Screening great ape museum specimens for DNA viruses

Authors

Michelle Hämmerle[1,2], Meriam Guellil[1,2], Olivia Cheronet[1,2], Susanna Sawyer[1,2], Irune Ruiz-Gartzia[3], Esther Lizano[4,3], Aigerim Rymbekova[1,2], Pere Gelabert[1,2], Paolo Bernardi[1], Sojung Han[1,2], Lovro Trgovec-Greif[5,6], Thomas Rattei[5], Verena J. Schuenemann[1,2,7,8], Tomas Marques-Bonet[3,4,9,10], Katerina Guschanski[11,12], Sebastien Calvignac-Spencer[13,14], Ron Pinhasi[1,2], Martin Kuhlwilm[1,2]

Affiliations

1. Department of Evolutionary Anthropology, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria
2. Human Evolution and Archaeological Sciences (HEAS), University of Vienna, Austria
3. Institute of Evolutionary Biology (UPF-CSIC), PRBB, Dr. Aiguader 88, 08003 Barcelona, Spain
4. Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain
5. Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
6. Doctoral School of Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria
7. Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland
8. Department of Environmental Sciences, University of Basel, Basel, Switzerland
9. National Center for Genomic Analysis (CNAG), Baldiri i Reixac 4, 08028 Barcelona, Spain
10. Institució Catalana de Recerca i Estudis Avançats (ICREA) and Universitat Pompeu Fabra. Pg. Luís Companys 23, 08010, Barcelona, Spain
11. Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh, UK
12. Department of Ecology and Genetics, Animal Ecology, Uppsala University, SE-75236 Uppsala, Sweden
13. Helmholtz Institute for One Health, Helmholtz-Centre for Infection Research (HZI), 17489 Greifswald, Germany
14. Faculty of Mathematics and Natural Sciences, University of Greifswald, 17489 Greifswald, Germany

corresponding author: Martin Kuhlwilm (martin.kuhlwilm@univie.ac.at)

Abstract

Natural history museum collections harbour a record of wild species from the past centuries, providing a unique opportunity to study animals as well as their infectious agents. Thousands of great ape specimens are kept in these collections, and could become an important resource to study the evolution of DNA viruses, whose genetic material is likely to be preserved in dry museum specimens. Here, we screened 209 great ape museum specimens for 99 different DNA viruses, using hybridization capture coupled with short-read high-throughput sequencing. We report a capture design for great ape DNA viruses, sequencing data obtained using this approach, as well as findings regarding the presence of viruses, and several viral genomes obtained from historical specimens.

Background & Summary

The extensive collections of fossils and specimens preserved by natural history museums document the diversity of life forms, ecosystem dynamics, and the transformative processes that have shaped our planet over millions of years. They are a critical resource for scientific research. By employing high-throughput DNA sequencing techniques tailored to ancient or historical specimens, scientists can reconstruct ancient genomes[1] opening a new dimension in the field of museomics. The integration of genomic technologies with traditional museum collections presents unprecedented opportunities to address consequential questions in ecology and evolutionary biology[2,3]. Comparing genomic data from museum specimens with modern populations allows for the reconstruction of a more comprehensive tree of life and facilitates assessments of environmental change impacts on genetic diversity and population structure. This information is essential not only for understanding biodiversity and the forces that have shaped it in the past but also for developing effective conservation strategies to preserve genetic diversity and mitigate anthropogenic disturbances to natural ecosystems[4].

An intriguing avenue of investigation for museomics is the study of infectious diseases[5,6]. Natural history museum specimens have rarely been used for this purpose, but archaeological specimens have demonstrated the immense potential of ancient microbial genetic material in shedding light on the evolution and spread of infectious agents[7]. This is notably true for research on viruses with a DNA genome (DNA viruses). Hundreds to thousands of years old DNA viral genomes reconstructed from archeological specimens have unveiled key aspects of their recent evolutionary histories. For example, researchers have shown complete lineage turn-over of hepatitis B viruses (HBV; family *Hepadnaviridae*) in European human populations around the Bronze Age[8,9]. Genomic pathways to increased host adaptation and virulence have also been clarified for variola viruses (*Poxviridae*) in medieval human populations[10], for Marek's disease virus (*Herpesviridae*) in 19th/20th century poultry[11], or myxoma virus (*Poxviridae*) in rabbits[12].

In parallel, efforts geared toward a better understanding of the determinants of health in our closest relatives, the nonhuman great apes (orangutans – *Pongo pygmaeus*, *P. tapanuliensis* and *P. abelli*, gorillas – *Gorilla gorilla* and *G. beringei*, chimpanzees – *Pan troglodytes* – and bonobos – *Pan paniscus*), have revealed that these species host a broad range of DNA viruses. This includes enzootic viruses belonging to the families *Adenoviridae*, *Anelloviridae*, *Circoviridae*, *Hepadnaviridae*, *Herpesviridae*, *Papillomaviridae*, *Parvoviridae* and *Polyomaviridae*, as well as emerging viruses such as members of the family *Poxviridae*[13,14]. Co-phylogenetic analyses have shown that DNA viruses have often been stably associated with their hominid hosts for extensive periods of time. While co-divergence with their hosts is frequent, rare cross-species transmission events between hominids have contributed to shaping all hominid DNA viromes. For example, the herpes simplex virus 2 (*Herpesviridae*) that causes genital herpes in humans likely arose from the cross-species transmission of a virus infecting members of the gorilla lineage, several million years ago[15]. Conversely, patterns of genomic variation suggest that HBV was transmitted from humans to nonhuman great apes over the last few thousand years[9]. As we learn more about nonhuman great ape viromes, we will uncover more of the complex origins of their DNA viruses and those infecting humans.

