	Human	Iclod (RO)
5	3	Human	Oldenburg (DE)
1	1	Human	Windeby (DE)
19	10	Human	Slivin (PL)
5	2	Human	Friesack (DE)
6	4	Human	Heidelberg (DE)
1	1	Human	Lingenfeld (DE)
2	2	Human	Trebur (DE)
48	48	Human	Melaten (DE)
1	1	Reindeer	Hamburg (DE)
13	13	Sheep	Dali (IN)
20	20	Human	Basel (CH)
2	2	Human	Bolivia

12	12	Human	Schleswig (DE)
3	3	Human	Kosenovka (UA)
7	5	Human	Uecker (DE)
13	6	Human	Werle (DE)
12	6	Human	Melaten (DE)
9	9	Human	Lübeck (DE)
18	18	Human	Ur (IQ)
3	1	Human	Soham (GB)
1469	1095		

10.

Chapter I: *Yersinia pestis* strains from Latvia show depletion of the *pla* virulence gene at the end of the second plague pandemic

Supplementary material

Title:

Yersinia pestis strains from Latvia show depletion of the *pla* virulence gene at the end of the second plague pandemic.

Authors:

Julian Susat, Joanna Bonczarowska, Elīna Pētersone-Gordina, Alexander Immel, Almut Nebel, Guntis Gerhards, Ben Krause-Kyora

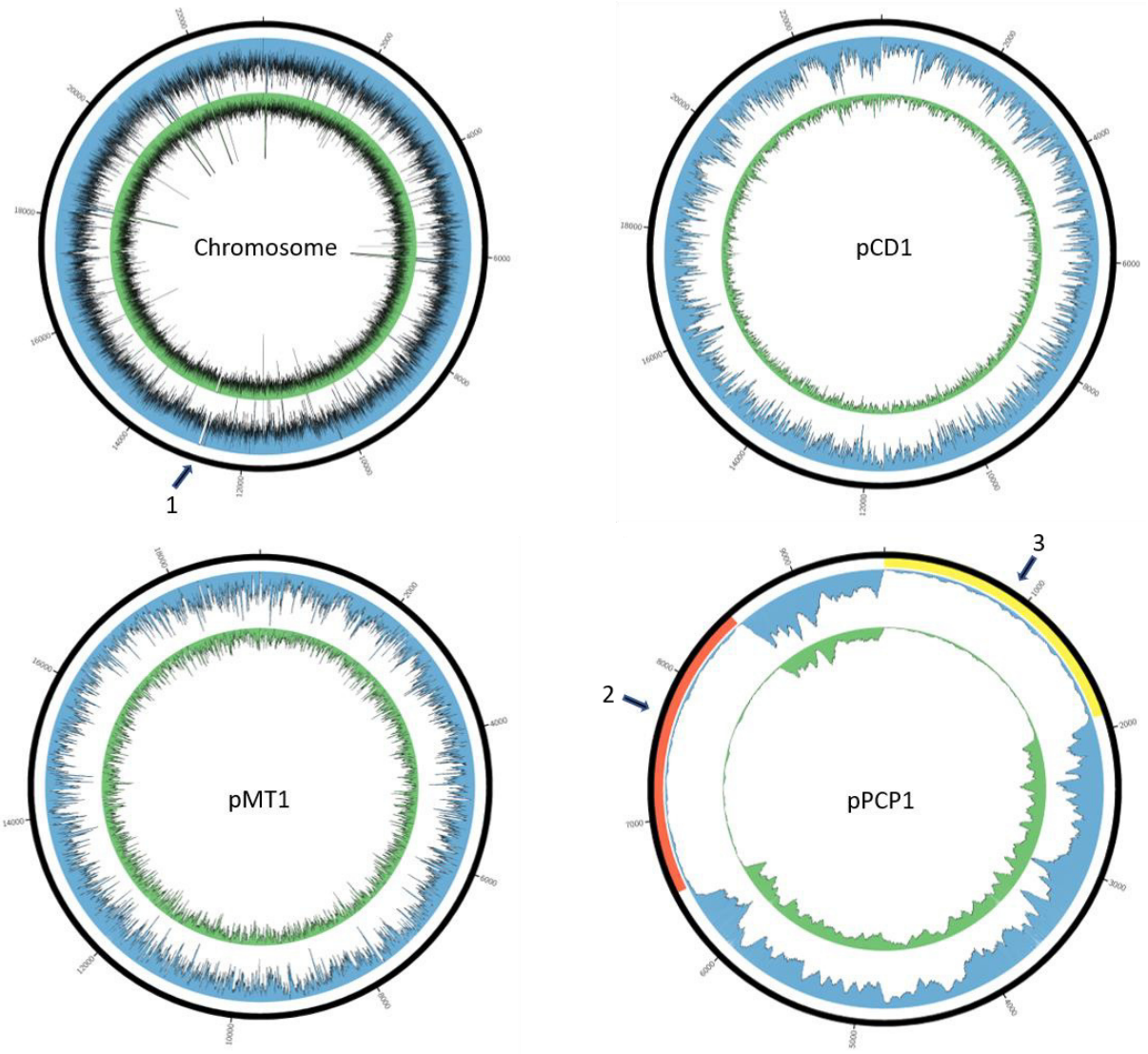


Figure S1. Average coverage plots for the chromosome and the three plasmids of the two Riga strains G488 (blue) and G701 (green) compared with the CO92 reference. The average depth of coverage was calculated for 200-bp regions for the chromosome, 3-bp regions for pCD1 and 5-bp regions for pMT1. For a true representation of coverage depth, each plot was given a different threshold for the maximum value. Chromosome 10x, pCD1 40x, pMT1 25x and pPCP1 100x. All figures were generated using Circos. 1: filamentous prophage YpfΦ which is not covered by any sequences in our sample, 2: *pla* gene region 6428-8530 including the *pla* gene (YPPCP1.07), a putative transcriptional regulator (YPPCP1.08c) and a hypothetical protein (YPPCP1.09c), 3: insertion sequence IS100 including a transposase (YPPCP1.01) and an ATP-binding protein (YPPCP1.01). Due to the presence of this region in several parts of the genome, filtering led to a drop in coverage in this region.

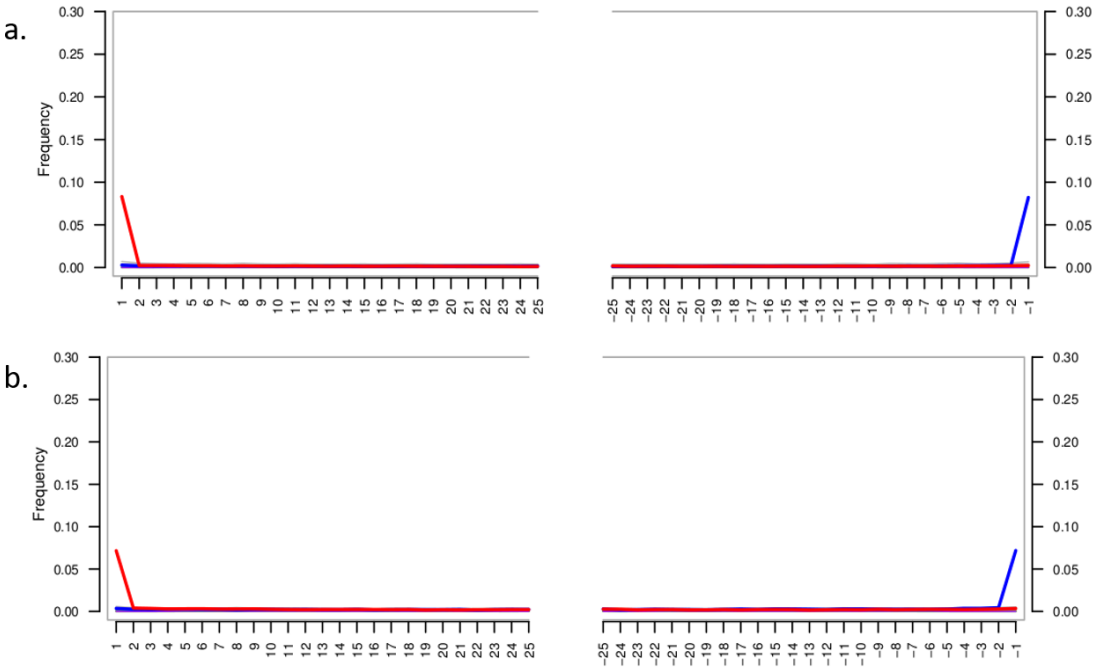


Figure S2. aDNA damage patterns for G448 (a) and for G701 (b). Both samples show the expected degradation patterns, increased C > T and G > A substitutions at the 5´ and 3´ends of the reads, after half-UDG library preparation.

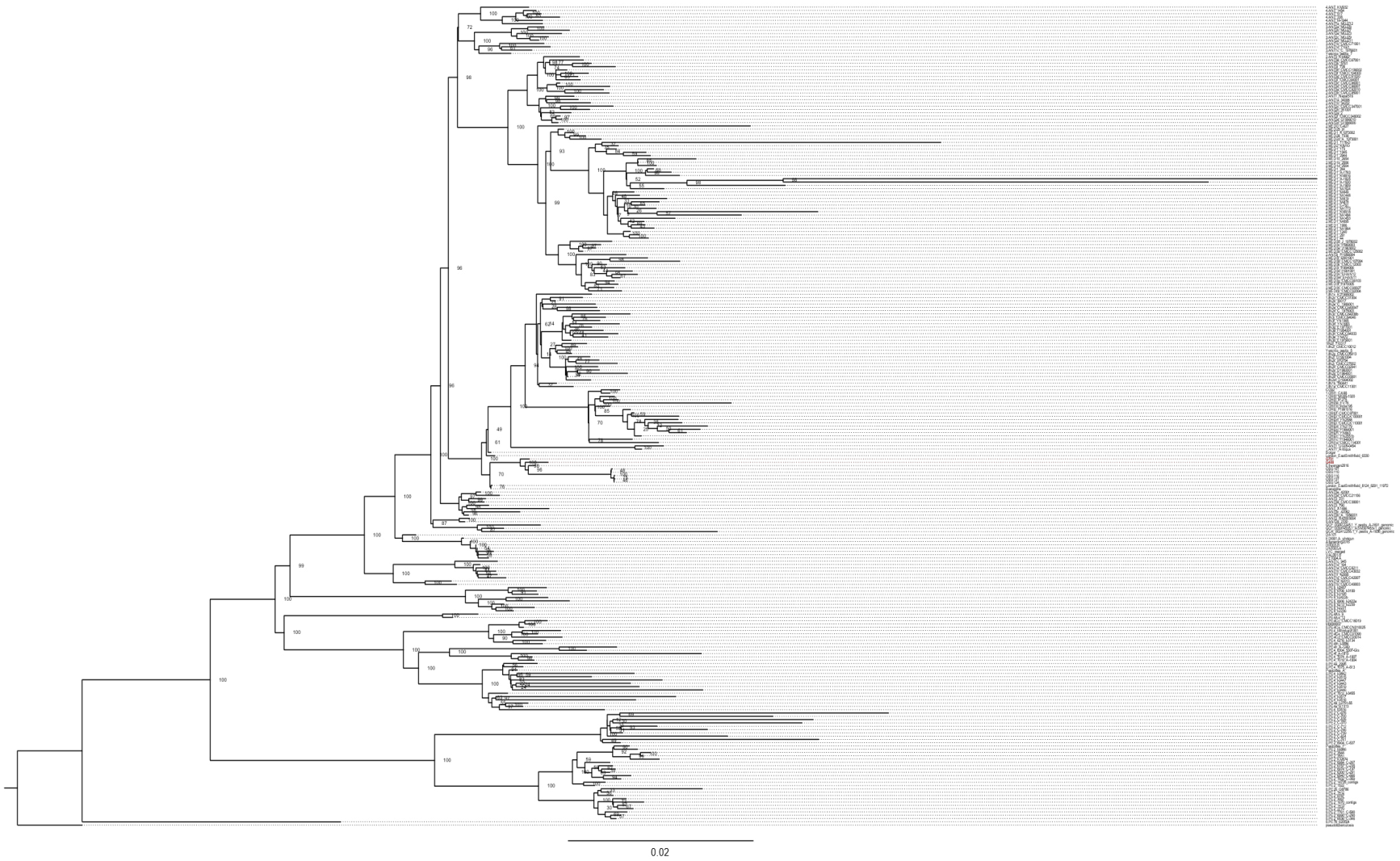


Figure S3. Maximum-likelihood tree generated with RAxML. Bootstrap values are shown for 500 replicates.

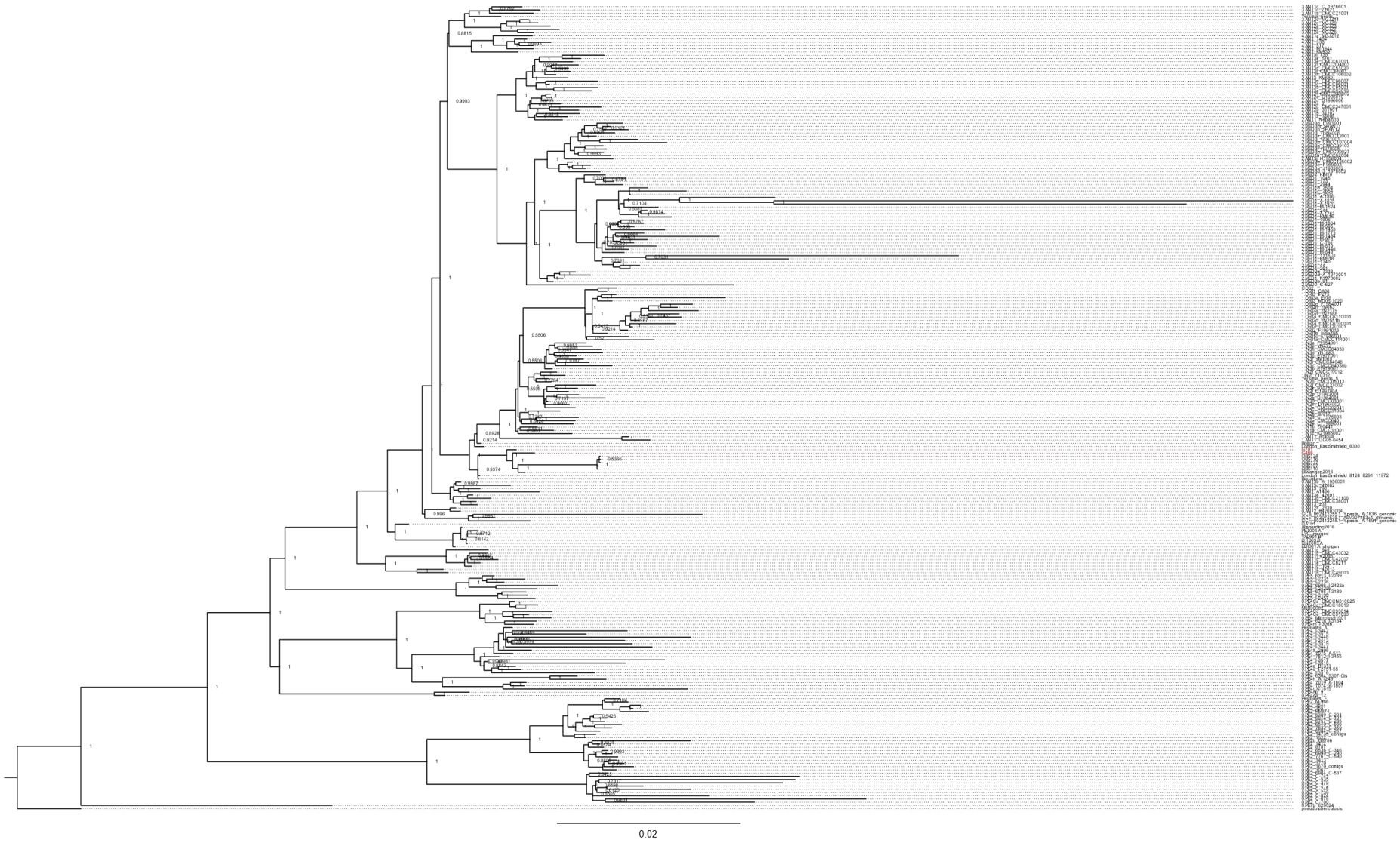


Figure S4. Bayesian tree generated with MrBayes. Posterior probability is shown for one million generations.

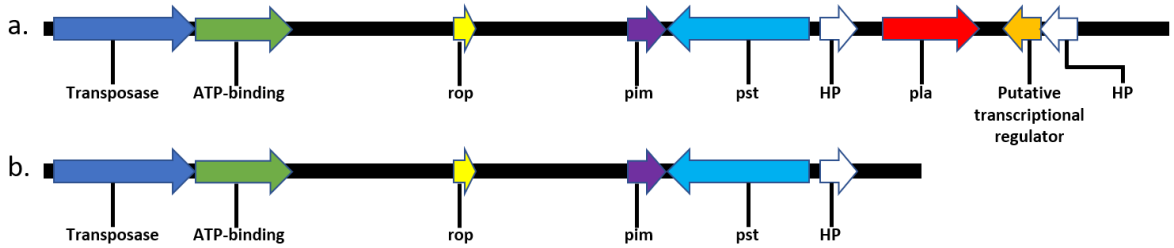


Figure S5. Schematic linear diagram of the pPCP1 plasmid with the *pla* region (*pla+*, a) and without the *pla* region (*pla-*, b). *rop*: replication regulation protein, *pim*: transcriptional regulator, *pst*: pesticin, *pla*: plasminogen activator (PLA) protease, *HP*: hypothetical protein.

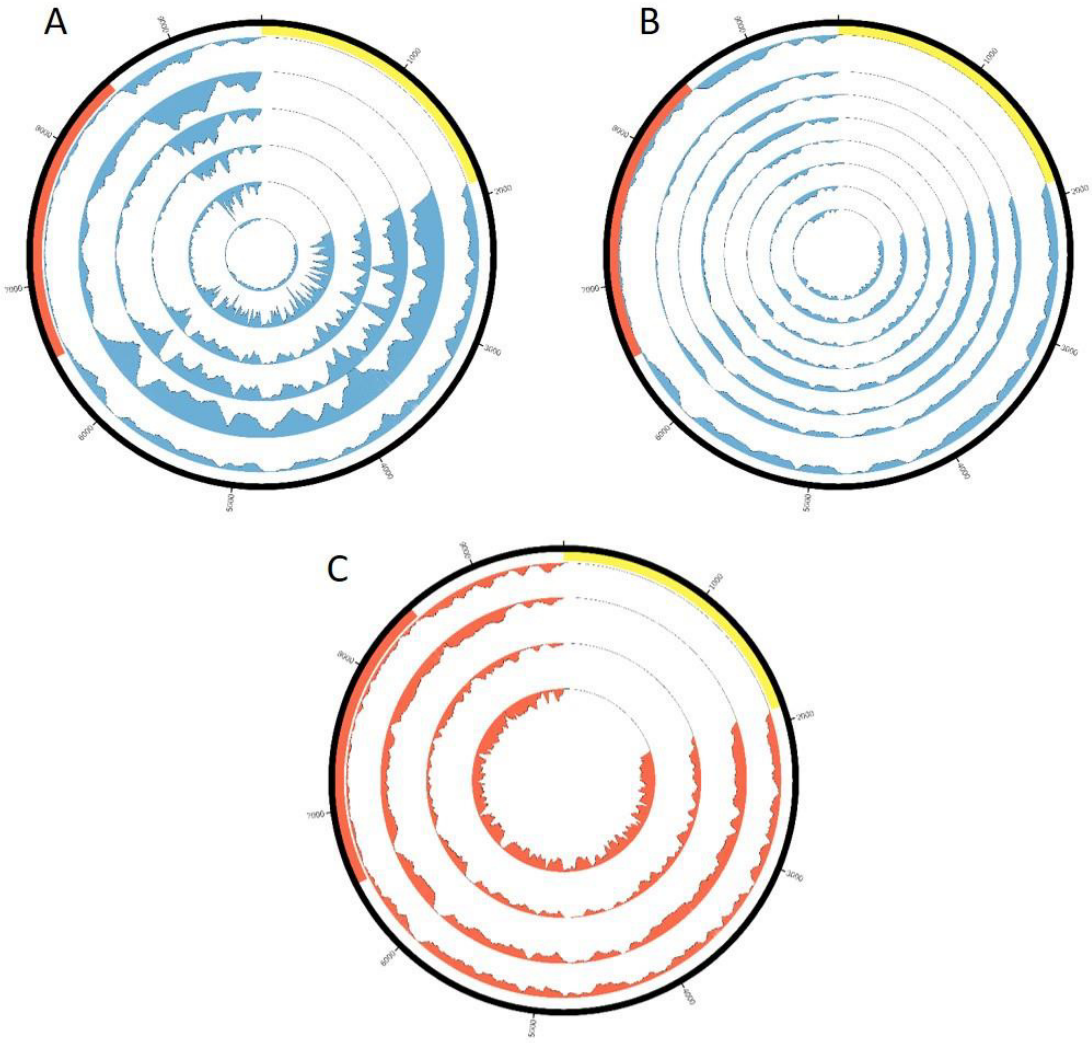


Figure S6. Coverage plots for the pPCP1 plasmid of the 14 strains from the second pandemic that could be analyzed for the depletion of the *pla* region. A and B depict the coverage of *pla*-depleted strains exhibiting gap-spanning reads (see blue colored strains Fig. 1). C depicts the coverage of the strains where no gap-spanning reads could be found (see red colored strains Fig. 1).

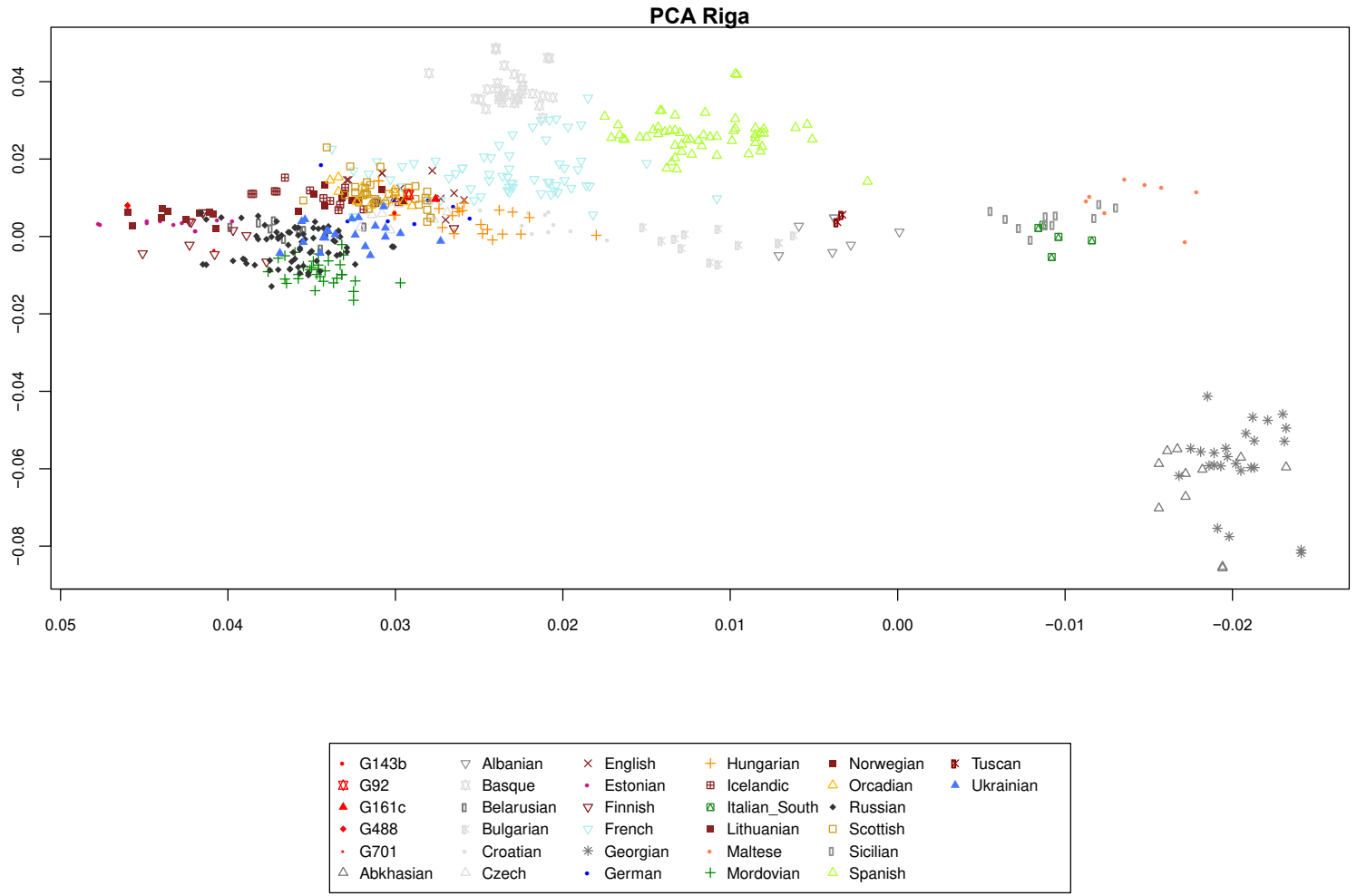


Figure S7. Principal component analysis (PCA) of the five Riga individuals. Riga individuals shown in red were projected onto a basemap of 59 modern-day West Eurasian populations. PC1 is shown on the X-axis and PC2 on the Y-axis.

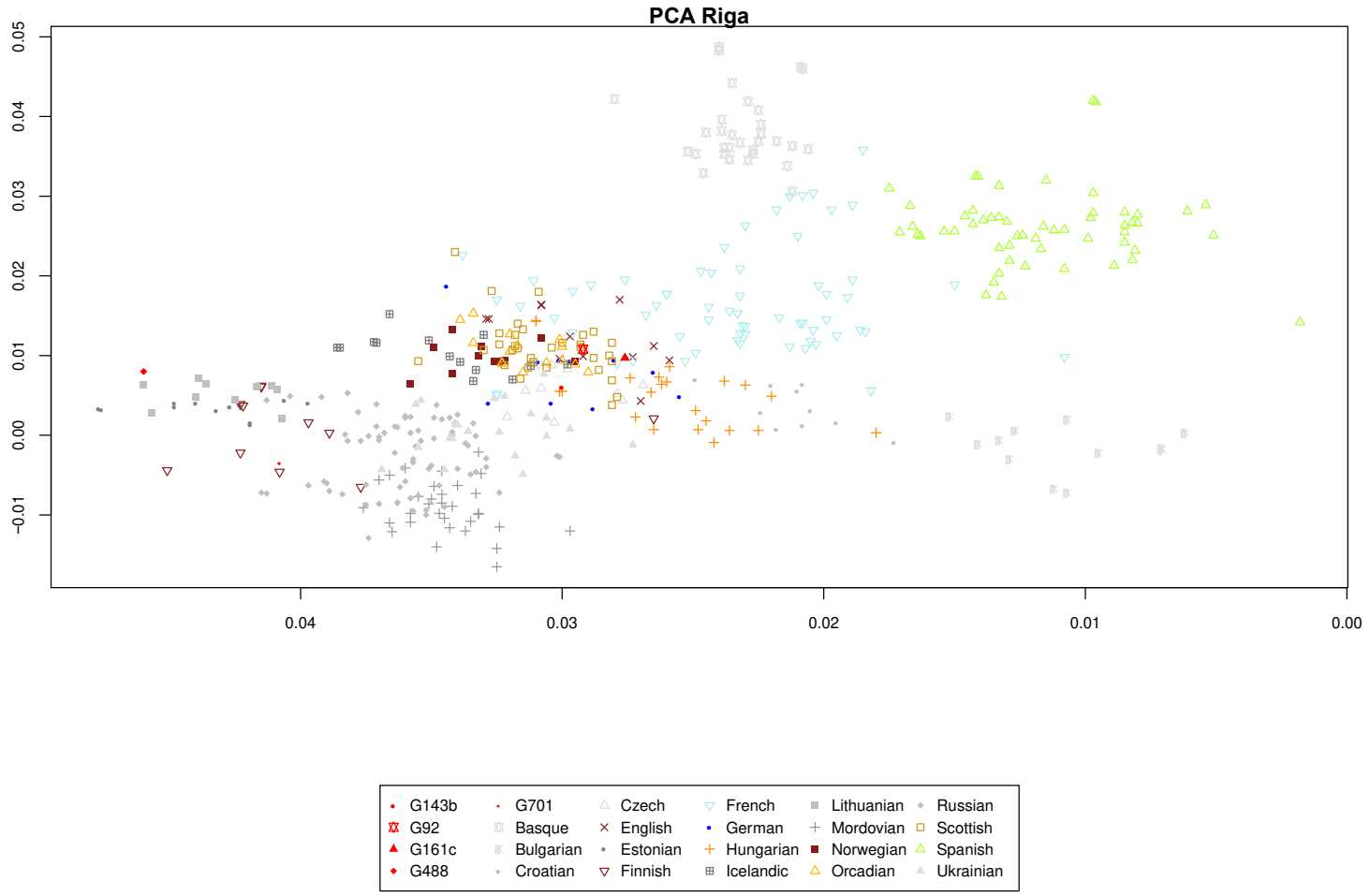


Figure S8. Detail of the principal component analysis (PCA) of the five Riga individuals. Riga individuals shown in red were projected onto a basemap of 59 modern-day West Eurasian populations. PC1 is shown on the X-axis and PC2 on the Y-axis.

