

Genomics of Ancient Pathogens: First Advances and Prospects

Alexandra B. Malyarchuk^{1,a*}, Tatiana V. Andreeva^{1,2}, Irina L. Kuznetsova^{2,3},
Svetlana S. Kunizheva^{2,3}, Maria S. Protasova², Lev I. Uralsky^{2,3}, Tatiana V. Tyazhelova²,
Fedor E. Gusev², Andrey D. Manakhov^{2,3}, and Evgeny I. Rogaev^{2,3,4,a*}

¹Center for Genetics and Genetic Technologies, Faculty of Biology,
Lomonosov Moscow State University, 119234 Moscow, Russia

²Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russia

³Center for Genetics and Life Science, Sirius University of Science and Technology, 354340 Sochi, Russia

⁴Department of Psychiatry, UMass Chan Medical School, Shrewsbury, MA 01545, USA

^ae-mail: a_malyarchuk98@mail.ru

^be-mail: rogaev@vigg.ru

Received December 14, 2021

Revised January 8, 2022

Accepted January 21, 2022

Abstract—Paleogenomics is one of the urgent and promising areas of interdisciplinary research in the today's world science. New genomic methods of ancient DNA (aDNA) analysis, such as next generation sequencing (NGS) technologies, make it possible not only to obtain detailed genetic information about historical and prehistoric human populations, but also to study individual microbial and viral pathogens and microbiomes from different ancient and historical objects. Studies of aDNA of pathogens by reconstructing their genomes have so far yielded complete sequences of the ancient pathogens that played significant role in the history of the world: *Yersinia pestis* (plague), *Variola virus* (smallpox), *Vibrio cholerae* (cholera), HBV (hepatitis B virus), as well as the equally important endemic human infectious agents: *Mycobacterium tuberculosis* (tuberculosis), *Mycobacterium leprae* (leprosy), and *Treponema pallidum* (syphilis). Genomic data from these pathogens complemented the information previously obtained by paleopathologists and allowed not only to identify pathogens from the past pandemics, but also to recognize the pathogen lineages that are now extinct, to refine chronology of the pathogen appearance in human populations, and to reconstruct evolutionary history of the pathogens that are still relevant to public health today. In this review, we describe state-of-the-art genomic research of the origins and evolution of many ancient pathogens and viruses and examine mechanisms of the emergence and spread of the ancient infections in the mankind history.

DOI: 10.1134/S0006297922030051

Keywords: human populations, ancient DNA, paleopathology, paleogenomics, pathogen, plague, cholera, leprosy, syphilis, smallpox, tuberculosis

INTRODUCTION

During the recent decades, technical possibilities for sequencing ancient DNA (aDNA), including DNA of ancient pathogens, increased significantly. More advanced methods are emerging for genomic library preparation, DNA enrichment, sequencing; tools are also developing for bioinformatics analysis of the genomic data. The number of studies on identification and characterization on the level of molecular genetics of the pathogenic microorganisms extracted from ancient remains is

increasing. Thus, a new field of research has emerged – molecular paleoepidemiology [1] that examines occurrence of infections and spread of pathogens in human populations.

For a long time, the study of infectious diseases was carried out by the traditional paleopathological evaluation of bones of ancient skeletons from archaeological excavations. However, this approach has significant limitations, because many infections do not leave visible traces on the bones, and often no other material is available for examination [2].

First advances in the study of DNA of ancient bacteria and viruses are associated with introduction of the polymerase chain reaction method (PCR). This method makes it possible to detect the presence of infectious agents, but information about the pathogen evolution is

Abbreviations: aDNA, ancient DNA; CPXV, cowpox virus; NGS, next generation sequencing; TIR, terminal inverted repeats; VACV, smallpox vaccine virus; VARV, Variola virus.

* To whom correspondence should be addressed.

scarce, because the genetic material of microorganisms is analyzed in one or in a small number of short fragments amplified from DNA isolated from the remains of ancient people [3, 4]. Moreover, aDNA extracted from archaeological material is usually present in it in small amounts, is strongly fragmented, and contains chemical modifications [5, 6] that makes amplification difficult.

Development of the next generation sequencing (NGS) has led to elucidation of many sequences of whole genomes of modern organisms, including microbes and viruses. Thus, the number of reference genome sequences has increased, including individual strains. For paleogenetic studies, such breakthrough made it possible not only to unequivocally establish the presence of an infectious pathogen in the total DNA isolated from an ancient sample, but also to more accurately determine phylogeny of the isolated strain. Genome of the known bacterial causative agent of plague *Yersinia pestis* published in 2011 was the first successful attempt to present the complete genome of an ancient pathogen. After this report, many studies on the genomes of various pathogens have been published.

By now, the largest amount of genomic data is obtained for *Yersinia pestis* (the plague agent), *Variola virus* (smallpox), *Vibrio cholerae* (cholera), and also for three equally important human endemic infectious agents: *Mycobacterium tuberculosis* (tuberculosis), *Mycobacterium leprae* (leprosy), and *Treponema pallidum* (syphilis).

MAIN APPROACHES OF PALEOGENETICS

Work with aDNA, including aDNA of microbial origin extracted from historic and archaeological specimens,

is associated with some problems and limitations caused by its specific features. Archaeological specimens usually contain mixtures of endogenous (lifetime) and exogenous (posthumous) microbial DNA which can include commensal bacteria (e.g., microbial colonies in the tartar), epidemic pathogens (e.g., *Y. pestis* in the pulp cavity of teeth), and bacteria from the environment (e.g., soil microorganisms involved in biodegradation). Moreover, aDNA in the specimen can be contaminated in the course of analysis, beginning from archaeological excavations and processing of the ancient material through contact with the researchers (e.g., skin microbes), suboptimal storage conditions (e.g., overgrowth of bacteria and fungi), and laboratory sources of contamination, such as reagents (e.g., DNA polymerases and other enzymes, buffers with admixtures of bacterial DNA) [7]. The target component of the isolated aDNA comprises, on average, 1-10%, and the pathogen fraction that can be isolated is less than 0.5% of the total metagenome. The exceptions are samples obtained from permafrost, or DNA preparations isolated from auditory ossicles. Therefore, in the study of ancient pathogens and other microorganisms it is important to confirm authenticity of the findings, i.e., to prove that the isolated DNA belongs to the ancient sample (bacteria or virus), but not to contamination. Therefore, the analysis of aDNA and validation of the findings must strictly follow certain methodological standards.

By now, ancient microbial DNA has been successfully isolated from various samples: bones, teeth, coprolites, museum specimens, calcified dental deposit, soil, etc. (table)

However, human skeletal remains are still the main object for extracting DNA of ancient pathogens. Modern

Sources of bacterial ancient DNA and examples of microorganisms and pathogens, which have been isolated and characterized

Sample	Pathogen
Bones	<i>Mycobacterium tuberculosis</i> , <i>Mycobacterium leprae</i> , <i>Brucella melitensis</i> , <i>Brucella abortus</i> , <i>Treponema pallidum</i>
Teeth (pulp)	various pathogens transmitted through human blood
Teeth (tartar)	oral cavity microbiome bacteria (<i>Streptococcus mutans</i> , <i>Porphyromonas gingivalis</i> , etc.)
Calcinated inclusions in lungs	<i>M. tuberculosis</i>
Mummified remains	various pathogens and commensal bacteria of human microbiome
Coprolites	source of human intestinal microbiota
Museum medical organ specimens embedded in paraffin	<i>Bacillus anthracis</i> , <i>Vibrio cholerae</i> , <i>Variola virus</i> (VARV)
Amber	Gram-negative proteobacterium <i>Brevundimonas</i>
Samples of fossil ice and permafrost	paleoviruses, e.g., <i>Pithovirus sibericum</i>

methodological approaches for extracting aDNA and preparing it for the genome sequencing with large-scale parallel sequencing on the “Illumina” platform are presented in the review [8]. In this section, we will consider examples of specific methods, which are used for assessment and authentication of DNA of the ancient pathogens (taxonomy identification, verification, and authentication of ancient microbes using data of high-throughput DNA sequencing).

Examining taxonomy of ancient pathogens and microorganisms. Traditional methods for determining taxonomic lineage, such as characterization of homology by the method of DNA–DNA re-association, for which it is necessary to isolate and cultivate individual microorganisms, cannot be used in the study of ancient bacteria. Therefore, initially ancient pathogens were genetically analyzed by the PCR method using primers specific for the marker genes of infectious diseases, for instance of the *rpoB* gene (the causative agent of plague, *Y. pestis*) [3, 4]. The main difficulty in interpretation of the PCR results is associated with the presence in the ancient specimens of mixtures of bacterial organisms that can lead to nonspecific and even false-positive results, as it took place in the case of DNA of soil bacteria containing sequences with a high homology to DNA of *M. tuberculosis* and *Y. pestis* [9, 10].

Another widely used approach for identification of microorganisms is based on comparing nucleotide sequences of the gene encoding 16S rRNA [11]. Since the sequence of the 16S rRNA gene is similar in different microbial taxa, universal primers are used for its PCR-amplification. Analysis of the nucleotide sequences of 16S rRNA makes it possible with high degree of certainty to determine both phylogenetic position of microorganisms and reliably identify them at the level of genus, and sometimes at the level of species. Possibility of rapid and inexpensive sequencing of tens or hundreds of thousands of hypervariable target sequences of the 16S rRNA gene, concurrently with the use of new generation sequencing methods has been successfully applied in the large “Human Microbiome Project” [12]. This approach was also used in the studies on aDNA, e.g., for determination of the taxonomy of DNA obtained from a mixture of skeletal remains belonging to different animal species [13]. However, this approach cannot be applied directly to studying ancient bacteria, because the target hypervariable regions of the gene encoding 16S rRNA are longer than the average length of the obtained fragments of the bacterial aDNA molecule.

At present, taxonomy of metagenomic data is studied using approaches of bioinformatics, which are focused on matching the reads with the sequences of 16S rRNA gene (for example, QIIME [14] and Mothur [15]), or with the panels of single-copy genes (for example, MetaPhlan2 [16, 17], MIDAS [18], and PhyloSift [19]), the k-mer-based whole genome matching (e.g., Kraken [20] and

CLARK [21, 22]), as well as the k-mer hybrid matching with the alignment extension (MALT [23, 24]). Targeted amplification of the 16S rRNA gene is popular for profiling complex microbial communities, whereas such programs as QIIME and Mothur are useful for processing the amplicon data and for sorting and taxonomic profiling based on the large databases of 16S rRNA genes, such as Greengenes [25]. DADA2 [26] and Deblur [27] are new methods for sorting 16S rRNA gene reads, which have been introduced in QIIME v2.0 and provide more accurate grouping of the reads than the methods introduced in QIIME v1. These programs require several copies of each sequence, which, although common in the 16S rRNA gene amplicon datasets, are unlikely to occur in the ancient read sets selected from metagenomic datasets by the shotgun-sequencing method because of the insufficient coverage. In the study by Velsko et al. [28] it was shown that the targeted amplification of the gene of 16S rRNA from the strongly fragmented ancient samples of DNA could lead to taxonomic errors because of the long hypervariable regions and the extensive length polymorphisms [29]. Moreover, it is difficult to assemble the gene of 16S rRNA from metagenomes because of its highly conserved regions [30].