Explorations of non-human great ape viromes are ongoing and should proceed in the framework of long-term studies of wild populations, but they could be efficiently complemented by analyses of the thousands of specimens kept in natural history museums around the world. Not only do these collections provide immediate access to animal tissues (as opposed to the noninvasive samples that constitute the bulk of biological sampling from modern populations), they also open a direct window on 200 years of increased anthropogenic disturbance characterized by dwindling, increasingly fragmented populations

102 of these animals[16]. These processes have likely altered nonhuman DNA viromes significantly,
103 and possibly resulted in the extinction of viral lineages[17]. Moreover, emerging viruses
104 constitute a threat to wild great ape populations[14,18], and they are often of human origin[19].
105 Therefore, studying the DNA viromes of nonhuman great ape museum specimens will
106 provide both information on viruses that still circulate in contemporary populations and
107 document hidden evolutionary trajectories taken by extinct viruses. This may also allow to
108 monitor changes in viromes of wild populations with the potential to intervene should the
109 populations be exposed to novel viruses (e.g. from humans or other species) they have not
110 co-evolved with.

111 Here, we present high-throughput sequencing data from 209 nonhuman great ape museum
112 specimens which were sampled and analysed in search for DNA viruses. These specimens
113 cover all nonhuman great ape species, most subspecies and large parts of their recent
114 geographical distribution. A majority originated in the wild, but our sample set also includes
115 captive individuals from European zoos (13%, n=28). Obtaining data from ancient specimens
116 can be challenging due to DNA degradation and damage, as well as potential contamination
117 during storage and handling[1]. While museum specimens are younger than typical
118 archaeological remains, high variability in DNA quality has been observed[20], and handling in
119 specific clean laboratory facilities, following strict protocols is necessary[21], and was applied
120 for this study.

121 In addition, we also developed and implemented an efficient customised enrichment
122 strategy, in-solution hybridization capture with RNA baits[22] to account for the extremely low
123 abundance of viral genetic material. Although metagenomic detection of viruses is possible,
124 it is particularly challenging for DNA viruses that, with a few exceptions, usually do not
125 replicate as intensely as RNA viruses[23]. Hybridization capture offers considerable flexibility,
126 because it allows us to multiplex baits targeting many different viruses, as well as enrich even
127 significantly divergent targets (up to 58% divergence tolerance according to some authors[24]).
128 However, probably many DNA viruses evolve relatively slowly[25]. We designed and used a bait
129 set that covered the genomes of 99 viral lineages representing 13 families. For capture, 8-10
130 samples were pooled to allow for a more cost-effective screening and sequenced on an
131 Illumina platform.

132 Analyses of the resulting sequences hints at the presence of multiple DNA viruses in this
133 sample set, including the detection of >500 reads of monkeypox virus (MPXV; family
134 *Poxviridae*) in three orangutans, as well as HBV in two chimpanzees and one gorilla,
135 respectively. For these specimens, the reconstruction of near complete viral genomes was
136 possible. Our work demonstrates the feasibility of viral DNA enrichment and detection from
137 museum specimens of great apes.
138
139

140 Methods

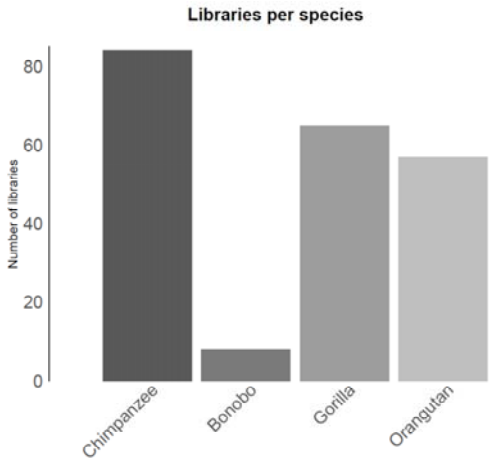
141

142 *Samples*

143 For this project, we collected a total of 209 great ape specimens (Fig. 1), from which 214
144 sequencing libraries were produced. Of these libraries, 66 were from gorillas, 84 from
145 chimpanzees (10 of these as *Pan* sp., but most likely *Pan troglodytes* ssp.), 8 from bonobos,
146 and 56 from orangutans, including different subspecies. We note that for five specimens two
147 separate extracts and sequencing libraries were prepared. The samples were obtained from
148 specimens housed in European natural history museums, namely in Germany in Berlin (n =
149 28), Bonn (n = 28), Dresden (n = 26), Frankfurt (n = 63) and Stuttgart (n = 6), in the Czech
150 Republic in Prague (n = 24), and in Austria in Salzburg (n = 22) and Vienna (n = 12).
151 Approximately 92% of libraries were obtained from teeth (n = 196), 17 libraries from soft
152 tissues (Table S1), and two specimens were phalanges. According to the museum metadata,

the oldest specimen was from 1838, and the most recent (a captive individual) was from 2014. Some individuals were held in captivity, mostly in zoos, whereas others were wild-caught. More detailed information concerning the individuals and the libraries can be found in Table S1.

