



The ecogenomics of dsDNA bacteriophages in feces of stabled and feral horses

V.V. Babenko^{a,1}, A. Millard^{b,1}, E.E. Kulikov^c, N.N. Spasskaya^d, M.A. Letarova^c, D.N. Konanov^a, I.S. Belov^c, A.V. Letarov^{c,e,*}

^a FSC Physico-chemical Medicine FMBA, Russia

^b Dept Genetics and Genome Biology, University of Leicester, UK

^c Winogradsky Institute of Microbiology RC Biotechnology RAS, Moscow, Russia

^d Zoology Museum, Faculty of Biology, Lomonosov Moscow State University, Russia

^e Faculty of Biology, Lomonosov Moscow State University, Russia



ARTICLE INFO

Article history:

Received 7 August 2020

Received in revised form 30 October 2020

Accepted 30 October 2020

Available online 10 November 2020

Keywords:

Intestinal viromes

dsDNA bacteriophages

Horse feces

Equus caballus

Bacteriophage metagenomics

Bacteriophage diversity

ABSTRACT

The viromes of the mammalian lower gut were shown to be heavily dominated by bacteriophages; however, only for humans were the composition and intervariability of the bacteriophage communities studied in depth. Here we present an ecogenomics survey of dsDNA bacteriophage diversity in the feces of horses (*Equus caballus*), comparing two groups of stabled horses, and a further group of feral horses that were isolated on an island. Our results indicate that the dsDNA viromes of the horse feces feature higher richness than in human viromes, with more even distribution of genotypes. No over-represented phage genotypes, such as CrAssphage-related viruses found in humans, were identified. Additionally, many bacteriophage genus-level clusters were found to be present in all three geographically isolated populations. The diversity of the horse intestinal bacteriophages is severely undersampled, and so consequently only a minor fraction of the phage contigs could be linked with the bacteriophage genomes. Our study indicates that bacteriophage ecological parameters in the intestinal ecosystems in horses and humans differ significantly, leading them to shape their corresponding viromes in different ways. Therefore, the diversity and structure of the intestinal virome in different animal species needs to be experimentally studied.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The existence of microbial populations inhabiting different niches of the human and other animal bodies was first observed by Antony van Leeuwenhoek in 17th century [1] and has since become a commonly accepted paradigm that is mentioned in almost any microbiology textbook. Remarkable progress has been made in this field over the last 15 years due to the introduction of the culture-independent tools for the analysis of the composition and function of the microbial component of human [2] or animal holobiont [3]. Much emphasis has been placed on the gut microbiome, representing the largest microbial community associated with human or other mammalian body. The intestinal microbiome is now considered as a “new organ”, exerting strong and

multifaceted influence over the physiology of the macro-host [4]. The gut microbiome is involved in the pathology of numerous conditions, including Crohn disease [5–6], obesity [7–8], some types of cancer [9] and even behavioral alterations [10–12].

It is well established that in all vertebrate animals the intestinal microbiome is associated with the corresponding virome the community of viruses infecting or produced by the microorganisms comprising the microbiome [13–16] and by the viruses of the macro-host and its eukaryotic symbionts and parasites, such as protozoa or helminthes. However, in most cases the eukaryotic viruses represent only minor components of the gut viromes, the bulk of which are comprised of bacteriophages [13–14,17]. Nevertheless, intestinal viral communities are also believed to be involved in multiple physiological effects and pathological processes via altering the composition and activity of the microbial community, and through direct interaction with the macro-host tissues and immune system [14,18–20].

Bacteriophage diversity, biogeography and dynamics in the human gut has been investigated in depth in numerous studies

* Corresponding author at: Winogradsky Institute of Microbiology RC Biotechnology RAS, pr. 60-letiya Oktyabrya 7 bld. 2, 117312, Moscow, Russia

E-mail address: letarov@gmail.com (A.V. Letarov).

¹ These authors contributed equally to this work.

using metagenomic approaches [16–17,21]; see also reviews [17,22–23]. In the majority of studies the viral community of the feces was used as a proxy of the intestinal viromes. The “normal” composition of the human fecal bacteriophage community has been established and the “core phageome” composition defined as bacteriophage genotypes present in more than 50% individuals worldwide was evaluated [24]. The first identified core phage lineage, named CrAssphage, that is highly prevalent in some samples (up to 90% of the viral reads) was initially identified using bioinformatic approaches [25] and was later cultured and shown to be a large podovirus infecting *Bacteroides* [26].

