AEM Accepts, published online ahead of print on 7 February 2014
Appl. Environ. Microbiol. doi:10.1128/AEM.03242-13
Copyright © 2014, American Society for Microbiology. All Rights Reserved.

AEM Accepts published online ahead of print

1 **Title: Viruses in a 14th-century coprolite**

2 **Running title:** Viruses in a 14th-century coprolite

3

4 Sandra Appelt[1,*], Laura Fancello[1,*], Matthieu Le Bailly[2], Didier Raoult[1], Michel Drancourt[1],

5 Christelle Desnues[†,1]

6

7 [1] Aix Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm 1095, 13385

8 Marseille, France.

9 [2] Franche-Comté University, CNRS UMR 6249 Chrono-Environment, 25 030 Besançon, France.

10 [*] These authors have contributed equally to this work

11 † Corresponding author:

12 Christelle Desnues, Unité de recherche sur les maladies infectieuses et tropicales émergentes

13 (URMITE), UM63, CNRS 7278, IRD 198, Inserm 1095, Faculté de médecine, Aix Marseille

14 Université, 27 Bd Jean Moulin, 13385 Marseille, France. Tel: (+33) 4 91 38 46 30, Fax: (+33) 4

15 91 38 77 72.

16 Email: christelle.desnues@univ-amu.fr

17 **Number of words in Abstract:** 133 words

18 **Number of words in Main Text:** 2538 words

19 **Number of words in Methods:** 954 words

20 **Figures**: 4, Supplementary Figures: 3

21 **Tables**: 0, Supplementary Tables: 6

22 **Keywords:** coprolite, paleomicrobiology, metagenomics, bacteriophages, viruses, ancient DNA

AEM Accepts published online ahead of print

23 **Abstract**

24 Coprolites are fossilized fecal material that can reveal information about ancient intestinal and
25 environmental microbiota. Viral metagenomics has allowed systematic characterization of viral
26 diversity in environmental and human-associated specimens, but little is known about the viral
27 diversity in fossil remains. Here, we analyzed the viral community of a 14th-century coprolite
28 from a closed barrel in a Middle Age site in Belgium using electron microscopy and
29 metagenomics. Viruses that infect eukaryotes, bacteria and archaea were detected, and we
30 confirmed the presence of some of them by *ad hoc* suicide PCR. The coprolite DNA viral
31 metagenome was dominated by sequences showing homologies to phages commonly found in
32 modern stools and soil. Although their phylogenetic compositions differed, the metabolic
33 functions of the viral communities have remained conserved across centuries. Antibiotic
34 resistance was one of the reconstructed metabolic functions detected.

AEM Accepts published online ahead of print

35 **Introduction**

36 Viral metagenomics is a sequencing-based analysis of all of viral genomes isolated from a
37 sample. It has promoted the characterization of viral community diversity. Viral metagenomics
38 has already been successfully applied to the exploration of modern environmental specimens
39 sampled from marine water, freshwater, stromatolites and thrombolites and soil (1-4) and to
40 modern human-associated specimens collected from the liver, blood, nasopharyngeal aspirates
41 and stool (5-9). The DNA viromes generated from modern stools have been demonstrated to be
42 dominated by bacteriophages (10, 11) and to be less diverse than environmental samples (8, 12).

43 Viral metagenomics does not require culturing viruses or *a priori* knowledge of the
44 sequences that will be targeted, which allows for the identification of new, unknown or
45 unexpected viruses and for the global assessment of the virome. Viral metagenomics is thus
46 particularly suitable for paleomicrobiological studies, as little is known about which viruses are
47 characteristic of ancient specimens. Indeed, the majority of ancient DNA (aDNA) studies are
48 based on the analysis of human and bacterial aDNA (13-15), and viral persistence and its
49 detectability in ancient specimens remains unclear. Electron microscopy has previously revealed
50 that viral particles can persist for over 400 years, but their viability was lost (16). Moreover, PCR
51 amplifications yielded positive results for viral aDNA in ancient specimens such as mummified
52 soft tissues, bones and teeth. The amplification products varied between 100 and 570 bp in size,
53 which indicated that viral aDNA can be detected for at least 1,500 years (17-20).

54 Here, we used electron microscopy and, for the first time, viral metagenomics to
55 characterize the viral community of an ancient stool specimen. A viral DNA metagenome was
56 generated from a 14th-century coprolite sample that was recovered from a Middle Age site in
57 Namur (Belgium).

AEM Accepts published online ahead of print

58

59 **Material and Methods**

60 **Virus-like particle isolation, transmission electron microscopy (TEM) and DNA extraction.**

61 First, 5.8 grams of the interior of the coprolite were aseptically removed and solubilized

62 overnight at 4°C under continuous rotation in 40 mL of phosphate saline buffer (PBS), pH 7.4

63 (bioMérieux, Marcy-l'Etoile, France), which had previously been passed through a 0.02-µm

64 filter. The coprolite solution was centrifuged for 10 min at 500 g; then, the upper layer was

65 removed and filtered in stages using sterile Whatman filters (pore sizes: 0.8 µm, 0.45 µm, and

66 0.22 µm, (Whatman Part of GE Healthcare, Dassel, Germany)). Twenty-five milliliters of the

67 coprolite filtrate were used to precipitate and purify viral particles onto a cesium chloride density

68 gradient using ultracentrifugation, and DNAse treatment was then performed (21). A 40-µl

69 aliquot of the purified viral particles was stained with 1.5% ammonium molybdate (Euromedex)

70 and observed by transmission electron microscopy using a Philips Morgagni 268D electron

71 microscope (FEI Co., Eindhoven, Netherlands). To isolate the nucleic acids from the purified

72 viral particles, the formamide procedure previously described by Thurber et al. (21) was used. A

73 standard 18S rDNA PCR was performed to verify the absence of human DNA contamination.