The museum identifiers of the specimens are as follows: Senckenberg Forschungsinstitut und Naturmuseum Frankfurt/M.: 10325, 1110, 1111, 1112, 1113, 1114, 1115, 1118, 1119, 1120, 1121, 1126, 1132, 1134, 1576, 1579, 15792, 15817, 16180, 17826, 17961, 24510, 2495, 2538, 2638, 2639, 2654, 3221, 4103, 4104, 4106, 4107, 4108, 4109, 45713, 5277, 5532, 59140, 59147, 59158, 59296, 59297, 59298, 59299, 59301, 59303, 59304, 6716, 6779, 6782, 6785, 6992, 89780, 89781, 92953, 94796, 94797, 94799, 96255, 96550, 97029, 97143, ZIH9; Naturhistorisches Museum Vienna: NMW 1779, NMW 20516, NMW 25124, NMW 3081, NMW 3105/ST 663, NMW 3106, NMW 3107, NMW 3111/ST 665, NMW 3119, NMW 3948, NMW 7136, NMW 793/ST 1647; Přírodovědecké muzeum Prague: NMP 09605, NMP 10588, NMP 10784, NMP 22891, NMP 22892, NMP 22893, NMP 23283, NMP 23284, NMP 23295, NMP 23296, NMP 23297, NMP 24474, NMP 24475, NMP 46815, NMP 46816, NMP 46816-b, NMP 47007, NMP 47656, NMP 49711, NMP 50432, NMP 94205, NMP 94564, NMP 94957, NMP 95098; Senckenberg Naturhistorische Sammlungen Dresden: MTD B11877, MTD B12034, MTD B12062, MTD B12099, MTD B12101, MTD B12177, MTD B12178, MTD B1384-A.S. 1289, MTD B14244, MTD B15789, MTD B1607-A.S. 1690, MTD B247-A.S. 216, MTD B249-A.S. 214, MTD B251-A.S. 231, MTD B253-A.S. 221, MTD B266-A.S. 211, MTD B281-A.S. 239, MTD B287-A.S. 200, MTD B288-A.S. 198, MTD B3686, MTD B4188, MTD B4786, MTD B4788, MTD B4789, MTD B4793, MTD B61-A.S. 244; Museum für Naturkunde Berlin: ZMB_Mam_108652, ZMB_Mam_11637, ZMB_Mam_11638, ZMB_Mam_12799, ZMB_Mam_14644, ZMB_Mam_17011, ZMB_Mam_24838, ZMB_Mam_30755, ZMB_Mam_31617, ZMB_Mam_31621, ZMB_Mam_37523, ZMB_Mam_45130, ZMB_Mam_48173, ZMB_Mam_83519, ZMB_Mam_83522, ZMB_Mam_83547, ZMB_Mam_83606, ZMB_Mam_83607, ZMB_Mam_83617, ZMB_Mam_83642, ZMB_Mam_83643, ZMB_Mam_83647, ZMB_Mam_83648, ZMB_Mam_83653, ZMB_Mam_83675, ZMB_Mam_83681, ZMB_Mam_83682, ZMB_Mam_83685; Museum Koenig Bonn: ZFMK_MAM1938-0136, ZFMK_MAM1957-0003, ZFMK_MAM1957-0004, ZFMK_MAM1962-0131, ZFMK_MAM1963-0660, ZFMK_MAM1965-0544, ZFMK_MAM1965-0545, ZFMK_MAM1965-0546, ZFMK_MAM1965-0547, ZFMK_MAM1965-0550, ZFMK_MAM1976-0410, ZFMK_MAM1994-0482, ZFMK_MAM1997-0070, ZFMK_MAM1997-0076, ZFMK_MAM2012-0036, ZFMK_MAM2015-0479, ZFMK_MAM2019-0404, ZFMK_MAM2019-0405, ZFMK_MAM2019-0407, ZFMK_MAM2019-0408, MAM2019-0410, ZFMK_MAM2019-0415, ZFMK_MAM2019-0416, ZFMK_MAM2019-0417, ZFMK_MAM2019-0418, ZFMK_MAM2019-0419, ZFMK_MAM2019-0420, ZFMK_MAM2019-0421; Haus der Natur Salzburg: HNS-Mam-S-0073, HNS-Mam-S-0075, HNS-Mam-S-0076, HNS-Mam-S-0077, HNS-Mam-S-0078, HNS-Mam-S-0079, HNS-Mam-S-0082, HNS-Mam-S-0084, HNS-Mam-S-0085, HNS-Mam-S-0086, HNS-Mam-S-0519, HNS-Mam-S-0524, HNS-Mam-S-0525, HNS-Mam-S-0530, HNS-Mam-S-0531, HNS-Mam-S-0532, HNS-Mam-S-0533, HNS-Mam-S-0534, HNS-Mam-S-0535, HNS-Mam-S-0536, HNS-Mam-S-0550, HNS-Mam-S-0742; Staatliches Museum für Naturkunde Stuttgart: SMNS-Z-MAM-001687, SMNS-Z-MAM-001750, SMNS-Z-MAM-002012, SMNS-Z-MAM-045995, SMNS-Z-MAM-046000, SMNS-Z-MAM-048948.