Therefore, the main characteristics of healthy human viromes have been established as follows: a diverse community, that is highly stable in time [17,21], highly individual with larger inter-individual distances compared to different time points [16–17,27] even if the dietary interventions were applied [21]. Human viromes are suggested to be dominated by temperate bacteriophages [27] although the prevalence of the contigs containing integrases or site-specific recombinases genes was found to vary greatly (0–68%) between individual viral metagenomes [28].

Despite significant progress in the understanding of bacteriophage ecology in the human gut, data on other animal species are scarce. Although a significant number of metagenomic datasets from various species have been published [29–32], the vast majority of these studies focus on the detection and interpretation of animal viruses, and bacteriophages have not been given significant attention. Only a few studies give emphasis to bacteriophage diversity in these samples, although this is limited to identifying differences between healthy and diseased states in rhesus monkeys [33–34] or specifically focusing on the diversity of ssDNA viruses [35].

In this work we focus on the ecogenomics of dsDNA viruses present in horse feces. The equine intestinal microbiome plays an essential role in animal nutrition, allowing the horse to digest cellulose, which is the major component of the grass consumed [36]. In contrast to ruminants where microbial cellulose digestion occurs in the forestomach (rumen), in horses the cellulolytic microbial community develops in the cecum and large intestine that have a cumulative volume of ~100 L with a food retention time of approximately 48–72 h [37]. The large intestine content is not subjected to any subsequent digestion (such as in ruminants) and is pushed by peristalsis into the rectum where it is subjected to partial dehydration to form the feces [38]. The average time intervals between food intake or between defecation acts in horses are much shorter than the indicated retention time [39]. Therefore, the horse large intestine functions as a natural chemostat with highly stable physical and chemical conditions and fairly constant flow through.

Adult horses do not show any coprophagy, but at the same time they do not avoid contact with feces of other individuals or fecally contaminated objects [39], potentially enhancing the exchange of bacteria and viruses between the individual viromes.

Only a few studies have been dedicated to horse intestinal bacteriophages. A limited Sanger-based metagenomic analysis of a single sample allowed to estimate the richness of the viral community, as about 1200 bacteriophage genotypes [40]. A more recent metagenomic study compared fecal microbiomes and viromes of cattle and horses held on the same farm [41]. However, the raw data from this study are not published and the amount of data for each virome was limited with no information concerning the specific characteristics of the bacteriophage communities of the samples reported.

There have also been several studies of the horse gut bacteriophage community based on other approaches. In a limited electron-microscopy study of horse feces almost all VLPs identified were classified as tailed phages, with 69 morphologically distinct types reported out of ~200 particles measured, indicating a high

level of diversity (see [13] for review of earlier work). A comprehensive study of coliphage diversity and dynamics [42] in the feces of four horses held in the same location suggested high prevalence of virulent coliphages. The *E. coli* host population was found to be highly divergent, and represented by hundreds of strains simultaneously present in the same sample. The overlap of the sensitivity of these strains to co-occurring bacteriophages was limited [42–43] with ~1–5% of the total *E. coli* counts being suitable hosts for any particular phage isolate. The data of Golomidova et al. [42,44] and the results of the longitudinal study of G7C-related bacteriophages persistence and evolution within the ecosystem of a horse stable [45] indicated the flow of the coliphage genotypes between the animals. However, *E. coli* and its phages make only a minor fraction of the total equine microbiome and virome respectively. Currently, to the best of our knowledge, no methods exist that allow the translation of the findings made using this model system to the total community.

Here we present the ecogenomics of horse fecal dsDNA viromes of three separate horse populations including two groups of stabled horses and one herd of feral horses isolated on an island. Our data indicate that equine intestinal viromes are highly diverse communities dominated by the tailed bacteriophages. Although the site of sampling or/and the life conditions of distinct populations have marked influence over the composition of individual viromes, it was possible to identify the equine intestinal core virome.

