

# Title

## Neolithic and Medieval virus genomes reveal complex evolution of Hepatitis B

### Authors

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

### Affiliations

<sup>1</sup>Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany.

<sup>2</sup>Max Planck Institute for the Science of Human History, Jena, Germany.

<sup>3</sup>Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland

<sup>4</sup>Systematic Proteomics & Bioanalytics, Institute for Experimental Medicine, Kiel University, Kiel, Germany.

<sup>5</sup>Institute of Pre- and Protohistoric Archaeology, Kiel University, Kiel, Germany.

<sup>6</sup>Broad Institute, Stanley Center for Psychiatric Research, Cambridge, MA, USA.

<sup>7</sup>Institute for Pre- and Protohistory and Near Eastern Archaeology, Heidelberg University, Heidelberg, Germany.

<sup>8</sup>Department of Physical Anthropology, Institute of Forensic Medicine, University of Bern, Bern, Switzerland.

<sup>9</sup>State Office for Heritage Management and Archaeology Saxony-Anhalt and State Museum of Prehistory, Halle, Germany.

<sup>10</sup>Danube Private University, Krems, Austria.

<sup>11</sup>Department of Biomedical Engineering, University Hospital Basel, University of Basel, Basel, Switzerland.

<sup>12</sup>Integrative Prehistory and Archaeological Science, University of Basel, Basel, Switzerland.

<sup>13</sup>Clinic for Internal Medicine I, University Hospital of Schleswig-Holstein, Kiel, Germany.

†Corresponding author E-mails: [b.krause-kyora@ikmb.uni-kiel.de](mailto:b.krause-kyora@ikmb.uni-kiel.de), [krause@shh.mpg.de](mailto:krause@shh.mpg.de)

\*These authors contributed equally to this work

## Abstract

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

## Introduction

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

HBV has a circular, partially double-stranded DNA genome of about 3.2kbp that encodes four overlapping open reading frames (P, pre-S/S, pre-C/C, and X). Based on the genomic sequence diversity, HBVs are currently classified into 8 genotypes (A-H) and numerous subgenotypes that show distinct geographic distributions (Castelhano et al., 2017). All genotypes are hypothesised to be primarily the result of recombination events (Littlejohn et al., 2016; Simmonds and Midgley, 2005). To a lesser extent, HBV evolution is also driven by the accumulation of point mutations (Schaefer 2007, Araujo 2015).

Despite being widespread and well-studied, the origin and evolutionary history of HBV is still unclear and controversial (Littlejohn et al., 2016, Souza et al., 2014). HBVs in non-human primates (NHP), for instance in chimpanzees and gorillas, are phylogenetically closely related to, and yet distinct from, human HBV isolates, supporting the notion of an Africa origin of the virus (Souza et al., 2014). Molecular-clock based analyses dating the origin of HBV have resulted in conflicting estimates with some as recent as about 400 years ago (Zhou and Holmes, 2007, Souza et al., 2014). These observations have raised doubts about the suitability of molecular dating approaches for reconstructing the evolution of HBV (Bouckaer et al., 2103, Souza et al., 2014). Moreover, ancient DNA (aDNA) research on HBV-infected mummies from the 16<sup>th</sup> century AD revealed a very close relationship between the ancient and modern HBV genomes (Kahila Bar-Gal et al., 2012, Patterson Ross et al., 2018), indicating a surprising lack of temporal genetic changes in the virus during the last 500 years (Patterson Ross et al., 2018). Therefore, diachronic aDNA HBV studies, in which both the changes in the viral genome over time as well as the provenance and age of the archaeological samples, are needed to better understand the origin and evolutionary history of the virus.

Here, we report the analysis of three complete HBV genomes recovered from human skeletal remains from the prehistoric Neolithic and Medieval Periods in Central Europe. Our results show that HBV already circulated in the European population more than 7000 years ago.