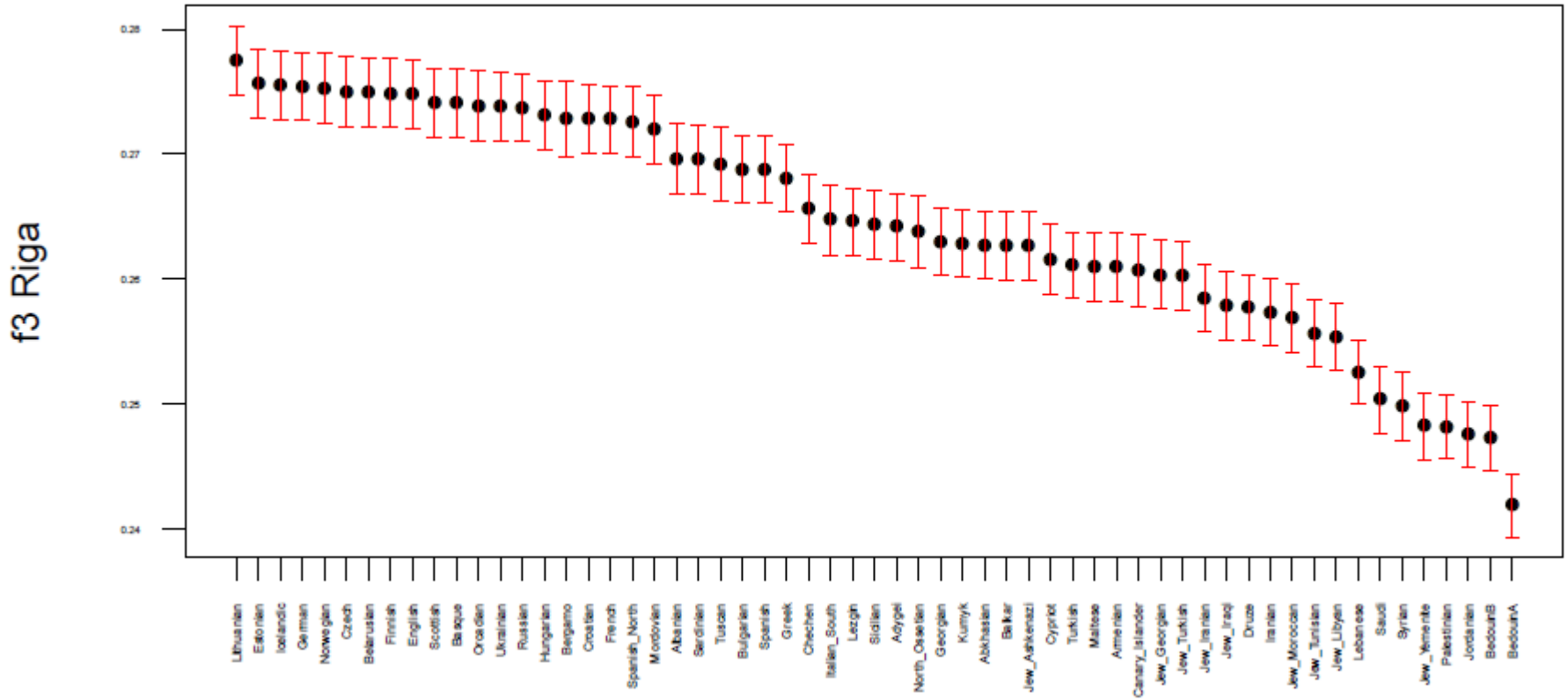


Figure S9. F3 statistics. f3 outgroup statistics *f3(Riga; test, Mbuti)* showing the amount of shared genetic drift between the Riga individuals analysed in this study and each of the 59 modern-day West Eurasian populations (*test*) as used in the PCA and admixture analysis.

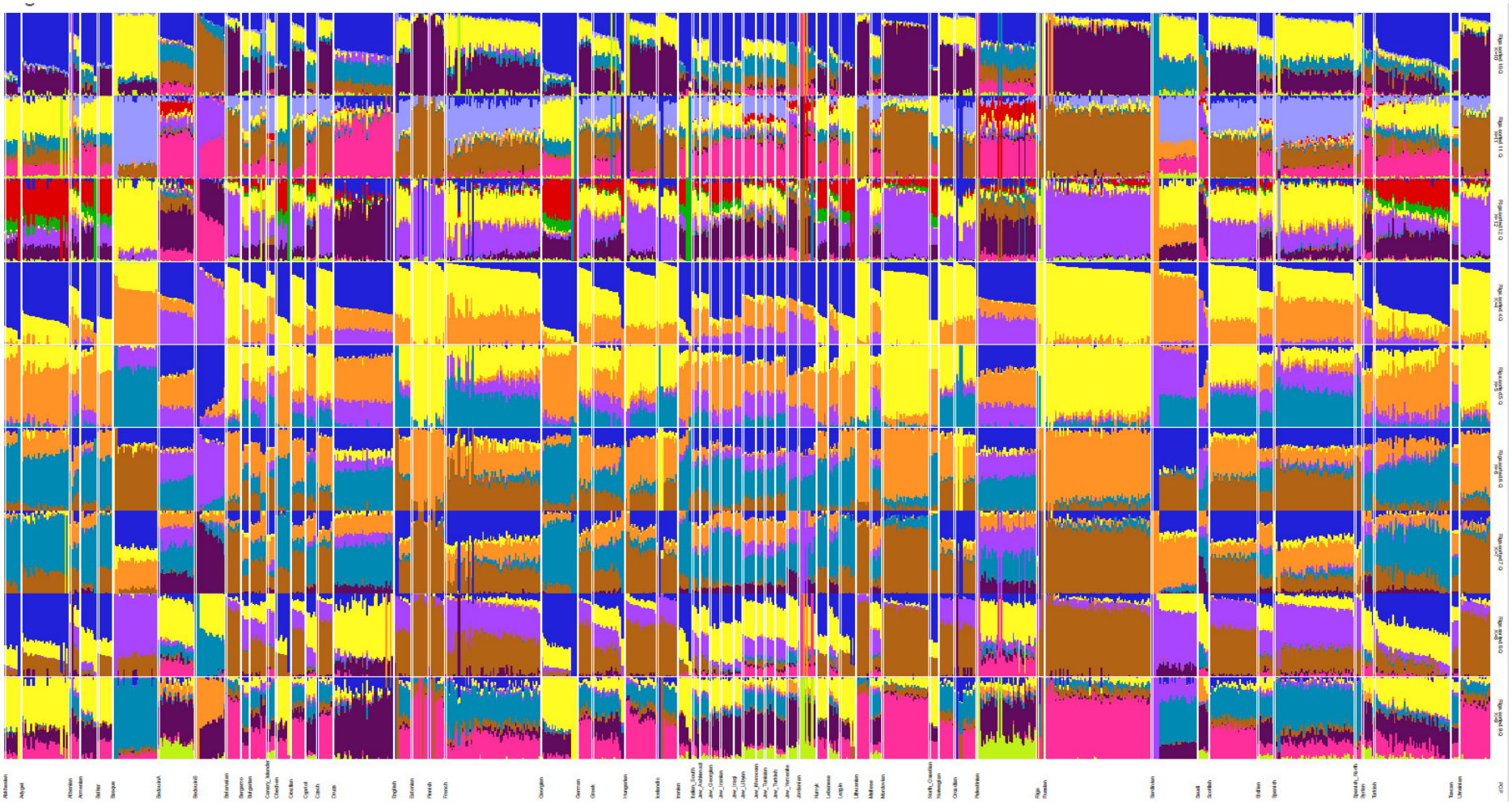


Figure S10. Admixture plot. Admixture analysis of the Riga individuals together with 59 modern-day West Eurasian populations (the same as in the PCA) shown for K=4 to K=12 components.

Table S1. Number of *Y. pestis* reads detected during the initial screening using MALT

Sample	Number of reads mapping to NC_003143.1
G103	8
G488	1489
G645	199
G701	1312

Table S2. *Yersinia* genomes used in the multi fasta reference mapping for authentication of a *Y. pestis*-positive sample

Species name	Strain	NCBI accession number
Y. pestis	CO92	NC_003143.1
Y. pseudotuberculosis	IP 32953	NC_006155.1
Y. enterocolitica	Subsp. enterocolitica 8081	NC_008800.1
Y. aldovae	ATCC 35236	NZ_ACCB01000210.1
Y. bercovie	ATCC 43970	NZ_AALC02000229.1
Y. frederiksenii	ATCC 33641	NZ_AALE02000161.1
Y. intermedia	ATCC 29909	NZ_AALF02000123.1
Y. kristensenii	ATCC 33638	NZ_ACCA01000153.1
Y. mollaretii	ATCC 43969	NZ_AALD02000179.1
Y. rohdei	ATCC 43380	NZ_ACCD01000141.1
Y. ruckeri	ATCC 29473	NZ_ACCC01000174.

Table S3. Score calculation for authentication of *Y. pestis*-positive samples

Sample	YPS	Max(YS)	Reads total in sample	Score
G103	16	133	11662509	-0.0100
G488	1881	181	8277430	0.2053
G645	211	163	11091584	0.0043
G701	1591	113	7930905	0.1863

Table S4. 228 modern strains used in the SNP-based phylogeny

Strain ID	Accession No.	Origin	Reference
0.ANT1a_42013	ADPG00000000	China	Cui et al. 2013
0.ANT1b_CMCC49003	ADQX00000000	China	Cui et al. 2013
0.ANT1c_945	ADPV00000000	China	Cui et al. 2013
0.ANT1d_164	ADOW00000000	China	Cui et al. 2013
0.ANT1e_CMCC8211	ADRD00000000	China	Cui et al. 2013
0.ANT1f_42095	ADPJ00000000	China	Cui et al. 2013
0.ANT1g_CMCC42007	ADQV00000000	China	Cui et al. 2013
0.ANT1h_CMCC43032	ADQW00000000	China	Cui et al. 2013
0.ANT2_B42003004	AAYU00000000	China	Cui et al. 2013
0.ANT2a_2330	ADOY00000000	China	Cui et al. 2013
0.ANT3_231	JMUF00000000	Former Soviet Union	Eroshenko et al. 2017
0.ANT3_790	CP006806	Former Soviet Union	Zhgenti et al. 2015
0.ANT3_A-1486	LYMP	Former Soviet Union	Eroshenko et al. 2017
0.ANT3a_CMCC38001	ADQU00000000	China	Cui et al. 2013
0.ANT3b_A1956001	ADPX00000000	China	Cui et al. 2013
0.ANT3c_42082	ADPH00000000	China	Cui et al. 2013
0.ANT3d_CMCC21106	ADQP00000000	China	Cui et al. 2013
0.ANT3e_42091b	ADPI00000000	China	Cui et al. 2013
0.ANT5_A-1691	LYMQ	Former Soviet Union	Eroshenko et al. 2017
0.ANT5_A-1836	LYOL00000000	Former Soviet Union	Eroshenko et al. 2017
0.PE2_1412	CP006783	Former Soviet Union	Zhgenti et al. 2015
0.PE2_1413	CP006762	Former Soviet Union	Zhgenti et al. 2015
0.PE2_14735	AYLS00000000	Former Soviet Union	Zhgenti et al. 2015
0.PE2_1522	CP006758	Former Soviet Union	Zhgenti et al. 2015
0.PE2_1670	CP006806	Former Soviet Union	Zhgenti et al. 2015
0.PE2_3067	CP006754	Former Soviet Union	Zhgenti et al. 2015
0.PE2_3544	LZNH	Former Soviet Union	Kutyrev et al. 2018
0.PE2_3551	MBSJ00000000	Former Soviet Union	Kutyrev et al. 2018

0.PE2_3770	CP006751	Former Soviet Union	Zhgenti et al. 2015
0.PE2_8787	CP006748	Former Soviet Union	Zhgenti et al. 2015
0.PE2_C-197	LIYX	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-235	LIYY	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-267	LIYZ	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-290	LIYU	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-291	LIZC	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-346	LIZE	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-359	LIZB	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-370	MIDX00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_C-535	MIDY00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_C-537	LIYP	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-590	LIYQ	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-666	LIZF	Former Soviet Union	Kislichkina et al. 2015
0.PE2_C-678	MIDZ00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_C-700	MIEA00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_C-712	MTZW00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_C-739	MTZX00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_C-741	LPTX	Former Soviet Union	Kutyrev et al. 2018
0.PE2_C-746	MTZY00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_C-824	MTZZ00000000	Former Soviet Union	Kislichkina et al. 2017
0.PE2_KM874	LZTG	Former Soviet Union	Kutyrev et al. 2018

0.PE2_M-986	LYMO	Former Soviet Union	Kutyrev et al. 2018
0.PE2_PEST-F	NC_009381	Former Soviet Union	Cui et al. 2013
0.PE2b_G8786	ADSG00000000	Former Soviet Union	Cui et al. 2013
0.PE4_5307-Gis	LIYS	Former Soviet Union	Kislichkina et al. 2015
0.PE4_A-1804	LIYW	Former Soviet Union	Kislichkina et al. 2015
0.PE4_A-1807	LIYT	Former Soviet Union	Kislichkina et al. 2015
0.PE4_A-513	LIZA	Former Soviet Union	Kislichkina et al. 2015
0.PE4_I-3134	LIYR	Former Soviet Union	Kislichkina et al. 2015
0.PE4_I-3442	NHYH00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3443	MIED00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3446	NHYI00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3447	MIEE00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3455	LIYV	Former Soviet Union	Kislichkina et al. 2015
0.PE4_I-3515	NHYJ00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3516	NHMW00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3517	NHMX00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3518	NHMY00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_I-3519	NHMZ00000000	Former Soviet Union	Kislichkina et al. 2018a
0.PE4_M0000002	ADST00000000	China	Cui et al. 2013
0.PE4_Microtus91001	NC_005810	China	Cui et al. 2013
0.PE4a_B1313	LYMS	Former Soviet Union	Kutyrev et al. 2018
0.PE4a_I-2751-55	LYCL00000000	Former Soviet Union	Kutyrev et al. 2018
0.PE4a_I-2998	LYMR	Former Soviet Union	Kutyrev et al. 2018

0.PE4Aa_12	ADOV00000000	China	Cui et al. 2013
0.PE4Ab_9	ADPT00000000	China	Cui et al. 2013
0.PE4Ba_PestoidesA	ACNT00000000	Former Soviet Union	Cui et al. 2013
0.PE4Ca_CMCCN010025	ADRT00000000	China	Cui et al. 2013
0.PE4Cc_CMCC18019	ADQO00000000	China	Cui et al. 2013
0.PE4Cd_CMCC93014	ADRM00000000	China	Cui et al. 2013
0.PE4Ce_CMCC91090	ADRJ00000000	China	Cui et al. 2013
0.PE4h_A-1249	LYMN	Former Soviet Union	Eroshenko et al. 2017
0.PE4m_I-3086	LZNY	Mongolia	Kutyrev et al. 2018
0.PE4t_A-1815	LPTY	Former Soviet Union	Eroshenko et al. 2017
0.PE5_I-2231	PVLX00000000	Mongolia	Kislichkina et al. 2018b
0.PE5_I-2236	PVLZ00000000	Mongolia	Kislichkina et al. 2018b
0.PE5_I-2239	LIZD	Mongolia	Kislichkina et al. 2015
0.PE5_I-2422a	LIZG	Mongolia	Kislichkina et al. 2015
0.PE5_I-2422b	QANK00000000	Mongolia	Kutyrev et al. 2018
0.PE5_I-2457	PVMB00000000	Mongolia	Kislichkina et al. 2018b
0.PE5_I-3189	LIYO	Mongolia	Kislichkina et al. 2015
0.PE5_I-3190	PVLY00000000	Mongolia	Kislichkina et al. 2018b
0.PE7b_620024	ADPM00000000	China	Cui et al. 2013
1.ANT1_Antiqua	NC_008150	Congo	Cui et al. 2013
1.ANT1_UG05-0454	AAYR00000000	Uganda	Cui et al. 2013
1.IN1a_CMCC11001	ADQK00000000	China	Cui et al. 2013
1.IN1b_780441	ADPS00000000	China	Cui et al. 2013
1.IN1c_K21985002	ADSS00000000	China	Cui et al. 2013
1.IN2a_CMCC640047	ADRA00000000	China	Cui et al. 2013
1.IN2b_30017	ADPC00000000	China	Cui et al. 2013
1.IN2c_CMCC31004	ADQR00000000	China	Cui et al. 2013
1.IN2d_C1975003	ADPZ00000000	China	Cui et al. 2013
1.IN2e_C1989001	ADQB00000000	China	Cui et al. 2013
1.IN2f_710317	ADPP00000000	China	Cui et al. 2013
1.IN2g_CMCC05013	ADQF00000000	China	Cui et al. 2013
1.IN2i_CMCC10012	ADQG00000000	China	Cui et al. 2013
1.IN2j_CMCC27002	ADQQ00000000	China	Cui et al. 2013
1.IN2k_970754	ADPW00000000	China	Cui et al. 2013
1.IN2l_D1991004	ADRX00000000	China	Cui et al. 2013
1.IN2m_D1964002b	ADRV00000000	China	Cui et al. 2013
1.IN2n_CMCC02041	ADQC00000000	China	Cui et al. 2013

1.IN2o_CMCC03001	ADQD00000000	China	Cui et al. 2013
1.IN2p_D1982001	ADRW00000000	China	Cui et al. 2013
1.IN2q_D1964001	ADRU00000000	China	Cui et al. 2013
1.IN3a_F1954001	ADSC00000000	China	Cui et al. 2013
1.IN3b_E1979001	AAYV00000000	China	Cui et al. 2013
1.IN3c_CMCC84038b	ADRF00000000	China	Cui et al. 2013
1.IN3d_YN1683	ADTD00000000	China	Cui et al. 2013
1.IN3e_YN472	ADTH00000000	China	Cui et al. 2013
1.IN3f_YN1065	ADTC00000000	China	Cui et al. 2013
1.IN3g_E1977001	ADRY00000000	China	Cui et al. 2013
1.IN3h_CMCC84033	ADRE00000000	China	Cui et al. 2013
1.IN3i_CMCC84046	ADRG00000000	China	Cui et al. 2013
1.ORI1_CA88	ABCD00000000	USA	Cui et al. 2013
1.ORI1_CO92	NC_003143	USA	Cui et al. 2013
1.ORI1a_CMCC114001	ADQL00000000	China	Cui et al. 2013
1.ORI1b_India195	ACNR00000000	India	Cui et al. 2013
1.ORI1c_F1946001	ADSB00000000	China	Cui et al. 2013
1.ORI2_F1991016	ABAT00000000	China	Cui et al. 2013
1.ORI2a_YN2179	ADTE00000000	Myanmar	Cui et al. 2013
1.ORI2c_YN2551b	ADTF00000000	China	Cui et al. 2013
1.ORI2d_YN2588	ADTG00000000	China	Cui et al. 2013
1.ORI2f_CMCC87001	ADRH00000000	China	Cui et al. 2013
1.ORI2g_F1984001	ADSD00000000	China	Cui et al. 2013
1.ORI2h_YN663	ADTI00000000	China	Cui et al. 2013
1.ORI2i_CMCCK100001a	ADRR00000000	China	Cui et al. 2013
1.ORI2i_CMCCK110001b	ADRS00000000	China	Cui et al. 2013
1.ORI3_IP275	AAOS00000000	Madagascar	Cui et al. 2013
1.ORI3_MG05-1020	AAYS00000000	Madagascar	Cui et al. 2013
1.ORI3a_EV76	ADSA00000000	Madagascar	Cui et al. 2013
2.ANT1_Nepal516	ACNQ00000000	Nepal	Cui et al. 2013
2.ANT1a_34008	ADPD00000000	China	Cui et al. 2013
2.ANT1b_34202	ADPE00000000	China	Cui et al. 2013
2.ANT2a_2	ADOX00000000	China	Cui et al. 2013
2.ANT2b_351001	ADPF00000000	China	Cui et al. 2013
2.ANT2c_CMCC347001	ADQS00000000	China	Cui et al. 2013
2.ANT2d_G1996006	ADSE00000000	China	Cui et al. 2013
2.ANT2e_G1996010	ADSF00000000	China	Cui et al. 2013
2.ANT2f_CMCC348002	ADQT00000000	China	Cui et al. 2013
2.ANT3_KM682	LPVG	Former Soviet Union	Kutyrev et al. 2018
2.ANT3a_CMCC92010	ADRL00000000	China	Cui et al. 2013
2.ANT3b_CMCC95001	ADRN00000000	China	Cui et al. 2013
2.ANT3c_CMCC96001	ADRO00000000	China	Cui et al. 2013
2.ANT3d_CMCC96007	ADRP00000000	China	Cui et al. 2013

2.ANT3e_CMCC67001	ADRB00000000	China	Cui et al. 2013
2.ANT3f_CMCC104003	ADQH00000000	China	Cui et al. 2013
2.ANT3g_CMCC51020	ADQY00000000	China	Cui et al. 2013
2.ANT3h_CMCC106002	ADQI00000000	China	Cui et al. 2013
2.ANT3i_CMCC64001	ADQZ00000000	China	Cui et al. 2013
2.ANT3j_H1959004	ADSI00000000	China	Cui et al. 2013
2.ANT3k_5761	ADPL00000000	Russia	Cui et al. 2013
2.ANT3l_735	ADPR00000000	Russia	Cui et al. 2013
2.MED0_C-627	MBSI00000000	Former Soviet Union	Kutyrev et al. 2018
2.MED1_1045	CP006794	Former Soviet Union	Zhgenti et al. 2015
2.MED1_1116-D	LPXS	Former Soviet Union	Kutyrev et al. 2018
2.MED1_1240	LZNI	Former Soviet Union	Kutyrev et al. 2018
2.MED1_139	QAPA00000000	Former Soviet Union	Kutyrev et al. 2018
2.MED1_173	LQAZ	Former Soviet Union	Kutyrev et al. 2018
2.MED1_1906	LYOM	Former Soviet Union	Kutyrev et al. 2018
2.MED1_244	LZND	Former Soviet Union	Kutyrev et al. 2018
2.MED1_261	LZNG	Former Soviet Union	Kutyrev et al. 2018
2.MED1_2944	CP006792	Former Soviet Union	Zhgenti et al. 2015
2.MED1_44	LZNF	Former Soviet Union	Kutyrev et al. 2018
2.MED1_A-1763	LQAW	Former Soviet Union	Kutyrev et al. 2018
2.MED1_A-1809	LYMF	Former Soviet Union	Eroshenko et al. 2017
2.MED1_A-1825	LYCM	Former Soviet Union	Kutyrev et al. 2018
2.MED1_A-1920	LYCO	Former Soviet Union	Kutyrev et al. 2018
2.MED1_C-791	LQAU	Former Soviet Union	Kutyrev et al. 2018
2.MED1_K1973002	AAYT00000000	China	Cui et al. 2013
2.MED1_KIM10	NC_004088	Iran	Cui et al. 2013
2.MED1_KM816	LPXU	Former Soviet Union	Kutyrev et al. 2018

2.MED1_KM918	LPQY	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-1448	LYCN	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-1453	LQAY	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-1484	LQAV	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-1524	LYCP	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-1773	LYMG	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-1864	LOHR	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-519	LQAX	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-549	LQBA	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-595	LYOH	Former Soviet Union	Kutyrev et al. 2018
2.MED1_M-978	LPXT	Former Soviet Union	Kutyrev et al. 2018
2.MED1b_2506	ADPA00000000	China	Cui et al. 2013
2.MED1c_2654	ADPB00000000	China	Cui et al. 2013
2.MED1d_2504	ADOZ00000000	China	Cui et al. 2013
2.MED2b_91	ADPU00000000	China	Cui et al. 2013
2.MED2c_K11973002	AAYT00000000	China	Cui et al. 2013
2.MED2d_A1973001	ADPY00000000	China	Cui et al. 2013
2.MED2e_7338	ADPQ00000000	China	Cui et al. 2013
2.MED3a_J1963002	ADSP00000000	China	Cui et al. 2013
2.MED3b_CMCC125002b	ADQN00000000	China	Cui et al. 2013
2.MED3c_I1969003	ADSK00000000	China	Cui et al. 2013
2.MED3d_J1978002	ADSQ00000000	China	Cui et al. 2013
2.MED3f_I1970005	ADSL00000000	China	Cui et al. 2013
2.MED3g_CMCC99103	ADRQ00000000	China	Cui et al. 2013
2.MED3h_CMCC90027	ADRI00000000	China	Cui et al. 2013
2.MED3i_CMCC92004	ADRK00000000	China	Cui et al. 2013
2.MED3j_I2001001	ADSO00000000	China	Cui et al. 2013
2.MED3k_CMCC12003	ADQM00000000	China	Cui et al. 2013
2.MED3l_I1994006	ADSN00000000	China	Cui et al. 2013
2.MED3m_SHAN11	ADTA00000000	China	Cui et al. 2013
2.MED3n_SHAN12	ADTB00000000	China	Cui et al. 2013
2.MED3o_I1991001	ADSM00000000	China	Cui et al. 2013
2.MED3p_CMCC107004	ADQJ00000000	China	Cui et al. 2013

3.ANT1a_7b	ADPN00000000	China	Cui et al. 2013
3.ANT1b_CMCC71001	ADRC00000000	China	Cui et al. 2013
3.ANT1c_C1976001	ADQA00000000	China	Cui et al. 2013
3.ANT1d_71021	ADPO00000000	China	Cui et al. 2013
3.ANT2a_MGJZ6	ADSX00000000	Mongolia	Cui et al. 2013
3.ANT2b_MGJZ7	ADSY00000000	Mongolia	Cui et al. 2013
3.ANT2c_MGJZ9	ADSZ00000000	Mongolia	Cui et al. 2013
3.ANT2d_MGJZ11	ADSU00000000	Mongolia	Cui et al. 2013
3.ANT2e_MGJZ3	ADSW00000000	Mongolia	Cui et al. 2013
4.ANT_1454	LZNC	Former Soviet Union	Kutyrev et al. 2018
4.ANT_338	LZNX	Former Soviet Union	Kutyrev et al. 2018
4.ANT_517	LYMH	Former Soviet Union	Kutyrev et al. 2018
4.ANT_KM932	LZNE	Former Soviet Union	Kutyrev et al. 2018
4.ANT_M-1944	LYOK	Former Soviet Union	Kutyrev et al. 2018
4.ANT1a_MGJZ12	ADSV00000000	Mongolia	Cui et al. 2013