2. Results

2.1. The sampling strategy, workflow and sequencing results

To characterize the virome composition and diversity in horse feces, we collected samples from three populations of horses in Russia. These included two groups of stabled animals and one feral population. The stable 1 (S1) population was kept at the equestrian center in the city of Moscow. The stable 2 (S2) population was located in the country side ~90 km from Moscow. The lifestyle and diet of these two populations differed significantly (see material and methods). In addition, we sampled the feral horse population in Rostovsky national reserve, that have been isolated on the island for several decades (see material and methods for detail) – population F. In population S1 we sampled four animals, two of which were sampled twice. In the S2 population five animals were sampled a single time (October 2018). In the F population, six animals (two harem stallions and two mares belonging to each of the stallions) were sampled in both May and October 2018 (Table S1). In all cases the samples were processed immediately after their delivery to our laboratory facility.

The viromes were extracted from all samples: viral DNA was extracted and sequenced using IonTorrent technology. Sample metadata and sequencing output values are given in Table S1. It is important to note that the procedure applied for virome isolation (Fig. 1) did not include any gradient centrifugation or ultrafiltration steps that may selectively remove some types of viral particles. We also did not use DNA amplification to avoid the biased representation of sequences that can occur. Based on previously published research [40,46], the bulk of the horse intestinal virome is composed of tailed bacteriophages, so we decided to focus on dsDNA viruses.

To check for bacterial contamination both virome QC and sort-meRNA were used, and both methods suggest the samples were highly enriched for viral DNA with minimal bacterial contamination. A total of 8097 nonredundant viral contigs >5 kb were identified, and were used for all further analysis. Among these contigs,

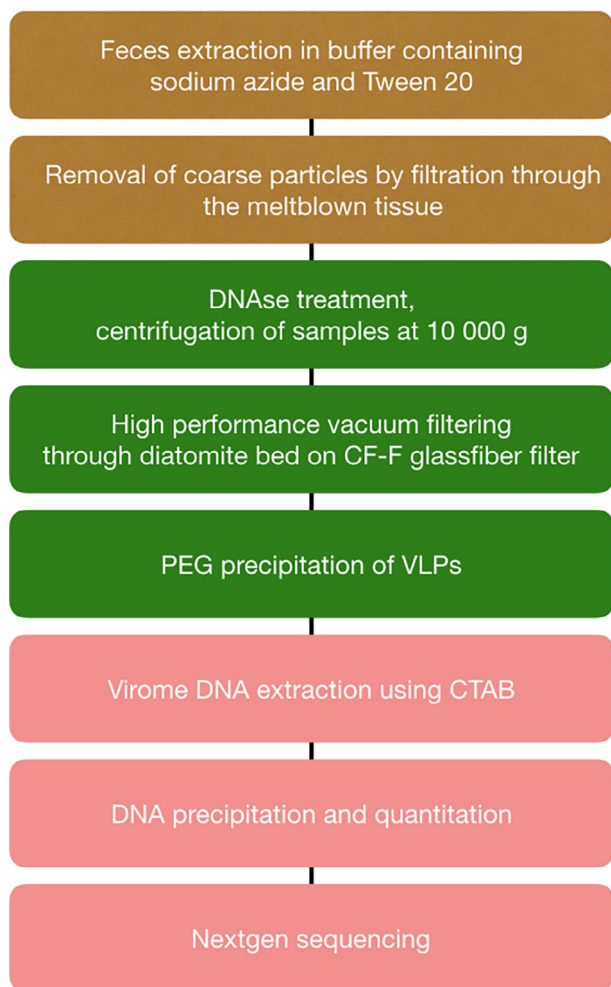


Fig. 1. Sample processing workflow.

we identified 46 contigs longer than 30 kb that may represent complete or almost complete phage genomes.

2.2. Complexity of the individual viromes

To estimate the alpha-diversity Shannon and Simpson indexes were calculated from the relative abundance using the R phyloseq package [47] and revealed that individual virome diversity tends to be higher in feral horses than in stabled (population S1). The population S2 falls in between, being closer to the feral populations (Fig. 2. and Table S2). The Shannon and Simpson indexes take into account both richness and evenness of the community composition [48] but they hardly can be translated into the absolute values of these parameters and are used mostly for comparison between different populations.