74 **Viral metagenomic library preparation and sequencing.** Nucleic acids were amplified in

75 duplicate reactions using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life

76 Sciences, Freiburg, Germany). Amplification products were pooled and ethanol purified.

77 A shotgun strategy was chosen for high-throughput pyrosequencing on a 454 Life Sciences

78 Genome FLX sequencer using titanium chemistry (Genome Sequencer RLX, Roche).

79 Sequencing was performed using 1/16 of a picotiter plate.

AEM Accepts published online ahead of print

80 **Preprocessing of sequencing data.** The reads were screened for quality using mothur (22). Only
81 reads longer than 50 bp and with an average quality score greater than 21 were kept. Reads with
82 more than two ambiguous base calls and/or reads with homopolymers longer than ten bases were
83 eliminated. Identical sequences artificially generated by the pyrosequencing technology were
84 also excluded using the "unique.seqs" mothur command. The preprocessed viral metagenome is
85 publicly available on the Metavir server (http://metavir-meb.univ-bpcler
86 mont.fr) with the identifier "NAMUR_viral" under the project "HumanCoprolite" and on the
87 NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession number
88 SRP033437.

89 **Annotation of Reads.** A BLASTN search against the non-redundant NCBI database (E-
90 value$<1e^{-05}$) was performed. Reads with no significant similarity to sequences stored in the
91 NCBI database were classified as "unknown reads." The virome taxonomic composition was
92 estimated using GAAS (23), which is based on a BLASTX search against the RefSeq Viral
93 Genomes database (E-value$<1e^{-05}$) and normalizes the number of reads matching each viral
94 genotype by the length of the genome.

95 Functional annotation was performed on the MG-RAST server (24) using the non-redundant
96 SEED database (E-value$<1e^{-05}$). A stringent search of virulence factors was also performed using
97 BLASTX on the Virulence Factor Database (25), with 60% as the minimum identity and a cutoff
98 E-value$<1e^{-05}$.

99 **Assembly and contig annotation.** The reads were assembled into contigs using the Newbler de
100 novo assembler (Roche) with at least 98% identity and 35 bp of overlap. Only contigs longer
101 than 400 bp were used in subsequent analyses.

AEM Accepts published online ahead of print

Known and unknown contigs were identified on the basis of the BLASTN search against the non-redundant NCBI database (E-value $<1e^{-05}$). The taxonomic and functional contig classification was based on a BLASTX search against the non-redundant NCBI database (E-value $<1^{-05}$). A specific search for contigs encoding antibiotic resistance genes was also performed using BLAST on the ARDB (Antibiotic Resistance Genes Database) with an E-value $<1e^{-05}$ (26). Significant hits were manually verified.

Phylogenetic trees. When possible, phylogenetic trees of the contigs encoding antibiotic resistance genes were built. The program Prodigal was used to search for open reading frames (ORFs) in these contigs (27). Homologs to the translated ORFs were searched against the non-redundant NCBI database using BLASTP. A multiple alignment was constructed using MUSCLE (28) and curated using Gblocks (29). The phylogenetic tree was then built using the PhyML algorithm (30) with a bootstrap value of 100. These tasks were all performed using the pipeline freely available on www.phylogenie.fr (31). The trees were visualized using MEGA v.4 (32).

Evidence of temperate bacteriophages. Contigs generated from the assembly were analyzed to search for indicators of temperate bacteriophages, as previously described (33). We searched for three indicators: 1) nucleotide identity to bacterial genomes (BLASTN, E-value $<1e^{-05}$, 90% minimum identity, 90% minimum query coverage), 2) presence of integrase-encoding genes using annotations from the COG and PFAM databases (E-value $<1e^{-05}$) and 3) significant similarity to prophage proteins available on the ACLAME database (BLASTX on the ACLAME prophages database, E-value $<1e^{-05}$). Data were graphically represented using the R package "VennDiagram."

AEM Accepts published online ahead of print

124 **Comparative metagenomics.** The coprolite-associated virome was taxonomically and
125 functionally compared to 21 published viromes of modern stools from healthy adult humans (12,
126 33), which had been generated using MDA amplification, as it is case of the coprolite-virome.
127 All viromes were taxonomically (Genbank database, E-value<1e^{-05}) and functionally annotated
128 (SEED database, E-value<1e^{-05}) using MG-RAST. Annotations were performed on reads using
129 amino acid-level comparisons. The taxonomic and functional virome profiles were compared
130 using principal component analysis on the MG-RAST server (normalized data, Bray-curtis
131 measure of distance). Species richness estimations were obtained from the MG-RAST server.
132 Functional diversity (measured by the Shannon-Wiener index) was calculated using the
133 "estimateDiversity" function of the ShotGunFunctionalizeR package on the SEED-based
134 functional metagenome annotations (E<1e^{-05}) (34).
135 **Specific PCR amplifications and sequencing.** Suicide PCR amplifications were performed to
136 confirm high-throughput pyrosequencing results. To perform suicide PCR, the primer pairs were
137 used only once in working areas, and no positive controls were used (35). For giant virus
138 detection, primer pairs targeting the nonfunctional B-family DNA polymerase were used.
139 Additional primer pairs were designed to specifically target ORFs identified in some viral
140 contigs assembled *de novo* from the virome and matching *Cyanophages*, *Mycobacterium* phages,
141 *Bacillus* phages, *Burkholderia* phages, *Celeribacter* phages and *Clostridium* phages
142 (Supplementary Table S1 and Supplementary Material, Section 4).