Table S5. 36 ancient strains used in the SNP-based phylogeny

Strain ID	Accession No.	Origin	Reference
DA101	PRJEB25891	Kyrgyzstan	Damgaard et al. 2018
Altenerding	PRJEB14851	Germany	Feldman et al. 2016
London_EastSmithfield_8124_8291_11972	SRR341961,SRR341962,SRR341963	United Kingdom	Bos et al. 2011
London_EastSmithfield_6330	SAMN00715799	United Kingdom	Bos et al. 2011
Ellwangen	PRJEB13664	Germany	Spyrou et al. 2016
Bolgar	PRJEB13664	Russia	Spyrou et al. 2016
Barcelona	PRJEB13664	Spain	Spyrou et al. 2016
OBS107	PRJEB12163	France	Bos et al. 2016
OBS110	PRJEB12163	France	Bos et al. 2016
OBS116	PRJEB12163	France	Bos et al. 2016
OBS124	PRJEB12163	France	Bos et al. 2016
OBS137	PRJEB12163	France	Bos et al. 2016
EDI001.A	SAMEA5661363	United Kingdom	Keller et al. 2019
LVC_merged	SAMEA5661382, SAMEA5661378, SAMEA5661380, SAMEA5661379	France	Keller et al. 2019
DIT003.B	SAMEA5661360	Germany	Keller et al. 2019
PET004.A	SAMEA5661385	Germany	Keller et al. 2019
UNT003.A	SAMEA5661389	Germany	Keller et al. 2019
VAL001.B	SAMEA5661384	Spain	Keller et al. 2019
LAI009	PRJEB29990	Russia	Spyrou et al. 2019
NAB003	PRJEB29990	Germany	Spyrou et al. 2019
MAN008	PRJEB29990	Germany	Spyrou et al. 2019
STA001	PRJEB29990	Germany	Spyrou et al. 2019
NMS002	PRJEB29990	Cambridge	Spyrou et al. 2019
BRA001	PRJEB29990	United Kingdom	Spyrou et al. 2019
BED024	PRJEB29990	United Kingdom	Spyrou et al. 2019
BED028	PRJEB29990	United Kingdom	Spyrou et al. 2019
BED030	PRJEB29990	United Kingdom	Spyrou et al. 2019
BED034	PRJEB29990	United Kingdom	Spyrou et al. 2019
STN002	PRJEB29990	Sweden	Spyrou et al. 2019
STN007	PRJEB29990	Sweden	Spyrou et al. 2019
STN008	PRJEB29990	Sweden	Spyrou et al. 2019
STN013	PRJEB29990	Sweden	Spyrou et al. 2019
STN014	PRJEB29990	Sweden	Spyrou et al. 2019
STN019	PRJEB29990	Sweden	Spyrou et al. 2019
STN020	PRJEB29990	Sweden	Spyrou et al. 2019
STN021	PRJEB29990	Sweden	Spyrou et al. 2019

Table S6. SNP description of diagnostic branch-1 positions in the newly sequenced *Y. pestis* genomes G701 and G488

SNP name	Chromosome position in CO92	CO92	G701	G488	Gene
p1	189227	C	C	C	pabA
p2	1871476	G	G	G	NC
p3	699494	A	G	G	alt(rpoD)
p4	2262577	T	G	G	YPO1990
p5	4301295	G	G	G	recQ
p6	3806677	C	T	T	b0125(hpt)
p7	3643387	G	G	G	YP03271

Table S7. References used in the competitive alignment for the *pla* gene

Reference	Organism	Type
NC_003143.1	Y. *pestis*	chromosome
NC_003134.1	Y. *pestis*	plasmid
NC_003132.1	Y. *pestis*	plasmid
NC_003131.1	Y. *pestis*	plasmid
LK931337.1	*Citrobacter koseri*	scaffold
LM995843.1	*Escherichia coli*	scaffold

Table S8. References used in the analysis of the putative gap-bridging reads

NC_003143.1
NC_003134.1
NC_003132.1
pla+_NC_003131.1
pla-_NC_003132.1

Table S9. Basic coverage information of the *pla+* and *pla-* plasmids and calculated ratio

Strain	Estimated coverage of *pla+*	Estimated coverage *pla-*	Ratio *pla+/pla-*
G701	3x	32x	1:11
G488	7x	83x	1:11

Table S10. Basic coverage information of the *pla+* and *pla-* plasmid in the 14 re-analyzed strains[*] and calculated ratio

Strain	Estimated coverage of *pla+*	Estimated coverage of *pla-*	Ratio *pla+/pla-*
BED024	9	36	1:4
BED028	22	67	1:3
BED030	46	130	1:3
BED034	12	37	1:3
BRA001	19	26	1:1
NMS002	10	31	1:3
STN002	8	17	1:2
STN007	10	40	1:4
STN008	5	18	1:4
STN013	6	19	1:3
STN014	28	46	1:2
STN019	10	28	1:3
STN020	16	41	1:3
STN021	12	30	1:3

[*]Data for these strains were generated using a *Y. pestis* capture[18]

11.

Chapter II: Late infection diagnosis: Rudolf Virchow´s skull collection indicates 5000-year-old hunter-gatherer already plagued by *Yersinia pestis*

Supplementary Material

Supplementary appendix

Title:

Late infection diagnosis: Rudolf Virchow's skull collection indicates 5000-year-old hunter-gatherer already plagued by *Yersinia pestis*

Authors:

Julian Susat, Harald Lübke, Alexander Immel, Ute Brinker, Aija Macane, John Meadows, Britta Steer, Andreas Tholey, Ilga Zagorska, Guntis Gerhards, Ulrich Schmölcke, Mārcis Kalniņš, Andre Franke, Elīna Pētersone-Gordina, Barbara Tessman, Mari Tõrv, Stefan Schreiber, Christian Andree, Valdis Bērziņš, Almut Nebel, Ben Krause-Kyora

Osteological context of reported individuals

The biological sex and age at death estimations, burial descriptions are given after Lübke and Brinker.[1,2]

RV 1852: Female, 12–18 years, buried on back, arms flexed from the elbow, with lower limbs slightly flexed to the left lateral side of the body. The head lay to the northeast, the feet to the southwest. The total length without skull, measured from the atlas vertebra to the toes, was 1.31 m. Next the lumbar spine, partly on the pelvis bone, a lump of fish scales and bones was found; the right hand was placed over it. The former almost entirely preserved skeleton, including the smallest finger bones, has not survived. The skull was shattered and has been restored in the 19th century; the associated mandible is almost completely preserved. Diagnostic findings: very slight attrition, sinusitis maxillaris, cribra orbitalia, ectocranial porosities, healed depressed lesion at the right frontal bone.

RV 2039: Male 20–30 years, buried on back, with the head directed towards NE and facing forwards, with the right arm and forearm slightly flexed in front of the body. Below the pelvis the burial was disturbed by another grave. The skull was fragmented post-mortem and has been restored in the 19th century; the associated mandible is almost completely preserved, but the postcranial remains have not survived. Diagnostic findings: dental calculus, especially in the lingual part of the mandible, periodontopathies, slight attrition, sinusitis maxillaris, cribra orbitalia, ectocranial porosities.

2017/1: Male 35–45 years, buried on back with the head pointing northwards and facing forwards, arms flexed from the elbow, hands pronated, with lower limbs flexed at the hip and knee to the left lateral side of the body. The skeleton has survived almost intact, but the skull was heavily fragmented. The man was about 166 cm tall, right-handed. Diagnostic findings: massive dental calculus, periodontopathies. minor caries, slight attrition, cribra orbitalia; changes in muscle attachment spots of the right humerus that go back to high physical strain, heavy loads or continually repeated movements in general; a healed lesion on a thoracic vertebra due to a disc herniation, probably caused by a single traumatic event happened in his youth; a few degenerative probably age-related skeletal changes (e.g. formation of osteophytes on the spine, slight arthrosis in the shoulder joints).

2018/1: Newborn between 38–40 prenatal weeks, buried on its stomach with the head directed towards NE, with the left arm adducted next to the left lateral side of the body, forearms and hands placed beneath the upper body, with lower limbs slightly flexed at the hip and knee to the right lateral side of the body. The skeleton is for the most part complete, but the upper thoracic area was disturbed, the skull was heavily fragmented. The length of skeleton is 51 cm.

Proteomic analysis

Sample preparation for bottom-up LC-MS/MS analysis

After adding 500 µl of 0.5 M EDTA in MilliQ water, the powdered teeth sample (50 mg) was homogenized with a pellet pestle and incubated overnight at 20°C to demineralize the bone matrix and extract the proteins.

The sample was then centrifuged for 5 minutes at 20°C and 14.000 g to collect the supernatant (EDTA-fraction) which was stored on ice. The pellet was resuspended in 300 µl of a Lysis buffer (4 % SDS, 0.1 M dithiothreitol (DTT), 0.1 M Tris-HCl), homogenized with a pellet pestle and incubated for 10 minutes at 95°C for further demineralization and reduction. Afterwards the sample was cooled down and centrifuged for 5 minutes at 20°C and 14.000 g to collect the supernatant (SDS-fraction).

To reduce the disulfide bonds (EDTA and SDS-fraction), 75 µl of a reduction buffer (0.2 M DTT in 50 mM ABC, 5 % ACN) was added and incubated for 30 minutes at 60°C and the sample was cooled down prior to alkylation. For alkylation both fractions were incubated at 20 minutes in the dark with 75 µl 0.8 M IAA in 50 mM ABC, 5 % acetonitrile (ACN).

Tryptic digestion was performed using the single-pot, solid phase enhanced sample preparation (SP3) protocol. The bead stock was prepared by adding 110 µl hydrophilic to the same amount of hydrophobic beads which are delivered in 0.05 % azide solution. After washing three times with 1.1 ml MilliQ water and collecting the beads with the help of magnets, 550 µl MilliQ water was added to get to a final bead concentration of 20 mg/ml. 750 µl of ethanol (96 %) and 50 µl of the bead solution was added to the EDTA-fraction and 450 µl of ethanol (96 %) and 50 µl of the bead solution was added to the SDS-fraction. Both was cautiously mixed by swinging lightly and incubated for 18 minutes at room temperature. The supernatant was then removed carefully by the use of a magnet and the samples were washed 3 times with 150 µl 80 % ethanol to remove

contaminants and salts from the previous steps. 150 µl digestion buffer (50 mM ammonium bicarbonate (ABC)) and 10 µl of 40 ng/µl trypsin (sequencing grade, Promega) was added. Digestion took place overnight at 37°C. The supernatant was collected with the help of magnet and the peptide solution was cleaned with house made C18 Stage tips.

The stage tips were conditioned with first 150 µl methanol, second 150 µl 80 % ACN, 0.5 % acetic acid and third 150 µl 0.5 % acetic acid. Afterwards the samples were loaded on the stage tips and washed with 150 µl 0.5 % acetic acid. The elution took place by first adding 40 µl 40 % ACN, 0.5 % acetic acid, 40 µl 60 % ACN, 0.5 % acetic acid and two times with 40 µl 80 % ACN, 0.5 % acetic acid. The samples were dried in a Speed Vac and stored at -20°C.

LC-MS/MS analysis

The samples were resuspended in 15 µl loading buffer (3 % ACN, 0.1 % TFA). Chromatographic separation was performed on an Ultimate 3000 UHPLC system (Thermo, Dreieich, Germany) equipped with an Acclaim PepMap C18 2UM 75UMx500MM NV FS column (Thermo, Dreieich, Germany) coupled online to a mass spectrometer. The eluents that were used were A: 0.05 % formic acid and B: 80 % ACN, 0.04 % formic acid.

Initial chromatographic conditions were isocratic 4 % B for 2 minutes followed by different gradients for the EDTA- and SDS-fractions, respectively, as the peaks in the EDTA run elute later than the peaks in the SDS run.

EDTA-fraction: from minute 2 to 92 (eluent B 5 % to 50 %), from minute 92 to 97 (eluent B 50 % to 95 %), from minute 97 to 107 (Eluent B 95 %) and from minute 107 to 120 (Eluent B 5 %).

SDS-fraction: from minute 2 to 30 (eluent B 10 % to 30 %), from minute 30 to 92 (eluent B 30 % to 55 %) and from minute 92 to 97 (eluent B 55 % to 95 %).

A constant flow rate of 300 nl / minute was used and 5 µl of the sample was injected per run.

LC-separation was coupled online to an Orbitrap Fusion Lumos mass spectrometer (Thermo, Dreieich, Germany) utilizing HCD ion activation at a collision energy of 30 %. A full scan MS acquisition was performed (resolution 120.000) with subsequent data dependent MS/MS (resolution 30.000) with a cycle time of 3 seconds.

Data analysis

The MS data files were searched against a set of FASTA-databases containing human proteins (Uniprot download 2019/03/13), *Y. pestis* proteins (Uniprot download 2019/03/14) and a cRAP list of common laboratory contaminants. The searches were performed using the Proteome Discoverer software (Version 2.2.0.388) and the SequestHT search algorithm. A false discovery rate (FDR) for peptides and proteins from 0,01 (strict) to 0,05 (relaxed) was applied. A maximum of two missed cleavage sites, a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0,02 Da were allowed. Dynamic modifications: oxidation (M, P); deamidation (N, Q). Static Modification: Carbamidomethylation (C).

Identified peptides were additionally interrogated with the PepQuery-algorithm.[3]

Table 1. Results of osteological and population genetic analysis

Burial	Material	osteological sex	Age at death	Genetic sex	Mt Haplotype	Y Haplotype
RV 1852	Petrous bone / tooth	Female	12-18	Female	U5a1d1	-
RV 2039	tooth	Male	20-30	Male	U5a2b2	BT
2017/01	Petrous bone / tooth	Male	35-45	Male	U4a1	K
2018/1	Petrous bone	-	38–40 prenatal weeks	Male	U5a1d1	CT

Table 2. 228 modern and 48 ancient strains used in the SNP-based phylogeny

Modern strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
0.ANT1a_42013	ADPG00000000	China	Cui et al. 2013[4]
0.ANT1b_CMCC49003	ADQX00000000	China	Cui et al. 2013[4]
0.ANT1c_945	ADPV00000000	China	Cui et al. 2013[4]
0.ANT1d_164	ADOW00000000	China	Cui et al. 2013[4]
0.ANT1e_CMCC8211	ADRD00000000	China	Cui et al. 2013[4]
0.ANT1f_42095	ADPJ00000000	China	Cui et al. 2013[4]
0.ANT1g_CMCC42007	ADQV00000000	China	Cui et al. 2013[4]
0.ANT1h_CMCC43032	ADQW00000000	China	Cui et al. 2013[4]
0.ANT2_B42003004	AAYU00000000	China	Cui et al. 2013[4]
0.ANT2a_2330	ADOY00000000	China	Cui et al. 2013[4]
0.ANT3_231	JMUF00000000	Former Soviet Union	Eroshenko et al. 2017[5]
0.ANT3_790	CP006806	Former Soviet Union	Zhgenti et al. 2015[6]
0.ANT3_A-1486	LYMP	Former Soviet Union	Eroshenko et al. 2017[5]
0.ANT3a_CMCC38001	ADQU00000000	China	Cui et al. 2013[4]
0.ANT3b_A1956001	ADPX00000000	China	Cui et al. 2013[4]
0.ANT3c_42082	ADPH00000000	China	Cui et al. 2013[4]
0.ANT3d_CMCC21106	ADQP00000000	China	Cui et al. 2013[4]
0.ANT3e_42091b	ADPI00000000	China	Cui et al. 2013[4]
0.ANT5_A-1691	LYMQ	Former Soviet Union	Eroshenko et al. 2017[5]
0.ANT5_A-1836	LYOL00000000	Former Soviet Union	Eroshenko et al. 2017[5]
0.PE2_1412	CP006783	Former Soviet Union	Zhgenti et al. 2015[6]
0.PE2_1413	CP006762	Former Soviet Union	Zhgenti et al. 2015[6]
0.PE2_14735	AYLS00000000	Former Soviet Union	Zhgenti et al. 2015[6]

0.PE2_1522	CP006758	Former Soviet Union	Zhgenti et al. 2015[6]
0.PE2_1670	CP006806	Former Soviet Union	Zhgenti et al. 2015[6]
0.PE2_3067	CP006754	Former Soviet Union	Zhgenti et al. 2015[6]
0.PE2_3544	LZNH	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE2_3551	MBSJ00000000	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE2_3770	CP006751	Former Soviet Union	Zhgenti et al. 2015[6]
0.PE2_8787	CP006748	Former Soviet Union	Zhgenti et al. 2015[6]
0.PE2_C-197	LIYX	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-235	LIYY	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-267	LIYZ	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-290	LIYU	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-291	LIZC	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-346	LIZE	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-359	LIZB	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-370	MIDX00000000	Former Soviet Union	Kislichkina et al. 2017[9]

0.PE2_C-535	MIDY00000000	Former Soviet Union	Kislichkina et al. 2017[9]
0.PE2_C-537	LIYP	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-590	LIYQ	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-666	LIZF	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE2_C-678	MIDZ00000000	Former Soviet Union	Kislichkina et al. 2017[9]
0.PE2_C-700	MIEA00000000	Former Soviet Union	Kislichkina et al. 2017[9]
0.PE2_C-712	MTZW00000000	Former Soviet Union	Kislichkina et al. 2017[9]
0.PE2_C-739	MTZX00000000	Former Soviet Union	Kislichkina et al. 2017[9]
0.PE2_C-741	LPTX	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE2_C-746	MTZY00000000	Former Soviet Union	Kislichkina et al. 2017[9]
0.PE2_C-824	MTZZ00000000	Former Soviet Union	Kislichkina et al. 2017[9]
0.PE2_KM874	LZTG	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE2_M-986	LYMO	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE2_PEST-F	NC_009381	Former Soviet Union	Cui et al. 2013[4]
0.PE2b_G8786	ADSG00000000	Former Soviet Union	Cui et al. 2013[4]

0.PE4_5307-Gis	LIYS	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE4_A-1804	LIYW	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE4_A-1807	LIYT	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE4_A-513	LIZA	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE4_I-3134	LIYR	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE4_I-3442	NHYH00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3443	MIED00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3446	NHYI00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3447	MIEE00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3455	LIYV	Former Soviet Union	Kislichkina et al. 2015[8]
0.PE4_I-3515	NHYJ00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3516	NHMW00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3517	NHMX00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3518	NHMY00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_I-3519	NHMZ00000000	Former Soviet Union	Kislichkina et al. 2018[10]
0.PE4_M0000002	ADST00000000	China	Cui et al. 2013[4]
0.PE4_Microtus91001	NC_005810	China	Cui et al. 2013[4]

0.PE4a_B1313	LYMS	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE4a_I-2751-55	LYCL00000000	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE4a_I-2998	LYMR	Former Soviet Union	Kutyrev et al. 2018[7]
0.PE4Aa_12	ADOV00000000	China	Cui et al. 2013[4]
0.PE4Ab_9	ADPT00000000	China	Cui et al. 2013[4]
0.PE4Ba_PestoidesA	ACNT00000000	Former Soviet Union	Cui et al. 2013[4]
0.PE4Ca_CMCCN010025	ADRT00000000	China	Cui et al. 2013[4]
0.PE4Cc_CMCC18019	ADQO00000000	China	Cui et al. 2013[4]
0.PE4Cd_CMCC93014	ADRM00000000	China	Cui et al. 2013[4]
0.PE4Ce_CMCC91090	ADRJ00000000	China	Cui et al. 2013[4]
0.PE4h_A-1249	LYMN	Former Soviet Union	Eroshenko et al. 2017[5]
0.PE4m_I-3086	LZNY	Mongolia	Kutyrev et al. 2018[7]
0.PE4t_A-1815	LPTY	Former Soviet Union	Eroshenko et al. 2017[5]
0.PE5_I-2231	PVLX00000000	Mongolia	Kislichkina et al. 2018[11]
0.PE5_I-2236	PVLZ00000000	Mongolia	Kislichkina et al. 2018[11]
0.PE5_I-2239	LIZD	Mongolia	Kislichkina et al. 2015[8]
0.PE5_I-2422a	LIZG	Mongolia	Kislichkina et al. 2015[8]
0.PE5_I-2422b	QANK00000000	Mongolia	Kutyrev et al. 2018[7]
0.PE5_I-2457	PVMB00000000	Mongolia	Kislichkina et al. 2018[11]
0.PE5_I-3189	LIYO	Mongolia	Kislichkina et al. 2015[8]
0.PE5_I-3190	PVLY00000000	Mongolia	Kislichkina et al. 2018[11]
0.PE7b_620024	ADPM00000000	China	Cui et al. 2013[4]
1.ANT1_Antiqua	NC_008150	Congo	Cui et al. 2013[4]
1.ANT1_UG05-0454	AAYR00000000	Uganda	Cui et al. 2013[4]
1.IN1a_CMCC11001	ADQK00000000	China	Cui et al. 2013[4]
1.IN1b_780441	ADPS00000000	China	Cui et al. 2013[4]

1.IN1c_K21985002	ADSS00000000	China	Cui et al. 2013[4]
1.IN2a_CMCC640047	ADRA00000000	China	Cui et al. 2013[4]
1.IN2b_30017	ADPC00000000	China	Cui et al. 2013[4]
1.IN2c_CMCC31004	ADQR00000000	China	Cui et al. 2013[4]
1.IN2d_C1975003	ADPZ00000000	China	Cui et al. 2013[4]
1.IN2e_C1989001	ADQB00000000	China	Cui et al. 2013[4]
1.IN2f_710317	ADPP00000000	China	Cui et al. 2013[4]
1.IN2g_CMCC05013	ADQF00000000	China	Cui et al. 2013[4]
1.IN2i_CMCC10012	ADQG00000000	China	Cui et al. 2013[4]
1.IN2j_CMCC27002	ADQQ00000000	China	Cui et al. 2013[4]
1.IN2k_970754	ADPW00000000	China	Cui et al. 2013[4]
1.IN2l_D1991004	ADRX00000000	China	Cui et al. 2013[4]
1.IN2m_D1964002b	ADRV00000000	China	Cui et al. 2013[4]
1.IN2n_CMCC02041	ADQC00000000	China	Cui et al. 2013[4]
1.IN2o_CMCC03001	ADQD00000000	China	Cui et al. 2013[4]
1.IN2p_D1982001	ADRW00000000	China	Cui et al. 2013[4]
1.IN2q_D1964001	ADRU00000000	China	Cui et al. 2013[4]
1.IN3a_F1954001	ADSC00000000	China	Cui et al. 2013[4]
1.IN3b_E1979001	AAYV00000000	China	Cui et al. 2013[4]
1.IN3c_CMCC84038b	ADRF00000000	China	Cui et al. 2013[4]
1.IN3d_YN1683	ADTD00000000	China	Cui et al. 2013[4]
1.IN3e_YN472	ADTH00000000	China	Cui et al. 2013[4]
1.IN3f_YN1065	ADTC00000000	China	Cui et al. 2013[4]
1.IN3g_E1977001	ADRY00000000	China	Cui et al. 2013[4]
1.IN3h_CMCC84033	ADRE00000000	China	Cui et al. 2013[4]
1.IN3i_CMCC84046	ADRG00000000	China	Cui et al. 2013[4]
1.ORI1_CA88	ABCD00000000	USA	Cui et al. 2013[4]
1.ORI1_CO92	NC_003143	USA	Cui et al. 2013[4]
1.ORI1a_CMCC114001	ADQL00000000	China	Cui et al. 2013[4]
1.ORI1b_India195	ACNR00000000	India	Cui et al. 2013[4]
1.ORI1c_F1946001	ADSB00000000	China	Cui et al. 2013[4]
1.ORI2_F1991016	ABAT00000000	China	Cui et al. 2013[4]
1.ORI2a_YN2179	ADTE00000000	Myanmar	Cui et al. 2013[4]
1.ORI2c_YN2551b	ADTF00000000	China	Cui et al. 2013[4]
1.ORI2d_YN2588	ADTG00000000	China	Cui et al. 2013[4]
1.ORI2f_CMCC87001	ADRH00000000	China	Cui et al. 2013[4]
1.ORI2g_F1984001	ADSD00000000	China	Cui et al. 2013[4]
1.ORI2h_YN663	ADTI00000000	China	Cui et al. 2013[4]
1.ORI2i_CMCCK100001a	ADRR00000000	China	Cui et al. 2013[4]
1.ORI2i_CMCCK110001b	ADRS00000000	China	Cui et al. 2013[4]
1.ORI3_IP275	AAOS00000000	Madagascar	Cui et al. 2013[4]
1.ORI3_MG05-1020	AAYS00000000	Madagascar	Cui et al. 2013[4]

1.ORI3a_EV76	ADSA00000000	Madagascar	Cui et al. 2013[4]
2.ANT1_Nepal516	ACNQ00000000	Nepal	Cui et al. 2013[4]
2.ANT1a_34008	ADPD00000000	China	Cui et al. 2013[4]
2.ANT1b_34202	ADPE00000000	China	Cui et al. 2013[4]
2.ANT2a_2	ADOX00000000	China	Cui et al. 2013[4]
2.ANT2b_351001	ADPF00000000	China	Cui et al. 2013[4]
2.ANT2c_CMCC347001	ADQS00000000	China	Cui et al. 2013[4]
2.ANT2d_G1996006	ADSE00000000	China	Cui et al. 2013[4]
2.ANT2e_G1996010	ADSF00000000	China	Cui et al. 2013[4]
2.ANT2f_CMCC348002	ADQT00000000	China	Cui et al. 2013[4]
2.ANT3_KM682	LPVG	Former Soviet Union	Kutyrev et al. 2018[7]
2.ANT3a_CMCC92010	ADRL00000000	China	Cui et al. 2013[4]
2.ANT3b_CMCC95001	ADRN00000000	China	Cui et al. 2013[4]
2.ANT3c_CMCC96001	ADRO00000000	China	Cui et al. 2013[4]
2.ANT3d_CMCC96007	ADRP00000000	China	Cui et al. 2013[4]
2.ANT3e_CMCC67001	ADRB00000000	China	Cui et al. 2013[4]
2.ANT3f_CMCC104003	ADQH00000000	China	Cui et al. 2013[4]
2.ANT3g_CMCC51020	ADQY00000000	China	Cui et al. 2013[4]
2.ANT3h_CMCC106002	ADQI00000000	China	Cui et al. 2013[4]
2.ANT3i_CMCC64001	ADQZ00000000	China	Cui et al. 2013[4]
2.ANT3j_H1959004	ADSI00000000	China	Cui et al. 2013[4]
2.ANT3k_5761	ADPL00000000	Russia	Cui et al. 2013[4]
2.ANT3l_735	ADPR00000000	Russia	Cui et al. 2013[4]
2.MED0_C-627	MBSI00000000	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_1045	CP006794	Former Soviet Union	Zhgenti et al. 2015[6]
2.MED1_1116-D	LPXS	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_1240	LZNI	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_139	QAPA00000000	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_173	LQAZ	Former Soviet Union	Kutyrev et al. 2018[7]