One of the main differences between the various programs is the approach for calculating relative abundance. Using a set of single-copy marker genes, MetaPhlan2 and MIDAS attempt to represent proportion of the cells of every species found in the sample. This is different from the k-mer methods, such as CLARK-S and MALT, which report proportion of the total DNA assigned to every species. The genome size can vary significantly between the bacterial species, and the species with larger genomes may appear more abundant in a sample because they possess the larger proportion of DNA, even if the number of the cells is low. To determine approximately the number of cell copies, the relative number of species obtained by the k-mer identification methods can be normalized to the predicted genome size, even if the exact strain is unknown, since the genome size essentially correlates within a species. When considering metagenomic community profiles, one should keep in mind the difference between the relative abundance reported by metagenomic profilers normalizing (MetaPhlan2 and MIDAS) and not normalizing (Clark-S, MALT, QIIME) the number of cell copies.

CLARK-S is best suited for obtaining the largest number of specific reads, or for detecting the relative contents of all DNA fragments (if, for example, it is necessary to attempt to assemble a genome from all reads assigned to the same species). However, it is impossible to detect true species with a low abundance (especially viruses and bacteriophages) using CLARK-S, because of the high frequency of false-positive identifications with the abundances lower than 0.1%.

MALT is a unique tool because it can provide functional classification of the reads, as well as a taxonomic

classification, but it has limitations in the assignment, when the database under consideration contains a large number of closely related species. Moreover, both CLARK-S and MALT exhibit a high percent of false-positive determinations in low-abundant taxa.

The least accurate QIIME/UCLUST method includes many false-positive results, even when the low-abundant taxa are discarded. According to the work of Velsko et al. [28], it was the only program, performance of which was markedly different between the ancient and modern specimens.

The influence of DNA damage models on the taxonomic distribution varies between the different programs, but majority of the programs tested [28] are prone to incorrect distribution caused by DNA damage only in a small number of the reads. Overall, the results show that for the highly abundant species taxonomic errors are similar between the models of the modern and ancient metagenomic datasets.

DNA damage affects differently the rates of false-positive results of the five programs tested, but displacement of the database affects profiling of the most common species much more strongly. CLARK-S, MetaPhlan2, MIDAS, and QIIME/UCLUST have the higher number of false-positive determinations in the ancient than in the modern datasets, however, these differences are not significant in the MetaPhlan2. This can be a challenge for comparing microbial profiles of the ancient and modern metagenomes, especially in order to determine the presence and distribution of species with low abundance. If we neglect the higher percent of false-positive results in the ancient specimens compared to the modern ones, it is possible to draw incorrect conclusion that many species with low frequency of occurrence were lost during the evolution. Therefore, it is very desirable to have similar false-positive rates for the ancient and modern datasets, and such programs, as MALT and MetaPhlan2 are more suitable for this purpose than CLARK-S, MIDAS, and QIIME/UCLUST, because the false-positive rates in the ancient and modern specimens are comparable both before and after filtration.

Recently, Hübner et al. [31] have developed a new bioinformatics tool HOPS for heuristic screening of pathogens based on the mapDamage algorithm. This algorithm analyzes the data obtained by MALT for identification and authentication of bacterial pathogens in the ancient metagenomic specimens and for extracting this information for further analysis. Combined with the hybridization capture for obtaining the larger number of the ancient eukaryote sequences, HOPS makes it possible to assess the aDNA identity based on the DNA damage profile, even if the number of available sequences is very low (≤ 50 reads per species).

The results of HOPS can be envisaged using the program MaltExtract Interactive Plotting Application

(MEX-IPA, by J. Fellows Yates; <https://github.com/jfy133/MEX-IPA>) for visualizing profiles of the aDNA damage of the target taxa.

Databases used for analyzing ancient pathogens and microorganisms. Currently, many various databases are available, which are used for metataxonomic analysis of ancient specimens. These databases include those oriented to one locus, such as the gene of 16S rRNA, or other genes presenting orthologs of these genes in all species of microorganisms contained in the database. There are several large-scale popular databases of 16S rRNA genes, each of which containing millions of aligned reference sequences: e.g., Ribosomal Database Project (RDP) – <http://rdp.cme.msu.edu> (more than 3 million sequences), the Greengenes database – <https://greengenes.secondgenome.com>, and the SILVA project – <https://www.arb-silva.de> (more than 2 million sequences).

Moreover, there are databases based on several loci consisting of a small set of bacterial housekeeping genes. Among them there is, in particular, the MetaPhlan3 database, which contains over a million unique marker genes selected for determination of specific microbial clades (bacterial, archaeal, and viral – over 17,000 reference genome sequences obtained from over 7000 taxa at the species level).

The third type of databases is the one including complete genomes. They are advantageous as they make it possible to identify the known microorganisms and pathogens, which are present in the specimen even in small amount, because any sequenced fragment of DNA of a definite microbial species is potentially present in the target database. Therefore, the full-genomic databases can be most useful in the studies for detecting ancient pathogens, when only trace amounts of DNA from pathogens are expected to be present in the sample. The currently available databases are listed below:

- Genomes OnLine Database (GOLD; <https://gold.jgi.doe.gov>), DOE Joint Genome Institute, USA.
- NCBI Reference Sequence Database (RefSeq; <https://www.ncbi.nlm.nih.gov/refseq>), National Center for Biotechnology Information, U. S. National Library of Medicine, USA.
- NCBI GenBank Sequence Database (GenBank; <https://www.ncbi.nlm.nih.gov/genbank/>), National Center for Biotechnology Information, U. S. National Library of Medicine, USA.
- Metagenome analyzer (MEGAN; <https://computomics.com/services/megan6.html>), Center for Bioinformatics, Tübingen University, Germany.
- Microbial Database for Activated Sludge (MiDAS; <https://www.midasfieldguide.org/global>), Center for Microbial Communities, Department of Chemistry and Bioscience, Aalborg University, Denmark.
- Virus Pathogen Resource (ViPR; <https://www.viprbrc.org/>), collaboration between the University of Chicago and J. Craig Venter Institute, USA.

- Greengenes 16S rRNA gene database (Greengenes, <https://greengenes.secondgenome.com>), Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, California, USA.
- Global Initiative on Sharing Avian Influenza Data (GISAID; <http://gisaid.org>).
- Virus-Host DB (<https://www.genome.jp/virushostdb/>), Institute for Chemical Research, Kyoto University, Japan.
- Viral Bioinformatics Resource Center (VBRC; <https://4virology.net/organisms/dsdna-viruses/poxviridae/>), University of Victoria, Canada.
- Nextstrain (<https://nextstrain.org/>), collaboration between researchers in Seattle, USA and Basel, Switzerland.

Assessment of DNA damage. DNA is strongly fragmented as a result of depurination with the subsequent hydrolysis of the major DNA chain, and as a result, molecules of ancient DNA are short fragments, more often consisting of 30-60 bp, and this trait can be used as a marker for discriminating the ancient sequences of DNA from contamination by modern DNA characterized by the longer sequences. Presence of cytosine deamination, specific replacements C→T and G→A in the fragments under analysis also may be considered as a trait of the authentic aDNA [6]. The supposed level of cytosine deamination can be assessed through analyzing misincorporation profile (GC>TA nucleotide shift) in the results of whole-genome sequencing, when comparing the analyzed sequence with the reference genome used for the sequence alignment. Currently, for more reliable quantitative assessment of patterns of aDNA damage of pathogens bioinformatics programs are used, such as mapDamage [32].

Assessment of DNA stability in a causative agent. At the assessment of the damage of aDNA from pathogens, it is necessary to take into account different preservation time of DNA in various microorganisms. Thus, it is known that the cell wall of Gram-positive bacteria, e.g., of causative agents of tuberculosis and leprosy (*M. tuberculosis* and *M. leprae*), is several times thicker than the cell wall of Gram-negative agents; presence of certain chemical components in the cell wall of these bacteria possibly protects against the damaging effect of water that can play a key role in the preservation of bacterial DNA. This is taken into account when DNA damage is assessed in these pathogens; for example, analysis of the ancient *M. leprae* genomes showed a decrease in the level of deamination [33].

Modeling of the origin of the ancient DNA of microorganisms is one of the tools for assessment of authenticity. For analysis of ancient bacteria from such sources as tartar and paleofeces, it is recommended to assess biological significance of the composition of the microbiome community before further analysis. For example, in the study by Warinner et al. [34] the total DNA isolated from a

specimen of tartar was modeled using the SourceTracker software tool as a mixture of DNA originating from the plaque-producing bacteria, skin bacteria, soil bacteria, etc. The algorithm estimated proportion of the tartar sample originated from each source based on comparison with the reference datasets of current metagenomic data selected for each of these sources before further analysis of the microbiome structure and function [34]. The authors determined that the archaeological specimens of tartar exhibiting inconsistent SourceTracker profiles were not suitable for determination of the oral cavity ancient microbiocenosis, because it could be altered as a result of DNA decomposition or contamination [34].

Enrichment of genomic libraries. Since availability of the massive parallel DNA sequencing technology increased, the methods have been developed for DNA library enrichment to study ancient microbes. In many cases, it is not necessary to sequence the whole genome of a pathogen – sometimes it is sufficient to obtain nucleotide sequences of several specific loci. In such cases, the libraries for sequencing are enriched with the target fragments of DNA by hybridization, using the single-stranded DNA or RNA sequences with high homology (Hybridization capture) as probes. An ancient strain of the *Y. pestis* pathogen responsible for the outbreak of the first plague epidemic (VI-VIII centuries) was identified using this approach [35].

BACTERIAL INFECTIONS

Leprosy (Hansen's disease) is considered to be one of the oldest known human diseases, the main causative agent of which is the *M. leprae* bacterium (discovered in XIX century, this bacterium became the first known causative agent of a disease). A small number of the most severe form of leprosy is associated with *Mycobacterium lepromatosis*, a bacterium from the sister taxon [36]. Leprosy is a chronic disease transmitted by airborne droplets from person to person; the disease is characterized by damage of the skull bones (the rhinomaxillary syndrome) and deformities of the hands and feet. Diagnosis of leprosy in the ancient skeletal remains is difficult, because similar changes in the skeleton can be caused by other diseases, such as syphilis or psoriatic arthritis [37]. Since *M. leprae* is an obligate pathogen, its presence in the remains of ancient people is thought to be clear evidence of the infection [38].

Based on the historic and epidemiological data, geographic distribution of leprosy in ancient times was analyzed, and it was identified that the leprosy outbreaks were associated with the periods of globalization, such as the era of the Great Migration of Peoples or establishing of the Roman Empire, when large regions were associated with regular contacts between populations of different

territories or with trade routes [39-43]. In Russia, leprosy is known from the time of Kievan Rus', as established by the paleopathological investigations of the skeletal remains from the X century cemeteries in the territory of Kiev located at the crossroads of the Great Silk Road in Europe and was part of the zone of leprosy spread in the ancient times [44].

For the first time the scheme of the leprosy causative agent evolution was proposed by Monot et al. [45] based on the comparative genetic analysis of numerous modern samples from various geographic regions. The study has shown that all current cases of leprosy may be assigned to one clone, and four main strains of *M. leprae* were described. Based on the spread of these strains, a hypothesis was suggested that the disease emerged in East Africa or Middle East and spread subsequently in the course of migrations of people. Europeans or North Africans introduced leprosy to West Africa and America during the last 500 years. Additional studies using the genetic data on the *M. leprae* bacterium isolated from the ancient remains (of 1500 years old) of a leprosy patient from Egypt, increased the number of *M. leprae* varieties survived to the present days to sixteen [46].

The genomes of ten strains of *M. leprae* were obtained in the study by Schuenemann et al. [47] from the skeletal remains of the patients with leprosy from various regions of medieval Europe (Italy, Hungary, Czech Republic, Great Britain, and Denmark) dating back to the IV-XIV centuries and genetic diversity and population structure of *M. leprae* were analyzed. Phylogeny reconstructed on the base of all published medieval and current genomes revealed high diversity of the *M. leprae* strains in the medieval Europe. These strains belong to four main phylogenetic branches of *M. leprae*, including the most basic of them, rarely found anywhere else [47]. It has been shown also that two ancient strains of *M. leprae* and 52 current strains from USA are phylogenetically close indicating that Europe was likely a key region for the early spread of leprosy [48-50].