143 **Results**

144 The specimen was excavated in 1996 and collected from the interior of a closed barrel, which
145 was commonly used during this period as a pit or latrine (36). The barrel was buried at a depth of
146 3.80 m. The 121.4-g coprolite specimen was dark brown and well preserved under anaerobic

AEM Accepts published online ahead of print

147 taphonomic conditions. Extensive precautions were undertaken to avoid contaminating the
148 coprolite specimen in our laboratory environment; no positive control was used (15) and suicide
149 PCR protocols were applied (35). All negative controls, used in a 1:4 control:specimen ratio,
150 were consistent with current recommendations for paleomicrobiological and paleoparasitological
151 studies (13, 15, 37, 38) and remained negative. Virus-like particles (VLPs) purified from the
152 internal region of the coprolite, after the external layer was removed, were morphologically
153 diverse and varied in size and shape. Oval particles of different lengths (up to 200 nm) and
154 diameters (up to 100 nm), as well as rod-shaped structures (up to 250 nm in length), were
155 observed (Fig. 1A). We identified a VLP with a dense core and a diameter of approximately 150
156 nm, apparently surrounded by an envelope-like structure (Fig. 1B). Viral particles exhibiting
157 characteristics typical of the *Siphoviridae* bacteriophage family (icosahedral head, long tail) were
158 also observed (Figs. 1C-1E).

159 High-throughput sequencing generated 30,654 reads corresponding to approximately 10.8
160 million bp. After quality trimming and duplicate removal, 29,811 reads remained
161 (Supplementary Table S2). The preprocessed read lengths ranged between 77 bp and 574 bp and
162 had an average GC content of 47% (Supplementary Fig. S2). Finally, 41.93% of the reads were
163 assembled into 1,464 contigs that ranged from 421 to 12,500 bp (Supplementary Tables S2). In
164 total, 22.15% of all reads and 17.28% of all contigs were significantly similar to known
165 sequences from public databases (Fig. 2A and Supplementary Table S2). Genome-length
166 normalized counts of viral reads showed that about 85.21% and 0.81% of viral similarities were
167 to double-stranded DNA viruses and single-stranded DNA viruses, respectively (Fig. 2B).
168 Among the double-stranded DNA viral reads we mostly observed *Siphoviridae* (58.89%),
169 *Myoviridae* (8.79%) and *Podoviridae* (5.95%). Overall, we found reads to viral families that can

AEM Accepts published online ahead of print

170 infect eukaryotes (*Ascoviridae*, *Poxviridae*, *Iridoviridae*, *Adenoviridae*, *Mimiviridae*,
171 *Herpesviridae*, *Baculoviridae*, *Polydnaviridae* and *Phycodnaviridae*), archaea (*Lipothrixiviridae*,
172 *Tectiviridae* and *Bicaudaviridae*) and bacteria (*Siphoviridae*, *Myoviridae* and *Podoviridae*) (Fig.
173 2B).

174 Few reads showed similarities with eukaryotic viruses and among them those belonging
175 to *Phycodnaviridae* were the most abundant (0.81%) (Fig. 2B). We also identified a contig
176 encoding a hypothetical protein of invertebrate iridescent virus 3 (IIV-3). IIV-3 is a member of
177 the *Iridoviridae* family, genus *Chloriridovirus*, with a large particle size (180 nm) that infects
178 mosquitoes (Supplementary Table S3). Metagenomic results were confirmed by *ad hoc* suicide
179 PCR (35). In the presence of negative controls, a 167-bp fragment of a *Mimiviridae*-like
180 nonfunctional B-family DNA polymerase was amplified and sequenced, revealing 84% identity
181 to that of the Moumouvirus of the *Mimiviridae* family (GenBank Accession No. GU265560.1).

182 Only a small proportion of viral reads were related to viral families infecting archaea.
183 These families corresponded to *Lipothrixiviridae* (0.04%), *Tectiviridae* (0.11%) and
184 *Bicaudaviridae* (0.02%) (Fig. 2B). One contig was found to have similarity to an environmental
185 *Halophage eHP-6*, an unclassified bacteriophage that infects *Haloarchaea* (Supplementary Table
186 S3).

187 In contrast, the majority of the identifiable viral reads showed homology to genomes of
188 viruses infecting bacteria (bacteriophages), especially those of the genus *Bacillus* (14.08%). We
189 identified reads with homology to genomes of bacteriophages infecting as many as 37 different
190 bacterial genera, including bacterial genera commonly associated with the human gut, such as
191 *Enterobacteria* phages (11.54%), *Lactobacillus* phages (2.23%) and *Lactococcus* phages
192 (2.14%) (Fig. 3). Other findings included reads with similarity to bacteriophages that infect

AEM Accepts published online ahead of print

193 typical soil-dwelling bacteria: *Geobacillus* phages (7.53%), *Streptomyces* phages (3.98%) and
194 *Delftia* phages (0.11%). Several reads were found to show homology to bacteriophages whose
195 bacterial hosts belong to genera that also include human pathogens, such as *Mycobacterium*
196 phages (7.89%), *Vibrio* phages (0.29%), *Pseudomonas* phages (4.01%), *Streptococcus* phages
197 (5.06%), *Staphylococcus* phages (5.07%), *Listeria* phages (3.48%), *Burkholderia* phages (3.38%)
198 and *Clostridium* phages (3.83%) (Fig. 3). The presence of sequences homologous to genes of
199 some of these bacteriophages (*Bacillus*, *Clostridium*, *Mycobacterium* and *Burkholderia* phages)
200 was further supported by contig reconstruction, *ad hoc* PCR amplification and sequencing
201 (Supplementary Tables S1 and S3). Moreover, contigs were found to harbor ORFs with
202 similarity to genes from bacteriophages that are likely to infect hosts known to live in aquatic
203 environments. In particular, we detected contigs matching the tail fiber protein coding gene of
204 Cyanophage S-TIM5, the tape measure protein coding gene of *Planctomyces limnophilus* DSM
205 3776, the gene coding for an unnamed protein product of *Synechococcus* phage S-CB53 and an
206 hypothetical protein coding gene from an uncultured phage identified in a viral metagenomic
207 study of water from the Mediterranean Sea. An ORF encoding a putative phage tail fiber protein
208 of *Celeribacter* phage P12053L was also identified on one of these contigs, amplified by specific
209 PCR, and the 280-bp amplicon was verified by Sanger sequencing. However, the weak
210 similarities shown to some of these highly shared bacteriophage genes and the database bias
211 towards genomes of marine viruses makes it difficult to state if these specific aquatic
212 bacteriophages or other populations of bacteriophages are present in the sample. At last, a 1,939-
213 bp contig matched an unidentified phage previously described in a viral metagenomic study
214 performed on modern human stools (33) (Supplementary Table S3). Only a scaffold is available
215 for the unidentified phage, and the matched protein is annotated as a hypothetical protein;