198
199 **Figure 1. Number of libraries per great ape species included in this study.**
200
201 ***DNA extraction and library preparation***
202 All steps from grinding to indexing except the qPCR were performed in laboratories designed
203 and dedicated only to ancient DNA (aDNA) research while wearing protective clothing and
204 following aDNA laboratory best practices. DNA was directly extracted from soft tissue. Bone
205 and teeth were treated with a sandblaster and ground to bone powder using a *MixerMill*
206 (Retsch). For each specimen, 50 mg of powder was collected. DNA was extracted using an
207 established protocol used for aDNA[26]. Single-stranded DNA libraries were prepared[27],
208 followed by a clean-up with the *QIAGEN MinElute PCR Purification Kit*. We performed a
209 quantitative PCR for calculating the cycle number in the indexing PCR. Indexing was
210 performed in quadruplicates using *NEBNext Q5U*, followed by a clean-up using the
211 *NucleoMag® NGS Clean-up and Size Select* kit. Indexes are listed in Table S1. We assessed
212 quantity and quality using an *Invitrogen QubitTM 4 Fluorometer* and an *Agilent 4150*
213 *TapeStation*. The grinding step for all samples was performed in the Vienna Ancient DNA
214 laboratory of the University of Vienna, while the DNA extraction, library and QC steps were
215 performed for a subset of libraries (n=97, Table S1) in the ancient DNA laboratory at the
216 Universitat Pompeu Fabra in Barcelona, following the exact same protocols and best
217 practices.
218 To maximise economic efficiency, we constructed pools of 8-10 libraries, except of four,
219 where the specimen condition might have been influenced by a pathological condition of the
220 individual. If DNA concentration was below 5ng/µl per sample, we performed another
221 amplification, to preserve sufficient amounts of library. If the concentration was above 25
222 ng/µl, a dilution was required. Between 8 and 10 libraries were pooled by equal
223 concentration, whereby it was crucial to avoid pooling those with overlapping P5 or P7
224 adapters. In total, 210 libraries were pooled in 24 pools. We used the same concentration
225 threshold for un-pooled libraries.
226
227 ***Hybridization capture***
228 As aDNA or historical DNA does not only contain host and host-associated microbiome DNA,
229 but often an overwhelming abundance of bacterial and environmental DNA, shotgun
230 sequencing is usually economically infeasible for viruses[28], and target-specific approaches
231 can help in enrichment of sequences of interest[21]. We designed a capture set containing 99
232 different viruses from 13 families, whereby 49 are human-infecting, 18 great ape-infecting,
233 and the remaining were isolated from other primate species. The design was based on
234 reference genomes available in NCBI, and commercially produced by *myBaits®*, where a
235 BLAST search against the human reference genome was performed to exclude sequences

236 with any hits to it; baits were designed to be 80 nucleotides long with a 2X tiling. The viruses
237 and the NCBI reference sequence are reported in Table S2. We followed the protocol for the
238 capture provided by the manufacturer (Version 5.03, as ordered in July 2021). Briefly, after
239 adding the blockers and the hybridization mix with the RNA baits, the libraries were
240 incubated for approximately 40 hours at 60°C, to increase the efficiency of the capture
241 reaction and to allow for higher sequence deviation from the baits. DNA was eluted from the
242 beads in 30 µl Buffer E, and the supernatant was kept. A qPCR of the capture product was
243 performed in order to estimate the yield, and another PCR to amplify the capture product.
244 Libraries were pooled to 20ng total DNA, and single-end sequencing was performed on the
245 *Illumina NovaSeq 6000 (SP SR100 XP workflow)* at the Vienna BioCenter. The targeted
246 amount of sequencing reads per library was around one million reads.
247
248 ***Bioinformatic processing***
249 Adapters were trimmed from the fastq files with trimmomatic[29] (version 0.39)**,** and BBmap
250 (version 39.01) clumpify was used to remove duplicates introduced by PCR amplification[30].
251 To determine the metagenomic composition of the sequenced libraries, taxonomic
252 classification via Kraken2 using the standard database was performed[31]. This database
253 contains the strains included in the capture design. Heatmaps were plotted via a customized
254 python3 script for the target taxa at species and genus levels.
255 Where Kraken2 assigned more than 500 reads to one of the reference genomes included in
256 the capture kit, we performed a mapping with BWA[32] (version 0.7.17), using bwa aln (with
257 parameters "-n 0.04 -l 1000"). Mapping coverage along the reference genome was visualized
258 using aDNA-BAMplotter[33], and inspected individually. In case of unequal coverage along the
259 genome, we performed a literature search for alternative genomes obtained from great
260 apes. Summary statistics were calculated using these best reference genomes, including edit
261 distance, mapping quality, mapping quality ratio, and the percentage for 1-, 2- , and 10-fold
262 coverage, using a customized python3 script. Other figures were created with the R package
263 ggplot2[34].
264

265 Data Records

266 A custom hybridization capture design for great ape DNA viruses is reported in this
267 publication. We report the bait design for 99 virus strains of potential relevance to great
268 apes. The final design can be found as Supplementary Material SM2, the underlying NCBI
269 identifiers are reported in Table S2, and the NCBI sequences used in fasta format as
270 Supplementary Material SM1. Raw sequencing data after capture for the 214 individual
271 libraries has been uploaded to the Short Read Archive under the accession ID PRJEB75038.
272