Fotakis et al. [48] were the first to successfully isolate DNA from the tartar of a Norwegian person of the XVI century, and the *M. leprae* genome was reconstructed for the further whole-genome sequencing on the Illumina platform in the single-end reads mode. The authors have obtained aDNA of *M. leprae* where 76% of the genome was represented with the five-fold coverage. The phylogenetic analysis using 164 earlier published ancient and modern genomes of *M. leprae* has shown a pronounced genetic similarity of this ancient Norwegian genome of *M. leprae* with the genomes of other modern and ancient strains from the Northern Europe. Thus, it was shown that the human tartar could serve an alternative material for detecting and genomic analysis of *M. leprae* in the ancient skeletal remains instead of bones and teeth, which is especially important when this disease cannot be diagnosed with paleopathological methods.

The origin of leprosy on the Pacific Ocean islands is under discussion. It was thought that leprosy was introduced to the islands by Europeans during colonization in the XIX century, but some paleopathological data suggest that the disease appeared earlier [51]. Results of the recent study on nine genomes of *M. leprae* from the Pacific Ocean islands [52] allow assuming that the bacterium could be brought to Oceania during the first human migrations about 3000 years ago, and re-introduced during the subsequent migrations. This supports the hypothesis of the earlier introduction of leprosy into the Pacific Ocean region [53].

Isolation of patients with leprosy and, respectively, organization of the separate burials at leper colonies made it possible to conduct a unique comparative genetic analysis of the remains of the patients with leprosy with the remains from the ordinary burials from the same area and the same time period. Thus, population analysis was performed of the remains from the cemetery at the St. Jørgen leper colony (Denmark, XIII-XVI centuries) and the genes associated with the immune status were studied. In the medieval populations there were shown significant associations of the human leukocytic antigen allele DRB1*15:01 and the combination of haplotypes DRB1*15:01 and DQB1*06:02 with lepromatous leprosy [54]. In the current populations the allele DRB1*15:01 also is a factor in the susceptibility to leprosy. Note that combinations of the alleles DRB1*15:01-DQB1*06:02 presents a haplotype widely spread in modern Europeans [55, 56], which is associated with a genetic risk factor for the development of ulcerative colitis, sarcoidosis, and multiple sclerosis [57-59], but protects against the Type I diabetes [60].

These different associations of diseases highlight the well-known pleiotropy of HLA variants, which influence the population frequency of certain haplotypes and contribute to the genetic diversity in the HLA region as a whole. More generally, these results are new evidences supporting the hypothesis that the ancient epidemics, such as leprosy, influenced the frequency of alleles of the genes associated with inflammations in the current populations [61-63]. Looking ahead, the reconstruction of human adaptation to the major epidemics in the past will be of help for understanding genetic factors involved in our current predisposition to infectious diseases.

Cholera is an acute disease caused by the Gram-negative bacteria *Vibrio cholerae*, which colonize small intestine and produce cholera toxin [64]. Before the XIX century, cholera was endemic in Asia, but in 1817, globalization caused the disease spread from India to other regions and the first pandemic. In total, during the XIX-XX centuries there were six waves of cholera epidemic, and the seventh wave began at the end of XX century and continues up to now; cholera cases are recorded in the endemic regions of the South-Eastern Asia, Africa, and Latin America [65]. All waves of epidemic are caused by contaminated water.

By now, only few studies on genomes of cholera vibrios from ancient specimens have been performed. In particular, in 2014 a group of researchers used NGS methods for reconstructing the genome of *V. cholerae* from the preserved stomach of a victim of the second cholera pandemic (the outbreak in Philadelphia in 1849). This strain was shown to be 95-97% homologous to the reference strain of the classic cholera O395 biovar, and was different from it in 203 single nucleotide substitutions, absence of three genomic loci, and presence of the tandem repeats of the prophage containing the cholera toxin gene, potentially affecting virulence [66]. It is difficult to detect *Vibrio cholerae* in archaeological specimens, because DNA of this vibrio seems to be absent in the skeletal remains and tooth enamel, which are the best sources for extracting DNA. However, investigation of the “cholera” cemetery remains of the fifth wave of the cholera epidemic in Argentina, revealed *Vibrio cholerae* in one sample of the sediments of the sacral openings using PCR and Sanger sequencing [67]. Until now, this is the only case of detecting aDNA of *V. cholerae* that presents new possibilities for the detecting pathogens in archaeological specimens, when sediments are included in the analysis.

Tuberculosis. Tuberculosis caused by the bacterium *M. tuberculosis* has been affecting people for thousands of years and is the most lethal of all bacterial infections [68]. Although the genome of Koch’s bacillus has been actively studied, there is still no consensus on the disease origin and time of its emergence [69]. Presence of the pathogen in ancient remains is usually detected by PCR with primers for peroxidase, catalase (*katG*), and phospholipase C (*mpt40*) genes, which are found exclusively in *M. tuberculosis* complex bacteria (although *mpt40* is absent in *Mycobacterium bovis*, *Mycobacterium caprae*, and possibly in some other tuberculosis strains), and thus are used as specific markers for the presence of *M. tuberculosis* DNA [70].

Based on the molecular studies of the modern genomes of *M. tuberculosis* bacterium, it was proposed that the most recent common ancestor of *M. tuberculosis* accompanied humans during the migrations from Africa for about 70,000 years. At the same time, the studies using the ancient genomes as calibration points gave the much younger age of the common ancestor, less than 6000 years [70, 71]. For example, the results of Bayesian phylogenetic analysis of the recently reconstructed *M. tuberculosis* genome from the XVII century remains of a Swedish bishop Peder Winstrup [for today this is the bacterial genome of the best quality with the highest (141-fold coverage)], of the five earlier published ancient genomes, and of the 255 modern genomes of this causative agent imply emergence of the *M. tuberculosis* complex in the Neolithic period [72].

M. tuberculosis strains from a thousand-years-old human remains found in the coastal region of Peru occurred to be most closely related to the strains of

Mycobacterium pinnipedii affecting seals and sea lions [70]. The authors supposed that the pinnipeds could transmit mycobacterium from an unknown host species from the African coast to the coast of Peru, where the hunting for meat and fur resulted in transmission of mycobacterium to humans [73]. There are also paleopathological evidences of the tuberculosis-caused damages in human remains from the Pacific Ocean islands, which could appear even before contacts with Europeans [74].

The recent discovery that the TYK2 P1104A polymorphism homozygosity result in a higher risk of development of clinical forms of tuberculosis became the first proof of a monogenic predisposition to tuberculosis that allows us to study the combined evolution of human and lethal pathogen *M. tuberculosis* [75]. Using Bayesian analysis methods on a large dataset (1013 ancient human genomes belonging to seven historic epochs), it was shown that the current P1104A variant descended from a common ancestor of the Western Eurasia inhabitants about 30,000 years ago [76]. Moreover, a noticeable fluctuation of the P1104A frequency was found during the last 10,000 years of the European history after the large-scale migrations of Anatolian Neolithic farmers and Eurasian steppe herders to Europe. The analysis also revealed a sharp decrease in the polymorphism frequency after the Bronze Age that could be due to a strong negative selection started about 2000 years ago, with a relative 20% decrease in the adaptation of homozygotes; this indicator is one of the highest in the human genome.

Thus, the sources of emergence and spread of tuberculosis in human populations are still not fully understood, therefore, further studies, in particular using paleogenetic methods, would be very relevant.

Syphilis is an ancient infectious disease caused by the spirochete *Treponema pallidum*. An epidemic of this disease broke out in Europe at the end of the XV century, and since then outbreaks of syphilis occur in different regions of the world (every year up to millions of cases), despite the discovery of antibiotics. The origin of syphilis remains controversial, there are two main hypotheses: “Columbian” (American) and “pre-Columbian” (European). There is also a hypothesis, which suggests a common pre-Columbian origin of syphilis and other pathologies caused by the *Treponema* genus bacteria [77].

According to the American hypothesis, syphilis emerged in the New World and was brought to Europe by the first American expedition, because the outbreak of this disease coincided with the return of Columbus. Rapid spread of syphilis among Europeans was associated with their lack of immunity to the new disease.

According to the so-called “pre-Columbian” hypothesis, syphilis existed in the Old-World population before the contact with America and could spread in Eurasia and Africa in the prehistoric times [78]. This hypothesis assumes that syphilis could not be detected in

the ancient European specimens because of the similarity of its traits with other pathologies, e.g., leprosy [79]; otherwise, syphilis could become more virulent and widespread by the XV century, in particular, due to social factors.

The third hypothesis of syphilis emergence is a version of the “pre-Columbian” hypothesis. It assumes that all treponema-caused diseases (syphilis, yaws, pint, bejel) are different manifestations of the same pathology depending on climate and other conditions. However, currently there are no reliable confirmations of this hypothesis, although possible recombinations between the different strains of treponema were suggested [80].

For the first time, the DNA of *Treponema pallidum* was detected by PCR in human bones in 1999 [81], but further attempts were unsuccessful [82]. For a long time, the researchers considered treponema an unsuitable object for paleogenetic studies: the number of spirochetes in the body is small, and it is difficult to detect them by genetic method even during the lifetime of patients [83]. Phylogenetic studies performed on the specimens from living patients and on the samples collected in 1912 and cultured under laboratory conditions made it possible to reconstruct the *T. pallidum* evolution and detect the dominant cluster, variety of which was formed in the XX century after the discovery of antibiotics [84].

The possibility of full-genome studies became a breakthrough in the study of syphilis, because the short sections amplified by PCR could not provide sufficient information about a pathogen with slow evolution [77]. The first successful studies based on the NGS methods were on treponemas from bones of primates. Then the works began to appear associated with the extracting of the treponema DNA from archaeological specimens. In 2018, the first comparative analysis of the complete genomes of *T. pallidum* from ancient human skeletons of XVII-XIX centuries from colonial Mexico was performed, which revealed probable recombinations between the different treponema subspecies [85].

Analysis of the DNA sequences from 25 skeletal specimens from Rio de Janeiro revealed the presence of syphilis in Brazil during the colonial period [86]. Molecular methods confirmed the paleopathological hypothesis about the cases of syphilis in the 150-year-old remains of a 29-week-old fetus in France [87]. The recent study, which showed more complicated than expected geographical distribution of the early treponema-caused epidemics, was based on the reconstruction of four *T. pallidum* genomes from the medieval Europe (XIV-XVII centuries). The studied ancient genomes have shown genetic diversity of *T. pallidum* that existed during the Middle Ages in the Old World, in particular, syphilis-causing treponemas were found, as well as subspecies associated with other treponematoses. However, these results do not contradict the hypothesis about the introduction of new treponema strains from the New World by

participants of the European expeditions and potential changes in the genomes as a result of recombinations between them and European treponemas [80].

Plague. *Yersinia pestis* is a causative agent of plague, the disease which took the lives of about 60% of the Old-World population during three destructive pandemics: the first, or the Justinian plague (VI-VIII centuries), the second, or the Black Death (XIV-XVIII centuries), and the third, or the modern pandemic (XVIII-XIX centuries), which began in China [88-90]. Now plague is considered a problem of the past, although hundreds of cases of the disease are still recorded annually in African countries [91].