85 Although the ancient forms show a relationship to modern isolates they appear to represent  
86 distinct lineages that have no close modern relatives and are possibly extinct today.  
87

## Results and Discussion

We detected evidence for presence of ancient HBV in three human tooth samples as part of a metagenomic screening for viral pathogens that was performed on shotgun sequencing data from 45 skeletons using the metagenomic alignment software MALT (Vagene et al., 2018). The remains of the individuals were excavated from the Neolithic sites of Karsdorf (Linearbandkeramik [LBK], 5056–4959 cal BC) and Sorsum (Tiefstichkeramik group of the Funnel Beaker culture, 3335-3107 cal BC), the medieval cemetery of Petersberg/Kleiner Madron (1020-1116 cal AD), all located in Germany (Fig. A, Supplementary Figure S1-S3). After the three aDNA extracts had appeared HBV-positive in the initial virus screening, they were subjected to deep-sequencing without any prior enrichment resulting in 367 to 419 million reads per sample (table 1). Analysis of the human DNA recovered from Karsdorf (3-fold genomic coverage) revealed that the sample clusters tightly with other contemporary early Neolithic individuals from the LBK (Supplementary Figure S12). The genetic makeup of the early LBK agriculturalists was previously found quite distinct from the preceding western hunter gatherers of Europe. The genetic shift between both populations was interpreted as a result of early farmers migrating from Western-Anatolia into Central Europe introducing agriculture (Lazaridis et al. 2014, Haak et al. 2015). The almost 2000 years younger Sorsum individual (1.2-fold genomic coverage) is most closely related with individuals from the contemporary Funnel Beaker culture that inhabited Northern Germany at the end of the 4th millenium BCE (Supplementary Figure S12). This population was previously shown to be quite admixed, as a result of a spatial and temporal overlap of early Neolithic farmers and remaining western hunter gatherers for almost 2000 years (Bollongino et al. 2013, Haak et al. 2015). The Petersberg individual (2.9-fold genomic coverage) showed, genetic affinities with modern day central European populations, typical for individuals from that time. All three ancient human individuals are therefore in agreement with the archaeological evidence and radiocarbon dates for their respective time of origin. Together with typical aDNA damage patterns (Supplementary Figures S4-S5), the human population genetic investigation supports the ancient origin of the obtained datasets.

For successful HBV genome reconstruction, we mapped all metagenomic sequences to 16 HBV reference genomes (8 human genotypes (A-H) and 8 NHP genotypes from Africa and Asia) that are representative of the current HBV strain diversity (Supplementary Table S6). The mapped reads were used for a *de novo* assembly, resulting in contigs from which one ancient HBV consensus sequence per sample was constructed. The consensus genomes are 3161 (46-fold coverage), 3182 (47-fold coverage), and 3183 (105-fold coverage) nucleotides

in length, which falls in the length range of modern HBV genomes and suggests that we successfully reconstructed the entire ancient HBV genomes (table 1, Supplementary Figure S6-S8). Further, when we conducted liquid chromatography-mass spectrometry (LC-MS) based bottom-up proteomics on tooth material from the three individuals, we identified in the Karsdorf and Petersberg samples a peptide that is part of the very stable HBV core protein, supporting the presence and active replication of HBV in the individuals' blood (Supplementary Figure S18).

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

Owing to continuous recombination over time, different gene segments or modules of the ancestral genomes can show up in various subsequent virus generations. Such precursors have been postulated (Simmonds and Midgley, 2005) and their existence is supported by the results of our recombination analysis (Supplementary Figure S14-S17, source data 1). Some fragments of the Karsdorf sequences appeared to be very similar to modern human (G, E) and African NHP genotypes, and the Sorsum genome partially showed a high similarity to the human genotypes G, E and B. (Supplementary Figure S14-S15, S17, source data 1). Given the close relationship between the two Neolithic virus genomes, it is also conceivable that the older HBV from Karsdorf could have been a distant source for the younger Sorsum virus (Supplementary Figure S14-S15, S17, source data 1). The closer relationship between the Neolithic and the NHP strains compared to other human strains is noteworthy and may have involved reciprocal cross-species transmission at one or possibly several times in the past (Simmonds and Midgley, 2005, Souza et al., 2014, Rasche et al. 2016).

Taken together, our results demonstrate that HBV already existed in Europeans 7000 years ago and that its genomic structure closely resembled that of modern hepatitis B viruses. Both Neolithic viruses fall between the present-day modern human and the known NHP diversity. Therefore, it can be hypothesized that although the two Neolithic HBV strains are no longer

observed today and thus may reflect two distinct clades that went extinct, they could still be closely related to the remote ancestors of the present-day genotypes, which is supported by signs of ancient recombination events. More ancient precursors, intermediates and modern strains of both humans and NHPs need to be sequenced to disentangle the complex evolution of HBV. As this evolution is characterized by recombination and point mutations and may further be complicated by human-ape host barrier crossing (Simmonds and Midgley, 2005, Souza et al., 2014, Rasche et al. 2016), genetic dating is not expected to yield meaningful results. This is additionally supported by a TempEst analysis (Rambaut et al., 2016) that shows very little temporal signal (Supplementary Figure S10). It should, however, be noted that the oldest genome (Karsdorf) was found in an individual that belonged to a population of early farmers that had migrated in the previous few hundred years from the Near East into central Europe. One might speculate that the close proximity to recently domesticated animals, changes in subsistence strategy as well as the adopted sedentary lifestyle might have contributed to the spread of HBV within Neolithic human populations.