273 Technical Validation

274
275 ***Sequencing data***
276 The median number of raw reads per library was 1,122,746 reads, with a high variability of
277 yield (SD=3,556,238). After adapter trimming, a median of 956,296 reads was retained, with
278 a reduction due to removal of adapter-multimers between 0.11 and 93.38%. The rates of
279 duplicated reads ranged between 21.35% and 88.67%, leaving a median of 338,560 high-
280 quality unique reads across the 214 libraries (SD=971,276; Fig. 2A-B).
281

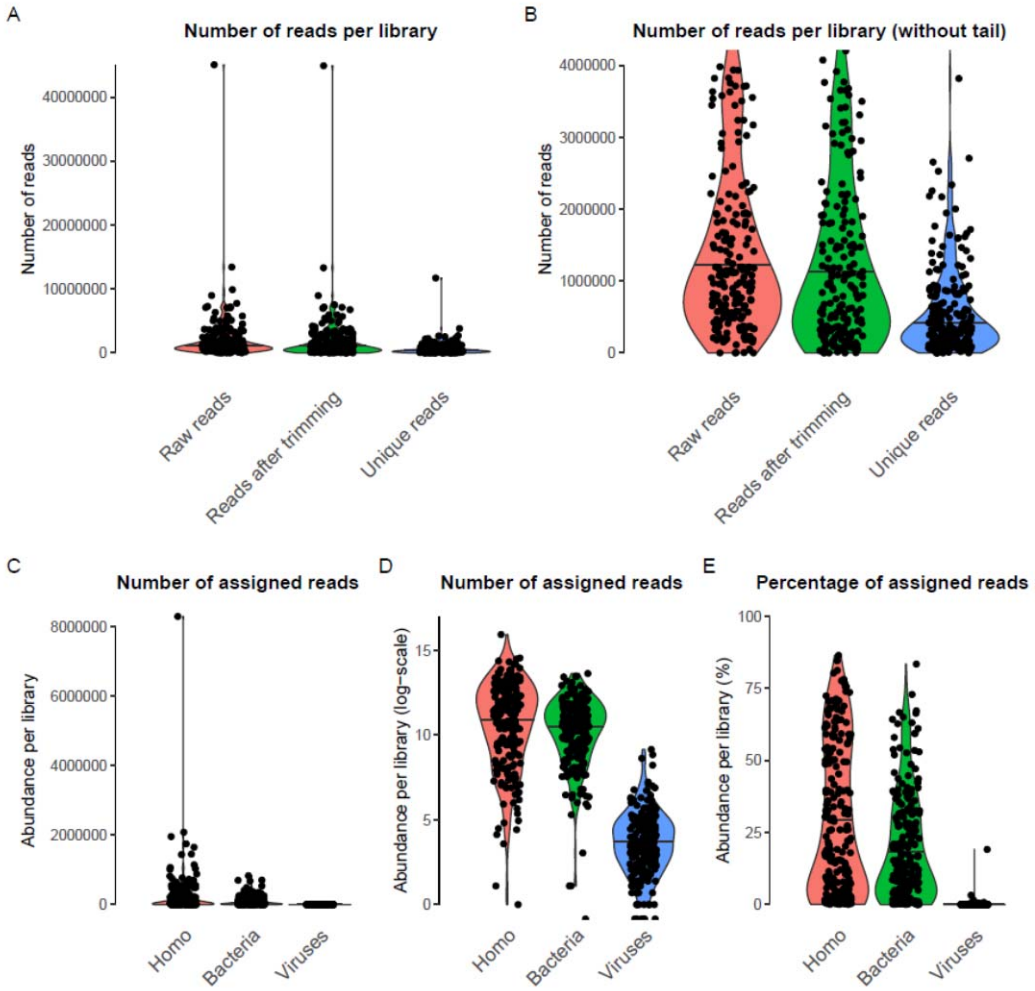


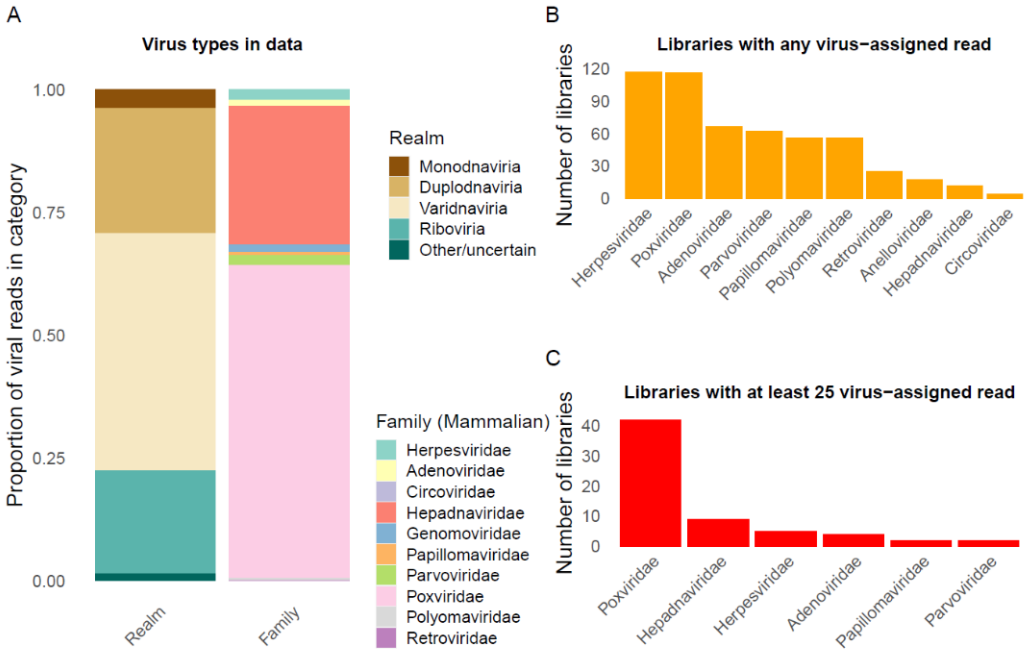
Figure 2: Distribution of sequencing reads per library. A) Numbers of raw reads, reads after trimming and unique reads (after BBmap clumpify) across 214 libraries. B) Same as in A) cut at 4 million reads (without tail). C) Number of reads assigned to *Homo*, Bacteria, or Viruses using Kraken2 per library. D) Same as C), but number of reads on log-scale. E) Percentage of assigned reads per library.