2.MED1_1906	LYOM	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_244	LZND	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_261	LZNG	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_2944	CP006792	Former Soviet Union	Zhgenti et al. 2015[6]
2.MED1_44	LZNF	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_A-1763	LQAW	Former Soviet Union	Kutyrev et al. 2018[7]
2.MED1_A-1809	LYMF	Former Soviet Union	Eroshenko et al. 2017[5]
2.MED1_A-1825	LYCM	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_A-1920	LYCO	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_C-791	LQAU	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_K1973002	AAYT00000000	China	Cui et al. 2013[4]
2.MED1_KIM10	NC_004088	Iran	Cui et al. 2013[4]
2.MED1_KM816	LPXU	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_KM918	LPQY	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-1448	LYCN	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-1453	LQAY	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-1484	LQAV	Former Soviet Union	Kutyrev et al. 2018[6]

2.MED1_M-1524	LYCP	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-1773	LYMG	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-1864	LOHR	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-519	LQAX	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-549	LQBA	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-595	LYOH	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1_M-978	LPXT	Former Soviet Union	Kutyrev et al. 2018[6]
2.MED1b_2506	ADPA00000000	China	Cui et al. 2013[4]
2.MED1c_2654	ADPB00000000	China	Cui et al. 2013[4]
2.MED1d_2504	ADOZ00000000	China	Cui et al. 2013[4]
2.MED2b_91	ADPU00000000	China	Cui et al. 2013[4]
2.MED2c_K11973002	AAYT00000000	China	Cui et al. 2013[4]
2.MED2d_A1973001	ADPY00000000	China	Cui et al. 2013[4]
2.MED2e_7338	ADPQ00000000	China	Cui et al. 2013[4]
2.MED3a_J1963002	ADSP00000000	China	Cui et al. 2013[4]
2.MED3b_CMCC125002b	ADQN00000000	China	Cui et al. 2013[4]
2.MED3c_I1969003	ADSK00000000	China	Cui et al. 2013[4]
2.MED3d_J1978002	ADSQ00000000	China	Cui et al. 2013[4]
2.MED3f_I1970005	ADSL00000000	China	Cui et al. 2013[4]
2.MED3g_CMCC99103	ADRQ00000000	China	Cui et al. 2013[4]
2.MED3h_CMCC90027	ADRI00000000	China	Cui et al. 2013[4]
2.MED3i_CMCC92004	ADRK00000000	China	Cui et al. 2013[4]
2.MED3j_I2001001	ADSO00000000	China	Cui et al. 2013[4]
2.MED3k_CMCC12003	ADQM00000000	China	Cui et al. 2013[4]
2.MED3l_I1994006	ADSN00000000	China	Cui et al. 2013[4]
2.MED3m_SHAN11	ADTA00000000	China	Cui et al. 2013[4]
2.MED3n_SHAN12	ADTB00000000	China	Cui et al. 2013[4]
2.MED3o_I1991001	ADSM00000000	China	Cui et al. 2013[4]
2.MED3p_CMCC107004	ADQJ00000000	China	Cui et al. 2013[4]
3.ANT1a_7b	ADPN00000000	China	Cui et al. 2013[4]
3.ANT1b_CMCC71001	ADRC00000000	China	Cui et al. 2013[4]
3.ANT1c_C1976001	ADQA00000000	China	Cui et al. 2013[4]

3.ANT1d_71021	ADPO00000000	China	Cui et al. 2013[4]
3.ANT2a_MGJZ6	ADSX00000000	Mongolia	Cui et al. 2013[4]
3.ANT2b_MGJZ7	ADSY00000000	Mongolia	Cui et al. 2013[4]
3.ANT2c_MGJZ9	ADSZ00000000	Mongolia	Cui et al. 2013[4]
3.ANT2d_MGJZ11	ADSU00000000	Mongolia	Cui et al. 2013[4]
3.ANT2e_MGJZ3	ADSW00000000	Mongolia	Cui et al. 2013[4]
4.ANT_1454	LZNC	Former Soviet Union	Kutyrev et al. 2018[7]
4.ANT_338	LZNX	Former Soviet Union	Kutyrev et al. 2018[7]
4.ANT_517	LYMH	Former Soviet Union	Kutyrev et al. 2018[7]
4.ANT_KM932	LZNE	Former Soviet Union	Kutyrev et al. 2018[7]
4.ANT_M-1944	LYOK	Former Soviet Union	Kutyrev et al. 2018[7]
4.ANT1a_MGJZ12	ADSV00000000	Mongolia	Cui et al. 2013[4]
0.ANT1a_42013	ADPG00000000	China	Cui et al. 2013[4]
0.ANT1b_CMCC49003	ADQX00000000	China	Cui et al. 2013[4]
0.ANT1c_945	ADPV00000000	China	Cui et al. 2013[4]
0.ANT1d_164	ADOW00000000	China	Cui et al. 2013[4]
0.ANT1e_CMCC8211	ADRD00000000	China	Cui et al. 2013[4]
Ancient strains			
Strain ID	**Accession No.**	**Origin**	**Publication**
DA101	PRJEB25891	Kyrgyzstan	Damgaard et al. 2018[12]
Altenerding	PRJEB14851	Germany	Feldman et al. 2016[13]
London_EastSmithfield_8124_8291_11972	SRR341961, SRR341962, SRR341963	United Kingdom	Bos et al. 2011[14]
London_EastSmithfield_6330	SAMN00715799	United Kingdom	Bos et al. 2011[14]
Ellwangen	PRJEB13664	Germany	Spyrou et al. 2016[15]
Barcelona	PRJEB13664	Spain	Spyrou et al. 2016[15]
Bolgar	PRJEB13664	Russia	Spyrou et al. 2016[15]
OBS107	PRJEB12163	France	Bos et al. 2016[16]

OBS110	PRJEB12163	France	Bos et al. 2016[16]
OBS116	PRJEB12163	France	Bos et al. 2016[16]
OBS124	PRJEB12163	France	Bos et al. 2016[16]
OBS137	PRJEB12163	France	Bos et al. 2016[16]
EDI001.A	SAMEA5661363	United Kingdom	Keller et al. 2019[17]
LVC_merged	SAMEA5661382, SAMEA5661378, SAMEA5661380, SAMEA5661379	France	Keller et al. 2019[17]
DIT003.B	SAMEA5661360	Germany	Keller et al. 2019[17]
PET004.A	SAMEA5661385	Germany	Keller et al. 2019[17]
UNT003.A	SAMEA5661389	Germany	Keller et al. 2019[17]
VAL001.B	SAMEA5661384	Spain	Keller et al. 2019[17]
RK1	ERS1892069	Russia	Valtuena et al. 2017[18]
GEN72	ERS1892068	Croatia	Valtuena et al. 2017[18]
Gyvakarai1	ERS1892064	Lithuania	Valtuena et al. 2017[18]
Kunilall	ERS1892065	Estonia	Valtuena et al. 2017[18]
1343UNTA85	ERS1892067	Germany	Valtuena et al. 2017[18]
6POST	ERS1892066	Germany	Valtuena et al. 2017[18]
Gökhem2	ERS434180	Sweden	Rascovan et al. 2019[19]
BED024	PRJEB29990	United Kingdom	Spyrou et al. 2019[20]
BED028	PRJEB29990	United Kingdom	Spyrou et al. 2019[20]
BED030	PRJEB29990	United Kingdom	Spyrou et al. 2019[20]
BED034	PRJEB29990	United Kingdom	Spyrou et al. 2019[20]
BRA001	PRJEB29990	United Kingdom	Spyrou et al. 2019[20]

LAI009	PRJEB29990	Republic of Tatarstan (Russia)	Spyrou et al. 2019[20]
LBG002	PRJEB29990	Germany	Spyrou et al. 2019[20]
MAN008	PRJEB29990	Germany	Spyrou et al. 2019[20]
NAB003	PRJEB29990	Germany	Spyrou et al. 2019[20]
NMS002	PRJEB29990	United Kingdom	Spyrou et al. 2019[20]
STA001	PRJEB29990	Germany	Spyrou et al. 2019[20]
STN002	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
STN007	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
STN008	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
STN013	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
STN014	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
STN019	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
STN020	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
STN021	PRJEB29990	Switzerland	Spyrou et al. 2019[20]
TRP002	PRJEB29990	France	Spyrou et al. 2019[20]
RT5	PRJEB24296	Russia	Spyrou et al. 2018[21]
G701	PRJEB36413	Latvia	Susat et al. 2020[22]
G488	PRJEB36413	Latvia	Susat et al. 2020[22]
Y. pseudotuberculosis			
NCTC10275	NZ_CP008943.1	Turkey	Direct data submission

Table 3. Subset of 41 strains used for molecular dating with BEAST2

Strain ID	Accession No.	Origin	Publication
0.PE7b_620024	ADPM00000000	China	Cui et al. 2013[4]
0.PE4Aa_12	ADOV00000000	China	Cui et al. 2013[4]
0.PE4Ba_PestoidesA	ACNT00000000	FSU	Cui et al. 2013[4]
0.PE2_PEST-F	NC_009381	FSU	Cui et al. 2013[4]
Yersinia_pestis_Angola	PRJNA224116	Laboratory strain Strain	Eppinger et al. 2010[23]
0.PE4Cc_CMCC18019	ADQO00000000	China	Cui et al. 2013[4]
0.ANT1e_CMCC8211	ADRD00000000	China	Cui et al. 2013[4]
0.ANT2a_2330	ADOY00000000	China	Cui et al. 2013[4]
0.ANT3e_42091b	ADPI00000000	China	Cui et al. 2013[4]
4.ANT1a_MGJZ12	ADSV00000000	Mongolia	Cui et al. 2013[4]
3.ANT2b_MGJZ7	ADSY00000000	Mongolia	Cui et al. 2013[4]
3.ANT1b_CMCC71001	ADRC00000000	China	Cui et al. 2013[4]
2.ANT3l_735	ADPR00000000	Russia	Cui et al. 2013[4]
2.ANT1a_34008	ADPD00000000	China	Cui et al. 2013[4]
2.MED3n_SHAN12	ADTB00000000	China	Cui et al. 2013[4]

2.MED1d_2504	ADOZ00000000	China	Cui et al. 2013[4]
2.MED2e_7338	ADPQ00000000	China	Cui et al. 2013[4]
1.ANT1_UG05-0454	AAYR00000000	Uganda	Cui et al. 2013[4]
1.IN1c_K21985002	ADSS00000000	China	Cui et al. 2013[4]
1.IN2l_D1991004	ADRX00000000	China	Cui et al. 2013[4]
1.IN3b_E1979001	AAYV00000000	China	Cui et al. 2013[4]
1.ORI2h_YN663	ADTI00000000	China	Cui et al. 2013[4]
1.ORI3_IP275	AAOS00000000	Madagascar	Cui et al. 2013[4]
2.ANT1b_34202	ADPE00000000	China	Cui et al. 2013[4]
3.ANT1c_C1976001	ADQA00000000	China	Cui et al. 2013[4]
Gökhem_2	ERS434180	Sweden	Rascovan et al. 2018[19]
RISE505	PRJEB10885	Russia	Rasmussen et al. 2015[24]
RISE509	PRJEB10885	Russia	Rasmussen et al. 2015[24]
Gyvakarai1	ERS1892064	Lithuania	Valtuena et al. 2017[18]
Kunilall	ERS1892065	Estonia	Valtuena et al. 2017[18]
1343UNTA85	ERS1892067	Germany	Valtuena et al. 2017[18]
6POST	ERS1892066	Germany	Valtuena et al. 2017[18]

DA101	PRJEB25891	Kyrgyzstan	Damgaard et al. 2018[12]
OBS124	PRJEB12163	France	Bos et al. 2016[16]
Ellwangen	PRJEB13664	Germany	Spyrou et al. 2016[15]
Barcelona	PRJEB13664	Spain	Spyrou et al. 2016[15]
London_EastSmithfield_8124_8291_11972	SRR341961,SRR341962,SRR341963	United Kingdom	Bos et al. 2011[14]
Bolgar	PRJEB13664	Russia	Spyrou et al. 2016[15]
Altenerding	PRJEB14851	Germany	Feldman et al. 2016[13]
CO92	NC_003143.1	-	Parkhill et al. 2001[25]
NCTC10275	NZ_CP008943.1	Turkey	Direct data submission

Table 4. BEAST2 statistics for molecular dating

Run	**No starting tree provided**	**BEAST can change the tree**
Number of runs combined	2	4
States	900001000	1078668000
Burnin	10%	10%
Sampling steps	1000	1000
Posterior(ESS)	738	265
TreeHeigth(ESS)	1197	1753
Prior(ESS)	636	551

Table 5. List of unique SNPs for the RV2039 Riņņukalns strain compared to the CO92 reference genome (NC_003143.1).

Position	Reference	Alternative	Mapping Quality	Coverage depth
45268	T	C	MQ=37.00	DP=8
102759	G	A	MQ=37.00	DP=9
145498	A	C	MQ=37.00	DP=17
218722	C	T	MQ=37.00	DP=250
218848	T	A	MQ=37.00	DP=11
228421	G	C	MQ=37.00	DP=12
233862	C	G	MQ=37.00	DP=17
258817	T	G	MQ=37.00	DP=12
361696	A	G	MQ=37.00	DP=23
450992	T	C	MQ=37.00	DP=11
512196	G	C	MQ=37.00	DP=33
512199	C	T	MQ=37.00	DP=43
512226	A	G	MQ=37.00	DP=61
513972	C	G	MQ=37.00	DP=20
521243	G	T	MQ=37.00	DP=4
655376	T	G	MQ=37.00	DP=8
699506	A	G	MQ=37.00	DP=34
704906	T	G	MQ=37.00	DP=4
807209	A	C	MQ=37.00	DP=6
842719	A	G	MQ=37.00	DP=4
939968	A	G	MQ=37.00	DP=6
974288	A	G	MQ=37.00	DP=22
980956	A	C	MQ=37.00	DP=6
1006899	T	C	MQ=37.00	DP=11
1117228	T	G	MQ=37.00	DP=5

1126669	A	C	MQ=37.00	DP=7
1202078	T	C	MQ=37.00	DP=5
1204122	A	C	MQ=37.00	DP=6
1236524	C	T	MQ=37.00	DP=10
1424267	G	C	MQ=37.00	DP=20
1459735	T	C	MQ=37.00	DP=6
1499519	A	C	MQ=37.00	DP=6
1633917	T	C	MQ=37.00	DP=4
1729234	A	C	MQ=37.00	DP=13
1729240	A	C	MQ=37.00	DP=13
1749746	G	A	MQ=37.00	DP=22
1749750	G	C	MQ=37.00	DP=22
1773706	G	C	MQ=37.00	DP=5
1790092	A	G	MQ=37.00	DP=66
1790107	G	A	MQ=37.00	DP=30
1790524	A	T	MQ=37.00	DP=24
1790527	C	T	MQ=37.00	DP=23
1817775	T	G	MQ=37.00	DP=5
2037692	A	G	MQ=37.00	DP=9
2088268	A	G	MQ=37.00	DP=10
2109643	C	A	MQ=37.00	DP=7
2119087	T	G	MQ=37.00	DP=15
2119652	A	C	MQ=37.00	DP=4
2137531	A	G	MQ=37.00	DP=5
2245900	T	A	MQ=37.00	DP=7
2256636	G	T	MQ=37.00	DP=4
2282213	T	G	MQ=37.00	DP=49
2342714	T	C	MQ=37.00	DP=6
2382275	A	G	MQ=37.00	DP=8

2434412	G	T	MQ=37.00	DP=5
2515437	T	C	MQ=37.00	DP=7
2543233	A	C	MQ=37.00	DP=6
2586505	A	G	MQ=37.00	DP=8
2676468	T	C	MQ=37.00	DP=75
2676480	T	C	MQ=37.00	DP=70
2676495	C	T	MQ=37.00	DP=17
2732131	A	G	MQ=37.00	DP=9
2763419	A	G	MQ=37.00	DP=13
2770489	C	A	MQ=37.00	DP=6
2838265	A	G	MQ=37.00	DP=9
2843387	C	G	MQ=37.00	DP=23
2949258	T	A	MQ=37.00	DP=10
2986142	G	A	MQ=37.00	DP=11
3018377	T	C	MQ=37.00	DP=22
3027937	T	G	MQ=37.00	DP=4
3139075	T	C	MQ=37.00	DP=49
3257101	A	G	MQ=37.00	DP=5
3351581	T	A	MQ=37.00	DP=9
3361639	T	G	MQ=37.00	DP=6
3361655	T	G	MQ=37.00	DP=6
3460356	A	T	MQ=37.00	DP=6
3516691	C	G	MQ=37.00	DP=6
3516692	T	A	MQ=37.00	DP=7
3519859	T	C	MQ=37.00	DP=5
3540608	A	G	MQ=37.00	DP=5
3617207	T	C	MQ=37.00	DP=8
3722960	T	A	MQ=37.00	DP=5
3722962	A	G	MQ=37.00	DP=6

3748300	G	C	MQ=37.00	DP=8
3752693	A	C	MQ=37.00	DP=5
3808191	A	G	MQ=37.00	DP=10
3809359	T	A	MQ=37.00	DP=9
3819379	T	C	MQ=37.00	DP=6
3849432	A	G	MQ=37.00	DP=5
3896808	T	G	MQ=37.00	DP=60
3964636	A	G	MQ=37.00	DP=14
4142158	T	C	MQ=37.00	DP=7
4201161	A	C	MQ=37.00	DP=216
4207518	A	T	MQ=37.00	DP=7
4211429	G	T	MQ=37.00	DP=4
4229928	T	C	MQ=37.00	DP=24
4252403	C	G	MQ=37.00	DP=10
4252429	T	C	MQ=37.00	DP=5
4267277	T	C	MQ=37.00	DP=7
4287308	A	G	MQ=37.00	DP=12
4338275	C	A	MQ=37.00	DP=9
4506792	A	G	MQ=37.00	DP=17
4571296	C	A	MQ=37.00	DP=5
4614434	A	G	MQ=37.00	DP=45
4645281	G	A	MQ=37.00	DP=4
4646816	A	G	MQ=37.00	DP=15

Table 6. Number of generated sequences for each grave

Sample	Number of reads generated
Burial 2017/01	11448118
RV1852	39991507
RV2039	1650917277
20208HL	1452296

Table 7. Basic mapping statistics for RV2039 after duplicate removal and quality filtering (MQ>=30).

Reference	Aligned reads	Coverage >=1x	Coverage >=2x	Coverage >=3x	Mean coverage whole reference
Chromosome NC_003143.1	600065	92.39%	86.75%	78.29%	5.85x
pMT1 NC_003134.1	2144	46.62%	21.28%	8.88%	0.83x
pPCP1 NC_003132.1	1050	78.05%	72.24%	66.14%	5.10x
pCD1 NC_003131.1	18005	92.04%	90.31%	87.88%	12.00x

Table 8. Peptides of *Y. pestis* identified by bottom-up proteomics/LC-MS/MS. All three peptides are proteotypic and were identified with medium confidence (5% FDR). The peptide ADIFELAGKTGVGMTSLPQNEQHLAAR was also confirmed with the peptide centric search algorithm PepQuery.

Peptide sequence	**Protein name (Accession number (UniProt))**	**Sequence coverage (%)**	**Modifications**
ADIFELAGKTGVGMTSLPQNEQHLAAR	Arginine N-succinyltransferase (Q9ZC67)	8	Deamidation: Q19, Q22
DELIAVEK	Aminoacyl-histidine dipeptidase (A0A384LI20)	2	none
GGKPAVVGEPDMPQSIADVAAELGAQLYR	Dihydrofolate synthase/ folylpolyglutamate synthase (Q0WDC2)	7	Deamidation: Q14, Q26 Oxidation: P4, M12

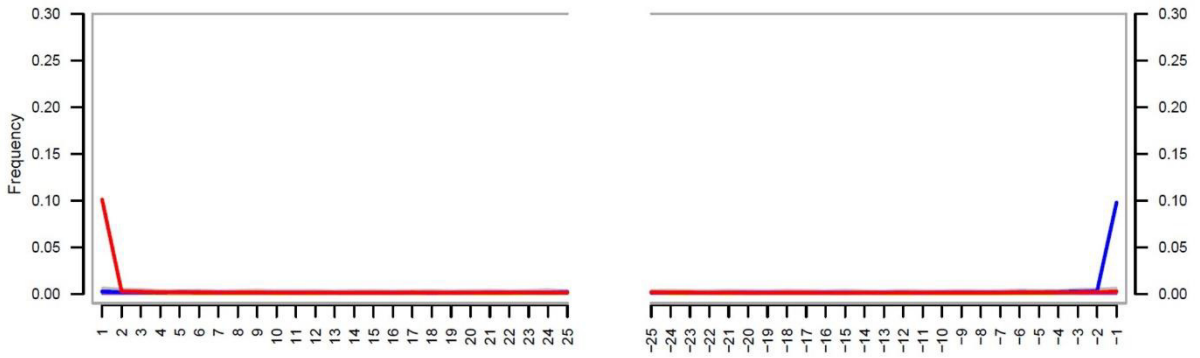


Figure 1. aDNA damage pattern after half-UDG library preparation for RV2039. The sample shows the expected degradation patterns, increased C > T and G > A substitutions at the 5´ and 3´ends of the reads, after half-UDG library preparation.

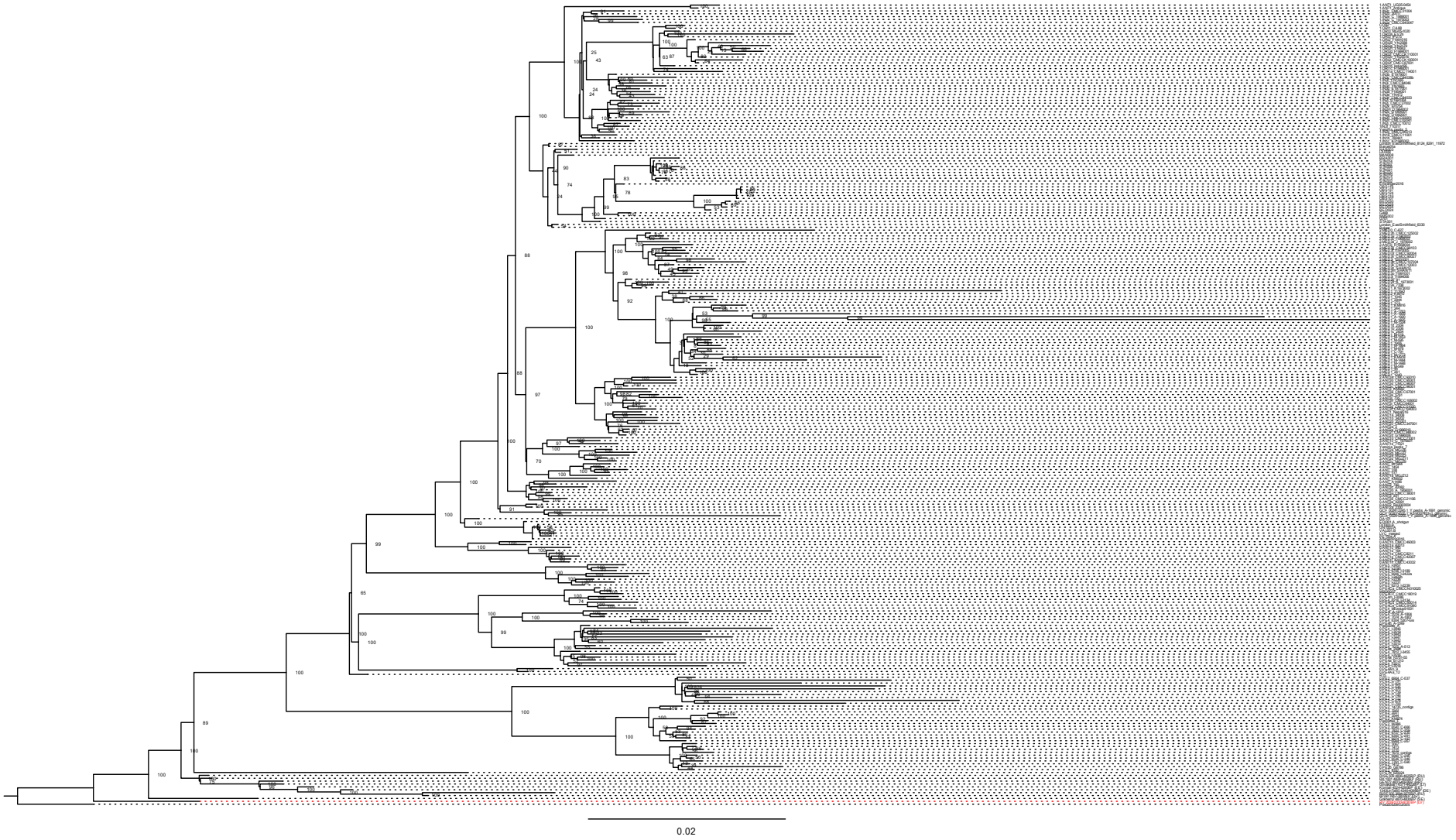


Figure 2. Maximum-likelihood tree for 277 strains generated with RAxML. Bootstrap values are shown for 1000 replicates.

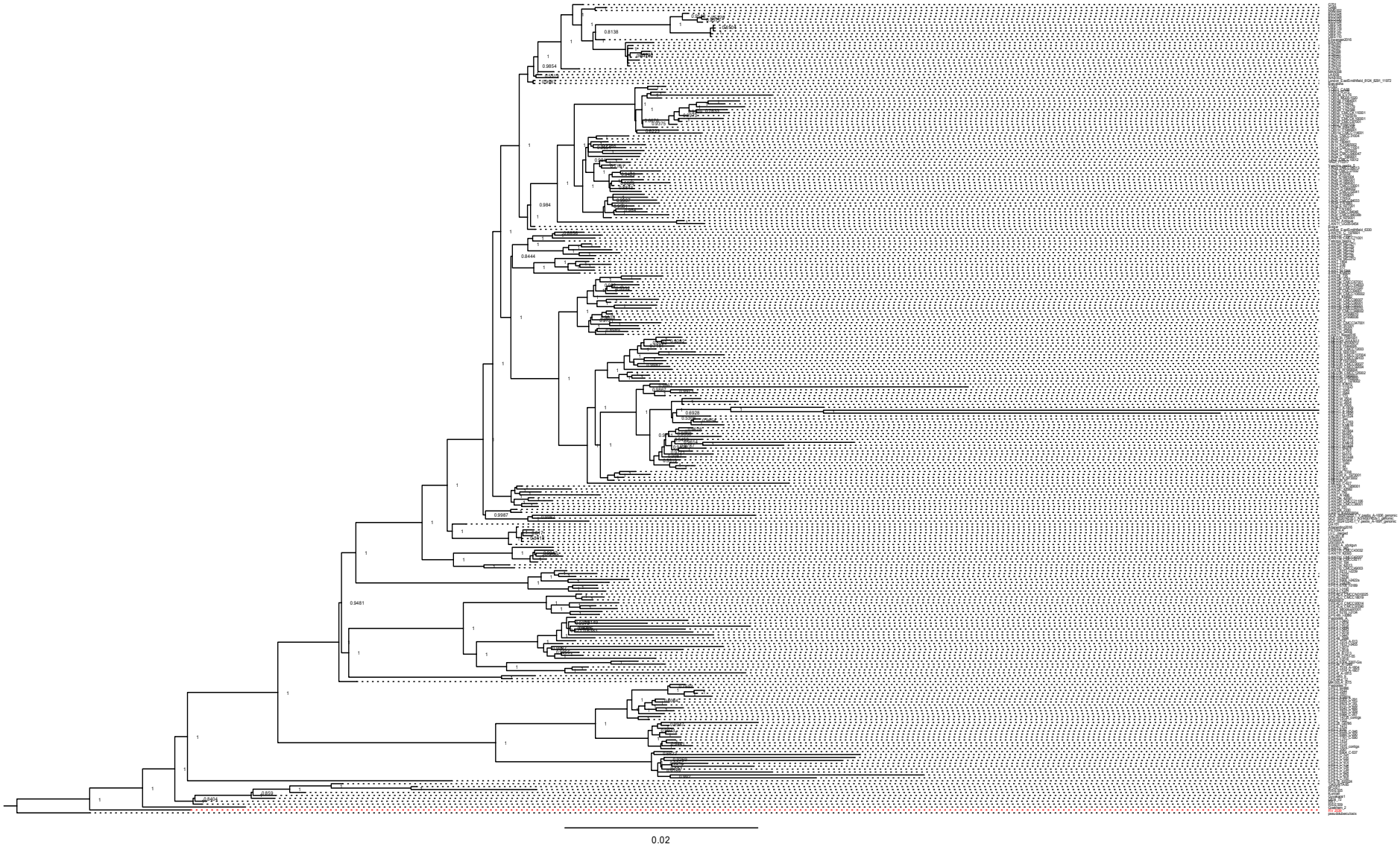


Figure 3. Bayesian tree for 277 strains generated with MrBayes. Posterior probability is shown for one million generations.