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

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

190 death as fulminant liver failure is rather rare in modern day patients. The HBV infection  
191 might have instead contributed to other forms of lethal liver failure such as cirrhosis or liver  
192 cancer.

193

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

196



## Materials and Methods

### *Human remains*

The LBK settlement of **Karsdorf**, Saxony-Anhalt, Germany, is located in the valley of the river Unstrut. Between 1996-2010 systematic excavations were conducted at Karsdorf that led to the discovery of settlements and graves from the Neolithic to the Iron Age (Behnke, 2007, 2011, 2012). The LBK is represented by 24 longhouses in north-west to south-east orientation that were associated with settlement burials (Veit, 1996). The investigated individual 537 is a male with an age at death of around 25-30 years (Supplementary Figure S1), dated to 5056–4959 cal BC (KIA 40357 – 6116 ± 32 BP) (Brandt et al., 2014, Nicklisch, 2017).

The gallery grave of **Sorsum**, Lower-Saxony, Germany, is typologically dated to the Tiefstichkeramik (group of the Funnelbeaker culture). Sorsum is exceptional as it was built into the bedrock. During the excavations (1956-1960) of the grave chamber around 105 individuals were recovered (Claus, 1983, Czarnetzki, 1966). Individual XLVII 11 analyzed in this study is a male (Supplementary Figure S2) and dates to 3335-3107 cal BC (MAMS 33641 – 4501 ± 19 BP).

The medieval cemetery on the **Petersberg/Kleiner Madron**, Bavaria, Germany, lies on a hill top at 850 meters asl and 400 meters above the floor of the Inn Valley. On the eastern part of the cemetery, which is under discussion here, members of a priory were buried that was most likely established in the late 10<sup>th</sup> century. Written sources document its existence from 1132 onwards (Meier, 1998). During systematic excavations (1997-2004) in the southeastern part of the churchyard 99 graves with a higher, but hardly determinable number of individuals were uncovered. The examined individual in grave 820 is a male with an age at death of around 65-70 years (Lösch, 2009 – Supplementary Figure S3) dating to 1020-1116 cal AD (MAMS 33642 – 982 ± 17 BP).

### *DNA extraction and sequencing*

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

of powder were used for extraction following a silica-based protocol (Dabney et al., 2013). Negative controls were included in all steps.

From each sample, double-stranded DNA sequencing libraries (UDG-half) were prepared according to an established protocol for multiplex high-throughput sequencing (Meyer and Kircher, 2010). Sample-specific indices were added to both library adapters via amplification with two index primers. Extraction and library blanks were treated in the same manner. For the initial screening, the library of the individual from Karsdorf was sequenced on 1/50 of a lane on the HiSeq 3000 (2x75 bp) at the MPI SHH in Jena and the libraries from Petersberg and Sorsum were sequenced on the Illumina HiSeq 4000 (2x75 bp) platform at the Institute of Clinical Molecular Biology, Kiel University, using the HiSeq v4 chemistry and the manufacturer's protocol for multiplex sequencing. Deep-sequencing for each of the three samples was carried out on 2 lanes on the Illumina HiSeq 4000 platform at the Institute of Clinical Molecular Biology, Kiel University.

#### ***Metagenomics data processing, screening, and analyses***

The datasets for the three ancient samples comprised paired-end reads. The adapter sequences were removed and overlapping paired-end reads were merged with ClipAndMerge which is a module of the EAGER pipeline (Peltzer et al., 2016). The metagenomic viral screening was carried out using MALT (Vagene et al., 2018) and the NCBI viral RefSeq database. All three samples showed HBV-specific reads. In order to obtain all HBV related sequencing reads we mapped against a multi-fasta reference containing one representative of each genotype (A-H) and eight ape strains using BWA (Li and Durbin, 2010) (Supplementary Table S6). Mapped reads were extracted from the BAM file, converted to FASTQ and a *de novo* assembly using SPAdes (Bankevich et al., 2012) was carried out. Resulting contigs for each K-value were checked and the k-value that spawned the longest contigs was selected as criteria for further analysis. The contigs were re-mapped with BWA against the multi-fasta reference. The resulting alignment was visually inspected in IGV v 2.3.92 (Thorvaldsdóttir et al., 2013) to archive information about contig order and direction. Based on that information, a consensus sequence was constructed from the contigs.