AEM Accepts published online ahead of print

216 however, this hypothetical protein is predicted to contain a conserved domain corresponding to
217 an N-acetylmuramoyl-L-alanine amidase. This domain is characteristic of autolysins that
218 degrade peptidoglycans and is typically observed in bacteriophage, prophage and bacterial
219 genomes.

220 Evidence of a temperate life cycle for the detected bacteriophage sequences was observed
221 using three indicators (33): 1) nucleotide identity to bacterial genomes (an indicator of prophage
222 formation), 2) the presence of integrase-encoding genes (markers of temperate bacteriophages);
223 3) similarity to prophage sequences available in the ACLAME database (see Materials and
224 Methods). We observed that 329 contigs (22.47%) significantly matched prophage proteins in
225 the ACLAME database; 52 contigs (3.55%) had significant nucleotide identity to bacterial
226 genomes (especially genomes of *Escherichia coli* strains) and 32 contigs (2.18%) harbored
227 integrase-encoding genes (Supplementary Figure S2). This strategy provides only a minimal
228 estimate of the number of temperate bacteriophages, as stated by Minot et al. (33). Overall, 375
229 contigs (25.61%) presented at least one of the three indicators and could be tentatively classified
230 as temperate bacteriophages (Supplementary Fig. S2).

231 The coprolite-associated DNA virome was compared to the viromes of 21 modern human
232 stool specimens (Fig. 4). At the taxonomic level, the coprolite virome did not group with modern
233 stool viromes, whereas it was functionally more similar to some of the modern stool samples
234 (Fig. 4). Overall, the coprolite virome displayed higher species richness (315.279) and seemed to
235 be more functionally diverse (average Shannon-Wiener index of 4.8693) than modern stool
236 viromes (average species richness of 77.824 and average Shannon-Wiener index of 4.1264)
237 (Supplementary Table S4). A more extensive functional analysis of the assembled contigs
238 revealed that most of the identifiable ORFs harbored by these contigs coded for genes involved

AEM Accepts published online ahead of print

239 in DNA metabolism (n=80), as is typical of viromes, and virulence genes (n=87). The most
240 abundant virulence genes were those involved in resistance to antibiotics and toxic compounds.
241 In particular, a contig encoding a chloramphenicol O-acetyltransferase gene that mediates
242 chloramphenicol resistance was observed. This gene was found to belong to *Chryseobacterium*
243 sp., and the BLAST-based annotation was further confirmed by a phylogenetic tree constructed
244 from the ORF of this contig (Supplementary Fig. S3). To further investigate the presence of
245 virulence genes, virome reads were examined using the Virulence Factor Database, which
246 includes both conventional factors directly involved in the pathogenesis and factors important to
247 establishing infection. A stringent search allowed the identification of 166 reads encoding
248 virulence factors. In particular, virulence factors of the bacterial genera *Escherichia* (n=42),
249 *Salmonella* (n=39) and *Shigella* (n=34) were observed (Supplementary Table S5). A pathway-
250 centric analysis based on COG annotation revealed that virulence (defense mechanisms) was
251 overrepresented in the coprolite compared to modern stools. Other differences included an
252 overrepresentation of lipid transport and metabolism, fatty acid biosynthesis and amino acid
253 transport and metabolism (Supplementary Table S6). Indeed, 12 ORFs on annotated contigs were
254 found to encode genes involved in lipid metabolism, in particular fatty acid biosynthesis (n=3),
255 glycerolipid and glycerophospholipid metabolism (n=3), isoprenoid metabolism (n=3) and
256 polyhydroxybutyrate metabolism (n=3). Annotated contigs also contained 36 ORFs encoding
257 functions related to the metabolism of amino acids, especially lysine, threonine, methionine and
258 cysteine (n=11) and arginine, urea and polyamines (n=10).
259
260 **Discussion**

AEM Accepts published online ahead of print

261 We report the first metagenomic analysis of an ancient human DNA virome. The use of
262 viral metagenomics allowed us to perform a systematic research of known and unknown viruses
263 without *a priori* targeting of expected viruses.

264 Because minimizing contamination is vital in paleomicrobiology, extensive precautions
265 established by previously published recommended protocols were implemented to avoid
266 contamination of the coprolite specimen (13, 15, 37, 38). The coprolite studied here was
267 recovered from a sealed barrel that was still intact at the time it was found, suggesting that the
268 coprolite was protected from contamination by environmental material for centuries. Only the
269 internal region of the coprolite was used in our experiments. We ascertained the presence of
270 viruses by three independent approaches, *i.e.*, electron microscopy, metagenomics and suicide
271 PCR. The PCR amplification product sequences were original, *i.e.*, they had not been previously
272 observed in our laboratory, and all negative controls remained negative.

273 The viral metagenome was generated using a multiple-displacement amplification of viral
274 genomes *via* the phi29 polymerase. This method is known to preferentially amplify circular and
275 single-stranded DNA (12). To potentially minimize this bias, a duplicate amplification reaction
276 was performed as previously suggested (21).