The last achievements in the field of genetic studies on *Yersinia pestis* and new data about the most known epidemics of plague in the human history are presented in the review by Kuznetsova et al. [92]. From the viewpoint of replenishing historical facts and studying evolution of the plague bacillus, the most important were the findings of *Y. pestis* in the remains of the Bronze Age, i.e., much earlier than the known pandemics. Moreover, accumulation of the genomic data allowed to establish possible causes for terminations of the historically known pandemics.

VIRAL INFECTIONS

As in the case of bacteria, the hypotheses about the evolution of viruses were proposed based on the data of the modern strain diversity. Accumulation of the genomic data from archaeological specimens made it possible to conduct combined analysis of the ancient and modern genomes. Thus, based on the evolutionary analysis of hepatitis B virus strains from the Neolithic epoch, it was suggested that this virus existed at least for 20,000 years [93, 94]. In addition to the studies on the virus strain diversity over time, another urgent area of study on the history of viral pandemics is analysis of medical protection of humans against diseases, including vaccination. The fight against the smallpox virus is one of such examples.

Smallpox. Smallpox is a highly contagious viral disease caused by the variola virus (VARV) belonging to the family Poxviridae, genus *Orthopoxvirus*. The genus *Orthopoxvirus* includes many immunologically related poxviruses, which strongly differ in their ability to infect different hosts. The variola virus VARV is unique among other *Orthopoxvirus* species, because it is an exclusively human pathogen.

The pathogenic VARV virus exists in two varieties: *Variola major*, which causes the more severe clinical form of the disease with the lethality of 20-40%, and up to 90% in some epidemics, and *Variola minor*, which is characterized by low lethality (1-3%) [95]. For many centuries, smallpox, or as it was called in ancient times “Black pox”,

caused several large-scale epidemics, in which the mortality rate reached 70% [96]. So, in the XX century smallpox killed more than 300 million people [97].

At present, smallpox is the only human disease, which has been eradicated by 1980 as a result of vaccination campaign, and this achievement remains one of the greatest triumphs of the modern medical science [95]. The smallpox genome is a linear double-stranded DNA containing 186-187 kbp with terminal inverted repeats (TIR) on the both ends of the sequence. Localization of the genes with different functions in the smallpox virus genome is not casual: the ~100 kbp central region of the genome contains the genes, which are common for other poxviruses and encode ~100 proteins involved, first of all, in the viral genome replication and expression and also in morphogenesis of the virions [97, 98]. Auxiliary genes encoding proteins involved in various aspects of the virus interaction with humans are grouped mainly in the peripheral regions of the genome adjacent to the TIR sequences. These viral proteins are responsible for resistance to antiviral host responses, including innate immunity and programmed cell death, and play major role in the virus evading immunity, as well as the virus pathogenicity and virulence [99, 100] (Fig. 1). In 2006, a large database was compiled with complete

genomic sequences of the modern variola virus VARV from some geographically separated isolates collected throughout the world [101]. The genomes of all sequenced modern strains of the VARV virus occurred to be very similar, which is explained by slow evolutionary changes in the viral DNA [102].

Comparative analysis of the VARV virus genomes and the genomes of other poxviruses [e.g., the cowpox virus (CPXV) whose ancestral CPXV-like strains have evolved into all modern orthopoxviruses] revealed that the smallpox virus strains, including the most virulent ones, lost many auxiliary genes after their divergence from the common ancestor, and this correlates with their narrow host range (exclusively human). These viral genes encode proteins involved in various intracellular and extracellular signaling pathways, such as proteins of the kelch-domain (1-4 such genes), poxvirus ankyrin proteins (ANK 5-8 such genes), proteins of the Bcl-2-domain, etc. [103] (Fig. 1).

Although the variola virus is actively studied by the methods of molecular genetics, up to now the time frame of emergence and spread of the VARV virus in human populations remain unknown. During the last decades, there were several attempts to reconstruct DNA of the smallpox virus from archaeological specimens: genome

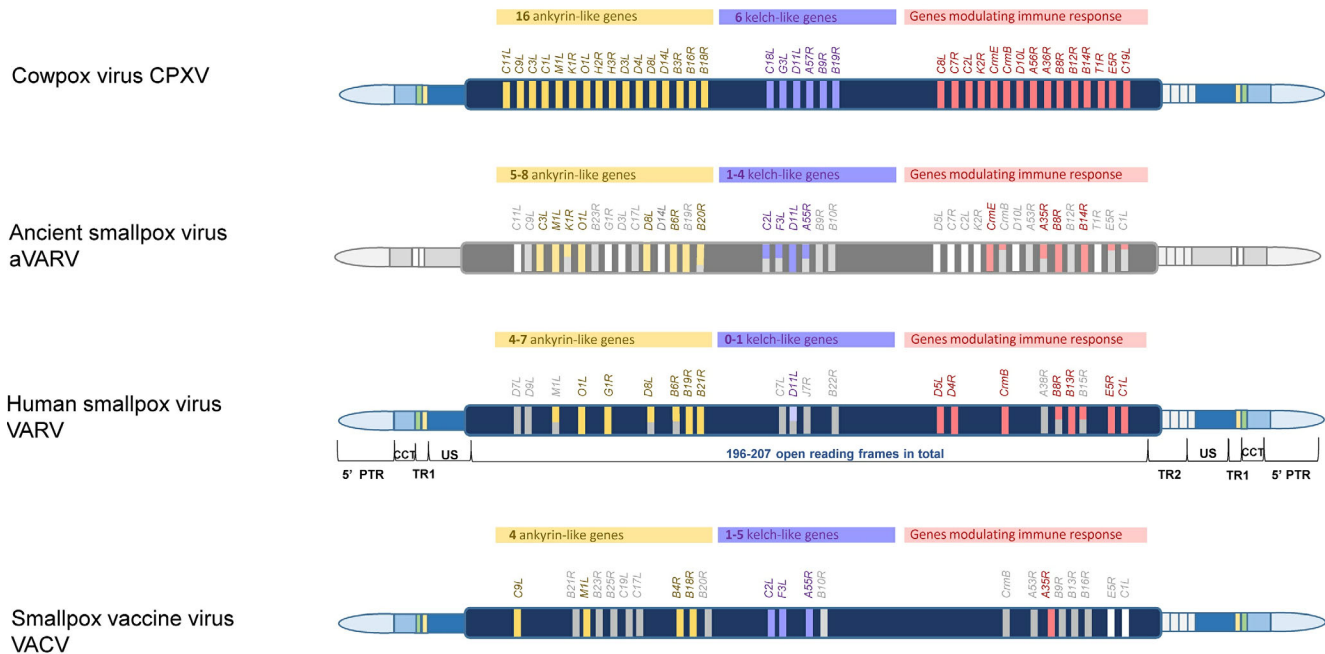


Fig. 1. Schematic representation of the smallpox virus genome structure. Abbreviations: PTR, palindrome terminal region; CCT, concatemer; TR, tandem repeats; US, unique sequence (repeat-free). The structures of genomes of the cowpox virus (CPXV) [104], the ancient smallpox virus (aVARV) from archaeological remains found in Northern Europe and dating back to the Viking Age (600-1050 AD) [105], modern variola virus (VARV) [103], smallpox vaccinia virus (VACV) are shown. The diagram displays the auxiliary viral genes that distinguish the presented genomes. The genes are sorted from left to right into categories depending on function (color-coded), and then by their position in the genome, relative to each other. Yellow, lilac, and red colors indicate genes that are present and functioning (presumably functioning, in the case of aVARV); grey color indicates genes that are absent or inactivated (in the case of aVARV, these are absent or presumably inactive genes); the partially colored genes indicate that they are functioning/inactivated in some aVARV/VARV/VACV; white color indicates genes for which there is no information in the case of aVARV [105, 106].

sequences of the ancient VARV viruses were reconstructed from the Siberian mummies of the XVII-XVIII centuries, from the mummy of a Lithuanian child of the XVII century, and from the Czech museum specimens of the XIX century [107-109]. The ancient DNA of the pathogen was successfully extracted from the soft tissues of the mummy, hybridization strategy was used for enrichment of the genome libraries followed by the massive parallel sequencing on the Illumina platform, and then it was compared with the reference genome of the modern variola virus VARV. During analysis of the ancient virus from the Lithuanian child of the XVII century, the authors also attempted to assemble *de novo* the genome of the ancient VARV strain. The final *de novo* genome was of 187,565 bp in the length, and the *de novo* consensus sequence had the identity of 97.5% of 185,578 bp with the reference sequence of the modern strain of VARV. The genome of the variola virus from the XVIII century museum specimen from England comprising the paraffin-embedded tissues of a child affected with smallpox was also reconstructed. In this case, the isolated DNA was used for creating libraries and subsequent full-genome sequencing on the Illumina platform without preliminary enrichment of the genomic libraries. The phylogenetic analysis of the ancient genome with all available modern and historic VARV genomes has revealed that the virus of the XVIII century belongs to the independent line of ancient strains, different from the virus genomes of the XVII century and that its evolutionary position is between the historical specimens of XVII century viruses from Lithuania and modern European strains of the XX century [110]. Based on the collected data it was concluded earlier that the common ancestor of all strains of the smallpox virus existed between the XIV and XVII centuries. However, the historic sources mentioning smallpox, such as written evidences about possible smallpox infections in India are dated back to 1500 years BC, and the Egyptian mummies with skin smallpox-like lesions are dated back to even earlier (3570 BC) [111].

Results of the recent study of 13 full-genome vaccinia virus sequences isolated from the archaeological remains found in Northern Europe dating back to the Viking era (600-1050 AD) pushed back the earliest case of smallpox by a millennium and made it possible to speak about the presence of this virus in Europe already at the end of the VI and beginning of the VII century [105]. The authors compared the genomic sequences of ancient viruses and modern VARV strains, and found differences in the number of functional genes belonging to the group of auxiliary poxvirus genes responsible for resisting antiviral responses. In the sequences of some genes mutations (namely: insertions, deletions, or point substitutions resulting to appearance of stop-codons or to the reading frame shift) were detected, which resulted in the gene inactivation. It was found out that at least three genes

(*CrmB*, *C1L*, and *E5R*) active in all samples of the modern VARV viruses, were inactivated in the ancient strains of smallpox viruses existed more than 1000 years ago.

However, fourteen genes, inactivated or absent in the modern strains of the VARV virus, seemed to be active in some or in all specimens of the ancient virus. Among them eight genes were found, which encoded different virulence factors, such as *C2L*, *F3L*, and *A55R*, for example, the gene *IL1B* encoding the cytokine IL-1- β of the interleukin 1 (IL-1) family involved in the regulation of immune reactions and inflammation [105] (Fig. 1). As it has been mentioned above, the loss of such genes leads to the changes in virulence and limitation of the host range of poxvirus [112]. The detected ancient smallpox variants of the Viking era were found to be a part of the previously unknown and now extinct viral clade, which evolved independently of the modern VARV. The findings have confirmed the theory of reductive evolution of poxviruses stating that the decrease in the number of genes of the supposed ancestral virus could play the decisive role in the species formation resulting in the limited ecological niche (one "host") of any newly emerged virus [105, 113].

Analysis of the genomes of ancient strains of smallpox virus indicated that there was high genetic diversity of the smallpox VARV viruses in Europe before the development of anti-smallpox vaccine. The selection pressure from the increasing vaccination level probably led to the disappearance of several ancient lines of the VARV virus, and the currently available smallpox VARV genomes represent only a fraction of the genetic diversity of VARV in the past [108, 110].