We assembled a comprehensive set of reference genomes using 5497 non-recombinant genomes available at hpvdb (<https://hbvdb.ibcp.fr/HBVdb/HBVdbDataset?seqtype=0>) and a previously defined set of 74 ape-infecting HBV genomes. In order to reduce the actual number of genomes used for subsequent inferences but retain the full range of known HBV diversity, we clustered all sequences using UClust v 1.1.579 (Edgar et al., 2010). We

extracted the centroid sequences based on a sequence identity of at least 97%, which resulted in 493 representative genomes. Those genomes together with all available ancient genomes were aligned using Geneious version 10.1.2 (Kearse et al., 2012) with a 65% similarity cost matrix, a gap open penalty of 12 and a gap extension penalty of 3. The multiple sequence alignment was stripped of any sites (columns) that had gaps in more than 95% of sequences. The complete alignment including all modern and ancient genomes is available as multi-fasta in source data 2. The alignment was used to construct a network with the software SplitsTree v4 (Huson and Bryant, 2006), creating a NeighborNet (Bryant and Moulton, 2004) with uncorrected P distances.

### ***Recombination analysis***

We performed recombination analysis using all modern full reference genomes (n=493) and five ancient genomes used for the network analysis (see above). The methods RDP, GENECONV, Chimera, MaxChi, BootScan, SiScan, 3Seq within RDP v4 (Martin et al., 2015) with a window size of 100 nt and the parameter set to circular genome with and without outgroup reference (results are provided in source data 1) and SimPlot v 3.5.1 (Lole et al. 1999, Supplementary Figure S14-S17) were applied to the data set.

### ***Human population genetic analyses***

Mapping of the adapter-clipped and merged FASTQ files to the human reference genome hg19 was done using BWA (Li and Durbin, 2010) using a reduced mapping stringency of “-n 0.01” and the mapping quality parameter “-q 30”. The mapped sequencing data was transformed into the *Eigenstrat* format (Price et al., 2006) and merged with a dataset of 1.233.013 SNPs (Haak et al., 2015, Mathieson et al., 2015). Using the software Smartpca (Patterson et al., 2006) the three samples and previously published ancient populations were projected onto a base map of genetic variation calculated from 32 West Eurasian populations (Supplementary Figure S11-S13).

### ***Sex determination***

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

### ***LC-MS based bottom-up proteomics***

Proteins were extracted from powdered tooth samples (50 mg) using a modified filter-aided sample preparation (FASP) protocol as previously described (Cappellini et al., 2013, Warinner et al., 2014). Samples were digested using trypsin and analyzed by LC-MS/MS. Protein identification was performed using the SequestHT (Thermo Scientific) search engine in a combined database comprising the full Swiss protein database (468,716 entries), a hepatitis B data base (7 entries) and a common contaminant list. Further details regarding the LC-MS/MS analysis and database search parameters are given in the supplementary information and Supplementary Figure S19.

## **Acknowledgements**

We are grateful to the following people and institutions for providing samples, support, and advice: Bodo Krause-Kyora, Hildegard Nelson (Referat A1 Archäologische Dokumentation, Niedersächsisches Landesamt für Denkmalpflege), Ulrike Weller (Sammlungsverwaltung Archäologie Landesmuseum Hannover) and Britta Steer for technical assistance with proteomics sample preparation.

This work was supported by the Collaborative Research Centre 1266 *Scales of Transformation*, the Excellence Cluster 306 *Inflammation at Interfaces*, the Medical Faculty of Kiel University, the Max Planck Society and the European Research Council (ERC) starting grant APGREID (to J.K.). Excavations and analysis of the archaeological site of Karsdorf were supported by the German Research Foundation (DFG) Grant of Kurt W. Alt (Al 287-7-1) and Harald Meller (Me 3245/1-1).

## **Declaration of interests**

All other authors declare that they have no conflicts of interest.