277 The majority of the generated metagenomic sequences were of unknown origin. The
278 taxonomic composition of the generated virome was estimated on the basis of the identifiable
279 viral sequences (known sequences) of the DNA virome. The known sequences corresponded to
280 DNA viruses that infect eukaryotes, bacteria and archaea. Eukaryotic and archaeal viral
281 sequences were detected only at low abundances, and their presence was supported by contig
282 recovery or confirmed by *ad hoc* suicide PCR. The majority of the identifiable sequences
283 recovered from the coprolite corresponded to bacteriophages, especially *Siphoviridae*. Contigs

AEM Accepts published online ahead of print

284 with significant similarity to characteristic bacteriophage genes were identified, such as genes
285 coding for structural proteins (tail fiber proteins, capsid proteins) as well as proteins involved in
286 replication (DNA polymerase) or DNA packaging (terminase). We could also identify reads with
287 significant similarity to virulence factors associated to pathogenic bacteria and genes from
288 bacteriophages that infect bacteria belonging to genera that include mammalian pathogens. These
289 findings are consistent with those obtained for the previously generated bacterial metagenome
290 associated to the coprolite (39). Comparative analyses to previously published viromes show that
291 modern human stool viromes do not group with the coprolite virome at the taxonomic level. All
292 previous works on viral communities associated with the stool of healthy individuals showed a
293 high prevalence of bacteriophages, in particular double-stranded DNA bacteriophages of the
294 *Siphoviridae* family (8, 11, 33) or single-stranded DNA bacteriophages of the *Microviridae*
295 family (10, 12), with high inter-individual variability. Accordingly, the coprolite virome shows a
296 high prevalence of *Siphoviridae* (8, 11, 33). Moreover, as in modern stools, we found evidence
297 for temperate bacteriophages (10, 33). However, we did not observe significant abundance of
298 single-stranded DNA viruses (10, 12) or the same most abundant prophages identified in other
299 modern stool viromes (33). At the functional level, no clear separation can be observed between
300 the coprolite virome and modern stool viromes, and functions might be conserved between the
301 coprolite and some modern stool samples. This finding is consistent with those of a recent study
302 that demonstrated that despite inter-individual taxonomic variability, the metabolic profile was
303 significantly conserved within viromes from the same ecological niche (40). This persistence of
304 metabolic functionalities across centuries may reinforce the crucial role of the viral community
305 in the human gastrointestinal tract.

AEM Accepts published online ahead of print

306 Finally, the coprolite virome is more functionally diverse and rich in virulence genes than
307 modern stool samples viromes. One contig encoding a gene for chloramphenicol resistance (the
308 chloramphenicol O-acetyltransferase), a broad-spectrum antibiotic that inhibits bacterial protein
309 synthesis, was identified. The presence of antibiotic resistance genes in viral metagenomes has
310 been reported in modern human stools (33). Indeed, bacteriophages constitute a reservoir of
311 resistance genes (41-43) and bacteriophage transduction represents one important mode of lateral
312 transfer of resistance genes between bacterial species. Phylogenetic studies have demonstrated
313 that the evolution and dissemination of resistance genes started well before the use of antibiotics
314 (44-46). Accordingly, direct evidence for the presence of antibiotic resistance genes in pre-
315 antibiotic era specimens was provided by *ad hoc* PCR amplifications using DNA extracted from
316 30,000-year-old permafrost sediments in Canada (47). Here, we demonstrate that bacteriophages
317 are an ancient reservoir of resistance genes associated with human samples that date back as far
318 as the Middle Ages. Moreover, we provide evidence for the lysogenic lifestyle of these
319 bacteriophages, which may support their role in the mobilization and lateral transfer of genes in
320 bacterial communities.
321 Overall, this study furthers our understanding of past viral diversity and distribution and
322 promotes the further exploration of ancient viral communities using coprolite specimens.
323
324 **Acknowledgments**
325 We thank Sonia Monteil Bouchard and Catherina Robert for technical assistance. C.D. and L.F.
326 were funded by a Starting Grant n°242729 from the European Research Council to CD.
327 The authors declare no competing interests.
328

AEM Accepts published online ahead of print

329 **References**

330 1. **Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM,**
331 **Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R,**
332 **Rayhawk S, Suttle CA, Rohwer F.** 2006. The marine viromes of four oceanic regions.
333 PLoS Biol **4:** e368
334 2. **Lopez-Bueno A, Tamames J, Velazquez D, Moya A, Quesada A, Alcami A.** 2009.
335 High diversity of the viral community from an Antarctic lake. Science **326:** 858-61
336 3. **Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M, Liu H,**
337 **Furlan M, Wegley L, Chau B, Ruan Y, Hall D, Angly FE, Edwards RA, Li L,**
338 **Thurber RV, Reid RP, Siefert J, Souza V, Valentine DL, Swan BK, Breitbart M,**
339 **Rohwer F.** 2008. Biodiversity and biogeography of phages in modern stromatolites and
340 thrombolites. Nature **452:** 340-3
341 4. **Kim KH, Chang HW, Nam YD, Roh SW, Kim MS, Sung Y, Jeon CO, Oh HM, Bae**
342 **JW.** 2008. Amplification of uncultured single-stranded DNA viruses from rice paddy
343 soil. Appl Environ Microbiol **74:** 5975-85
344 5. **Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G,**
345 **Khristova ML, Weyer J, Swanepoel R, Egholm M, Nichol ST, Lipkin WI.** 2009.
346 Genetic detection and characterization of Lujo virus, a new hemorrhagic fever-associated
347 arenavirus from southern Africa. PLoS Pathog **5:** e1000455
348 6. **Anderson NG, Gerin JL, Anderson NL.** 2003. Global screening for human viral
349 pathogens. Emerg Infect Dis **9:** 768-74
350 7. **Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N,**
351 **Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R,**