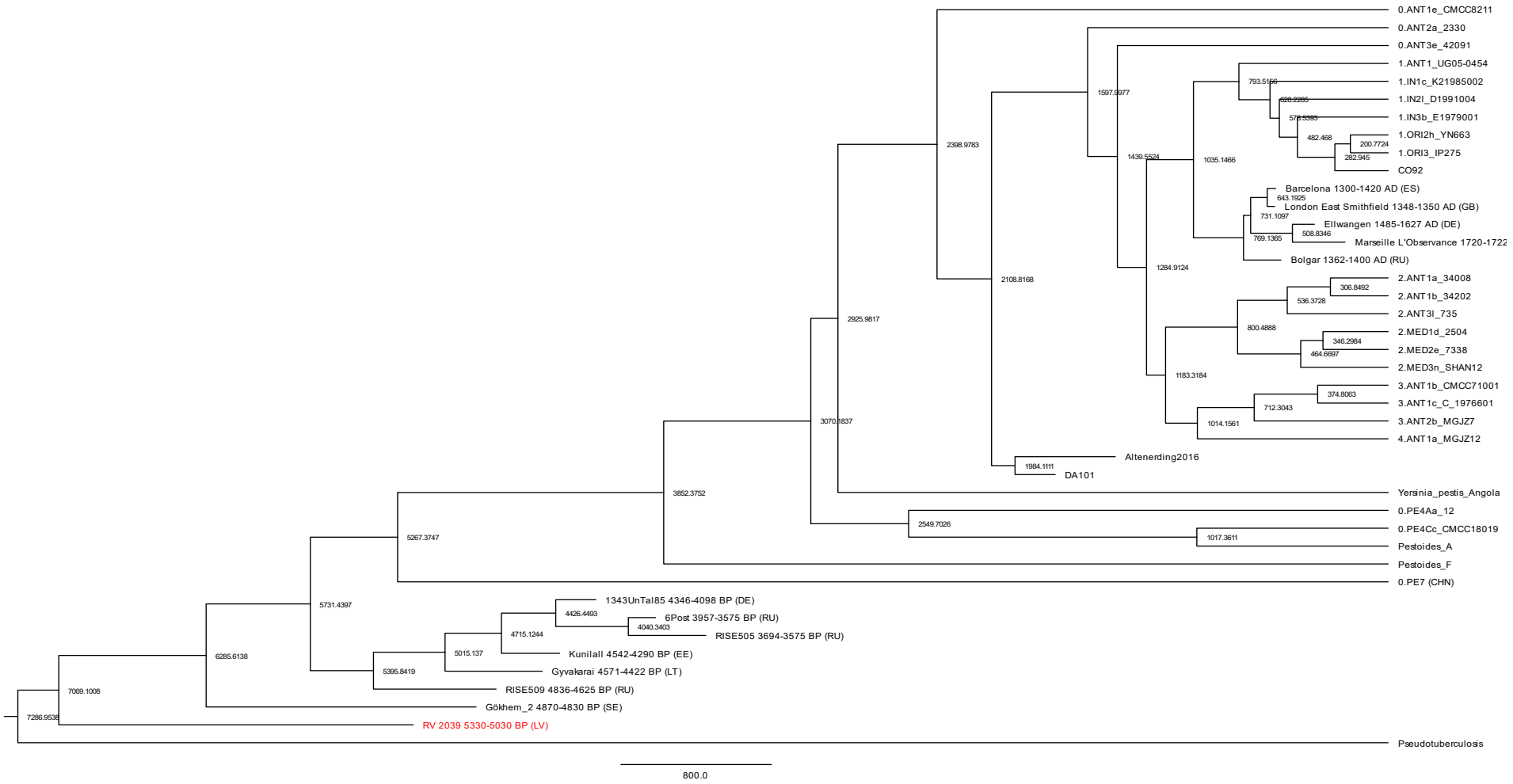


Figure 4. BEAST2 maximum clade credibility tree summarized from 970803 trees. BEAST2 was allowed to alter the starting tree. Branch lengths are given as years before the present (BP).

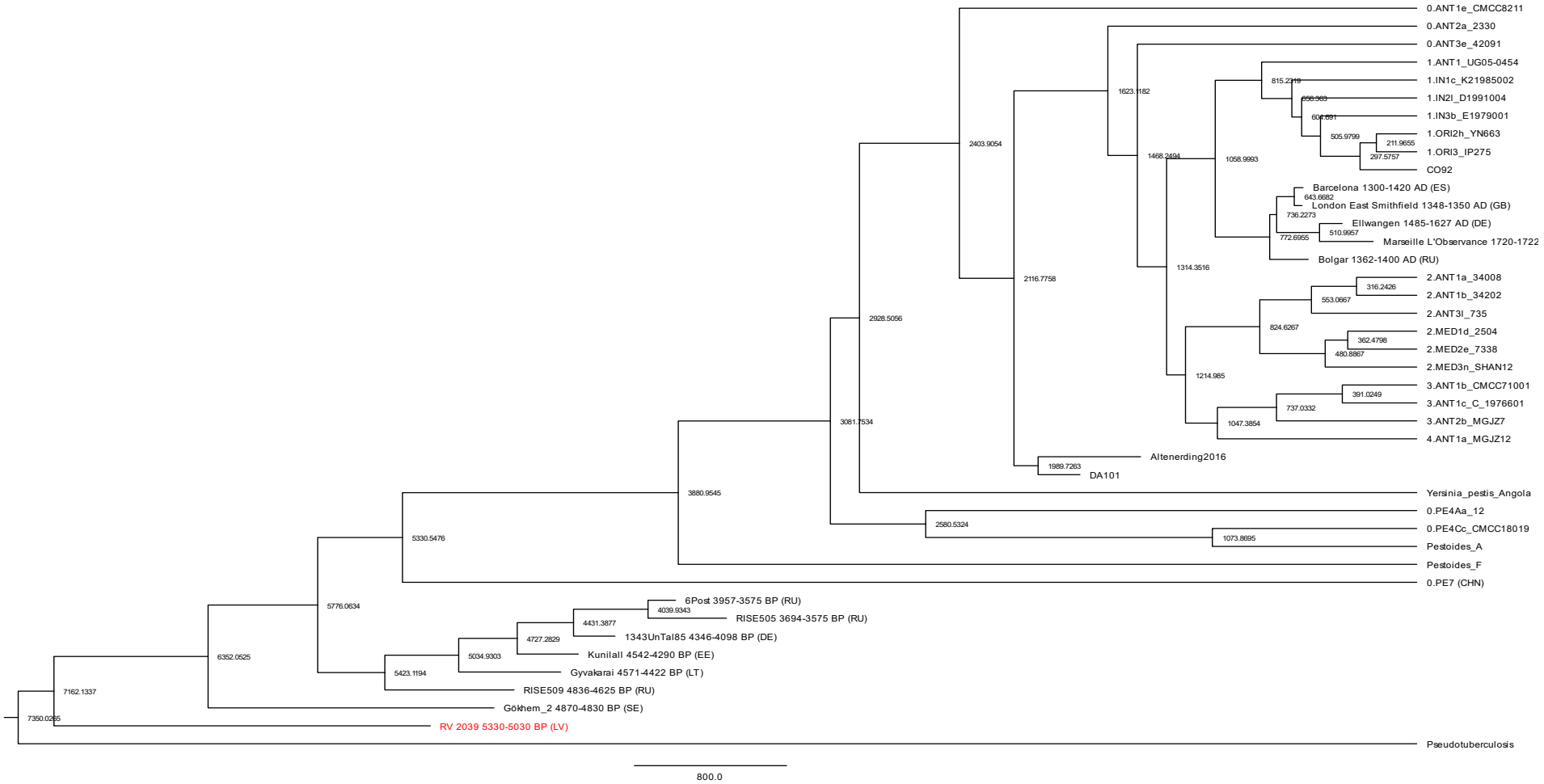


Figure 5. BEAST2 maximum clade credibility tree summarized from 810252 trees. No starting tree was provided. Branch lengths are given as years before the present (BP).

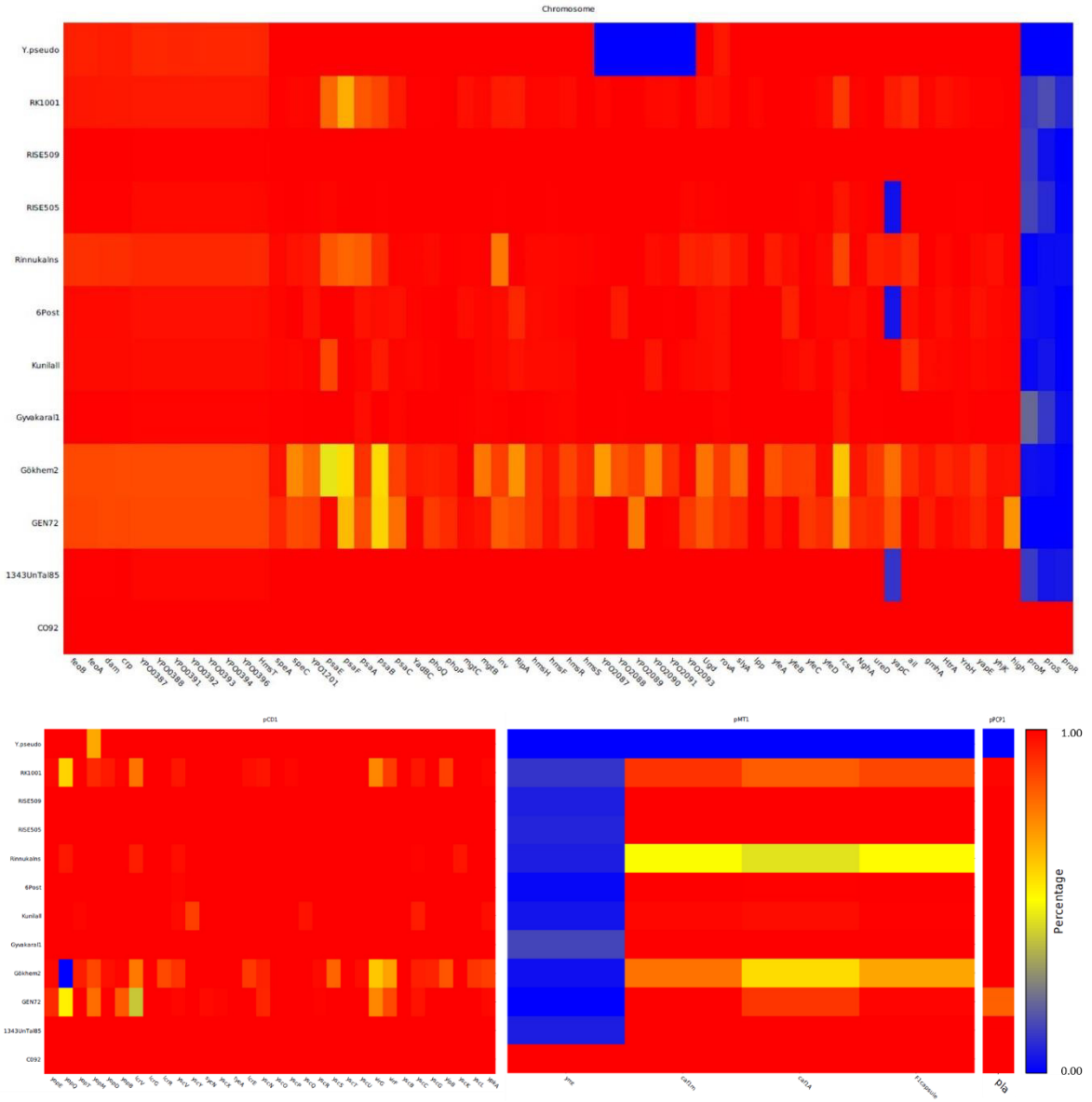


Figure 6. Heatmap depicting percent coverage of the *Yersinia pestis* pathogenicity factors.

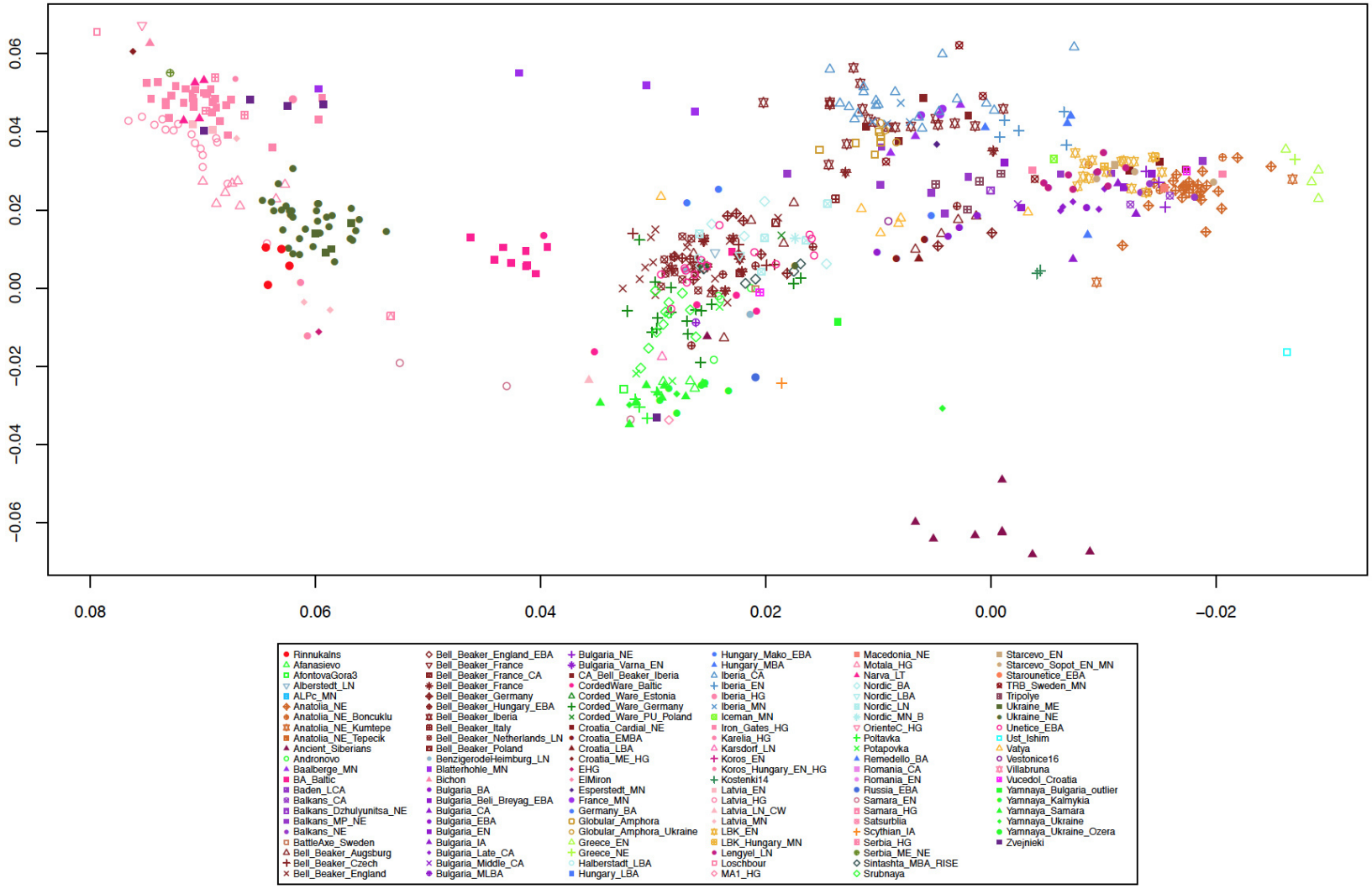


Figure 7. PCA of individuals from 123 ancient populations including Riņņukalns projected onto the first two principal components calculated from 59 present-day West-Eurasian populations (not shown for clarity). Riņņukalns individuals are depicted as red dots.

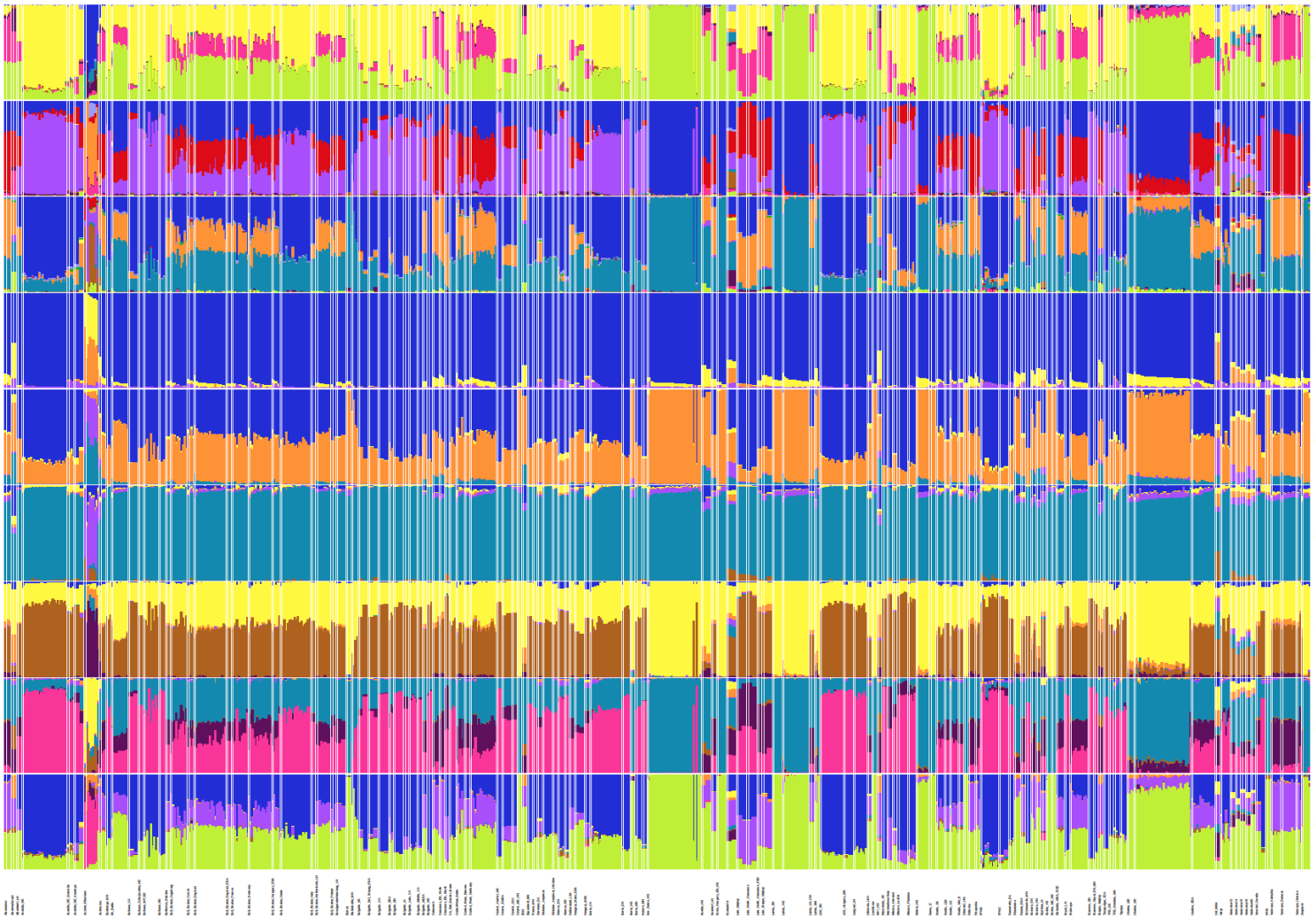


Figure 8. ADMIXTURE modelling of ancient populations including Riņņukalns with K=4 to K=12 genetic components.

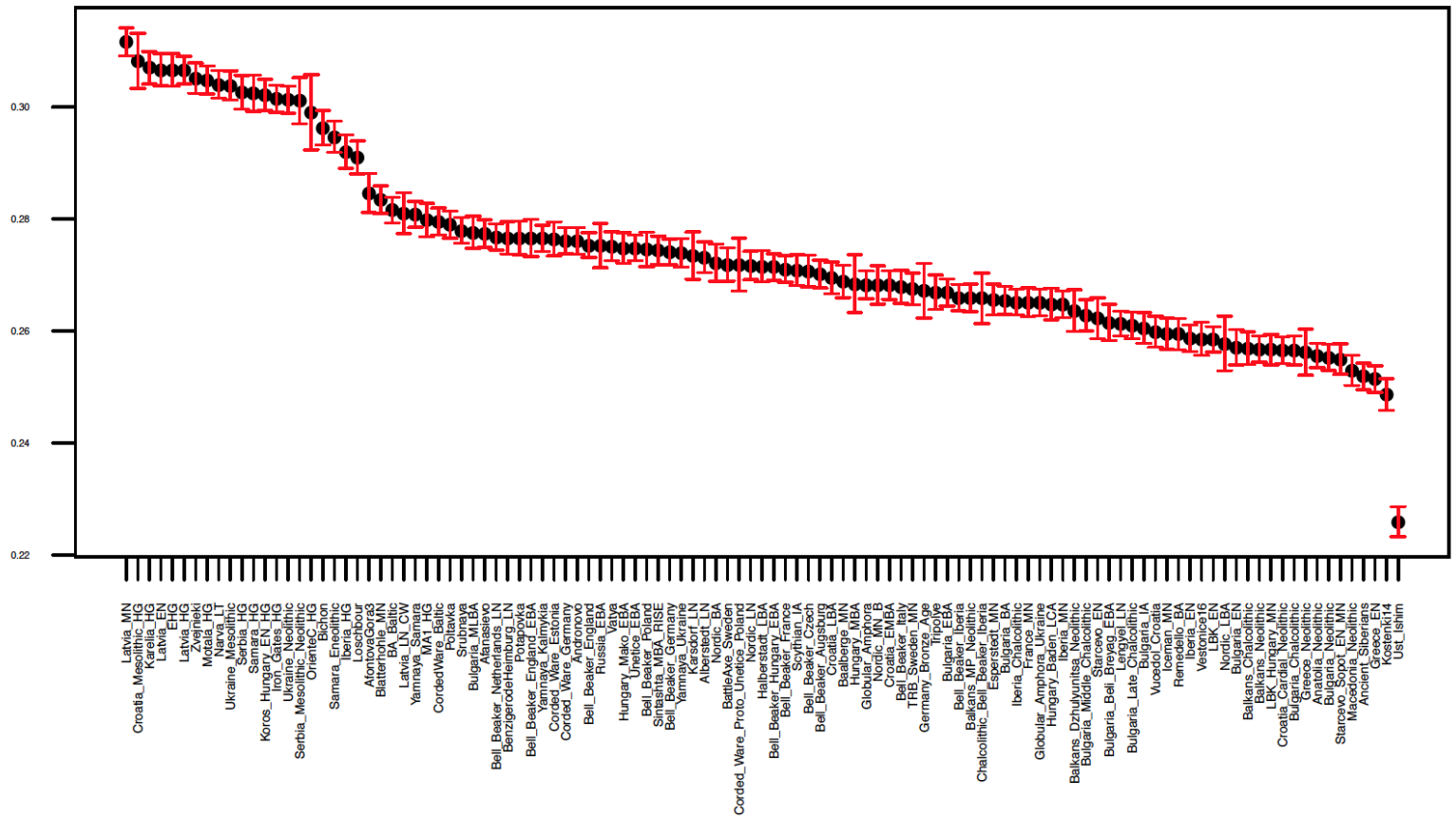
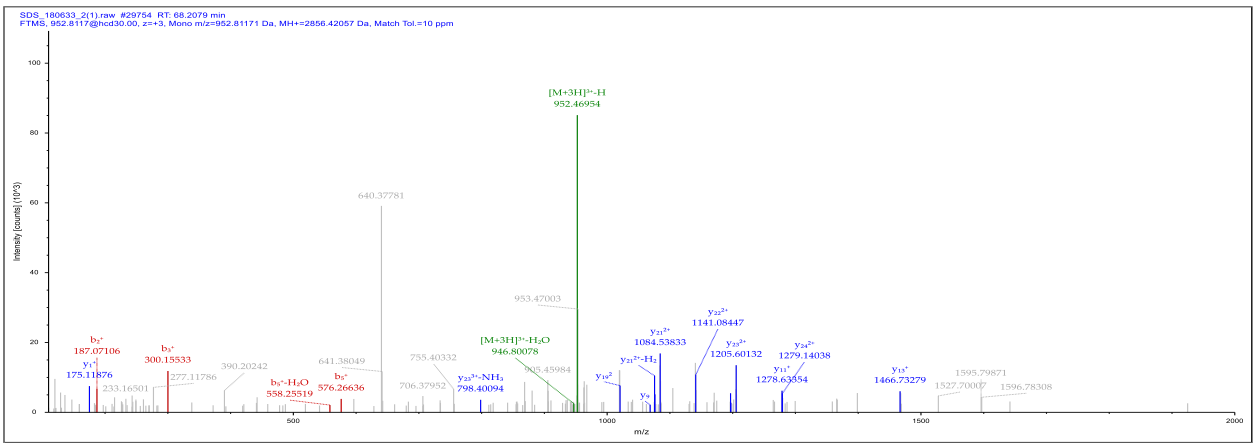
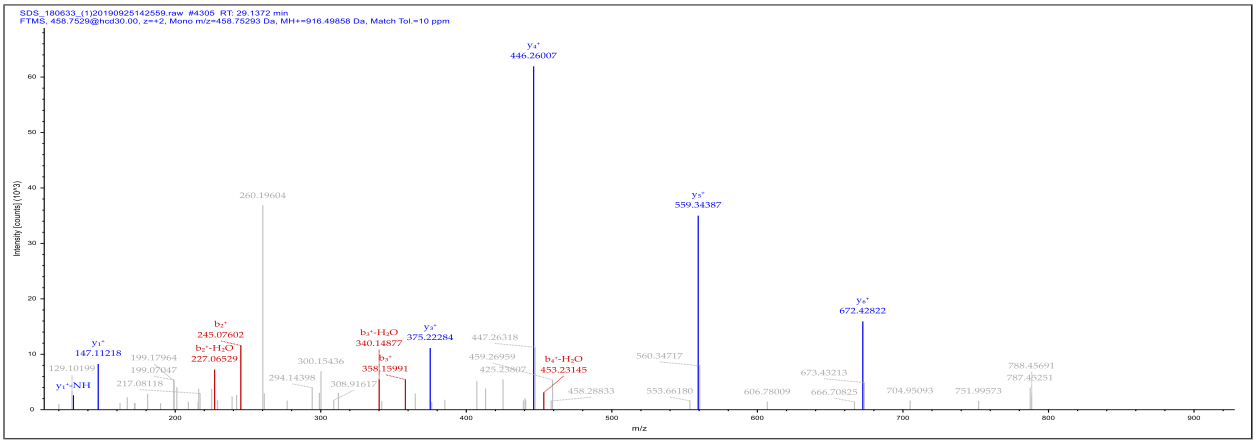


Figure 9. f3 outgroup statistics *f3(Riņņukalns; Test, Mbuti)* showing the amount of shared genetic drift between the Riņņukalns individuals and selected previously published ancient populations/individuals (as used in PCA and admixture analysis).