## **Accession numbers**

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

## References

- Araujo NM. 2015. Hepatitis B virus intergenotypic recombinants worldwide: An overview. *Infect Genet Evol.* 36:500-510. DOI: 10.1016/j.meegid.2015.08.024. PMID: 26299884
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19:455-477. doi: 10.1089/cmb.2012.0021. PMID: 22506599
- Behnke HJ. 2007. Süßes aus der Bronzezeit – Siedlungsgunst im Unstruttal bei Karsdorf, Burgenlandkreis. *Die Ausgrabungen 2004. Archäologie in Sachsen-Anhalt N.F.* 4. 63-86.
- Behnke HJ. 2011. Erste Siedler der Linienbandkeramik in der Karsdorfer Feldflur. *Ergebnisse der Ausgrabungen im Jahr 2005. Archäologie in Sachsen-Anhalt N.F.* 5. 184-199.
- Behnke HJ. 2012. Siedlungsgunst im Unstruttal bei Karsdorf, Burgenlandkreis: Ergebnisse der Grabungen 2006 und 2007. *Archäologie in Sachsen-Anhalt N.F.* 6. 35-70.
- Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, Sell C, Fajkosová Z, Powell A, Burger J. 2013. 2000 years of parallel societies in Stone Age Central Europe. *Science* 342:479-81. DOI: 10.1126/science.1245049. PMID 24114781
- Bouckaert R, Alvarado-Mora MV, Pinho JR. 2013. Evolutionary rates and HBV: issues of rate estimation with Bayesian molecular methods. *Antiviral Therapy* 18:497-503 DOI: 10.3851/IMP2656. PMID: 23792904
- Brandt G, Knipper C, Nicklisch N, Ganslmeier R, Klamm M, Alt KW. 2014. Settlement burials at the Karsdorf LBK site, Saxony-Anhalt, Germany - biological ties and residential mobility. In: A. Whittle, Penny Bickle (eds) *Early farmers. The view from archaeology and Science.* Oxford University Press, British Academy, 95-114.
- Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol.* 2004 Feb;21(2):255-65. DOI: <https://doi.org/10.1093/molbev/msh018>. PMID: 14660700
- Cappellini E, Gentry A, Palkopoulou E, Ishida Y, Cram D, Roos AM, Watson M, Johansson US, Fernholm B, Agnelli P, Barbagli F, Littlewood DTJ, Kelstrup ChD, Olsen JV, Lister AM, Roca AL, Dalén L, Gilbert MT. 2013. Resolution of the type material of the Asian elephant, *Elephas maximus* Linnaeus, 1758 (Proboscidea, Elephantidae). *Zoological Journal of the Linnean Society.* 170:222–232. DOI: 10.1111/zoj.12084
- Castelhano N, Araujo NM, Arenas M. 2017. Heterogeneous recombination among Hepatitis B virus genotypes. *Infection, Genetics and Evolution* 54:486-490. DOI: 10.1016/j.meegid.2017.08.015, PMID: 28827173
- Czarnetzki A. 1966. Die menschlichen Skelettreste aus vier neolithischen Steinkisten Hessens und Niedersachsens.
- Claus M. 1983. Das neolithische Felsenkammergrab auf dem Halsberg bei Sorsum, Stadt Hildesheim. *Die Kunde N. F.* 34/35, 1983/84, 91-122.
- Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga JL, Meyer M. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene

cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci USA*. 110:15758-63. DOI: 10.1073/pnas.1314445110. PMID: 24019490

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 26:2460-1. DOI: 10.1093/bioinformatics/btq461. PMID: 20709691

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 522:207-211. DOI: 10.1038/nature14317. PMID: 25731166

Huson DH, Bryant D. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*. 23:254-267. PMID: 16221896

Kahila Bar-Gal G, Kim MJ, Klein A, Shin DH, Oh CS, Kim JW, Kim TH, Kim SB, Grant PR, Pappo O, Spigelman M, Shouval D. 2012. Tracing hepatitis B virus to the 16th century in a Korean mummy. *Hepatology* 56:1671-80. DOI: 10.1002/hep.25852. PMID: 22610996

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

Key FM, Posth C, Krause J, Herbig A, Bos KI. 2017. Mining Metagenomic Data Sets for Ancient DNA: Recommended Protocols for Authentication. *Trends Genet*. 33:508-520. DOI: 10.1016/j.tig.2017.05.005. Epub 2017 Jul 5. PMID: 28688671

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet JM, Wahl J, Ayodo G, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GB, Cali F, Churnosov M, Cole DE, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon JM, Fedorova SA, Gallego Romero I, Gubina M, Hammer M, Henn BM, Hervig T, Hodoglugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Melegh B, Metspalu E, Molina J, Mountain J, Näkkäläjärvi K, Nesheva D, Nyambo T, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Vilems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, Krause J. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513:409-13. DOI: 10.1038/nature13673. PMID 25230663