AEM Accepts published online ahead of print

352 **Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T.** 2009. Direct
353 metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased
354 high-throughput sequencing approach. PLoS One **4:** e4219
355 8. **Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, Salamon P, Rohwer F.** 2003.
356 Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol
357 **185:** 6220-3
358 9. **Breitbart M, Rohwer F.** 2005. Method for discovering novel DNA viruses in blood
359 using viral particle selection and shotgun sequencing. Biotechniques **39:** 729-36
360 10. **Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI.** 2010.
361 Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature **466:**
362 334-8
363 11. **Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM,**
364 **Mueller J, Nulton J, Rayhawk S, Rodriguez-Brito B, Salamon P, Rohwer F.** 2008.
365 Viral diversity and dynamics in an infant gut. Res Microbiol **159:** 367-73
366 12. **Kim MS, Park EJ, Roh SW, Bae JW.** 2011. Diversity and abundance of single-stranded
367 DNA viruses in human feces. Appl Environ Microbiol **77:** 8062-70
368 13. **Pääbo S, Poinar H, Serre D, Jaenicke-Despres V, Hebler J, Rohland N, Kuch M,**
369 **Krause J, Vigilant L, Hofreiter M.** 2004. Genetic analyses from ancient DNA. Annu
370 Rev Genet **38:** 645-79
371 14. **Willerslev E, Cooper A.** 2005. Ancient DNA. Proc Biol Sci **272:** 3-16
372 15. **Drancourt M, Raoult D.** 2005. Palaeomicrobiology: current issues and perspectives. Nat
373 Rev Microbiol **3:** 23-35

AEM Accepts published online ahead of print

374 16. **Marennikova SS, Shelukhina EM, Zhukova OA, Yanova NN, Loparev VN.** 1990. Smallpox diagnosed 400 years later: results of skin lesions examination of 16th century Italian mummy. J Hyg Epidemiol Microbiol Immunol **34:** 227-31

377 17. **Bedarida S, Dutour O, Buzhilova AP, de Micco P, Biagini P.** 2011. Identification of viral DNA (Anelloviridae) in a 200-year-old dental pulp sample (Napoleon's Great Army, Kaliningrad, 1812). Infect Genet Evol **11:** 358-62

380 18. **Biagini P, Theves C, Balaresque P, Geraut A, Cannet C, Keyser C, Nikolaeva D, Gerard P, Duchesne S, Orlando L, Willerslev E, Alekseev AN, de Micco P, Ludes B, Crubezy E.** 2012. Variola virus in a 300-year-old Siberian mummy. N Engl J Med **367:** 2057-9

384 19. **Li HC, Fujiyoshi T, Lou H, Yashiki S, Sonoda S, Cartier L, Nunez L, Munoz I, Horai S, Tajima K.** 1999. The presence of ancient human T-cell lymphotropic virus type I provirus DNA in an Andean mummy. Nat Med **5:** 1428-32

387 20. **Sonoda S, Li HC, Cartier L, Nunez L, Tajima K.** 2000. Ancient HTLV type 1 provirus DNA of Andean mummy. AIDS Res Hum Retroviruses **16:** 1753-6

389 21. **Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F.** 2009. Laboratory procedures to generate viral metagenomes. Nat Protoc **4:** 470-83

391 22. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75:** 7537-41

AEM Accepts published online ahead of print

396 23. **Angly FE, Willner D, Prieto-Davo A, Edwards RA, Schmieder R, Vega-Thurber R,**
397 **Antonopoulos DA, Barott K, Cottrell MT, Desnues C, Dinsdale EA, Furlan M,**
398 **Haynes M, Henn MR, Hu Y, Kirchman DL, McDole T, McPherson JD, Meyer F,**
399 **Miller RM, Mundt E, Naviaux RK, Rodriguez-Mueller B, Stevens R, Wegley L,**
400 **Zhang L, Zhu B, Rohwer F.** 2009. The GAAS metagenomic tool and its estimations of
401 viral and microbial average genome size in four major biomes. PLoS Comput Biol **5:**
402 e1000593
403 24. **Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T,**
404 **Rodriguez A, Stevens R, Wilke A, Wilkening J, Edwards RA.** 2008. The
405 metagenomics RAST server - a public resource for the automatic phylogenetic and
406 functional analysis of metagenomes. BMC Bioinformatics **9:** 386
407 25. **Chen L, Xiong Z, Sun L, Yang J, Jin Q.** 2012. VFDB 2012 update: toward the genetic
408 diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Res **40:**
409 D641-5
410 26. **Liu B, Pop M.** 2009. ARDB--Antibiotic Resistance Genes Database. Nucleic Acids Res
411 **37:** D443-7
412 27. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal:
413 prokaryotic gene recognition and translation initiation site identification. BMC
414 Bioinformatics **11:** 119
415 28. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high
416 throughput. Nucleic Acids Res **32:** 1792-7
417 29. **Talavera G, Castresana J.** 2007. Improvement of phylogenies after removing divergent
418 and ambiguously aligned blocks from protein sequence alignments. Syst Biol **56:** 564-77