To estimate the richness in a more direct way we used the approach mimicking that of Torsvik et al. [49] to estimate bacterial population richness in a soil sample. These authors estimated the complexity of the bacterial DNA extracted from soil using DNA re-association kinetics measurements. Knowing the average bacterial genome size, these authors calculated the approximate number of unique bacterial genotypes present. Instead of experimental determination of the viral DNA re-association we computed the plots of the cumulative read recruitment against the cumulative length of the contigs ranked by abundance (CPM) in the given sample (Fig. 2).

The samples appear to differ by the abundance of the most prevalent viral genotypes. This is reflected by different slopes of the initial rise of the curves, though after this initial rise, the curves are almost parallel, indicating the similar law of the genotypes abundance distribution in the viromes of different animals belonging to different population. Noteworthy, in none of the samples could we detect an over-represented genotype. If we estimate an average phage genome as 50–100 kbp, the top represented 20–40 genotypes would account for 2 Mbp of the cumulative non-redundant DNA sequence. This value corresponds to 1.3–8.6% of the total amount of the viral DNA in different samples (Fig. 2). We were not able to reveal the law of the distribution of the genotypes fractions within the communities studied and, therefore, we did not find any reliable function to extrapolate the curves outside of the available data interval. However, to estimate the lower limit of the richness we used the function $f(x) = ax^b \log(cx + 1)$, where x is the cumulative length of the contigs, $f(x)$ is the fraction of reads recruited by the most covered contigs with the cumulative length x , and a , b and c are the parameters fitted to minimize the square deviation from the experimental curves. As shown in the Fig. 2A, within the existing data range the modeled curves run always higher than the experimental curves. Therefore, if the distribution law remains the same, the real x values corresponding to any (x) threshold chosen will be higher than the values predicted by the function. The calculated lowest estimates of the non-redundant length of the genomes sequences of the phage particles comprising 50% of total community for most (20 out of 24) of the curves were in the range 10^9 – 10^{11} b.p. This translates into 10^4 – 10^6 distinct bacteriophage genotypes without taking into consideration of possible overlap of the sequences in many different but still related viral genomes.

2.3. Composition of horse intestinal virome

Having established a high level of diversity of the equine gut dsDNA bacteriophage community, we asked how related are these numerous viral genotypes to known viruses. First, we attempted direct classification of the filtered reads using centrifuge software [50]. However, only ~0.1% of reads can be classified in this way. Out of them, 97% matched viruses and 94% could be classified as dsDNA containing viruses, 81% of which were assigned to the order *Caudovirales* (tailed phages). Of these 43% were *Siphoviridae*, 41% *Myoviridae* and 13% *Podoviridae* (see Supplementary file Fig S1 for the interactive Krona plot analysis). However, given only a minor fraction of the reads could be classified, all the subsequent analysis was performed on the assembled contigs.

Initially we used pVOGs [51] to annotate all predicted proteins on viral contigs, with a simple scoring matrix. Out of 8097 contigs, 7483 (92%) had at least one pVOG detected. Only seven contigs where pVOGs were detected, were found to have a pVOG not found in the order *Caudovirales*. Thus, further suggesting that the vast majority of contigs originated from tailed phages. We then analyzed the relatedness at the protein level using vCONTACT2, including RefSeq genomes plus other available phage genomes at the time (May 2019). The horse virome contigs were spread across 1156 viral clusters (VCs), but only 31 were found in VCs that contain a known bacteriophage reference sequence, allowing classification at the genus level (Fig. 3, Fig S2). A further 2873 virome contigs were singletons, once more highlighting the diversity of phages present.

Due to the inability to link the majority of contigs to any known phage at the subfamily or genus level, we manually inspected the 10 largest contigs that belonged to 10 different VC clusters. Proteins were analysed with both BLASTp and HHpred [52–53] along with gene order and orientation in the genomes. We confirmed that even for the large (35–65kbp) contigs the links to the known