Another direction in the study of ancient strains of the smallpox virus is the history and evolutionary relationships of different strains of anti-smallpox vaccine, which still remains unclear, because there are no reliable data on the origin and diversity of viruses used in the early vaccination programs. In the end of the XVIII century, the discoverer of anti-smallpox vaccine, Edward Jenner, showed that humans could be vaccinated against the smallpox infection using material of the CPXV that was much safer than the existent practice of "variolation" with the smallpox virus VARV [114]. Unlike the smallpox virus VARV, the vaccinia virus CPXV possesses the greatest genetic diversity and wide range of hosts, and has the entire set of genes, which are present in all other orthopoxviruses. It is assumed that the gene loss has played an important role in the *Orthopoxvirus* evolution and that the ancestral CPXV-like strains evolved into all modern orthopoxviruses as a result of "reductive evolution" [113]. However, the vaccinia virus (VACV), which was really used for eradicating smallpox, and which is contained in all modern smallpox vaccines, is different from any known modern strain of CPXV. The biological origin of VACV is unknown, although it was suggested that the virus similar to the equine poxvirus could be its ancestor, although the modern equine poxvirus (HPXV)

genome carries many additional genes [115]. This hypothesis was supported by the Jenner's report [114] that later he received his inoculation from the infected horses.

Genomic analysis of the vaccine-associated virus in historical specimens provided data on the historical VACV vaccine strains of smallpox virus circulating in Philadelphia in the XIX century during the active vaccination of American military personnel [116]. The authors investigated origin, diversity, and distribution of the early strains of anti-smallpox vaccine using whole-genome analysis of both the virion and the metagenome, isolated from the historical museum "vaccination kits". The performed phylogenetic analysis allowed the authors to assign the isolated historical vaccine strains to the *Orthopoxvirus* clade (OPXV). It was found that the closest to the historical vaccine strains of the virus were both the modern North American vaccine strains VACV and different strains of the vaccinia virus circulating in the natural environment, including the South American cattle (Brazil) and buffalos from the South-Eastern Asia (India), as well as the virus *Cantagalo* (CTGV), which is a separate strain of the vaccinia virus [116]. The whole-genome analysis was performed earlier of the historical strain of the vaccinia virus isolated from the glycerol-treated specimen containing the anti-smallpox vaccine "Mulford" of 1902 from Europe. The analysis revealed that the genome of this strain of the vaccinia virus VACV showed a 99.7% identity with the horse vaccinia virus HPXV, which confirmed the suggested role of the horse vaccinia virus as the origin of the historical anti-smallpox vaccine. The authors found also unique deletions in "Mulford" at both ends of the sequence of 10.7 kbp to the left and 5.5 kbp to the right of the adjacent TIR-repeats region, which are present in the modern VACV strains, but absent in the genome of the horse vaccinia virus HPXV and the vaccinia virus CPXV [117]. Thus, the studies on the historic strains of the anti-smallpox vaccine led to the conclusion that the European populations in the XIX century were immunized with the material from both the cow and horse smallpox. Possibly, the ancestral smallpox vaccine virus (aVACV) existed also, which could circulate in the populations of domestic animals (horses and cows) in Europe in the XIX century, but then disappeared and now did not have natural hosts, except occasional cases of infecting animals in some countries, such as Brazil and India.

At present, all existent strains of the anti-smallpox vaccine are divided into three main phylogenetic clusters based on the full-genome analysis of the modern and historical vaccine strains: the Eurasian cluster including the British strain Lister, the Chinese strain Tian-tan, and the Russian strain Tashkent; the South American cluster including Brazil field strains of the anti-smallpox vaccine presented by the *Cantagalo* virus; and the American, or Dryvax-cluster, which combines all viral clones isolated from the American vaccine Dryvax [118]. The authors

have shown that the central part of the VACV genome is highly conserved in all vaccine strains, whereas the major genetic diversity (deletions and inversions) is localized in the region of telomeres. It is shown that all modern strains of the smallpox vaccine VACV are characterized by three genetic features: two deletions of 10.7 and 5.5 kbp are located in the region of TIR repeats at the ends of the viral DNA, as well as mutation or deletion of the gene *DVX_213*, encoding the ankyric protein of the F-domain, homolog of the gene *B18R* in the smallpox virus (Bangladesh), loss of which could affect the VACV virus virulence. It should be noted that the modern strains of the smallpox vaccine virus VACV are not always grouped to simple phylogenetic trees corresponding to the known historic relations between these strains. It is more likely, that all existent strains of the smallpox vaccine VACV originated from a complex set of different vaccine viruses (virus pool), precursors of the modern VACV, which during the long period of vaccination in the XIX-XX centuries were many times passed, distributed, and randomly chosen for creating strains of the present-day vaccines (Fig. 2).

Thus, it may be concluded that reconstruction of the genomes of the ancient smallpox viruses using the large-scale genomic sequencing technologies significantly broadened our knowledge on the smallpox virus evolution. The new findings about interrelations between the human smallpox virus and other orthopoxviruses, including the extinct ancient strains of human and animal viruses, changed our notions about the evolutionary interrelationships among orthopoxviruses (OPV) and improved our understanding of the evolutionary history of smallpox. However, questions concerning the origin of smallpox virus, e.g., how VARV has evolved into a human-specific pathogen, remain to be answered by further researches.

NEAR-FUTURE PERSPECTIVES

The long-term co-evolution of human and causative agents of infectious diseases has led to changes in the genetic structure of human populations under the influence of infectious agents. Such changes were often accompanied by the selection of genetic variants in the loci associated with resistance to specific infectious agents, which was noted already in the first works devoted to the simultaneous analysis of 50 and more samples of aDNA [119, 120]. It is expected that the studies of previously unstudied ancient specimens will reveal similar evolutionary changes in the human populations in response to the action of infectious agents.

Especially interesting are the same-time collective burials, because in some cases they are possibly associated with the outbreaks of infectious diseases. Metagenomic analysis of DNA from such specimens

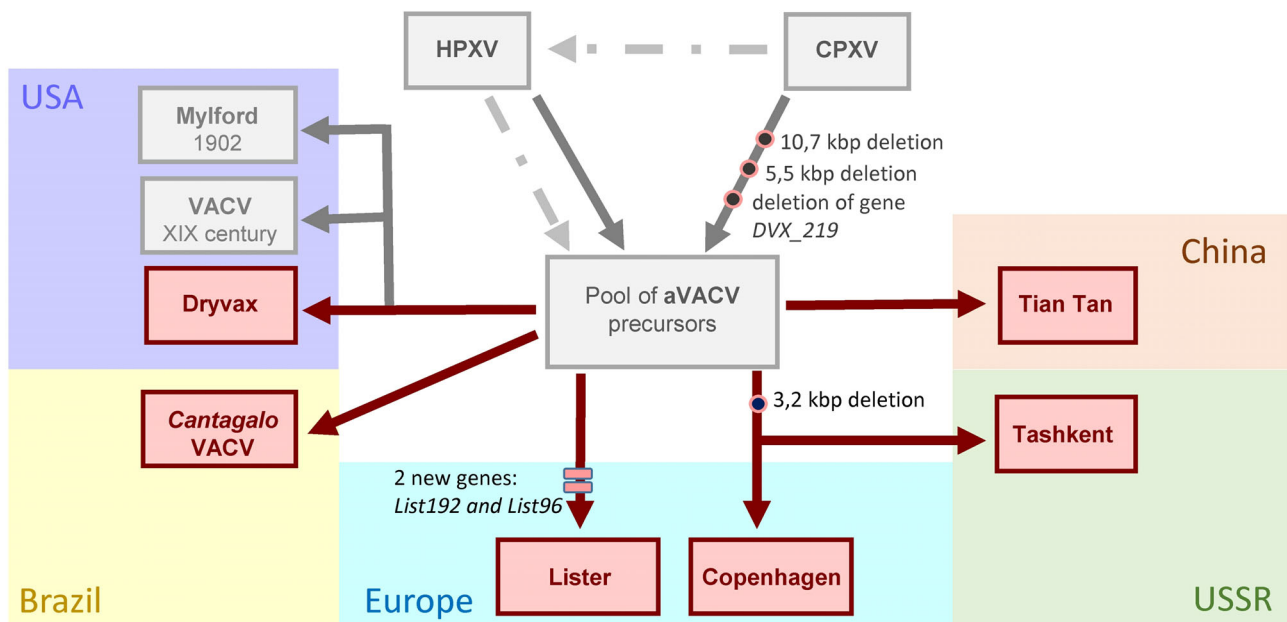


Fig. 2. Schematic model of the smallpox vaccine virus VACV evolution compiled on the basis of the genomes of modern smallpox vaccine VACV strains and historical smallpox vaccine strains obtained from various sources. Evolution processes are shown, which facilitated diversity of the strains of smallpox vaccine VACV viruses. The modern vaccine strains are shown in red color, the historic vaccine strains are shown in gray; aVACV presents the pool from different historic precursors of all present-day smallpox vaccine strains; HPXV is the horse smallpox virus, which is a hypothetical precursor of the historic vaccine virus; CPXV is the cowpox virus, which is a hypothetical precursor of the historical smallpox vaccine virus; the dotted line shows one of the probable evolutionary processes of the origin of horse smallpox virus HPXV from the CPXV virus.

could potentially reveal a pathogen, which caused the death.

Bioinformatics approaches are also being developed in parallel with the increase in the technical capabilities of sequencing. Majority of the analytical methods are based on comparison with the known reference sequences of modern prokaryotes, of which almost 200,000 have been obtained by now [121]. Nevertheless, even such large numbers of reference sequences can be analyzed within a few hours and even minutes [122, 123]. However, it should be considered that the divergence of genomic sequences of ancient specimens could be greater than that of the modern samples and that it could be additionally aggravated by chemical modifications of DNA. Nevertheless, these methods have been successfully applied also for analysis of ancient metagenomes [124]. The *de novo* assembly methods [125, 126] can be used for both modern and ancient metagenomes without using the reference databases, and this makes it possible to obtain more accurate sequence of the genome. Moreover, new studies have been conducted regularly: at present, over a thousand of ancient metagenomes have been published [127]. This would allow to perform a broader systematic analysis and not only spot studies of individual specimens in future.

The main technical difficulties in paleogenetic studies to be solved in the near future are worth mentioning.

As a rule, archaeological specimens are presented by bones and teeth, however, many known pathogens are present in other tissues, which are not preserved over time. For today, there are successful works for isolating DNA of ancient pathogens from mummified tissues and historical specimens, but there are only few such cases. Another obstacle for obtaining the complete genome of an ancient pathogen is the small amount of the pathogen in the archaeological material, which could be mitigated by the increasing the number of “reads” during sequencing. Moreover, an important problem for the correct interpretation of sequencing results is the similarity between the pathogenic forms of bacteria and non-pathogenic species, which are abundant in the soil. Analysis of soil and sediments from collective burials can be an approach to solve this problem. The genetic studies of ancient sediments published within last two years have demonstrated their great potential. Firstly, the samples from the sediments could contain DNA identical to the DNA of skeletal remains [128], as well as DNA of pathogens from the internal organs [67]. Secondly, this creates the possibility to analyze DNA of the environment of the period under study [129].

RNA molecules degrade even faster than DNA molecules, and this limits the study of the evolution of pathogens with RNA genomes (influenza virus, measles virus, HIV, and yellow fever virus). Nevertheless, the

recent studies have shown that under favorable conditions RNA can be preserved for millennia [130, 131]. Sequencing the genomes of the ancient RNA viruses of measles and cattle plague [132] is an important result that would allow us to better understand the history of many other RNA-based pathogens. It is also interesting that formalin fixation, which is detrimental for DNA, seems to have no such effect on RNA, but, on the contrary, preserves it, as it has been shown in the case of successful reconstruction of HIV from the archival tissues [133].