419 30. **Guindon S, Delsuc F, Dufayard JF, Gascuel O.** 2009. Estimating maximum likelihood
420 phylogenies with PhyML. Methods Mol Biol **537:** 113-37
421 31. **Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF,**
422 **Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O.** 2008. Phylogeny.fr: robust
423 phylogenetic analysis for the non-specialist. Nucleic Acids Res **36:** W465-9
424 32. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary
425 Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24:** 1596-9
426 33. **Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD.**
427 2011. The human gut virome: inter-individual variation and dynamic response to diet.
428 Genome Res **21:** 1616-25
429 34. **Kristiansson E, Hugenholtz P, Dalevi D.** 2009. ShotgunFunctionalizeR: an R-package
430 for functional comparison of metagenomes. Bioinformatics **25:** 2737-8
431 35. **Raoult D, Aboudharam G, Crubezy E, Larrouy G, Ludes B, Drancourt M.** 2000.
432 Molecular identification by "suicide PCR" of Yersinia pestis as the agent of medieval
433 black death. Proc Natl Acad Sci U S A **97:** 12800-3
434 36. **Rocha GCd.** 2003. Praça das Armas, Namur, Bélgica. Contribuição de um estudo
435 paleoparasitológico. Escola Nacional de Saúde Pública-Fiocruz, Rio de Janeiro, Rio de
436 Janeiro. 142 pp.
437 37. **Cooper A, Poinar HN.** 2000. Ancient DNA: do it right or not at all. Science 289: 1139
438 38. **Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S.** 2001. Ancient DNA. Nat Rev
439 Genet **2:** 353-9
440 39. **Appelt S, Armougom F, Le Bailly M, Robert C, Drancourt M.** 2013. Polyphasic
441 analysis of a Middle Ages coprolite microbiota, Belgium. unpublished data.

AEM Accepts published online ahead of print

AEM Accepts published online ahead of print

442 40. **Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, Tammadoni S,**
443 **Nosrat B, Conrad D, Rohwer F.** 2009. Metagenomic analysis of respiratory tract DNA
444 viral communities in cystic fibrosis and non-cystic fibrosis individuals. PLoS One **4:**
445 e7370

446 41. **Muniesa M, Garcia A, Miro E, Mirelis B, Prats G, Jofre J, Navarro F.** 2004.
447 Bacteriophages and diffusion of beta-lactamase genes. Emerg Infect Dis **10:** 1134-7

448 42. **Colomer-Lluch M, Jofre J, Muniesa M.** 2011. Antibiotic resistance genes in the
449 bacteriophage DNA fraction of environmental samples. PLoS One 6: e17549

450 43. **Mazaheri Nezhad Fard R, Barton MD, Heuzenroeder MW.** 2011. Bacteriophage-
451 mediated transduction of antibiotic resistance in enterococci. Lett Appl Microbiol **52:**
452 559-64

453 44. **Hall BG, Barlow M.** 2004. Evolution of the serine beta-lactamases: past, present and
454 future. Drug Resist Updat **7:** 111-23

455 45. **Garau G, Di Guilmi AM, Hall BG.** 2005. Structure-based phylogeny of the metallo-
456 beta-lactamases. Antimicrob Agents Chemother **49:** 2778-84

457 46. **Aminov RI, Mackie RI.** 2007. Evolution and ecology of antibiotic resistance genes.
458 FEMS Microbiol Lett **271:** 147-61

459 47. **D'Costa VM, King CE, Kalan L, Morar M, Sung WW, Schwarz C, Froese D, Zazula**
460 **G, Calmels F, Debruyne R, Golding GB, Poinar HN, Wright GD.** 2011. Antibiotic
461 resistance is ancient. Nature **477:** 457-61

462

463

AEM Accepts published online ahead of print

464 **Figure Legends**

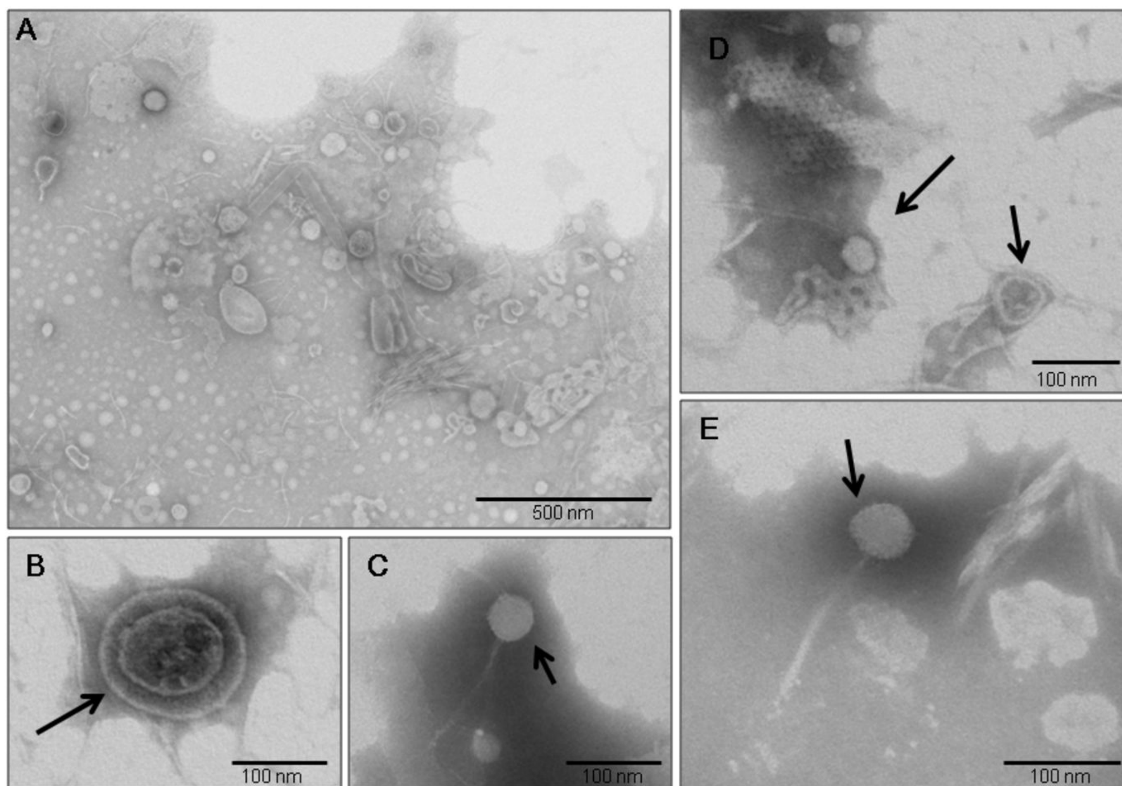
465 **FIGURE 1. Transmission electron microscopy of negatively stained viral particles.** (**A**)
466 Overview of stained viral particles, which vary in size and shape, isolated from the Middle Age
467 coprolite. (**B**) A representative virion and (**C-E**) viral-like particles with icosahedral
468 nucleocapsids and a long filament tail characteristic of *Siphoviridae* bacteriophages.