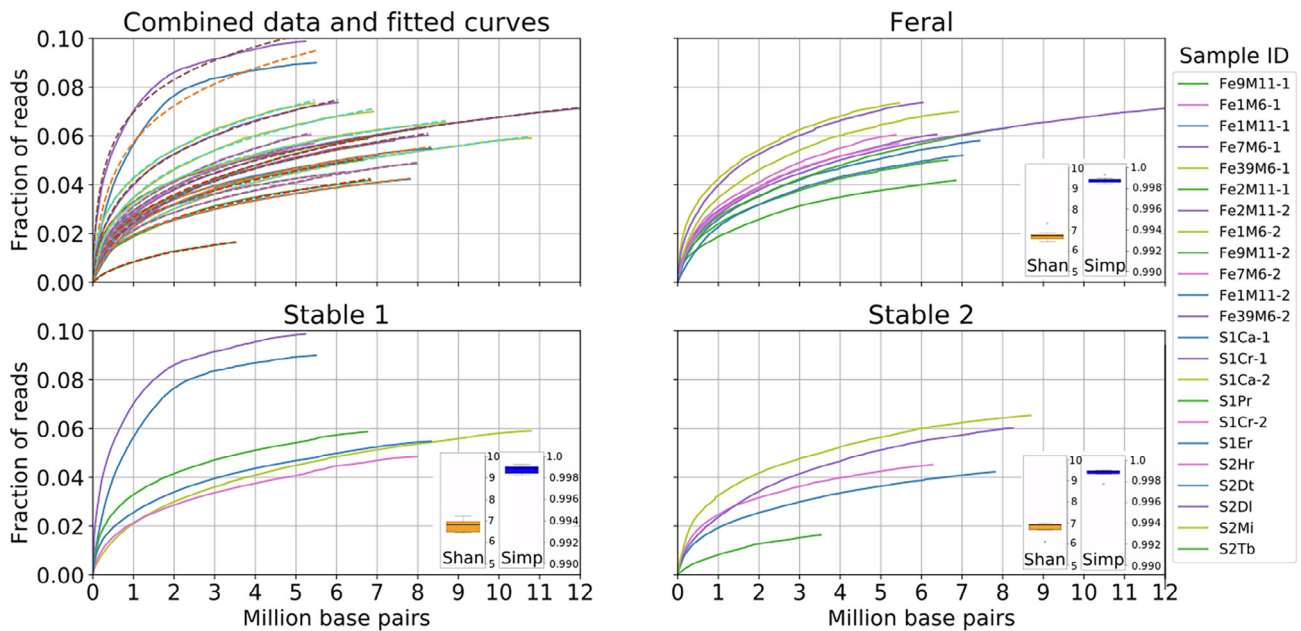


Fig. 2. Individual viromes richness. The curves represent the cumulative fraction of the reads recruited by the contigs plotted against the cumulative length of the contigs. Contigs were sorted by CPM, taken as a proxy for their relative abundance. The dotted lines on the panel with the all the samples combined indicate the modelling function fitted to each of the curves. The distribution of the Shannon and Simpson diversity indexes in the samples from each population are plotted as box and whisker plots as inserts of respective panels.



Fig. 3. Network graph produced using vCONTACT 2. Red dots – bacteriophage reference sequences, retrieved from GenBank, green dots – the horse viromes contigs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

viral genomes were barely detectable and lie beyond the genus or subfamily level (Table S3, Fig. S3) at which vCONTACT2 is able to operate. Only in one case (the contig 070k255_67966) were dis-

tant, but reliable relationships to the known *Gordonia* phage Gravy discovered, which was not in the vCONTACT2 database at the time of analysis. The results of the manual analysis further confirmed the viral diversity of the horse gut is to date very poorly sampled.

We also detected pVOGs that may be considered markers of the temperate lifestyle (transposase, integrase, recombinase, resolvase and excisionases; see Material and methods section for detail). At least one of such pVOGs was detected in 462 (5.7%) of the 5 K contigs (Table S5). Among the contigs detected in the individual samples, the highest prevalence of the temperate lifestyle markers was observed in the population S1 (7%), in the populations F and S2 the prevalence of the “temperate” contigs was about 4%. Taking into consideration that the mean length of viral contigs was 8.3 kbp, the average number of the temperate lifestyle markers per genome is three, and estimating the average length of a temperate phage genome as 40–100 kbp, we can estimate the prevalence of the phage genotypes carrying these markers as 10–30%.