CONCLUSIONS

Thus, ancient DNA is a valuable material for improvement our understanding of the time of emergence of many pathogens and pathways of their spread and evolution. We can see that paleoepidemiology is developing rapidly thanks to molecular genetics methods, and the results of new works often disprove old hypotheses or find confirmation of scenarios that were considered unlikely based on the classical approaches. Our current successes were unthinkable a few decades ago, and it may be assumed that the further scientific progress will provide even more information about the origin and diversity of ancient pathogens.

The COVID-19 pandemic has forced scientists of the world to think about origins and causes of new dangerous infectious diseases, as well as about pathways for the possible interspecies transmission of viral pathogens, in particular, of transmission of zoonotic virus to humans. From this viewpoint, a promising direction in the study of the ancient pathogen genomes is the search for and isolation of DNA and RNA of paleoviruses and bacteria preserved in the soft tissues of paleolithic animals found in the Arctic in Russia, as well as in the soil samples from permafrost regions. Studying of these genomes would help to better understand evolution of the virus formation, investigate molecular and evolutionary mechanisms of animal adaptation to viral and bacterial pathogens, and also to assess impact of epidemics on the genetic structure of human and animal populations.

Funding. This work was supported by the Ministry of Education and Science of the Russian Federation, system number 075-10-2020-116 (grant no. 13.1902.21.0023).

Ethics declarations. Authors declare no conflicts of interests. This article does not contain description of studies with the involvement of humans or animal subjects.

REFERENCES

- Zink, A. R., Reischl, U., Wolf, H., and Nerlich, A. G. (2002) Molecular analysis of ancient microbial infections, *FEMS Microbiol. Lett.*, **213**, 141-147, doi: 10.1111/j.1574-6968.2002.tb11298.x.
- Buikstra, J. E., and Roberts, C. (2012) *The Global History of Paleopathology: Pioneers and Prospects*, Oxford Univ. Press.
- Arriaza, B. T., Salo, W., Aufderheide, A. C., and Holcomb, T. A. (1995) Pre-Columbian tuberculosis in northern Chile: Molecular and skeletal evidence, *Am. J. Phys. Anthropol.*, **98**, 37-45, doi: 10.1002/ajpa.1330980104.
- Drancourt, M., Aboudharam, G., Signoli, M., Dutour, O., and Raoult, D. (1998) Detection of 400-year-old *Yersinia pestis* DNA in human dental pulp: An approach to the diagnosis of ancient septicemia, *Proc. Natl. Acad. Sci. USA*, **95**, 12637-12640, doi: 10.1073/pnas.95.21.12637.
- Pääbo S. (1989) Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification, *Proc. Natl. Acad. Sci. USA*, **86**, 1939-1943, doi: 10.1073/pnas.86.6.1939.
- Briggs, A. W., Stenzel, U., Johnson, P. L., Green, R. E., Kelso, J., et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal, *Proc. Natl. Acad. Sci. USA*, **104**, 14616-14621, doi: 10.1073/pnas.0704665104.
- Warinner, C., Herbig, A., Mann, A., Fellows Yates, J. A., Weiß, C. L., et al. (2017) A Robust framework for microbial archaeology, *Annu. Rev. Genomics Hum. Genet.*, **18**, 321-356, doi: 10.1146/annurev-genom-091416-035526.
- Latham, K. E., and Miller, J. J. (2018) DNA recovery and analysis from skeletal material in modern forensic contexts, *Forensic Sci. Res.*, **4**, 51-59, doi: 10.1080/20961790.2018.1515594.
- Müller, R., Roberts, C. A., and Brown, T. A. (2016) Complications in the study of ancient tuberculosis: presence of environmental bacteria in human archaeological remains, *J. Archaeol. Sci.*, **68**, 5-11, doi: 10.1016/j.jas.2016.03.002.
- Gilbert, M., Cuccui, J., White, W., Lynnerup, N., Titball, R. W., et al. (2004) Absence of *Yersinia pestis*-specific DNA in human teeth from five European excavations of putative plague victims, *Microbiology (Reading)*, **150**, 341-354, doi: 10.1099/mic.0.26594-0.
- Gilbert, J. A., Jansson, J. K., and Knight, R. (2014) The Earth microbiome project: successes and aspirations. *BMC Biol.*, **12**, 69, doi: 10.1186/s12915-014-0069-1.
- Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome, *Nature*, **486**, 207-214, doi: 10.1038/nature11234.
- Murray, D. C., Haile, J., Dortch, J., White, N. E., Haouchar, D., et al. (2013) Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages, *Sci. Rep.*, **3**, 3371, doi: 10.1038/srep03371.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., et al. (2010) QIIME allows analysis of high-throughput community sequencing data, *Nat. Methods*, **7**, 335-336, doi: 10.1038/nmeth.f.303.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., et al. (2009) Introducing Mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities, *Appl. Environ. Microbiol.*, **75**, 7537-7541, doi: 10.1128/AEM.01541-09.
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., et al. (2012) Metagenomic microbial commu-

- nity profiling using unique clade-specific marker genes, *Nat. Methods*, **9**, 811-814, doi: 10.1038/nmeth.2066.
17. Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., et al. (2015) MetaPhlan2 for enhanced metagenomic taxonomic profiling, *Nat. Methods*, **12**, 902-903, doi: 10.1038/nmeth.3589.
 18. Nayfach, S., Rodriguez-Mueller, B., Garud, N., and Pollard, K. S. (2016) An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography, *Genome Res.*, **26**, 1612-1625, doi: 10.1101/gr.201863.115.
 19. Darling, A. E., Jospin, G., Lowe, E., Matsen, F. A., Bik, H. M., et al. (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes, *PeerJ*, **2**, e243, doi: 10.7717/peerj.243.
 20. Wood, D. E., and Salzberg, S. L. (2014) Kraken: Ultrafast metagenomic sequence classification using exact alignments, *Genome Biol.*, **15**, R46, doi: 10.1186/gb-2014-15-3-r46.
 21. Ounit, R., Wanamaker, S., Close, T. J., and Lonardi, S. (2015) CLARK: Fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers, *BMC Genomics*, **16**, 236, doi: 10.1186/s12864-015-1419-2.
 22. Ounit, R., and Lonardi, S. (2016) Higher classification sensitivity of short metagenomic reads with CLARK-S, *Bioinformatics*, **32**, 3823-3825, doi: 10.1093/bioinformatics/btw542.
 23. Vågene, Å. J., Herbig, A., Campana, M. G., Robles García, N. M., Warinner, C., et al. (2018) Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico, *Nat. Ecol. Evol.*, **2**, 520-528, doi: 10.1038/s41559-017-0446-6.
 24. Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., et al. (2017) Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus, *Nature*, **544**, 357-361, doi: 10.1038/nature21674.
 25. DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, *Appl. Environ. Microbiol.*, **72**, 5069-5072, doi: 10.1128/AEM.03006-05.
 26. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., et al. (2016) DADA2: High-resolution sample inference from Illumina amplicon data, *Nat. Methods*, **13**, 581-583, doi: 10.1038/nmeth.3869.
 27. Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., et al. (2017) Deblur rapidly resolves single-nucleotide community sequence patterns, *mSystems*, **2**, e00191-16, doi: 10.1128/mSystems.00191-16.
 28. Velsko, I. M., Frantz, L. A. F., Herbig, A., Larson, G., and Warinner, C. (2018) Selection of appropriate metagenome taxonomic classifiers for ancient microbiome research, *mSystems*, **3**, e00080-18, doi: 10.1128/mSystems.00080-18.
 29. Ziesemer, K. A., Mann, A. E., Sankaranarayanan, K., Schroeder, H., Ozga, A. T., et al. (2015) Intrinsic challenges in ancient microbiome reconstruction using 16S rRNA gene amplification, *Sci. Rep.*, **5**, 16498, doi: 10.1038/srep16498.
 30. Yuan, C., Lei, J., Cole, J., and Sun, Y. (2015) Reconstructing 16S rRNA genes in metagenomic data, *Bioinformatics*, **31**, i35-i43, doi: 10.1093/bioinformatics/btv231.
 31. Hübner, R., Key, F. M., Warinner, C., Bos, K. I., Krause, J., et al. (2019) HOPS: Automated detection and authentication of pathogen DNA in archaeological remains, *Genome Biol.*, **20**, 280, doi: 10.1186/s13059-019-1903-0.
 32. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., and Orlando, L. (2013) mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters, *Bioinformatics*, **29**, 1682-1684, doi: 10.1093/bioinformatics/btt193.
 33. Schuenemann, V. J., Singh, P., Mendum, T. A., Krause-Kyora, B., Jäger, G., et al. (2013) Genome-wide comparison of medieval and modern *Mycobacterium leprae*, *Science*, **341**, 179-183, doi: 10.1126/science.1238286.
 34. Warinner, C., Rodrigues, J. F., Vyas, R., Trachsel, C., Shved, N., et al. (2014) Pathogens and host immunity in the ancient human oral cavity, *Nat. Genet.*, **46**, 336-344, doi: 10.1038/ng.2906.
 35. Wagner, D. M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., et al. (2014) *Yersinia pestis* and the plague of Justinian 541-543 AD: A genomic analysis, *Lancet Infect. Dis.*, **14**, 319-326, doi: 10.1016/S1473-3099(13)70323-2.
 36. Skinsnes O. K. (1973) Notes from the history of leprosy. I. Interpretive chronology of leprosy concept and practice, *Int. J. Lepr. Other Mycobact. Dis.*, **41**, 220-233.
 37. Ortner, D. J., and Putschar, W. G. J. (1985) *Identification of Pathological Conditions in Human Skeletal Remains*, Washington DC: Smithsonian Institution Press.
 38. Donoghue, H. D., Gladkowska-Rzeczycka, J., Marcsik, A., Holton, J., and Spigelman, M. (2002) *Mycobacterium leprae* in archaeological samples, in *The Past and Present of Leprosy. Archaeological, Historical, Palaeopathological and Clinical Approaches* (Roberts, Ch. A., Lewis, M., and Manchester, K., eds.) BAR International Series 1054, Bradford, pp. 271-286.
 39. Buzhilova, A. (2002) In *The Past and Present of Leprosy. Archaeological, Historical, Palaeopathological and Clinical Approaches* (Roberts, Ch. A., Lewis, M., and Manchester, K., eds.) BAR International Series 1054, Oxford, Archaeopress, pp. 123-133.
 40. Manchester, K. (1984) Tuberculosis and leprosy in antiquity: An interpretation, *Med. Hist.*, **28**, 162-173, doi: 10.1017/S0025727300035705.
 41. Lechat, M. F. (1999) The paleopathology of leprosy: An overview, *Int. J. Lepr. Other Mycobact. Dis.*, **67**, 460-470.
 42. Brothwell, D. R., Powers, R., and Hirst, S. M. (2000) In *Cannington Cemetery: Excavations 1962-3 of Prehistoric, Roman, post-Roman and Later Features at Cannington Park Quarry, Near Bridgewater, Somerset*, London: Britannia Monograph Series No. 17, Society for the Promotion of Roman Studies, pp. 195-256.
 43. Blondiaux, J., Dürr, J., Khouchaf, L., and Eisenberg, L. E. (2002) In *The Past and Present of Leprosy. Archaeological, Historical, Palaeopathological and Clinical Approaches* (Roberts, Ch. A., Lewis, M., and Manchester, K., eds.) BAR International Series 1054, Oxford, Archaeopress.
 44. Kozak, A. D. (2002) About existence of leprosy in the ancient Kiev, *OPUS: Interdiscipline Studies in Archaeology, IA Russian Academy of Sciences Publishers*, Moscow, Issues 1-2, [in Russian].
 45. Monot, M., Honoré, N., Garnier, T., Araoz, R., Coppée, J. Y., et al. (2005) On the origin of leprosy, *Science*, **308**, 1040-1042, doi: 10.1126/science/1109759.