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26:589-95. DOI: 10.1093/bioinformatics/btp698. PMID: 20080505

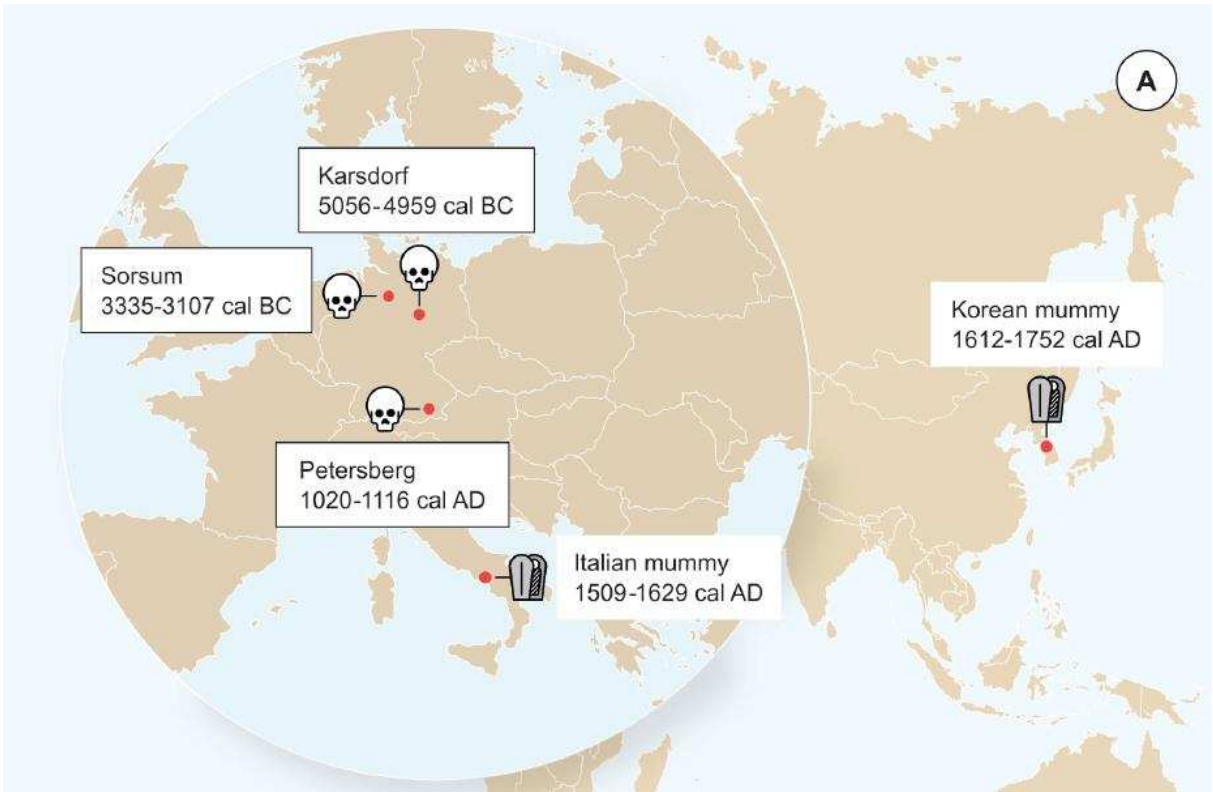
Littlejohn M, Locarnini S, Yuen L. 2016. Origins and Evolution of Hepatitis B Virus and Hepatitis D Virus. *Cold Spring Harb Perspect Med*. 6:a021360. DOI: 10.1101/cshperspect.a021360, PMID: 26729756