The sequence length of the raw reads was 100 bp, as defined by the sequencing setup, and after filtering and trimming, the median fragment length per library ranged from 40 to 100 bp. This suggests that fragmentation typical for ancient DNA[1] may not have been present across all libraries, as might be expected for DNA extracted from relatively recent museum specimens, and after the enrichment performed here.

Results from virus capture

Fragments from targeted virus fragments are likely absent or very rare in the libraries. Given the low expected abundance of any viral DNA, a large degree of amplification led to high duplication rates (see above). Furthermore, as we chose a lower hybridization temperature and a prolonged incubation time for the already highly amplified libraries, more unspecific binding resulted from our approach. We observed a median assignment of 21.95% to *Homo sapiens*, 13.91% to bacterial sequences and solely 0.1% to viruses (Fig. 2C-E). Mapping to human DNA likely reflected endogenous great ape DNA (and possibly human contamination),

304 and bacterial DNA likely resulted from post-mortem colonization of the specimen. While
305 many viral reads were assigned to bacteriophages (viruses infecting bacteria, not animals),
306 many libraries contained at least one read assigned to a viral family known to infect animals
307 (Table S3, Fig. 3A-B). Numerous libraries contained at least a small number of reads (25 or
308 more) assigned to a viral family (Fig. 3C). A surprising amount of libraries apparently
309 comprised poxvirus reads, which likely represented wrong assignments. When restricting to
310 virus strains that were included in the capture design, we observe fewer instances (n=30) of
311 libraries with potential virus fragments (Fig. 4), which likely reflects are more accurate
312 picture.
313



314
315 **Figure 3**. Summaries of reads assigned to different virus domains and families. A) Proportion of
316 virus-assigned reads based on kraken2 across 214 libraries, stratified by realm and by family (for
317 DNA viruses and *Retroviridae*). B) Number of libraries with any read assigned to virus families. C)
318 Number of libraries with at least 25 reads assigned to virus families.
319

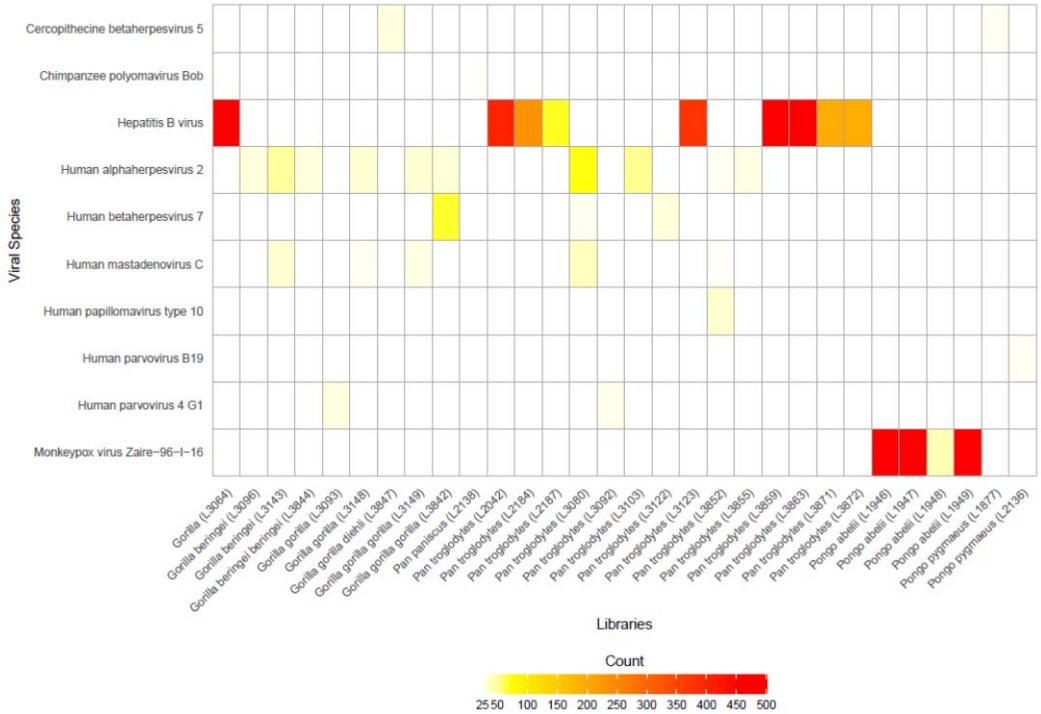


Figure 4: Heatmap of positive reads per viral species. The figure depicts all libraries that have at least 25 hits assigned to one of the 99 viruses in the capture kit (with merged numbers for Cytomegalovirus strains).