46. Monot, M., Honoré, N., Garnier, T., Zidane, N., Sherafi, D., et al. (2009) Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*, *Nat. Genet.*, **41**, 1282-1289, doi: 10.1038/ng.477.
47. Schuenemann, V. J., Avanzi, C., Krause-Kyora, B., Seitz, A., Herbig, A., et al. (2018) Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe, *PLoS Pathog.*, **14**, 1-17, doi: 10.1371/journal.ppat.1006997.
48. Fotakis, A. K., Denham, S. D., Mackie, M., Orbegozo, M. I., Mylopotamitaki, D., et al. (2020) Multi-omic detection of *Mycobacterium leprae* in archaeological human dental calculus, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **375**, 1812, doi: 10.1098/rstb.2019.0584.
49. Mendum, T. A., Schuenemann, V. J., Roffey, S., Taylor, G. M., Wu, H., et al. (2014) *Mycobacterium leprae* genomes from a British medieval leprosy hospital: Towards understanding an ancient epidemic, *BMC Genomics*, **15**, 270, doi: 10.1186/1471-2164-15-270.
50. Neukamm, J., Pfrengele, S., Molak, M., Seitz, A., Francken, M., et al. (2020) 2000-year-old pathogen genomes reconstructed from metagenomic analysis of Egyptian mummified individuals, *BMC Biol.*, **18**, 108, doi: 10.1186/s12915-020-00839-8.
51. Robbins, G., Tripathy, V. M., Misra, V. N., Mohanty, R. K., Shinde, V. S., et al. (2009) Ancient skeletal evidence for leprosy in India (2000 B.C.), *PLoS One*, **4**, e5669, doi: 10.1371/journal.pone.0005669.
52. Blevins, K. E., Crane, A. E., Lum, C., Furuta, K., Fox, K., et al. (2020) Evolutionary history of *Mycobacterium leprae* in the Pacific Islands, *Philos. Trans. R. Soc. B*, **378**, 20190582, doi: 10.1098/rstb.2019.0582.
53. Stone, A. C., Lewis, C. M. Jr., and Schuenemann, V. J. (2020) Insights into health and disease from ancient biomolecules, *Phil. Trans. R. Soc. B*, **375**, doi: 10.1098/rstb.2019.0568.
54. Krause-Kyora, B., Nutsua, M., Boehme, L., Pierini, F., Pedersen, D. D., et al. (2018) Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans, *Nat. Commun.*, **9**, 1569, doi: 10.1038/s41467-018-03857-x.
55. Schipper, R. F., Schreuder, G. M., D'Amaro, J., and Oudshoorn, M. (1996) HLA gene and haplotype frequencies in Dutch blood donors, *Tissue Antigens*, **48**, 562-574, doi: 10.1111/j.1399-0039.1996.tb02670.x.
56. Müller, C. R., Ehninger, G., and Goldmann, S. F. (2003) Gene and haplotype frequencies for the loci HLA-A, HLA-B, and HLA-DR based on over 13,000 German blood donors, *Hum. Immunol.*, **64**, 137-151, doi: 10.1016/s0198-8859(02)00706-1.
57. International Multiple Sclerosis Genetics Consortium, Hafler, D. A., Compston, A., Sawcer, S., Lander, E. S., et al. (2007) Risk alleles for multiple sclerosis identified by a genomewide study, *N. Engl. J. Med.*, **357**, 851-862, doi: 10.1056/NEJMoa073493.
58. Gregersen, J. W., Kranc, K. R., Ke, X., Svendsen, P., Madsen, L. S., et al. (2006) Functional epistasis on a common MHC haplotype associated with multiple sclerosis, *Nature*, **443**, 574-577, doi: 10.1038/nature05133.
59. Fischer, A., Grunewald, J., Spagnolo, P., Nebel, A., Schreiber, S., et al. (2014) Genetics of sarcoidosis, *Semin. Respir. Crit. Care Med.*, **35**, 296-306, doi: 10.1055/s-0034-1376860.
60. Hu, X., Deutsch, A. J., Lenz, T. L., Onengut-Gumuscu, S., Han, B., et al. (2015) Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk, *Nat. Genet.*, **47**, 898-905, doi: 10.1038/ng.3353.
61. Karlsson, E. K., Kwiatkowski, D. P., and Sabeti, P. C. (2014) Natural selection and infectious disease in human populations, *Nat. Rev. Genet.*, **15**, 379-393, doi: 10.1038/nrg3734.
62. Domínguez-Andrés, J., Kuijpers, Y., Bakker, O. B., Jaeger, M., Xu, C. J., et al. (2021) Evolution of cytokine production capacity in ancient and modern European populations, *Elife*, **10**, e64971, doi: 10.7554/eLife.64971.
63. Kerner, G., Patin, E., and Quintana-Murci, L. (2021) New insights into human immunity from ancient genomics, *Curr. Opin. Immunol.*, **72**, 116-125, doi: 10.1016/j.coi.2021.04.006.
64. Faruque, S. M., Albert, M. J., and Mekalanos, J. J. (1998) Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*, *Microbiol. Mol. Biol. Rev.*, **62**, 1301-1314, doi: 10.1128/MMBR.62.4.1301-1314.1998.
65. Piret, J., and Boivin, G. (2021) Pandemics throughout history, *Front. Microbiol.*, **11**, 631736, doi: 10.3389/fmicb.2020.631736.
66. Devault, A. M., Golding, G. B., Waglechner, N., Enk, J. M., Kuch, M., et al. (2014) Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849, *N. Engl. J. Med.*, **370**, 334-340, doi: 10.1056/NEJMoa1308663.
67. Ramirez, D. A., Saka, H. A., and Nores, R. (2021) Detection of *Vibrio cholerae* aDNA in human burials from the fifth cholera pandemic in Argentina (1886-1887 AD), *Int. J. Paleopathol.*, **32**, 74-79, doi: 10.1016/j.ijpp.2020.12.004.
68. Gagneux, S. (2018) Ecology and evolution of *Mycobacterium tuberculosis*, *Nat. Rev. Microbiol.*, **16**, 202-213, doi: 10.1038/nrmicro.2018.8.
69. Menardo, F., Duchene, S., Brites, D., and Gagneux, S. (2019) The molecular clock of *Mycobacterium tuberculosis*, *PLoS Pathol.*, **15**, e1008067, doi: 10.1371/journal.ppat.1008067.
70. Bos, K. I., Harkins, K. M., Herbig, A., Coscolla, M., Weber, N., et al. (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis, *Nature*, **514**, 494-497, doi: 10.1038/nature13591.
71. Kay, G. L., Sergeant, M. J., Zhou, Z., Chan, J. Z., Millard, A., et al. (2015) Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe, *Nat. Commun.*, **6**, 6717, doi: 10.1038/ncomms7717.
72. Sabin, S., Herbig, A., Vågene, Å. J., Ahlström, T., Bozovic, G., et al. (2020) A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex, *Genome Biol.*, **21**, 201, doi: 10.1186/s13059-020-02112-1.
73. Silva-Pereira, T. T., Ikuta, C. Y., Zimpel, C. K., Camargo, N., de Souza Filho, A. F., et al. (2019) Genome sequencing of *Mycobacterium pinnipedii* strains: genetic characterization and evidence of superinfection in a South American sea lion (*Otaria flavescens*), *BMC Genomics*, **20**, 1030, doi: 10.1186/s12864-019-6407-5.