- Lösch S. 2009. Paläopathologisch-anthropologische und molekulare Untersuchungen an mittelalterlichen und frühneuzeitlichen Bevölkerungsgruppen. Dissertation zum Erwerb des Doktorgrades der Humanbiologie an der Medizinischen Fakultät der Ludwig-Maximilians-Universität zu München.
- Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol.* 73:152-60. PMID: 9847317
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution* 1:vev003. DOI: 10.1093/ve/vev00. PMID: 27774277
- Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JM, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Guerra MA, Roodenberg J, Vergès JM, Krause J, Cooper A, Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, Reich D. 2015 Genome-wide patterns of selection in 230 ancient Eurasians. *Nature.* 528:499-503. DOI: 10.1038/nature16152. PMID: 26595274
- Meier T. 1998. Ein mittelalterliches Kloster mit Friedhof auf dem Kleinen Madron/Petersberg bei Flintsbach a. Inn. Landkreis Rosenheim, Oberbayern. *Das Archäologische Jahr in Bayern* 1998: 127-129.
- Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010:pdb.prot5448. DOI: 10.1101/pdb.prot5448. PMID: 20516186
- Nicklisch N. 2017. Spurensuche an Skeletten. Paläodemografische und epidemiologische Untersuchungen an neolithischen und frühbronzezeitlichen Bestattungen aus dem Mittelbe-Saale-Gebiet im Kontext populationsdynamischer Prozesse. *Forschungsberichte des Landesmuseums für Vorgeschichte Halle*, Band 11.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190. PMID: 17194218
- Patterson Ross Z, Klunk J, Fornaciari G, Giuffra V, Duchêne S, Duggan AT, Poinar D, Douglas MW, Eden JS, Holmes EC, Poinar HN. 2018. The paradox of HBV evolution as revealed from a 16th century mummy. *PLoS Pathog.* 4;14:e1006750. DOI: 10.1371/journal.ppat.1006750. PMID: 29300782
- Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. 2016. EAGER: efficient ancient genome reconstruction. *Genome Biol.* 17:60. DOI: 10.1186/s13059-016-0918-z. PMID: 27036623
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 38:904-909. PMID: 16862161
- Rambaut A, T. T. Lam, L. Max Carvalho, and O. G. Pybus, "Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen)," *Virus Evol.*, vol. 2, no. 1, p. vew007, Jan. 2016.