(A)



(B)



(C)

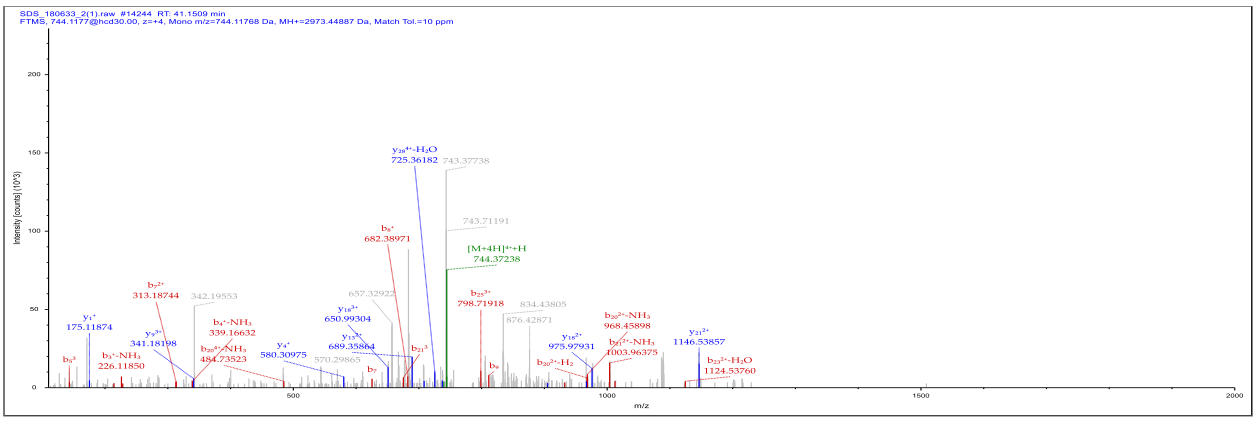


Figure 10. MS/MS spectra of three peptides identified by LC-MS/MS analysis of protein extracted from teeth powder after extraction with SDS. Information about identification confidence and protein modifications: see Suppl. Table 1. (A) Peptide ADIFELAGKTGVGMTSLPQNEQHLAAR from the protein Arginine N-succinyltransferase. (B) Peptide DELIAVEK from the protein Aminoacyl-histidine dipeptidase. (C) Peptide

GGKPAVVGEPDMPQSIADVAAELGAQLYR from the protein Dihydrofolate synthase/folylpolyglutamate synthase.

References

1. Lübke H, Brinker U, Meadows J, Bērziņš V, Zagorska New research on the human burials of Riņņukalns, Latvia. In: Grünberg J, GrmascB, Larsson L, Orschiedt J, Meller H (ed) Mesolithic Burials–Rites, Symbols and Social Organization of Early Postglacial Communities, Proceedings of the International Conference, Halle (Saale), 18th–21st September 2013. 2016: 241–256.
2. Brinker U, Bērziņš V, Ceriņa A, et al. (2020) Two burials in a unique freshwater shell midden: insights into transformations of Stone Age hunter-fisher daily life in Latvia. Archaeol. Anthropol. Sci. 2020;12: 97 [DOI: 10.1007/s12520-020-01049-7].
3. Wen, B.; Wang, X.; Zhang, B. (2019); PepQuery enables fast, accurate, and convenient proteomic validation of novel genomic alterations. *Genome Research*, 29: 485 – 493.
4. Cui Y, Yu C, Yan Y, et al. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. Proc Natl Acad Sci U S A 2013;110: 577-582.
5. Eroshenko GA, Nosov NY, Krasnov YM, et al. Yersinia pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. PLoS One. 2017;12.
6. Zhgenti E, Johnson SL, Davenport KW, et al. Genome assemblies for 11 Yersinia pestis strains isolated in the Caucasus region. Genome Announc 2015;3.
7. Kutyrev V V., Eroshenko GA, Motin VL, et al. Phylogeny and classification of Yersinia pestis through the lens of strains from the Plague Foci of commonwealth of independent states. Front Microbiol 2018;6: 1106.
8. Kislichkina AA, Bogun AG, Kadnikova LA, et al. Nineteen whole-genome assemblies of Yersinia pestis subsp. microtus, including representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica. Genome Announc 2015;3.
9. Kislichkina AA, Bogun AG, Kadnikova LA, et al. Eight whole-genome assemblies of Yersinia pestis subsp. microtus bv. caucasica isolated from the common vole (Microtus arvalis) plague focus in Dagestan, Russia. Genome Announc 2017;5.
10. Kislichkina AA, Bogun AG, Kadnikova LA, et al. Nine whole genome assemblies of Yersinia pestis subsp. microtus bv. Altaica strains isolated from the Altai Mountain natural plague focus (no. 36) in Russia. Genome Announc 2018;6.

11. Kislichkina AA, Bogun AG, Kadnikova LA, et al. Six whole-genome assemblies of Yersinia pestis subsp. microtus bv. ulegeica (phylogroup 0.PE5) strains isolated from Mongolian natural plague foci. Genome Announc 2018;6.
12. De Barros Damgaard P, Marchi N, Rasmussen S, et al. 137 ancient human genomes from across the Eurasian steppes. Nature 2018;557: 369-374.
13. Feldman M, Harbeck M, Keller M, et al. A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. Mol Biol Evol 2016;33: 2911-2923.
14. Bos KI, Schuenemann VJ, Golding GB, et al. A draft genome of Yersinia pestis from victims of the Black Death. Nature 2011;478; 506-510.
15. Spyrou MA, Tukhbatova RI, Feldman M, et al. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. Cell Host Microbe 2016;19; 874-881.
16. Bos KI, Herbig A, Sahl J, et al. Eighteenth century Yersinia pestis genomes reveal the long-term persistence of an historical plague focus. Elife 2016.
17. Keller M, Spyrou MA, Scheib CL, et al. Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). Proc Natl Acad Sci U S A 2019;116: 12363-12372.
18. Andrades Valtueña A, Mittnik A, Key FM, et al. The Stone Age Plague and Its Persistence in Eurasia. Curr Biol 2017;27: 3683-3691.
19. Rascovan N, Sjögren KG, Kristiansen K, et al. Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. Cell 2019;176: 295-305.
20. Spyrou MA, Keller M, Tukhbatova RI, et al. Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. Nat Commun 2019;10.
21. Spyrou MA, Tukhbatova RI, Wang CC, et al. Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. Nat Commun 2018;9.
22. Susat J, Bonczarowska JH, Pētersone-Gordina, et al. Yersinia pestis strains from Latvia show depletion of the pla virulence gene at the end of the second plague pandemic. Currently under review at Scientific reports.
23. Eppinger M, Worsham PL, Nikolich MP, et al. Genome sequence of the deep-rooted Yersinia pestis strain angola reveals new insights into the evolution and pangenome of the plague bacterium. J Bacteriol 2010;192: 1685-1699.

24. Rasmussen S, Allentoft ME, Nielsen K, et al. Early Divergent Strains of Yersinia pestis in Eurasia 5,000 Years Ago. Cell 2015;163: 571-582.
25. Parkhill J, Wren BW, Thomson NR, et al. Genome sequence of Yersinia pestis, the causative agent of plague. Nature. 2001;413: 523-527.

12.

Chapter III: Neolithic and medieval virus genomes reveal complex evolution of hepatitis B

Supplementary material

Figures and figure supplements

Neolithic and medieval virus genomes reveal complex evolution of hepatitis B

Ben Krause-Kyora *et al*

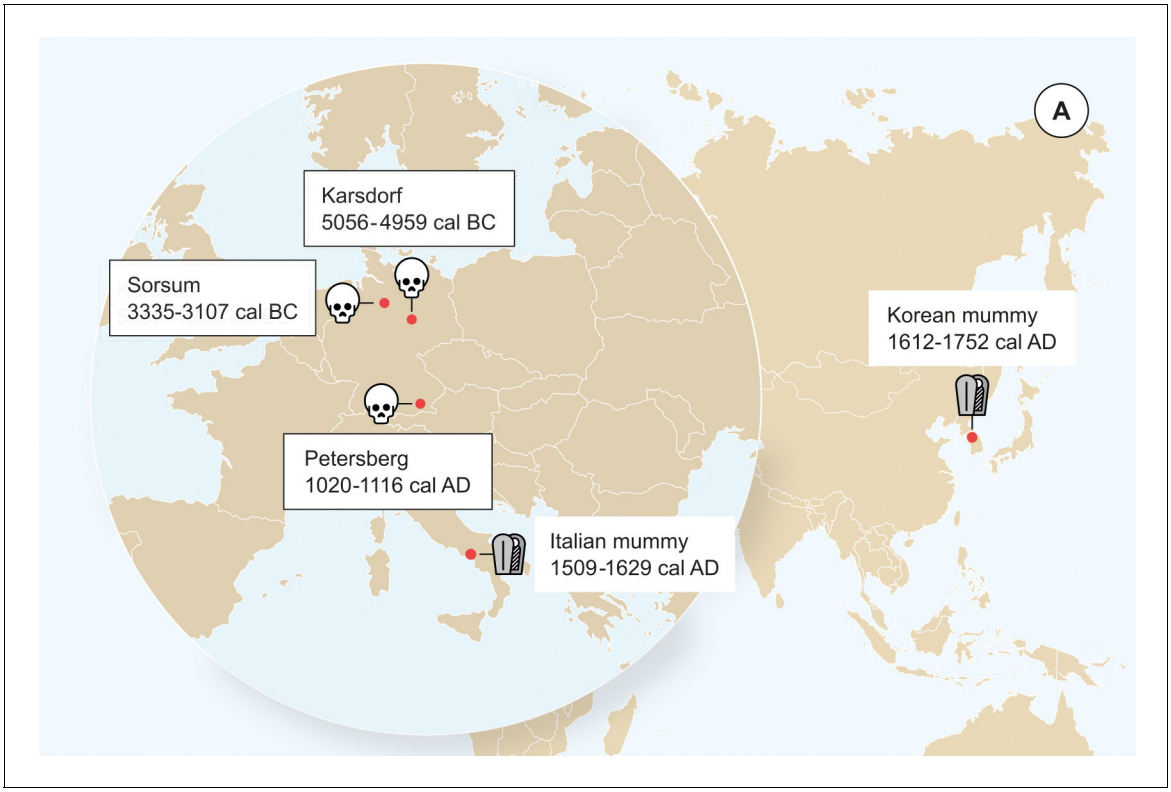


Figure 1. Origin of samples. Geographic location of the samples from which ancient HBV genomes were isolated. Radiocarbon dates of the specimens is given in two sigma range. Icons indicate the sample material (tooth or mummy). HBV genomes obtained in this study are indicated by black frame.
DOI: https://doi.org/10.7554/eLife.36666.002

Figure 1—figure supplement 1. Skull of the investigated Karsdorf individual 537 is from a male with an age at death of around 25–30 years.
DOI: https://doi.org/10.7554/eLife.36666.003

Figure 1—figure supplement 2. Mandible fragment of the Sorsum individual XLVII 11 analyzed in this study is from a male.
DOI: https://doi.org/10.7554/eLife.36666.004

Figure 1—figure supplement 3. Skull of the analyzed Petersberg individual from grave 820 is from a male with an age at death of around 65–70 years.
DOI: https://doi.org/10.7554/eLife.36666.005

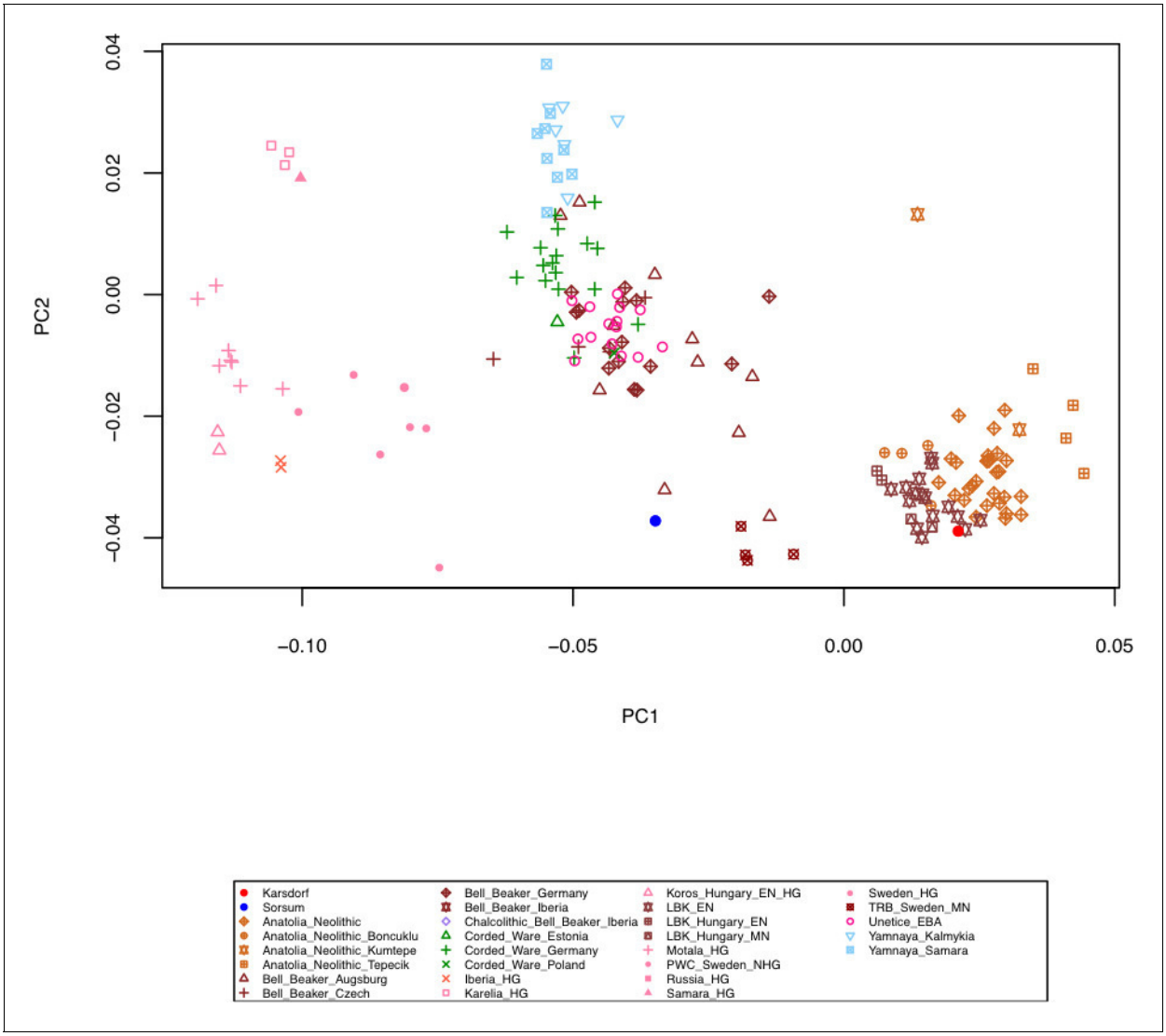


Figure 1—figure supplement 4. Principal Component Analysis (PCA) of the human Karsdorf and Sorsum samples together with previously published ancient populations projected on 27 modern day West Eurasian populations (not shown) based on a set of 1.23 million SNPs (***Mathieson et al., 2015***).
DOI: https://doi.org/10.7554/eLife.36666.006

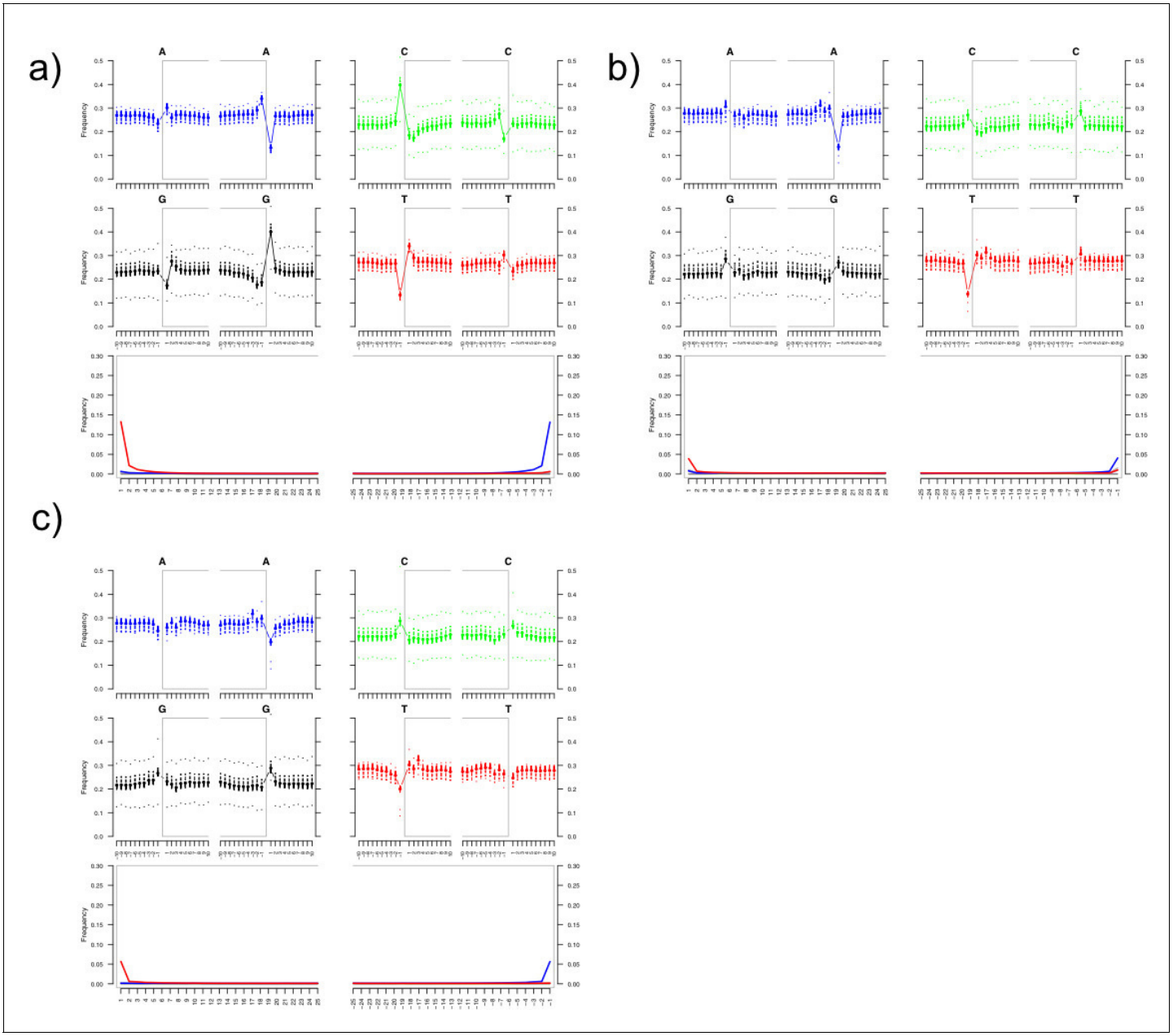


Figure 1—figure supplement 5. Damage plots showing deamination patterns of hg19-specific reads for the HalfUDG-treated libraries of (**a**) Karsdorf, (**b**) Sorsum, (**c**) Petersberg.

DOI: https://doi.org/10.7554/eLife.36666.007

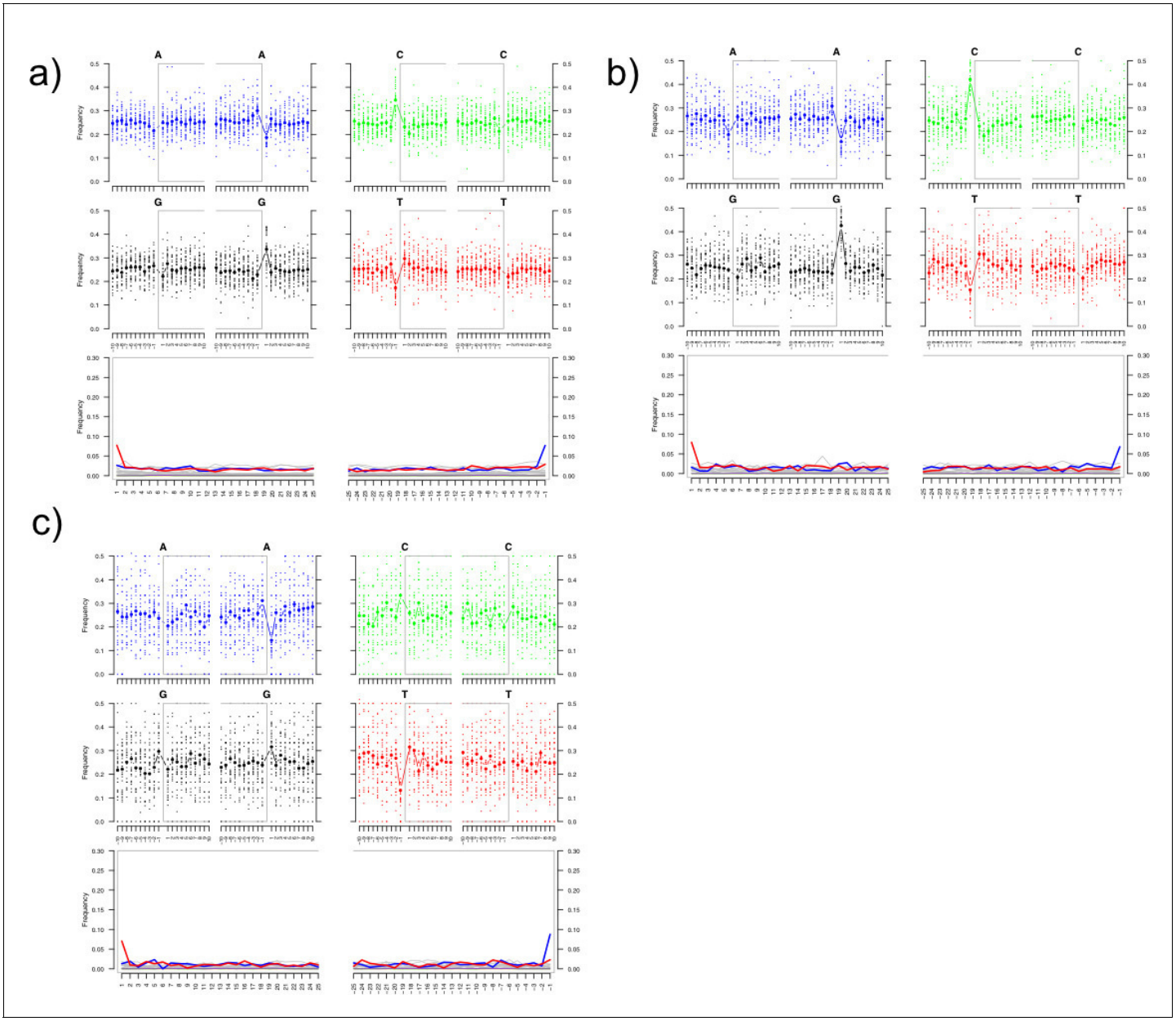


Figure 1—figure supplement 6. Damage plots showing deamination patterns of HBV-specific reads for the HalfUDG-treated libraries of (**a**) Karsdorf, (**b**) Sorsum, (**c**) Petersberg. References shown in ***Supplementary file 1*** were used to carry out the alignment.

DOI: https://doi.org/10.7554/eLife.36666.008

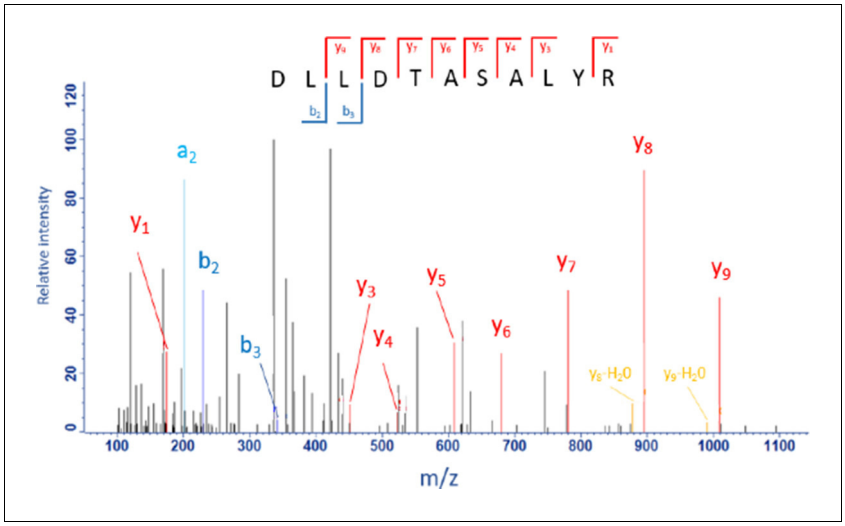


Figure 1—figure supplement 7. MS/MS spectrum of the proteotypic HBV-peptide DLLDTASALYR from the HBV-protein external core antigen (residues 58–68). ($[M + 2\ hr]^{2+}$): m/z = 1237.6429 Da. Mass accuracy of the precursor peptide = 0.56 ppm.