2.4. Variability of fecal viromes between individuals and between populations

To compare samples at the read level, we computed the Jaccard's distances (also referred to as mash distances [54]) between the datasets using mash (Fig. 4). The samples clustered according to the populations they were collected from. Interestingly, the feral horses cluster tighter than the stable animals within the populations S1 and S2. Simultaneously, we did not observe any cluster formation according to the social (harem) groups of population F. Interestingly, in population F samples collected at the different time points from the same animal were closer to each other than to any sample from the other animals. This remarkable stability of the individual viromes was observed despite the fact that between two sampling points the animals lived out a very hot summer. The hot summer resulted in severe water deprivation because the debit of the water hole was decreased about two times for several months because the hole was blocked by sand (it was cleaned by the rangers in late September). The clustering of

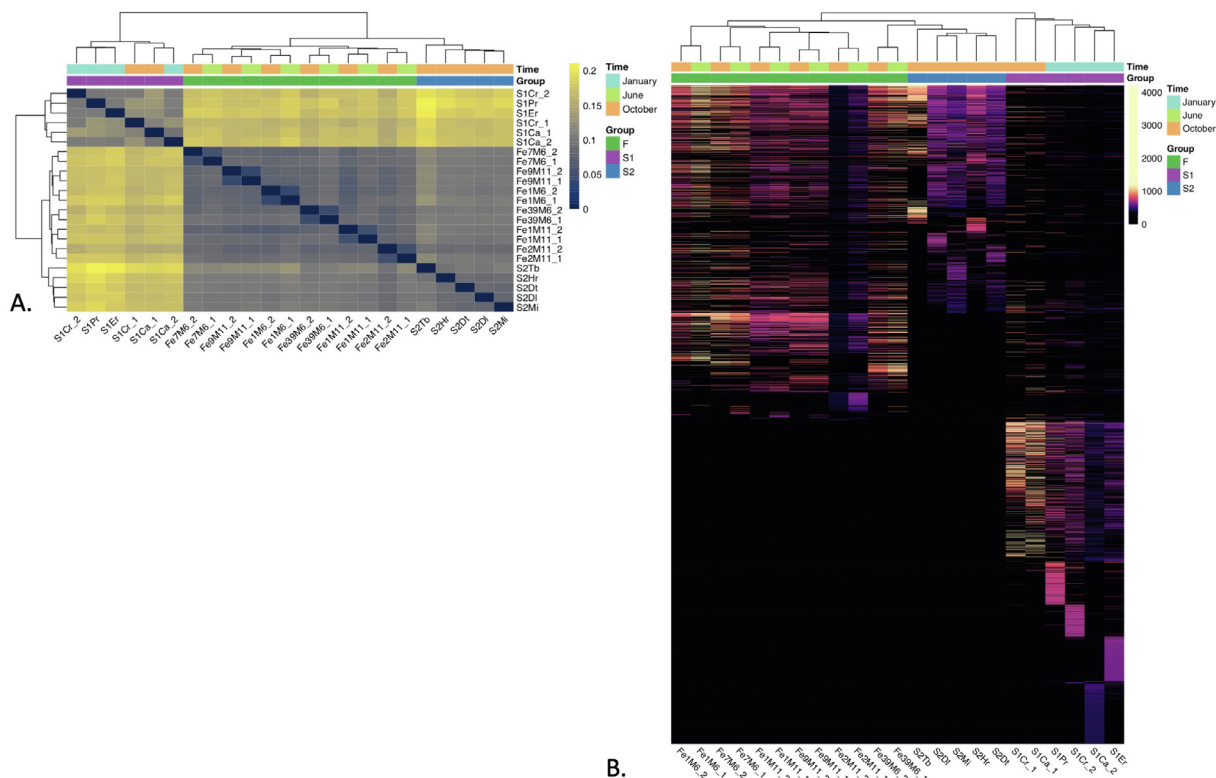


Fig. 4. Individuality of horse fecal viromes. A – Jaccard's distances between the samples. B. Abundance of contigs in different samples.

samples collected from the same animal at different time points was not observed in the population S1. However, in this population the period between sequential sampling was longer (Table S1). The population S2 appears much closer to the population F. This may reflect the fact that the diet of these two populations is much closer to each other than to the population S1.