74. McDonald, S. K., Matisoo-Smith, E. A., Buckley, H. R., Walter, R. K., Aung, H. L., et al. (2020) "TB or not TB": the conundrum of pre-European contact tuberculosis in the Pacific, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **375**, doi: 10.1098/rstb.2019.0583.
75. Abel, L., Fellay, J., Haas, D. W., Schurr, E., Srikrishna, G., et al. (2018) Genetics of human susceptibility to active and latent tuberculosis: Present knowledge and future perspectives, *Lancet Infect. Dis.*, **18**, e64-e75, doi: 10.1016/S1473-3099(17)30623-0.
76. Kerner, G., Laval, G., Patin, E., Boisson-Dupuis, S., Abel, L., et al. (2021). Human ancient DNA analyses reveal the high burden of tuberculosis in Europeans over the last 2,000 years, *Am. J. Hum. Genet.*, **108**, 517-524, doi: 10.1016/j.ajhg.2021.02.009.
77. Gogarten, J. F., D ux, A., Schuenemann, V. J., Nowak, K., Boesch, C., et al. (2016) Tools for opening new chapters in the book of *Treponema pallidum* evolutionary history, *Clin. Microbiol. Infect.*, **22**, 916-921, doi: 10.1016/j.cmi.2016.07.027.
78. Hackett, C. J. (1963) On the origin of the human treponematoses (pinta, yaws, endemic syphilis and venereal syphilis), *Bull. World Health Organ.*, **29**, 7-41.
79. Rothschild, B. M. (2005) History of syphilis, *Clin. Infect. Dis.*, **40**, 1454-1463, doi: 10.1086/429626.
80. Majander, K., Pfrengle, S., Kocher, A., Neukamm, J., du Plessis, L., et al. (2020) Ancient bacterial genomes reveal a high diversity of *Treponema pallidum* strains in early modern Europe, *Curr. Biol.*, **30**, 3788-3803, doi: 10.1016/j.cub.2020.07.058.
81. Kolman, C. J., Centurion-Lara, A., Lukehart, S. A., Owsley, D. W., and Tuross, N. (1999) Identification of *Treponema pallidum* subspecies pallidum in a 200-year-old skeletal specimen, *J. Infect. Dis.*, **180**, 2060-2063, doi: 10.1086/315151.
82. Von Hunnius, T. E., Yang, D., Eng, B., Wayne, J. S., and Saunders, S. R. (2007) Digging deeper into the limits of ancient DNA research on syphilis, *J. Archaeol. Sci.*, **34**, 2091-2100, doi: 10.1016/j.jas.2007.02.007.
83. Bouwman, A. S., and Brown, T. A. (2005) The limits of biomolecular palaeopathology: ancient DNA cannot be used to study venereal syphilis, *J. Archaeol. Sci.*, **32**, 703-713, doi: 10.1016/j.jas.2004.11.014.
84. Arora, N., Schuenemann, V. J., J ager, G., Peltzer, A., Seitz, A., et al. (2016) Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster, *Nat. Microbiol.*, **2**, 16245, doi: 10.1038/nmicrobiol.2016.245.
85. Schuenemann, V. J., Lankapalli, A. K., Barquera, R., Nelson, E. A., Ira ız Hern andez, D., et al. (2018) Historic *Treponema pallidum* genomes from Colonial Mexico retrieved from archaeological remains, *PLoS Negl. Trop. Dis.*, **12**, e0006447, doi: 10.1371/journal.pntd.0006447.
86. Guedes, L., Dias, O., Neto, J., Ribeiro da Silva, L. D. P., Mendonça de Souza, S. M. F., et al. (2018) First paleogenetic evidence of probable syphilis and treponematoses cases in the Brazilian Colonial Period, *Biomed Res. Int.*, **8304129**, doi: 10.1155/2018/8304129.
87. Meffray, A., Perrin, M., Richier, A., Schmitt, A., Ardagna, Y., et al. (2019) Molecular detection of *Treponema pallidum* subspecies pallidum in 150-year-old foetal remains, southeastern France, *J. Med. Microbiol.*, **68**, 761-769, doi: 10.1099/jmm.0.000978.
88. Stenseth, N. C., Atshabar, B. B., Begon, M., Belmain, S. R., Bertherat, E., et al. (2008) Plague: Past, present, and future, *PLoS Med.*, **5**, e3, doi: 10.1371/journal.pmed.0050003.
89. Scott, S., and Duncan, C. J. (2001) *Biology of Plagues: Evidence from Historical Populations*, Cambridge University Press, Cambridge, UK.
90. Frith, J. (2012) The history of plague – Part 1. The three great pandemics, *J. Mil. Veterans' Health*, **20**, 11-16.
91. Barbieri, R., Signoli, M., Chev e, D., Costedoat, C., Tzortzis, S., et al. (2020) *Yersinia pestis*: the natural history of plague, *Clin. Microbiol. Rev.*, **34**, e00044-19, doi: 10.1128/CMR.00044-19.
92. Kuznetsova, I. L., Andreeva, T. V., Malyarchuk, A. B., Kunizheva, S. S., Tyazhelova, T. V. et al. (2021) Genomics of ancient pathogens at the *Yersinia pestis* example in the historic context, *Ros. Arhaeologia* (in press).
93. Kocher, A., Papac, L., Barquera, R., Key, F. M., Spyrou, M. A., et al. (2021) Ten millennia of hepatitis B virus evolution, *Science*, **374**, 182-188, doi: 10.1126/science.abi5658.
94. M uhlemann, B., Jones, T.C., Damgaard, P. B., Allentoft, M. E., Shevnina, I., et al. (2018) Ancient hepatitis B viruses from the Bronze Age to the Medieval period, *Nature*, **557**, 418-423, doi: 10.1038/s41586-018-0097-z.
95. Fenner, F., Henderson, D. A., Arita, I., Je ek, Z., and Ladnyi, I. D. (1988) *Smallpox and Its Eradication*, World Health Organization, Geneva, Switzerland.
96. Hopkins, D. R. (2002) *The Greatest Killer: Smallpox in History*, University of Chicago Press, Chicago, IL.
97. Massung, R. F., Liu, L. I., Qi, J., Knight, J. C., Yuran, T. E., et al. (1994) Analysis of the complete genome of smallpox *Variola major* virus strain Bangladesh-1975, *Virology*, **201**, 215-240, doi: 10.1006/viro.1994.1288.
98. Massung, R. F., Loparev, V. N., Knight, J. C., Totmenin, A. V., Chizhikov, V. E., et al. (1996) Terminal region sequence variations in variola virus DNA, *Virology*, **221**, 291-300, doi: 10.1006/viro.1996.0378.
99. Shchelkunov, S. N., Massung, R. F., and Esposito, J. J. (1995) Comparison of the genome DNA sequences of Bangladesh-1975 and India-1967 variola viruses, *Virus Res.*, **36**, 107-118, doi: 10.1016/0168-1702(94)00113-q.
100. Shchelkunov, S. N., Totmenin, A. V., Loparev, V. N., Safronov, P. F., Gutorov, V. V., et al. (2000) Alastrim smallpox *Variola minor* virus genome DNA sequences, *Virology*, **266**, 361-386, doi: 10.1006/viro.1999.0086.
101. Esposito, J. J., Sammons, S. A., Frace, A. M., Osborne, J. D., Olsen-Rasmussen, M., et al. (2006) Genome sequence diversity and clues to the evolution of variola (smallpox) virus, *Science*, **313**, 807-812, doi: 10.1126/science.1125134.
102. Duch ene, S., Holmes, E. C., and Ho, S. Y. (2014) Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates, *Proc. Biol. Sci.*, **281**, 20140732, doi: 10.1098/rspb.2014.0732.
103. Senkevich, T. G., Yutin, N., Wolf, Y. I., Koonin, E. V., and Moss, B. (2021) Ancient gene capture and recent gene loss shape the evolution of orthopoxvirus–host interaction genes, *mBio*, **12**, e0149521, doi: 10.1128/mBio.01495-21.
104. Shchelkunov, S. N., Safronov, P. F., Totmenin, A. V., Petrov, N. A., Ryazankina, O. I., et al. (1998) The genomic sequence analysis of the left and right species-specific terminal region of a cowpox virus strain reveals unique sequences and a cluster of intact ORFs for immunomodulation.

- latory and host range proteins, *Virology*, **243**, 432-460, doi: 10.1006/viro.1998.9039.
105. Mühlemann, B., Vinner, L., Margaryan, A., Wilhelmson, H., de la Fuente Castro, C., et al. (2020) Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age, *Science*, **369**, eaaw8977, doi: 10.1126/science.aaw8977.
 106. Shchelkunov, S. N. (2012) Orthopoxvirus genes that mediate disease virulence and host tropism, *Adv. Virol.*, **2012**, 524743, doi: 10.1155/2012/524743.
 107. Pajer, P., Dresler, J., Kabícková, H., Písa, L., Aganov, P., et al. (2017) Characterization of two historic smallpox specimens from a Czech Museum, *Viruses*, **9**, 200, doi: 10.3390/v9080200.
 108. Duggan, A. T., Perdomo, M. F., Piombino-Mascalì, D., Marciniak, S., Poinar, D., et al. (2016) 17th century *Variola virus* reveals the recent history of smallpox, *Curr. Biol.*, **26**, 3407-3412, doi: 10.1016/j.cub.2016.10.061.
 109. Biagini, P., Thèves, C., Balaesque, P., Géraut, A., Cannet, C., et al. (2012) *Variola virus* in a 300-year-old Siberian mummy, *N. Engl. J. Med.*, **367**, 2057-2059, doi: 10.1056/NEJMc1208124.
 110. Ferrari, G., Neukamm, J., Baalsrud, H. T., Breidenstein, A. M., Ravinet, M., et al. (2020) *Variola virus* genome sequenced from an eighteenth-century museum specimen supports the recent origin of smallpox, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **375**, doi: 10.1098/rstb.2019.0572.
 111. Hopkins, D. R. (1983) *Princes and Peasants: Smallpox in History*, Univ. of Chicago Press, Chicago and London.
 112. Kochneva, G., Kolosova, I., Maksyutova, T., Ryabchikova, E., and Shchelkunov, S. (2005) Effects of deletions of kelch-like genes on cowpox virus biological properties, *Arch. Virol.*, **150**, 1857-1870, doi: 10.1007/s00705-005-0530-0.
 113. Hendrickson, R. C., Wang, C., Hatcher, E. L., and Lefkowitz, E. J. (2010) Orthopoxvirus genome evolution: the role of gene loss, *Viruses*, **2**, 1933-1967, doi: 10.3390/v2091933.
 114. Jenner, E. (1798) *An Inquiry into the Causes and Effects of the Variolae Vaccinae: A Disease Discovered in Some of the Western Counties of England, Particularly Gloucestershire, and Known by the Name of Cow Pox*, London: Printed for the author by Samson Low, No. 7, Berwick Street, Soho.
 115. Tulman, E. R., Delhon, G., Afonso, C. L., Lu, Z., Zsak, L., et al. (2006) Genome of horsepox virus, *J. Virol.*, **80**, 9244-9258, doi: 10.1128/JVI.00945-06.
 116. Duggan, A. T., Klunk, J., Porter, A. F., Dhody, A. N., Hicks, R., et al. (2020) The origins and genomic diversity of American Civil War Era smallpox vaccine strains, *Genome Biol.*, **21**, 175, doi: /10.1186/s13059-020-02079-z.
 117. Schrick, L., Tausch, S. H., Dabrowski, P. W., Damaso, C. R., Esparza, J., et al. (2017) An early American smallpox vaccine based on horsepox, *N. Engl. J. Med.*, **377**, 1491-1492, doi: 10.1056/NEJMc1707600.
 118. Qin, L., Favis, N., Famulski, J., and Evans, D. H. (2015) Evolution of and evolutionary relationships between extant vaccinia virus strains, *J. Virol.*, **89**, 1809-1824, doi: 10.1128/JVI.02797-14.
 119. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians, *Nature*, **528**, 499-503, doi: 10.1038/nature16152.
 120. Lindo, J., Huerta-Sánchez, E., Nakagome, S., Rasmussen, M., Petzelt, B., et al. (2016) A time transect of exomes from a Native American population before and after European contact, *Nat. Commun.*, **7**, 13175, doi: 10.1038/ncomms13175.
 121. Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., et al. (2021) RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation, *Nucleic Acids Res.*, **49**, D1020-D1028, doi: 10.1093/nar/gkaa1105.
 122. Pierce, N. T., Irber, L., Reiter, T., Brooks, P., and Brown, C. T. (2019) Large-scale sequence comparisons with sourmash, *F1000Res.*, **8**, 1006, doi: 10.12688/f1000research.19675.1.
 123. Wood, D. E., Lu, J., and Langmead, B. (2019) Improved metagenomic analysis with Kraken 2, *Genome Biol.*, **20**, 257, doi: 10.1186/s13059-019-1891-0.
 124. Cárdenas, Y. O. A., Neuenschwander, S., and Malaspinas, A.-S. (2021) Benchmarking metagenomics classifiers on ancient viral DNA: a simulation study, *bioRxiv*, doi: 10.1101/2021.04.30.442132.
 125. Wibowo, M. C., Yang, Z., Borry, M., Hübner, A., Huang, K. D., et al. (2021) Reconstruction of ancient microbial genomes from the human gut, *Nature*, **594**, 234-239, doi: 10.1038/s41586-021-03532-0.
 126. Borry, M., Hübner, A., Rohrlach, A. B., and Warinner, C. (2021) PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA *de novo* assembly, *PeerJ*, **9**, e11845, doi: 10.7717/peerj.11845.
 127. Fellows Yates, J. A., Andrades Valtueña, A., Vågene, Å. J., Cribdon, B., Vělsko, I. M., et al. (2021) Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir, *Sci. Data*, **8**, 31, doi: 10.1038/s41597-021-00816-y.
 128. Vernot, B., Zavala, E. I., Gómez-Olivencia, A., Jacobs, Z., Slon, V., et al. (2021) Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments, *Science*, **372**, eabf1667, doi: 10.1126/science.abf1667.
 129. Zavala, E. I., Jacobs, Z., Vernot, B., Shunkov, M. V., Kozlikin, M. B., et al. (2021) Pleistocene sediment DNA reveals hominin and faunal turnovers at Denisova Cave, *Nature*, **595**, 399-403, doi: 10.1038/s41586-021-03675-0.
 130. Ng, T. F., Chen, L. F., Zhou, Y., Shapiro, B., Stiller, M., et al. (2014) Preservation of viral genomes in 700-y-old caribou feces from a subarctic ice patch, *Proc. Natl. Acad. Sci. USA*, **111**, 16842-16847, doi: 10.1073/pnas.1410429111.
 131. Smith, O., Dunshea, G., Sinding, M. S., Fedorov, S., Germonpre, M., et al. (2019) Ancient RNA from Late Pleistocene permafrost and historical canids shows tissue-specific transcriptome survival, *PLoS Biol.*, **17**, e3000166, doi: 10.1371/journal.pbio.3000166.
 132. Düx, A., Lequime, S., Patrono, L. V., Vrancken, B., Boral, S., et al. (2020) Measles virus and rinderpest virus divergence dated to the sixth century BCE, *Science*, **368**:1367-1370, doi: 10.1126/science.aba9411.
 133. Gryseels, S., Watts, T. D., Kabongo Mpolesha, J. M., Larsen, B. B., Lemey, P., et al. (2020) A near full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue, *Proc. Natl. Acad. Sci. USA*, **117**, 12222-12229, doi: 10.1073/pnas.1913682117.