Henceforth, we focused on libraries with more than 500 assigned reads, which are the most likely to reflect true viral infection. We found six such instances, with assignments either to the family *Poxviridae* (n=3; all orangutan specimens) or the family *Hepadnaviridae* (n=3; 1 gorilla and 2 chimpanzee specimens). The only library with more viral than bacterial or human-mapping sequences is L1949 (19.13% viral reads), obtained from an orangutan specimen. We note that further rounds of enrichment capture might have provided higher on-target coverage[1,35], but at substantially higher costs.

Positive specimens

To validate the results from classifying reads, we attempted to reconstruct the corresponding viral genomes. Using reference-based mapping, we managed to obtain low-coverage monkeypox virus genomes from three libraries (up to 3.6-fold coverage on the reference genome sequence KJ642614) (Table 1). Further investigation of these specimens, including deeper sequencing to obtain high-coverage genomes, was published in another study[36], where we present their origin from a zoo outbreak in 1965. Similar efforts aimed at assembling HBV genomes from the three most promising HBV-positive specimens resulted in one high-coverage genome from a wild gorilla, and two from wild chimpanzees (Table 1). In other samples, putative viral sequences were at much lower abundance (Table S3).

Our data shows that a sizeable fraction of museum great ape specimens is amenable to the recovery of whole DNA virus genomes. This resource could contribute to a much better understanding of recent viral evolution in our closest living relatives.

Library ID (genus)	Individual ID (museum)	Reference	Mapped Reads	Mean ED	Mean DOC	1X	5X	10X
L1946	ZFMK_MAM1965-	KJ642614	7,931	0.51	3.83	93.04	33.25	2.58

(*Pongo*)	0546 (Bonn)	(monkeypox virus)						
L1947 (*Pongo*)	ZFMK_MAM1965-0545 (Bonn)	KJ642614 (monkeypox virus)	4,659	0.5	2.29	84.45	12.04	0.09
L1949 (*Pongo*)	ZFMK_MAM1965-0546 (Bonn)	KJ642614 (monkeypox virus)	2,352	0.61	1.13	63.83	1.03	0
L3064 (*Gorilla*)	96550 (Frankfurt)	AJ131567 (hepatitis B virus)	2,969	0.82	97.07	99.62	99.56	99.43
L3863 (*Pan*)	ZMB_Mam_83617 (Berlin)	ON706349 (hepatitis B virus)	517	1.82	18.19	94.44	73.60	53.80
L3859 (*Pan*)	ZMB_Mam_11638 (Berlin)	AF305327 (hepatitis B virus)	919	1.98	31.27	92.36	75.83	61.56

Table 1. Statistics for virus genomes identified in this study. Mapping metrics for the six viral genomes discovered in this study, including proportion of the reference genome covered by at least 1, 5 or 10 reads. Mapped reads = unique reads with MQ >30. DOC = Depth of coverage. ED = Edit distance.

Usage Notes

Data can be reprocessed using the tools described in the Methods section, namely kraken2 for metagenomic classification, bwa for mapping to reference genomes, following the documentation in the associated repository. Custom code for processing and visualisation is provided below.

Code Availability

The code used is available under https://github.com/admixVIE/Great-Ape-DNA-Virome.

Acknowledgements

The computational results of this work have been achieved using the Life Science Compute Cluster (LiSC) of the University of Vienna. We are grateful to the Zoologisches Forschungsmuseum A. Koenig, Leibniz-Institut zur Analyse des Biodiversitätswandels in Bonn, in particular Eva Bärmann, Jan Decher, and Christian Montermann at the Section Theriology; to Irina Ruf and Katrin Krohmann at the Mammalogy collection at Senckenberg Forschungsinstitut und Naturmuseum Frankfurt/M.; to Frank Zachos and Alexander Bibl at Naturhistorisches Museum Wien; to Stefan Merker at Zoology at Staatliches Museum für Naturkunde Stuttgart; to Petr Benda at Department of Zoology at National Museum (Natural History) Prague; to Clara Stefen and Jens Jakobitz at Mammalogie at Senckenberg Naturhistorische Sammlungen Dresden; to Robert Lindner at Haus der Natur – Museum für Natur und Technik in Salzburg; to Christiane Funk and Frieder Mayer at the mammalian collection at Museum für Naturkunde/Leibniz Institute for Evolution and Biodiversity research in Berlin.

This project has been funded by the Vienna Science and Technology Fund (WWTF) [10.47379/VRG20001] and by the Austrian Science Fund (FWF) [FW547002] to M.K. K.G. received support from the Swedish Research Council (VR) through grant 2020-03398.