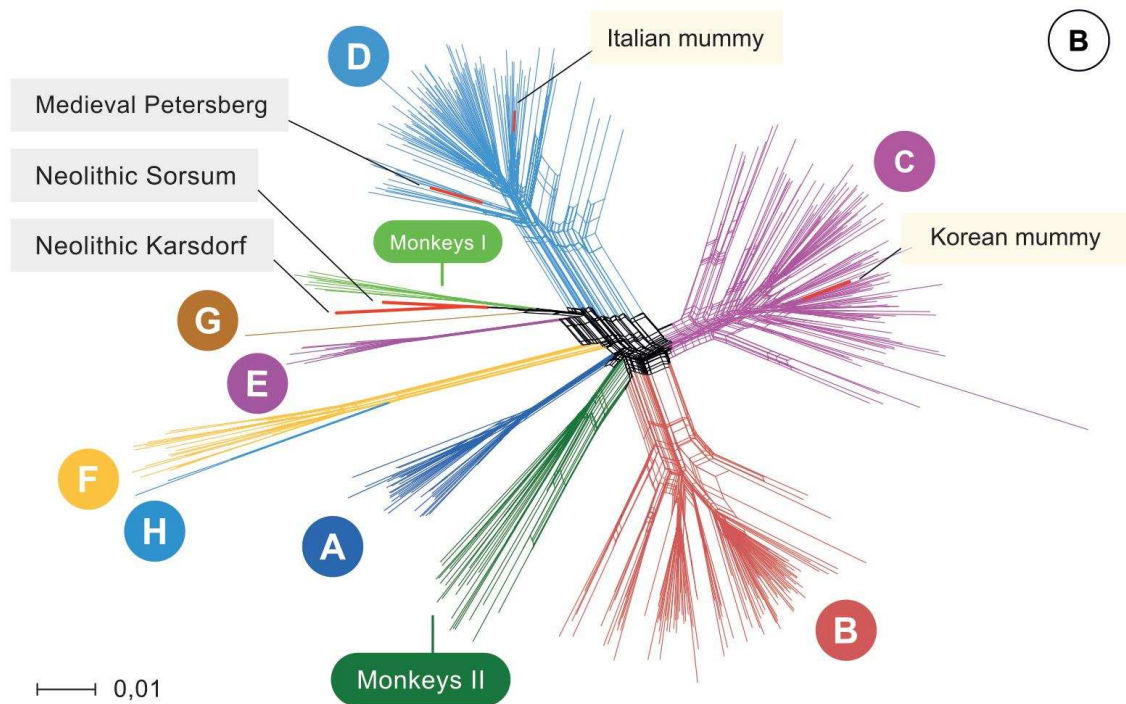
- Rasche A, Souza BFC, Drexler JF. 2016. Bat hepadnaviruses and the origins of primate hepatitis B viruses. *Curr Opin Virol*. 16:86-94. DOI: 10.1016/j.coviro.2016.01.015. PMID: 26897577
- Schaefer S. 2007. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J Gastroenterol*. 13:14-21. PMID: 17206751
- Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, Herbig A, Economou C, Benjak A, Busso P, Nebel A, Boldsen JL, Kjellström A, Wu H, Stewart GR, Taylor GM, Bauer P, Lee OY, Wu HH, Minnikin DE, Besra GS, Tucker K, Roffey S, Sow SO, Cole ST, Nieselt K, Krause J. 2013 Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science*. 341:179-183. DOI: 10.1126/science.1238286. PMID: 23765279
- Simmonds P, Midgley S. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. *J Virol*. 79:15467-76. DOI: 10.1128/JVI.79.24.15467-15476.2005, PMID: 16306618
- Skoglunda P, Storå J, Götherström A, Jakobsson M. 2013. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science*. 12:4477-4482. DOI: <https://doi.org/10.1016/j.jas.2013.07.004>
- Souza BF, Drexler JF, Lima RS, Rosário Mde O, Netto EM. 2014. Theories about evolutionary origins of human hepatitis B virus in primates and humans. *Braz J Infect Dis*. 18:535-43. DOI: 10.1016/j.bjid.2013.12.006, PMID: 24726560
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 14:178-192. DOI: 10.1093/bib/bbs017. PMID: 22517427
- Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, Spyrou MA, Andrades Valtueña A, Huson D, Tuross N, Bos KI, Krause J. 2018. *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat Ecol Evol*. 2:520-528. DOI: 10.1038/s41559-017-0446-6. PMID: 29335577
- Veit U. 1996. Studien zum Problem der Siedlungsbestattungen im europäischen Neolithikum, Münster, New York: Waxmann
- Warinner C, Rodrigues JF, Vyas R, Trachsel C, Shved N, Grossmann J, Radini A, Hancock Y, Tito RY, Fiddyment S, Speller C, Hendy J, Charlton S, Luder HU, Salazar-García DC, Eppler E, Seiler R, Hansen LH, Castruita JA, Barkow-Oesterreicher S, Teoh KY, Kelstrup CD, Olsen JV, Nanni P, Kawai T, Willerslev E, von Mering C, Lewis CM Jr, Collins MJ, Gilbert MT, Rühli F, Cappellini E. 2014. Pathogens and host immunity in the ancient human oral cavity. *Nat Genet*. 46:336-44. DOI: 10.1038/ng.2906. PMID: 24562188
- Warinner C, Herbig A, Mann A, Fellows Yates JA, Weiß CL, Burbano HA, Orlando L, Krause J. 2017. A Robust Framework for Microbial Archaeology. *Annu Rev Genomics Hum Genet*. 18:321-356. DOI: 10.1146/annurev-genom-091416-035526. PMID: 28460196
- WHO, Global hepatitis report, 2017
- Zhou Y, Holmes EC. 2007. Bayesian estimations of the evolutionary rate and age of hepatitis B virus. *J Mol Evol* 65:197-205. DOI:10.1007/s00239-007-0054-1, PMID: 17684696

**Figure Legends**



**Fig A. Origin of samples**

Geographic location of the samples from which ancient HBV genomes were isolated. Radiocarbon dates of the specimens is given in 2 sigma range. Icon indicate the sample material (tooth or mummy). HBV genomes obtained in this study indicated by black frame.



**Fig B Network**

Network of 495 modern, two published ancient genomes (light yellow box), and three ancient hepatitis B virus (HBV) obtained in this study (grey box). Colors indicating the 8 human HBV Genotypes (A-H), two monkey genotypes (Monkeys I, African apes and Monkeys II, Asian monkeys) and ancient genomes (red).

558 **Tables**

559

560 **Table 1.** Results of the genome reconstruction.

561

	# Merged reads	Length of HBV consensus sequence	mean HBV coverage	Gaps in the consensus sequence at nt position	# mapped reads HBV	# mapped reads human	mean human coverage	human genomes/ HBVgenomes
<b>Karsdorf</b>	386,780,892	3183	104X	2157-2175; 3107-3128; 3133-3183	10,718	122,568,310	2.96X	1 : 35.1
<b>Sorsum</b>	367,574,767	3182	47X	-	3,249	9,856,001	1.17X	1 : 40.2
<b>Petersberg</b>	419,413,082	3161	46X	880-1000; 1232-1329; 1331-1415; 1420-1581; 1585-1598	2,125	105,476,677	2.88X	1 : 16

562 # - number

563 nt - nucleotide