DOI: https://doi.org/10.7554/eLife.36666.009

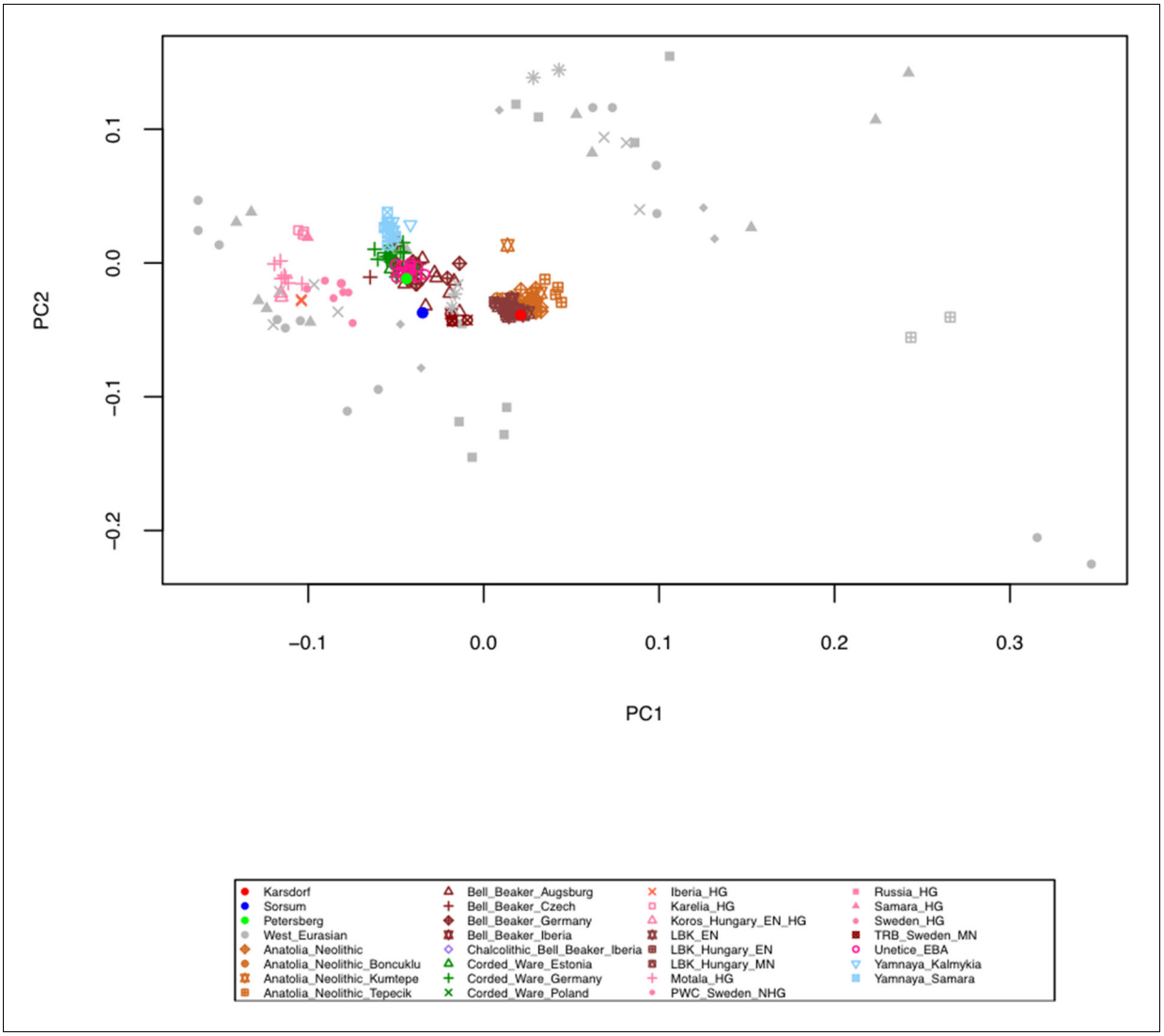


Figure 1—figure supplement 8. Principal Component Analysis (PCA) of the human Karsdorf and Sorsum samples together with previously published ancient populations projected on 27 modern day West Eurasian populations (shown in gray) based on a set of 1.23 million SNPs (***Mathieson et al., 2015***).
DOI: https://doi.org/10.7554/eLife.36666.010

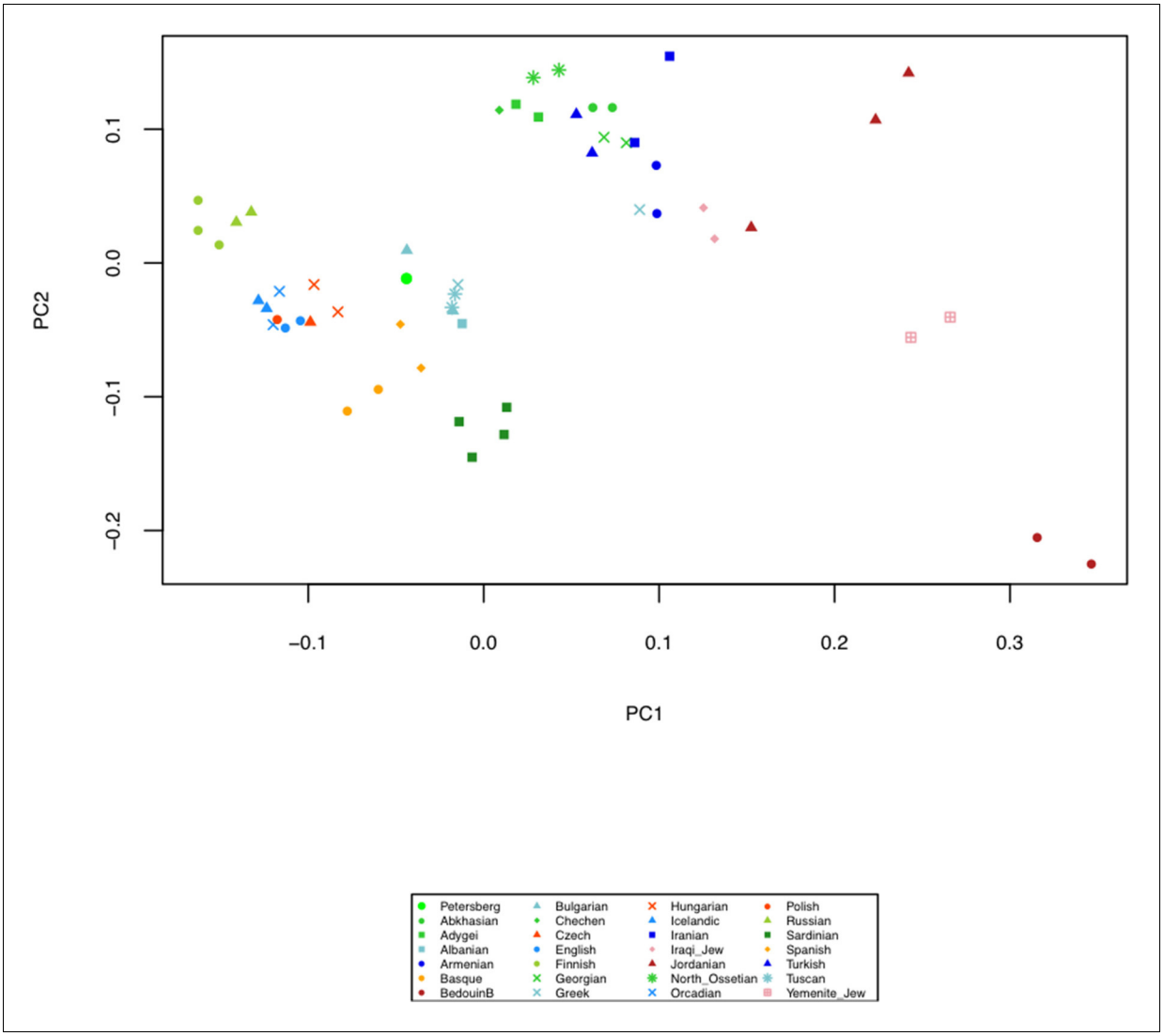


Figure 1—figure supplement 9. Principal Component Analysis (PCA) of the human Petersberg sample projected on 27 modern day West Eurasian populations based on a set of 1.23 million SNPs (***Mathieson et al., 2015***).

DOI: https://doi.org/10.7554/eLife.36666.011

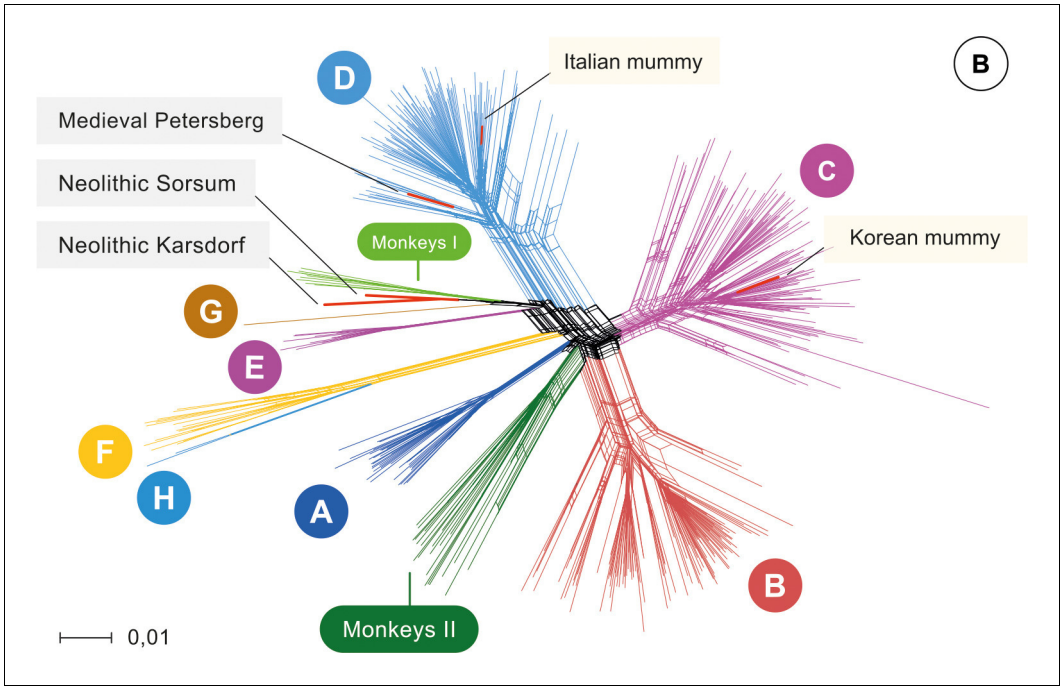


Figure 2. Network. Network of 493 modern, two published ancient genomes (light yellow box), and three ancient hepatitis B virus (HBV) obtained in this study (grey box). Colors indicate the eight human HBV genotypes (**A–H**), two monkey genotypes (Monkeys I, African apes and Monkeys II, Asian monkeys) and ancient genomes (red).
DOI: https://doi.org/10.7554/eLife.36666.013

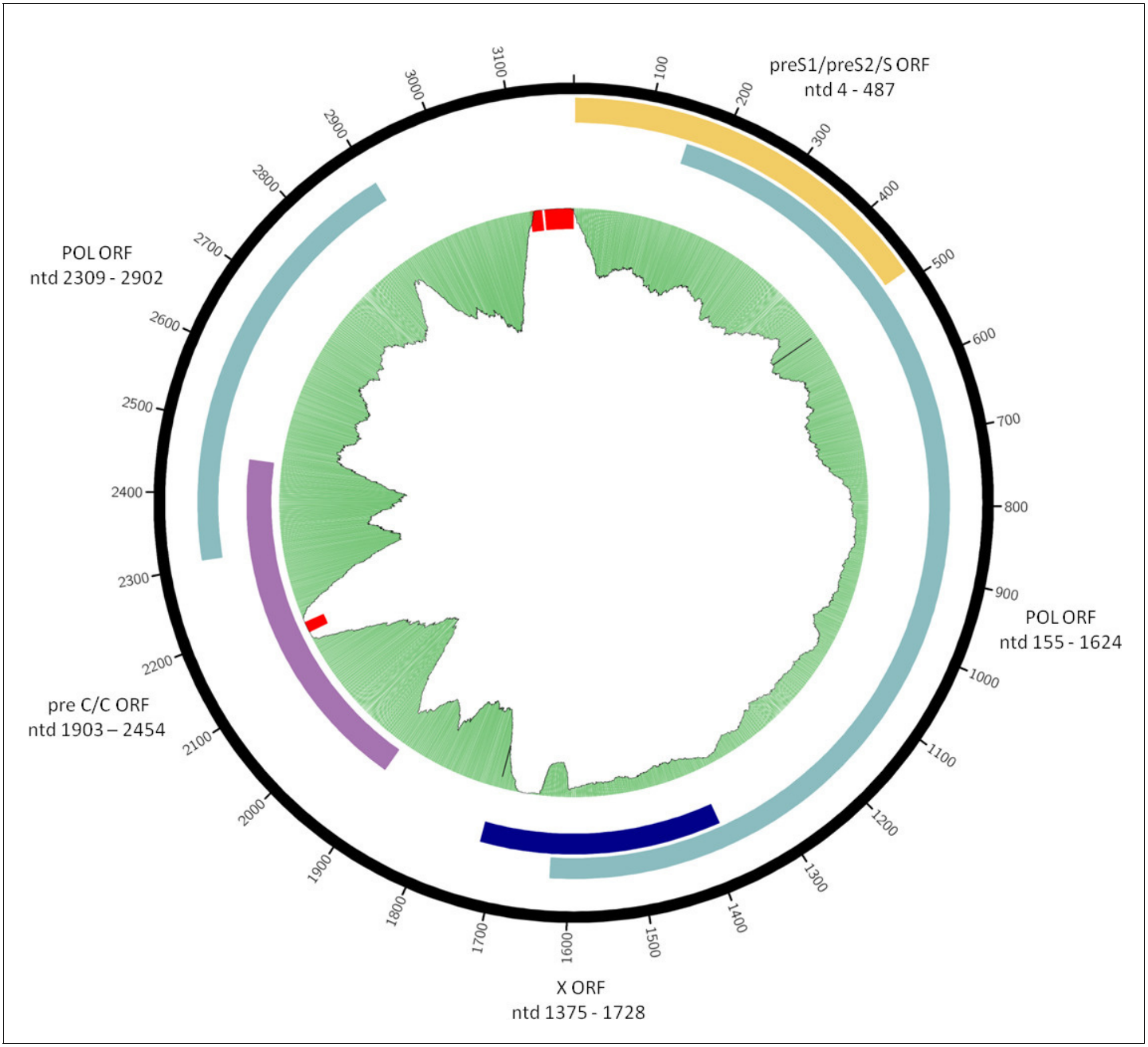


Figure 2—figure supplement 1. Consensus sequence of the Karsdorf HBV genome. Organization of overlapping open reading frames and approximate location of single-stranded portion of plus strand are indicated. Gaps in the sequence are marked in red. The green plot depicts the coverage of the re-mapping of raw reads against the consensus. Circular plots were generated using circos-0.69-6 and coverage information from the re-mapping.

DOI: https://doi.org/10.7554/eLife.36666.014

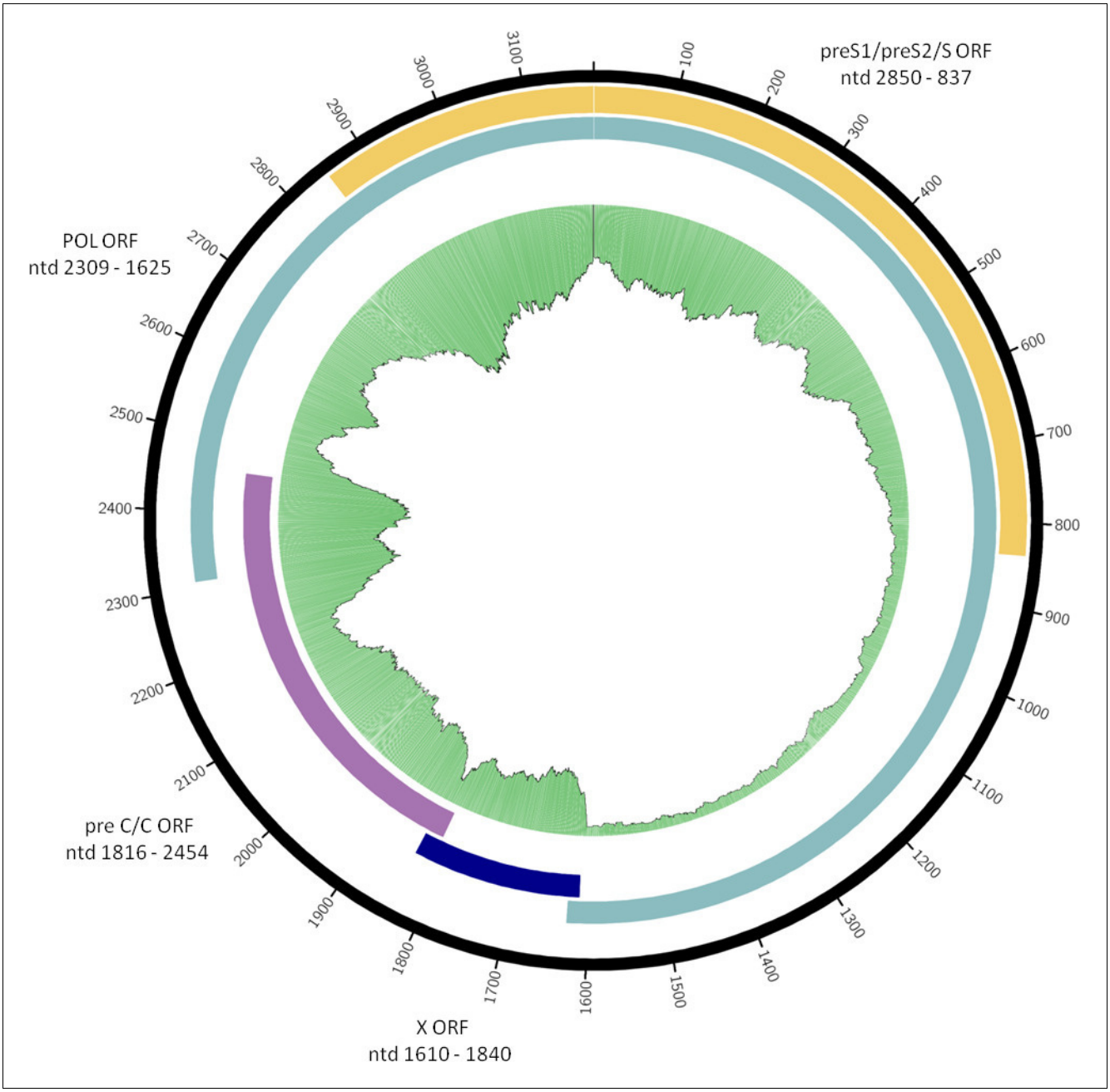


Figure 2—figure supplement 2. Consensus sequence of the Sorsum HBV genome. Organization of overlapping open reading frames and approximate location of single-stranded portion of plus strand are indicated. Gaps in the sequence are marked in red. The green plot depicts the coverage of the re-mapping of raw reads against the consensus. Circular plots were generated using circos-0.69-6 and coverage information from the re-mapping.

DOI: https://doi.org/10.7554/eLife.36666.015

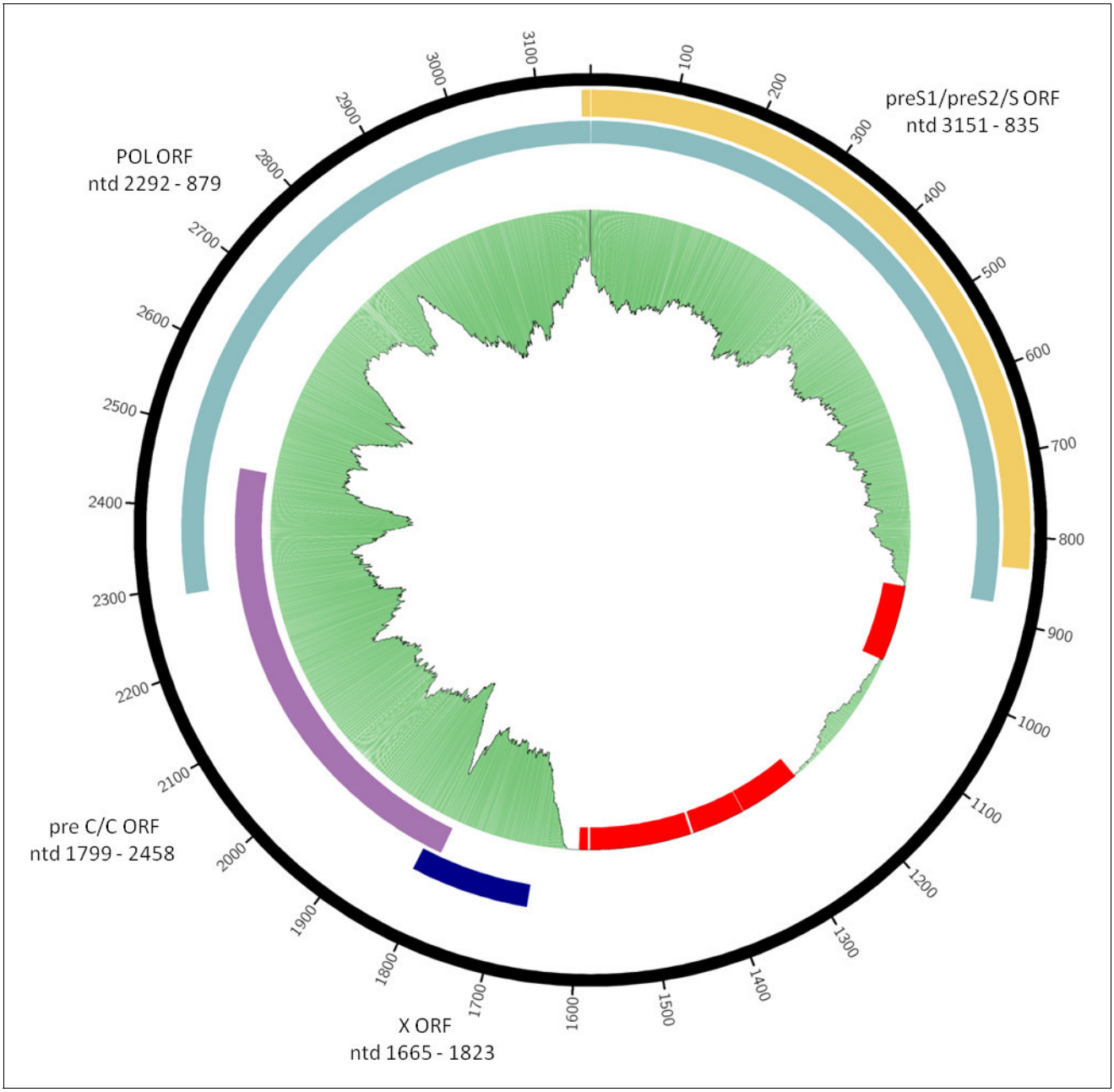


Figure 2—figure supplement 3. Consensus sequence of the Petersberg HBV genome. Organization of overlapping open reading frames and approximate location of single-stranded portion of plus strand are indicated. Gaps in the sequence are marked in red. The green plot depicts the coverage of the re-mapping of raw reads against the consensus. Circular plots were generated using circos-0.69-6 and coverage information from the re-mapping.

DOI: https://doi.org/10.7554/eLife.36666.016

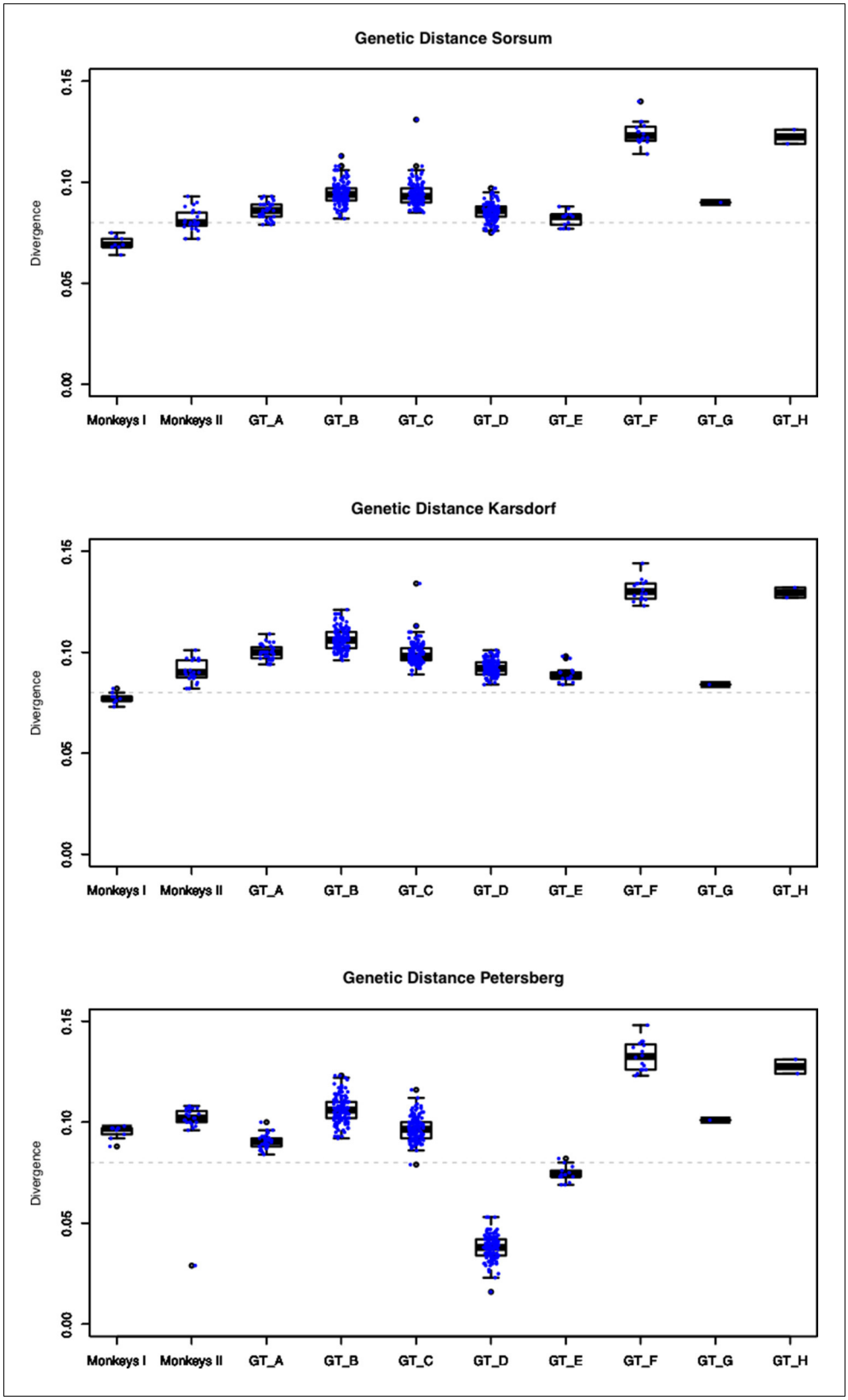


Figure 2—figure supplement 4. Genetic (hamming) distance of our three ancient HBV genomes compared to all 493 reference genomes. Gaps or non-called sites ('N') were ignored.

DOI: https://doi.org/10.7554/eLife.36666.017

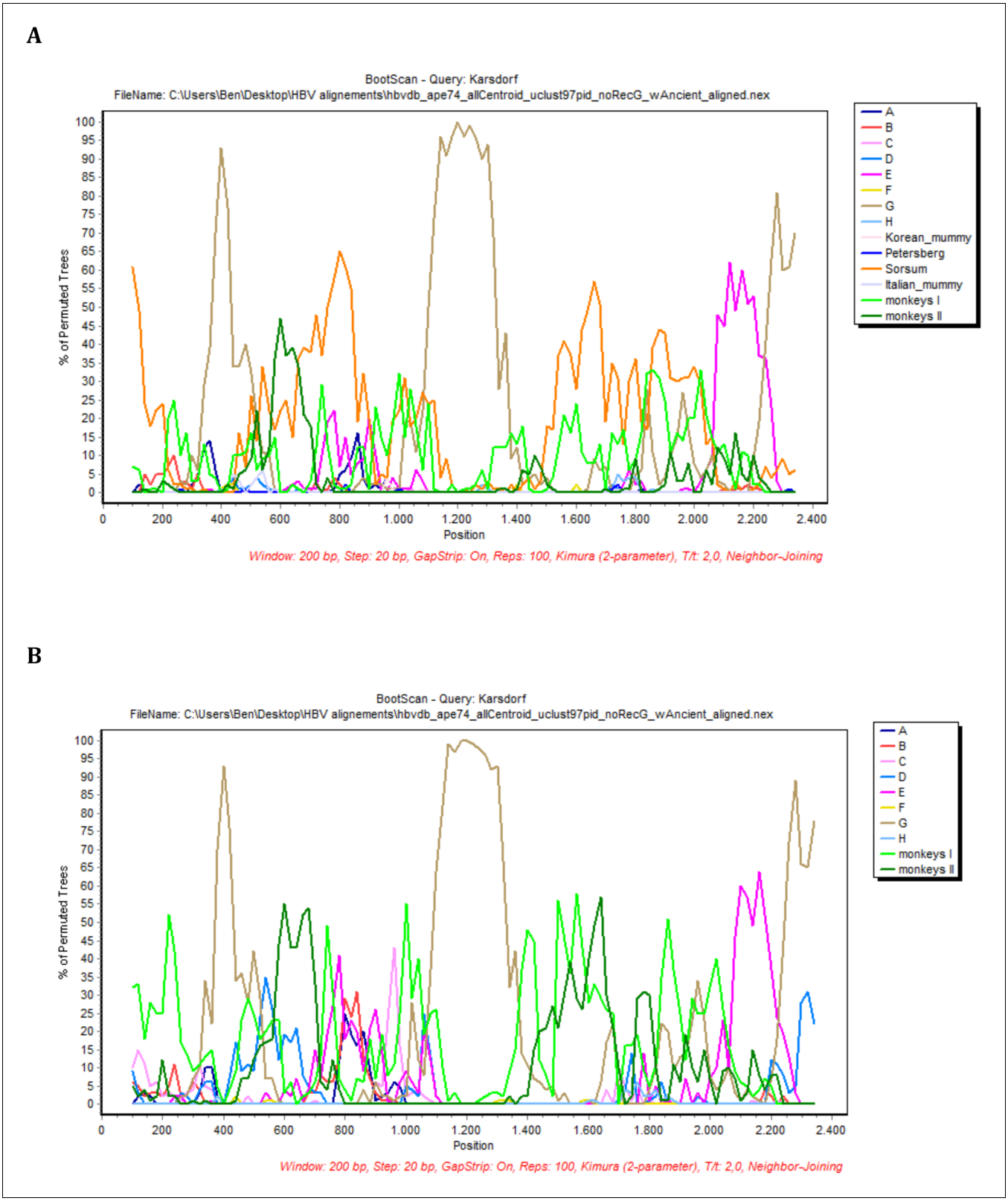


Figure 2—figure supplement 5. BootScan analysis of the sequence Karsdorf. In each case, sequence fragments of 200 bases incrementing by 20 bases, 100 bootstrap replicates, were compared with sequence groups of (**a**) the eight human genotypes, two primate genotypes, and four ancient genomes and (**b**) the eight human genotypes, two primate genotypes (color coded as described in the legend).