469 **FIGURE 2. (A) The proportion of known and unknown reads (in percent).** Reads were
470 defined as "unknown" if they lacked homology to the non-redundant NCBI database according
471 to a BLASTN search (E-value<1e-05) and as "known" otherwise. **(B) The relative abundance**
472 **of viral families.** The relative abundance of identified viral families was estimated using the
473 GAAS software.

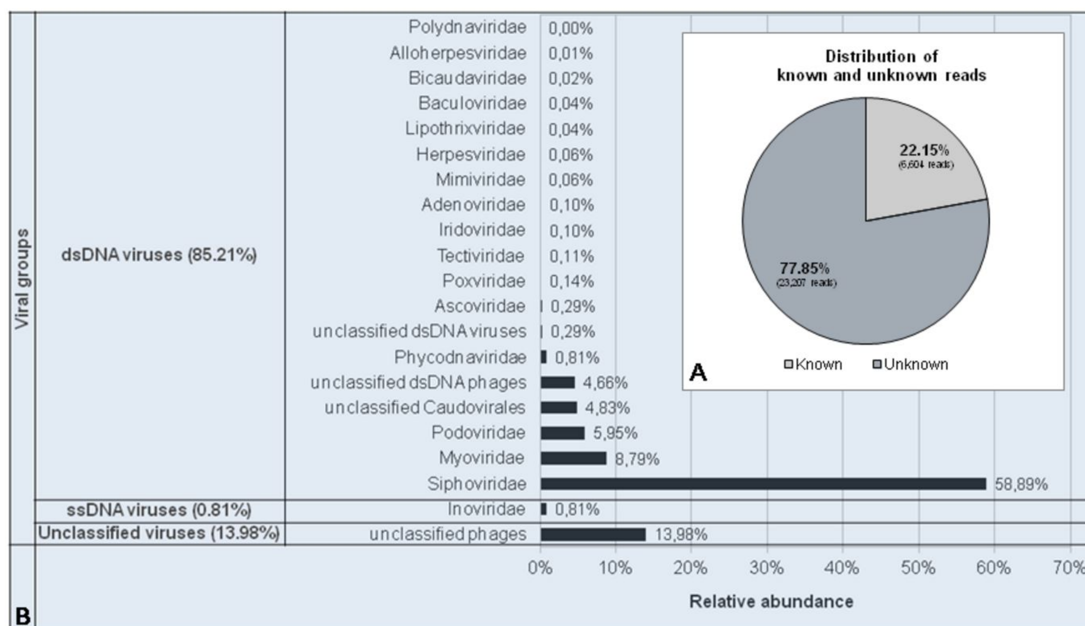
474 **FIGURE 3. Relative abundance of hits to known Bacteriophages.** The relative abundance of
475 hits to known bacteriophages were estimated using the GAAS software. The hosts of the
476 bacteriophages that were also identified in a previous study on the bacterial community
477 associated with this specimen (unpublished data (39) are marked with a red point.

478 **FIGURE 4. Comparison between the modern human stool viromes and the coprolite**
479 **virome.** Principal component analysis was used to compare the viral metagenomes associated
480 with the coprolite (highlighted in red) to those associated with modern human stool samples (S1-
481 S21) at the taxonomic (**A**) and functional (**B**) levels.

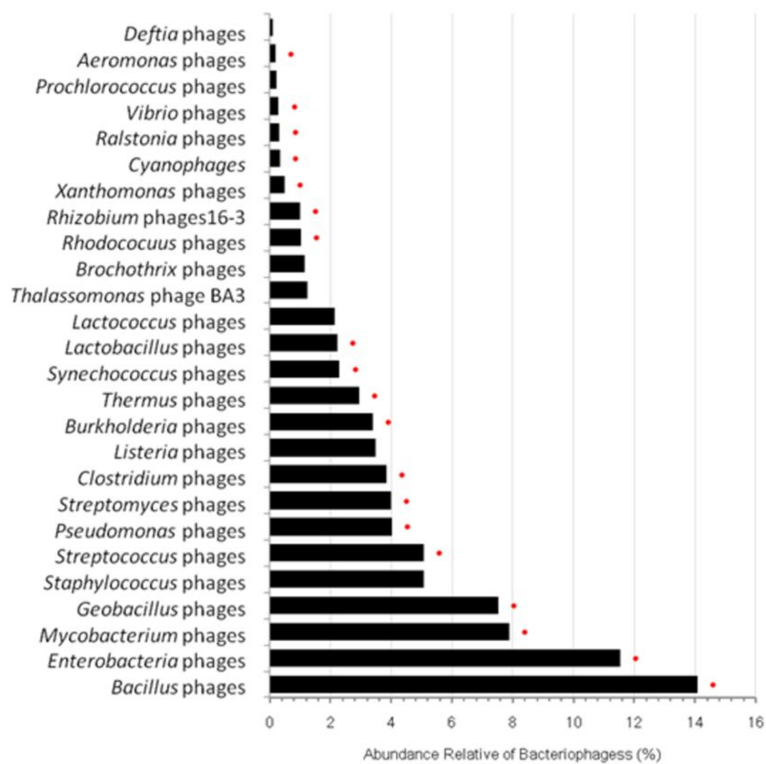
AEM Accepts published online ahead of print



AEM Accepts published online ahead of print



AEM Accepts published online ahead of print



AEM Accepts published online ahead of print