To compare the individual samples and populations at the contig level we mapped reads from each sample against viral contigs. Contigs were considered present in a sample if the contig had ≥ 1 x coverage $\geq 75\%$, when mapping at 95% identity. Contig abundance was normalized, for both contig length and depth of sequencing. Thus, we used “counts per million” (CPM) value as a proxy of contig abundance [55]. An average of 913 contigs (range 655–1105, Table S4) were detected per sample.

The heatmap of the contig abundance in the samples is shown in Fig. 4. One can see that many contigs are shared by the animals belonging to the same group but much fewer are shared between the animals. The existence of the core-virome, defined as the assortment of the viral lineages present in the majority of the sequenced samples, was recently demonstrated for human feces [24]. The crAssphage-like viruses that were shown to be highly abundant in some of the samples [24–25,56] also belong to the human feces core-virome. In order to reveal a possible horse core-virome we identified the contigs detected in all samples or in more than half of the animals (in the cases when two time points were available for an animal, the detection of a contig in at least one of them was counted as positive). These criteria were applied for each population to reveal the local core-viromes, and for all the samples to retrieve the universal core-virome.

Venn diagrams of the local core viromes relatedness are shown in Fig. 5. Only one contig was omnipresent in all samples, however 192 contigs were shared by all three populations, among them 14 contigs were simultaneously present in 50% of samples in each population (Fig. 5). Given the F population was completely isolated

from S1 and S2 by a distance of ca. 1500 km and protection by a national reserve regimen (indirect exchange of viruses between the animal of the populations S1 and S2 is also highly unlikely though could not be completely excluded), these 192 contigs can be considered as potential candidates for a “equine core viromes”. At the same time no contig exhibited abnormal coverage comparable to the values reported for human crAssphage [25]. The largest fraction of a single contig in the sum of all CPMs of all contigs of a sample was 0.006 (range 0.001–0.006). So, no equine analog of the over-represented and wide spread crAssphage group was detected in our dataset.

The distribution of VCs between the populations (Fig. 5B) revealed more of commonality between locations studied. Out of 1156 VCs, 262 (23%) VCs were detected in all three populations. This equates to 34–40% of VCs present in any of these populations (the VC was considered as present in a population if at least 1 contig belonging to this VC passed the detection criteria for at least 1 of the samples from this population). Interestingly, in each population many VCs were present in $\geq 50\%$ of samples (Fig. 5C). The fractions of such prolific VC are larger in the populations F and S2 (89% and 90%) compared to S1 (23%). The fraction of contigs present in $\geq 50\%$ of samples in each population are smaller (13%, 11% and 2% for the populations F, S2 and S1 respectively) with only 30 contigs present in 50% in all three populations simultaneously (Fig. 5D). Thus, viromes appear highly individual at the level of the viral genotypes, but they appear to consist of similar sets of bacteriophage genera.

2.5. Host-phage relationships

The high prevalence of common VCs in three geographically isolated populations may reflect the presence of similar bacterial groups in the gut of horses belonging to different populations. To test this hypothesis, we performed sequencing of bacterial 16S