Author contributions

M.K. conceived the topic of the study, supervised data analysis and wrote the manuscript. M.H. performed experiments and data analysis and wrote the manuscript. R.P. conceived the study and supervised the experimental work. M.G. supervised data analysis. S.C.-S. wrote the manuscript. S.S., E.L. and O.C. supervised experimental work. I. R.-G. performed experiments and contributed to the design of the dataset. P.B. performed experiments. A.R., P.B. and L.G. provided help in analysing data. P.G., T.R., V.J.S., T.M.-B. and K.G. provided help in writing the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Orlando, L. *et al.* Ancient DNA analysis. *Nat. Rev. Methods Prim.* **1**, 14 (2021).
2. Raxworthy, C. J. & Smith, B. T. Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol.* **36**, 1049–1060 (2021).
3. Holmes, M. W. *et al.* Natural history collections as windows on evolutionary processes. *Mol. Ecol.* **25**, 864–881 (2016).
4. Blair, M. E. Conservation museomics. *Conserv. Biol.* **n/a**, (2023).
5. Cook, J. A. *et al.* Integrating Biodiversity Infrastructure into Pathogen Discovery and Mitigation of Emerging Infectious Diseases. *Bioscience* **70**, 531–534 (2020).
6. DiEuliis, D., Johnson, K. R., Morse, S. S. & Schindel, D. E. Specimen collections should have a much bigger role in infectious disease research and response. *Proc. Natl. Acad. Sci.* **113**, 4–7 (2016).
7. Duchêne, S., Ho, S. Y. W., Carmichael, A. G., Holmes, E. C. & Poinar, H. The Recovery, Interpretation and Use of Ancient Pathogen Genomes. *Curr. Biol.* **30**, R1215–R1231 (2020).
8. Mühlemann, B. *et al.* Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* **557**, 418–423 (2018).
9. Arthur, K. *et al.* Ten millennia of hepatitis B virus evolution. *Science (80-.).* **374**, 182–188 (2021).
10. Mühlemann, B. *et al.* Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. *Science (80-.).* **369**, eaaw8977 (2020).
11. Fiddaman, S. R. *et al.* Ancient chicken remains reveal the origins of virulence in Marek's disease virus. *Science (80-.).* **382**, 1276–1281 (2023).
12. Alves, J. M. *et al.* Parallel adaptation of rabbit populations to myxoma virus. *Science (80-.).* **363**, 1319–1326 (2019).
13. Calvignac-Spencer, S., Düx, A., Gogarten, J. F., Leendertz, F. H. & Patrono, L. V. Chapter One - A great ape perspective on the origins and evolution of human viruses. in (eds. Kielian, M., Mettenleiter, T. C. & Roossinck, M. J. B. T.-A. in V. R.) vol. 110 1–26 (Academic Press, 2021).
14. Calvignac-Spencer, S., Leendertz, S. A. J., Gillespie, T. R. & Leendertz, F. H. Wild great apes as sentinels and sources of infectious disease. *Clin. Microbiol. Infect.* **18**, 521–527 (2012).
15. Wertheim, J. O. *et al.* Discovery of Novel Herpes Simplexviruses in Wild Gorillas,

Bonobos, and Chimpanzees Supports Zoonotic Origin of HSV-2. *Mol. Biol. Evol.* **38**, 2818–2830 (2021).

16. van der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K. & Dalén, L. Historical Genomes Reveal the Genomic Consequences of Recent Population Decline in Eastern Gorillas. *Curr. Biol.* **29**, 165-170.e6 (2019).
17. Beatrix, K. *et al.* Local Virus Extinctions following a Host Population Bottleneck. *J. Virol.* **89**, 8152–8161 (2015).
18. Fontsere, C. *et al.* The genetic impact of an Ebola outbreak on a wild gorilla population. *BMC Genomics* **22**, 735 (2021).
19. Tan, C. C. S., van Dorp, L. & Balloux, F. The evolutionary drivers and correlates of viral host jumps. *Nat. Ecol. Evol.* (2024) doi:10.1038/s41559-024-02353-4.
20. van der Valk, T., Lona Durazo, F., Dalén, L. & Guschanski, K. Whole mitochondrial genome capture from faecal samples and museum-preserved specimens. *Mol. Ecol. Resour.* **17**, e111–e121 (2017).
21. Spyrou, M. A., Bos, K. I., Herbig, A. & Krause, J. Ancient pathogen genomics as an emerging tool for infectious disease research. *Nat. Rev. Genet.* **20**, 323–340 (2019).
22. Furtwängler, A. *et al.* Comparison of target enrichment strategies for ancient pathogen DNA. *Biotechniques* **69**, 455–459 (2020).
23. Moustafa, A. *et al.* The blood DNA virome in 8,000 humans. *PLOS Pathog.* **13**, e1006292 (2017).
24. Wylie, T. N., Wylie, K. M., Herter, B. N. & Storch, G. A. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* **25**, 1910–1920 (2015).
25. Patterson Ross, Z. *et al.* The paradox of HBV evolution as revealed from a 16th century mummy. *PLOS Pathog.* **14**, e1006750 (2018).
26. Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci.* **110**, 15758–15763 (2013).
27. Kapp, J. D., Green, R. E. & Shapiro, B. A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. *J. Hered.* **112**, 241–249 (2021).
28. Gaudin, M. & Desnues, C. Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases. *Front. Microbiol.* **9**, (2018).
29. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
30. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner.
31. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
32. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
33. Guellil, M. MeriamGuellil/aDNA-BAMPlotter: aDNA-BAMPlotter. at https://doi.org/10.5281/zenodo.5676093 (2021).
34. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2009).
35. Fontsere, C. *et al.* Maximizing the acquisition of unique reads in non-invasive capture sequencing experiments. *Mol. Ecol. Resour.* 1755–0998.13300 (2020) doi:10.1111/1755-0998.13300.
36. Hämmerle, M. *et al.* Link between Monkeypox Virus Genomes from Museum Specimens and 1965 Zoo Outbreak. *Emerg. Infect. Dis. J.* **30**, 815 (2024).