DOI: https://doi.org/10.7554/eLife.36666.018

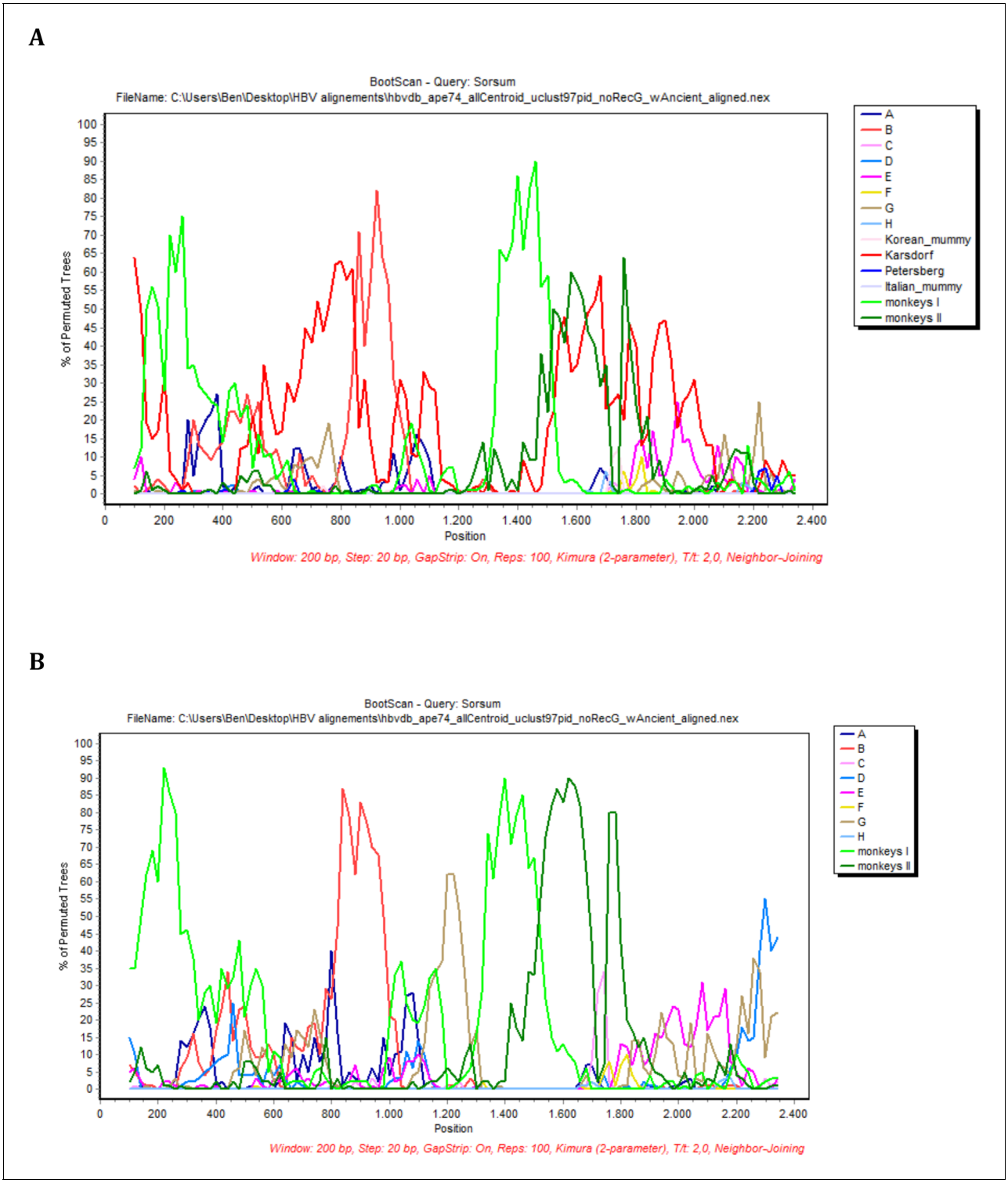


Figure 2—figure supplement 6. BootScan analysis of the sequence Sorsum. In each case, sequence fragments of 200 bases incrementing by 20 bases, 100 bootstrap replicates, were compared with sequence groups or 50% consensus sequences of (**a**) the eight human genotypes, two primate genotypes, and four ancient genomes and (**b**) the eight human genotypes, two primate genotypes (color coded as described in the legend).
DOI: https://doi.org/10.7554/eLife.36666.019

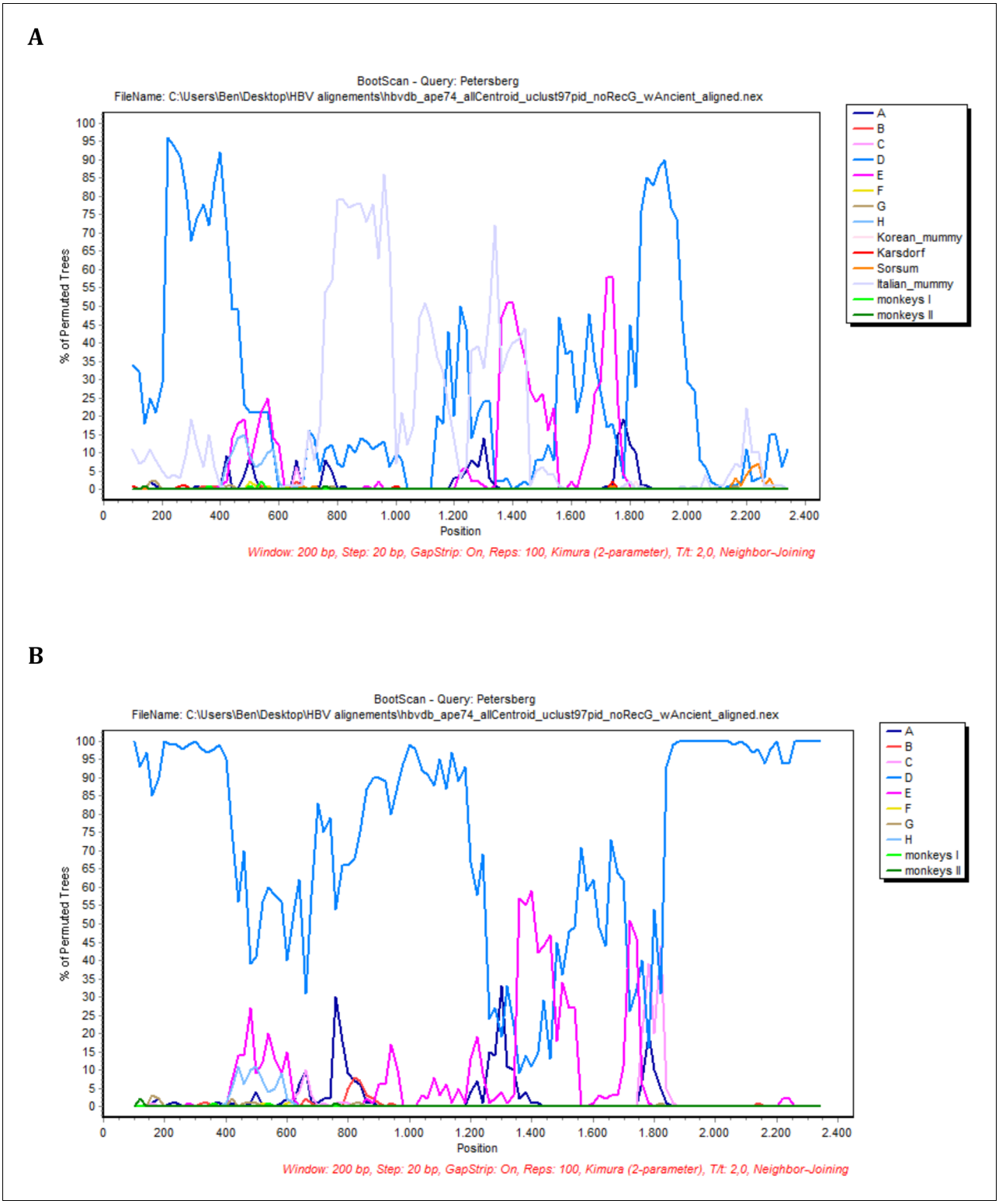


Figure 2—figure supplement 7. BootScan analysis of the sequence Petersberg. In each case, sequence fragments of 200 bases incrementing by 20 bases, 100 bootstrap replicates, were compared with sequence groups or 50% consensus sequences of (**a**) the eight human genotypes, two primate genotypes, and four ancient genomes and (**b**) the eight human genotypes, two primate genotypes (color coded as described in the legend).
DOI: https://doi.org/10.7554/eLife.36666.020

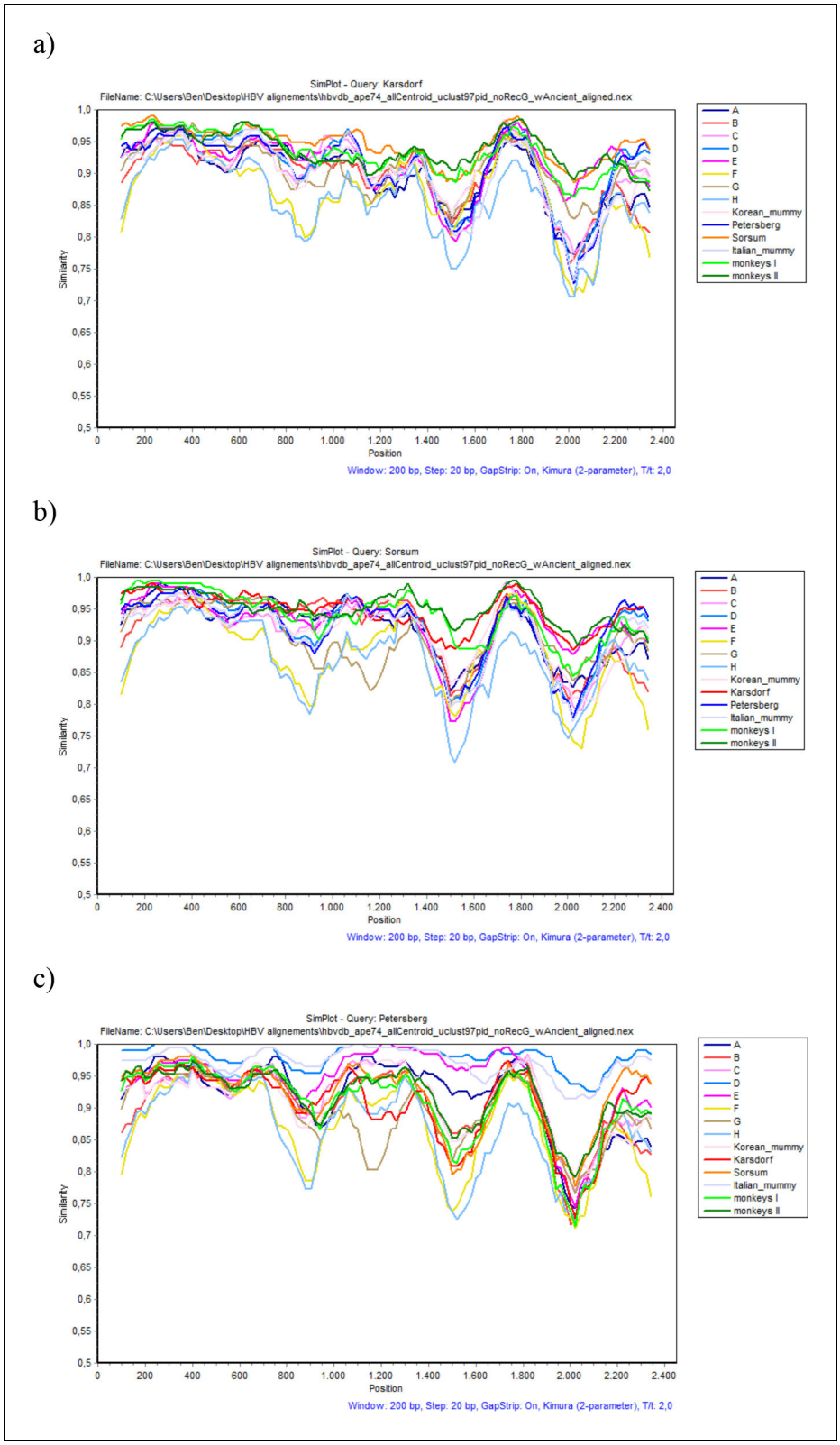


Figure 2—figure supplement 8. SimPlot analysis of (**a**) Karsdorf, (**b**) Sorsum and (**c**) Petersberg. In each case, sequence fragments of 200 bases incrementing by 20 bases, 100 bootstrap replicates, were compared with sequence groups of the eight human genotypes, four primate genotypes and four ancient genomes (color coded as described in the legend).

DOI: https://doi.org/10.7554/eLife.36666.021

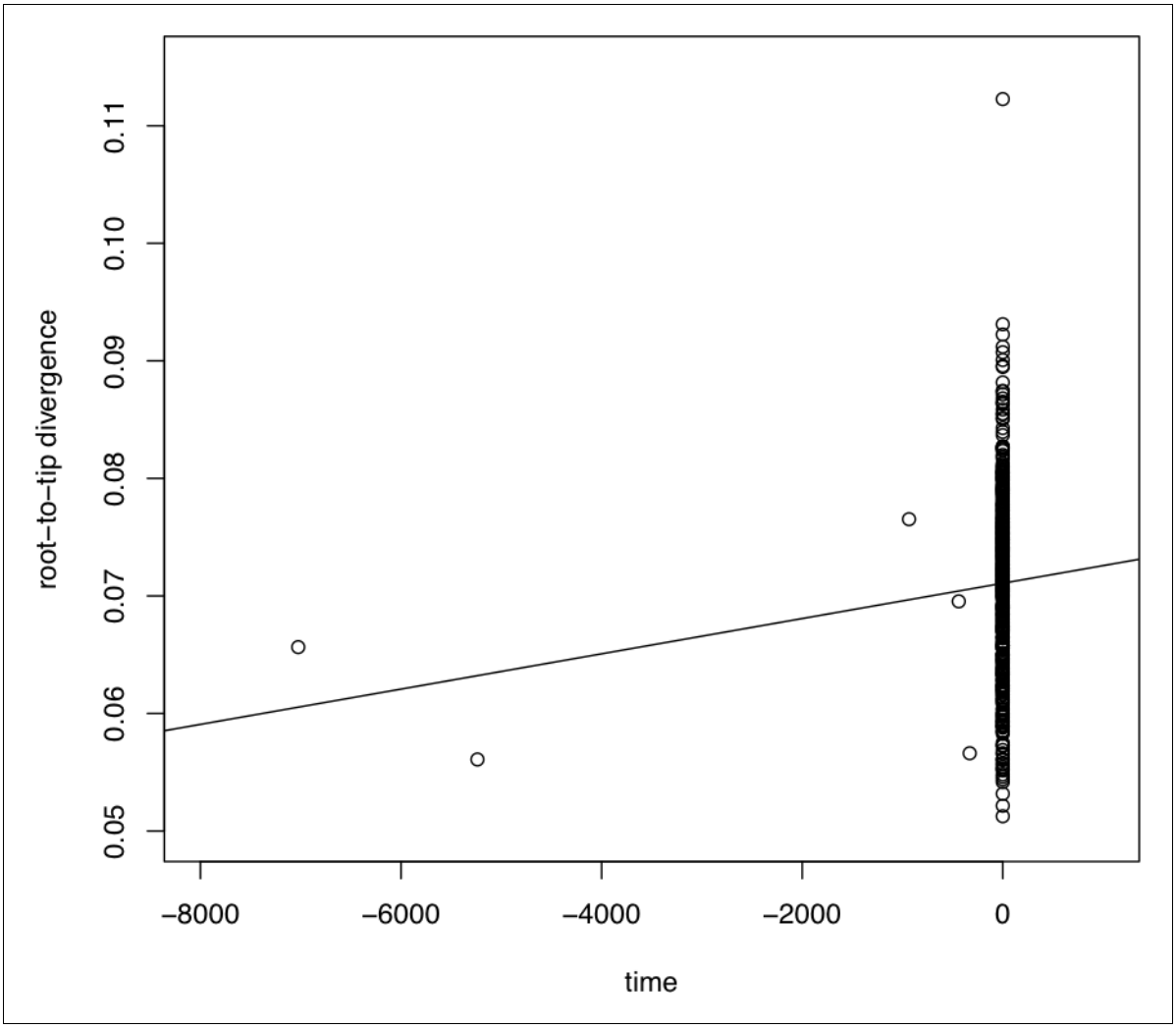


Figure 2—figure supplement 9. Plot of phylogenetic root-to-tip distance relative to sampling time (TempEst). Each dot represents one sample.
DOI: https://doi.org/10.7554/eLife.36666.022

13.

Chapter IV: Mass child burials in early Berlin. Victims of pathogens or general harsh circumstances?

Supplementary material

Table S1. List of samples from Petriplatz that were analyzed in this study. Graves are separated by thick black dashes. The sample ID contains of four numbers and a letter. The second letter is indicating the material as given in column three. The same sample ID can appear several times when different material was sampled. The fourth column depicts the detected pathogen and number of reads in the initial screening.

Grave	Sample ID	Material	Detected Pathogen (No of reads in the initial screening)
1	2675DT	tooth	B19(5)
1	2675AT	tooth	-
1	2675AP	petrous bone	-
1	2675CT	tooth	-
2	3009BT1	tooth	B19(33), Y. pestis (), HBV (D1)
2	3009BT2	tooth	
2	3009CT	tooth	-
2	3009CP	petrous bone	-
2	3009DT1	tooth	B19(1)
2	3009DT2	tooth	
2	3009ET	tooth	B19(9), HBV(16)
2	3009EP	petrous bone	
2	3009FT	tooth	-
2	3009FP	petrous bone	-
2	3009GT	tooth	HBV(1)
2	3009AT	tooth	
3	2993ET	tooth	B19(4)
3	2993AT	tooth	B19(3)
3	2993BT	tooth	HBV(25)
3	2993BP	petrous bone	
3	2993CT	tooth	-
3	2993CP	petrous bone	-
3	2993DT	tooth	B19(1)
3	2993DP	petrous bone	
4	3152BT	tooth	B19(2)
4	3152AT	tooth	-
4	3152AP	petrous bone	-
5	3129AT	tooth	B19(18)
5	3129CT	tooth	B19(33)
5	3129BT	tooth	HBV(63)
5	3129BP	petrous bone	
6	2884ET	tooth	B19(29)
6	2884AT	tooth	HBV(109)
6	2884AP	petrous bone	
6	2884BP	petrous bone	-
6	2884CT	tooth	B19(2)
6	2884DT	tooth	-
7	3060AP	petrous bone	-
7	3060CT	tooth	B19(81)
7	3060CT	tooth	

7	3060BT	tooth	B19(1)
7	3060BP	petrous bone	
7	3060BT2	tooth	
7	3060DT	tooth	B19(1)
7	3060DP	petrous bone	
8	2674AT	tooth	-
8	2674AP	petrous bone	-
8	2674BT2	tooth	-
8	2674BT	tooth	-
8	2674BP	petrous bone	-
8	2674CT	tooth	B19(3)
8	2674CP1	petrous bone	
8	2674CP2	petrous bone	

Table S2. List of references used in the in the multi fasta file for the HBV mapping.

Accession	Genotype	Host
X51970	A	Human
AB073846.1	B	Human
M12906	C	Human
M32138	D	Human
AB032431	E	Human
AB036910	F	Human
AB064310	G	Human
AY090454	H	Human
LT992440.1	A	Ancient Human Strain
LT992441.1	A	Ancient Human Strain
LT992442.1	B	Ancient Human Strain
LT992447.1	A	Ancient Human Strain
LT992448.1	A	Ancient Human Strain
LT992438.1	D	Ancient Human Strain
LT992439.1	D	Ancient Human Strain
LT992443.1		Ancient Monkey Strain
LT992444.1	D	Ancient Human Strain
LT992454.1	D	Ancient Human Strain
LT992459.1		Ancient Monkey Strain
JN315779.1	C	Ancient Human Strain
MG585269.1	D	Ancient Human Strain
Karsdorf		Ancient Monkey Strain
Sorsum		Ancient Monkey Strain
Petersberg	D	Ancient Human Strain
AB032433		Chimpanzee
AF222323		Chimpanzee
AJ131567		Gorilla
AY330911.1		Chimpanzee
AJ131571.1		Gibbon
U46935.1		Gibbon
FM209516.1		Gibbon
AF193863.1		Orangutan

Table S3. Basic mapping statistics of HBV positive samples depicting the number of merged reads, the number of mapped reads against the multi fasta reference file and the depth of coverage after remapping the raw data against the generated consensus (DeDup means after duplicate removal).

Sample	Merged reads	Mapped reads after DeDup	Mean depth coverage remapping remapping reads to consensus after DeDup
3009B	549840325	567	4.85x
3009E	15365549	4273	43.48x
3009G	15675347	166	1.36x
2993B	14727771	7871	78.93x
3129B	11684433	13994	100.44x(D) 42.06x(G)
2884A	26113493	15102	104.4x

Table S4. List of references used in the in the multi fasta file for the Parvovirus B19 mapping.

Accession	Genotype
FN669502.1	1
DQ357065.1	1
AF113323.1	1
AJ717293.1	2
DQ333427.1	2
HQ340602.1	2
AY083234.1	3
AJ249437.1	3
NC_001540.1	Bovine Parvovirus
DA251	1
DA337	1
DA336	1
DA66	1
NEO105	1
VK477	2
VK143	2
RISE569	2
VK6	2
VK154	2

Table S5. Basic mapping statistics of B19 positive samples depicting the number of mapped reads and the number of mapped reads. Depth of coverage after the remapping of the raw data is only available for the samples where a consensus could be generated (DeDup means after duplicate removal).

Sample	Merged reads	Mapped reads after DeDup	Mean depth coverage remapping remapping reads to consensus after DeDup
2675D	139722857	39	-
3009B	549840325	1476	-

3009D	129039531	20	-
3009E	15365549	374	-
2993E	130807633	131	-
2993A	12778918	137	-
2993D	11323639	0	-
3152B	117739289	223	-
3129A	122642356	349	-
3129C	122876767	2909	6.34x
2884E	135739375	713	-
2884C	15988129	1	-
3060C	525214336	8943	39.78x
3060B	9867678	2	-
3060D	18097766	75	-
2674C	15578512	1200	

Table S6. Chosen K-values for B19 and HBV

B19	
Sample	K-value
3129C	21
3060C	23
HBV	
KH180682 B507	29
KH180676 B196	99
KH180658 B186	25
KH180652+KH170209 B3009	31
KH180647 B267	103
KH180663 B223	27

Additional Information.

The HBV dataset containing the 495 representative genomes can be found on the attached CD.

Dataset 1: Representative dataset of complete parvovirus B19 genomes used for genotype assignment. The composition of the dataset was taken from (1).

KX752821, AJ781038, KM393165, KM393168, KM393167, KM393166, AB126268, AB126267, AB126266, KM065415, KC013340, KT310174, KR005643, DQ225150, DQ225149, KR005644, KR005641, KR005640, KR005642, NC_000883, AY386330, KM065414, KM393163, M13178, FN598218, FN598217, Z70560, Z68146, DQ408301, AB030693, AB030673, KM393164, AY028237, KC013329, KC013325, Z70528, KT268312, KC013305, 5 AF113323, M24682, Z70599, AY504945, AB126271, AB126264, AB126263, AB126262, AB126269, AF162273, FJ591158, KC013343, KC013324, KC013308, KM393169, AB030694, DQ293995, KC013316, DQ225151, KC013321, KC013344, KC013338, KC013331, KC013314, KC013312, KC013333, KC013346, KC013351, KC013313, KC013327, KC013332, AB126265, KC013322, DQ225148, AB126270, DQ357065, DQ357064, KF724387, AY903437, AY044266, DQ333426, AB550331, EF216869, AJ717293, KF724386, DQ333428, AY064476, AY064475, AY647977, AY083234, AY582124, DQ234779, DQ234778, DQ408305, DQ408302,

DQ408304, DQ408303, FJ265736, AY582125, DQ234775, DQ234771, DQ234769, NC_004295, AJ249437

Dataset 2: Dataset used for phylogenetic analyses. Contains 10 ancient and 77 modern sequence. The composition of the dataset was taken from (1) and expanded bay the ancient B19 strains published in (1).

AJ781038, KM393165, KM393168, KM393166, AB126267, AB126266, KM065415, KC013340, KT310174, DQ225150, DQ225149, KR005641, KR005640, AY386330, KM393163, M13178, FN598218, Z70560, Z68146, DQ408301, AB030673, KM393164, KC013329, KC013325, KT268312, KC013305, AF113323, Z70599, AY504945, AB126271, AB126262, AB126269, AF162273, FJ591158, KC013343, KC013324, KC013308, KM393169, AB030694, DQ293995, KC013316, DQ225151, KC013321, KC013312, KC013333, KC013346, KC013313, KC013327, AB126265, AB126270, DQ357065, DQ357064, KF724387, AY903437, AY044266, DQ333426, AB550331, EF216869, AJ717293, KF724386, DQ333428, AY064476, AY064475, AY647977, AY083234, AY582124, DQ234779, DQ234778, DQ408305, DQ408302, DQ408304, DQ408303, FJ265736, DQ234775, DQ234771, DQ234769, AJ249437, DA251, DA336, DA337, RISE569, NEO105, VK6, VK143, VK154, VK477, DA66

1. Mühlemann B, Margaryan A, De Barros Damgaard P, Allentoft ME, Vinner L, Hansen AJ, et al. Ancient human parvovirus B19 in Eurasia reveals its long-term association with humans. Proc Natl Acad Sci U S A. 2018;115(29):7557–62.