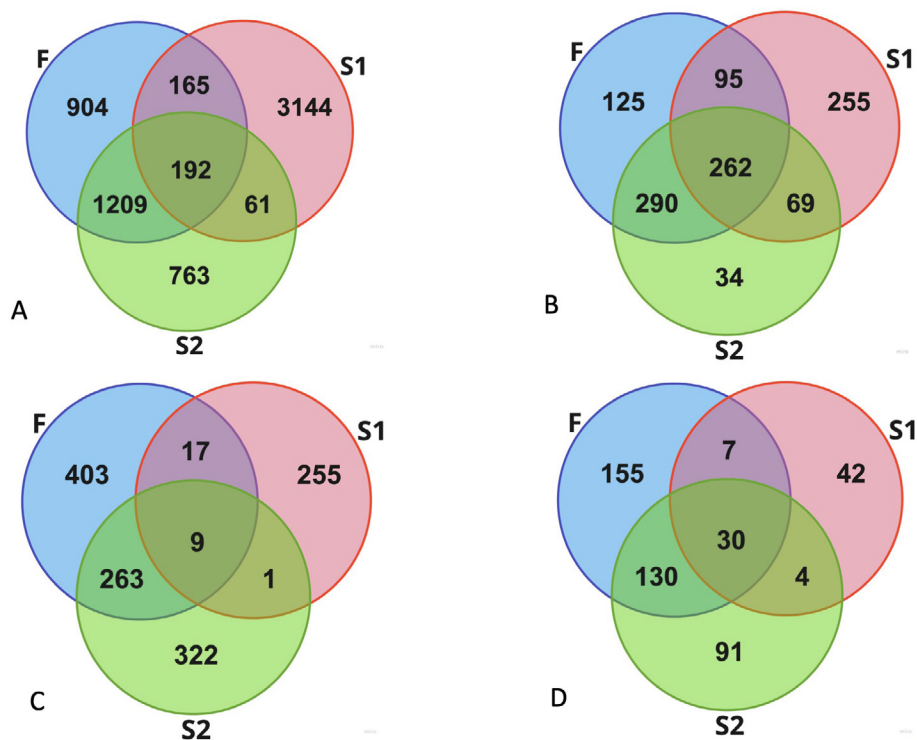


Fig. 5. Composition of viromes in the different horse populations. A – distribution of the contigs detected in three populations. B – distribution of the VC clusters between the populations. C – distribution of the VCs, detected in 50% or more of the samples at least in one of three populations. D – distribution of the contigs, detected in 50% or more of the samples at least in one of three populations.

rRNA genes libraries for all samples. We also predicted putative hosts for viral contigs using WiSH [57]. The prevalence of the bacterial genus level OTUs and the prevalence of the contigs predicted to belong to bacteriophages infecting these host groups is shown in Fig. 6 and Table S6.

The samples clustered according to 16S pattern still reflects the location (Fig. 6). At the same time, the distribution of the prevalence of the phage genotypes predicted to infect different hosts was close to uniform. Noticeably the list of the genus-level bacterial OTUs and the list of bacterial genera predicted to be the hosts of the bacteriophage contigs only partially overlap. However, the non-overlapping OTUs have low prevalence as inferred from 16S sequencing data or from the statistics of the prevalence of the contigs allocated to the particular host.

3. Discussion

Although the viral component of the intestinal microbiome is now widely believed to be an important factor in both shaping the microbial community of the gut and mediating its interactions with the macro-host [14,18–20]. The main component of the gut viromes – the community of tailed bacteriophages has only been comprehensively studied in humans. The results of our study provide a basic understanding of the composition of the dsDNA viromes of one more mammal species – *Equus caballus*. Given the long history of domestication [58], until recently the integral involvement of domestic horses in almost all spheres of business and military activity, and the multifaceted influence of the relationship with this species on human culture [58]. It would not be an exaggeration to say that this animal species is the second most important in the development of our civilization after *Homo sapiens*.

The equine intestinal bacterial community has been extensively studied, and it is now considered to influence horse organism

homeostasis and health, to the same extent as bacteria in the human gut (reviewed in [36,59], see also [60–61]). The equine gut bacterial community is involved in the pathology of specific diseases such as equine metabolic syndrome and laminitis [62]. With horse behavior also suggested to be influenced by the gut bacterial community [63].

The existing electron-microscopy based observations [42] and sequencing data [41] suggest that the main component of the horse viromes are the tailed bacteriophages (order Caudovirales). Therefore, in this work we focused on the dsDNA virome, although in humans in some individuals the ssDNA bacteriophages comprise a significant fraction [17] and the investigation of the diversity of these viruses in horses may be also of interest in future.

Our data confirms that tailed bacteriophages (order Caudovirales) comprise the majority of total dsDNA viromes, so for brevity we use here below the term “virome” to describe the community of dsDNA containing bacteriophages.

An individual horse virome appears highly diverse, including more than 2000 viral genotypes (the extrapolation of the curves Fig. 2 gives estimates of more than 10^4 – 10^6 viral genotypes per sample). The viromes richness did not differ much between the samples analyzed. In contrast to human viromes where the contig number per sample had varied for more than three orders of magnitude [17], in our samples the variation was limited to a factor of less than two.

At the same time the evenness of the viral genotype abundance was much higher in horses as could be seen by a comparison of Shannon and Simpson diversity indexes and also inferred from the reads recruitment analysis (Fig. 2). No analogue of human crAssphage that is hyper-dominant in some human gut samples, was observed. The most prolific 20–40 viral genotypes, between them only accounted for 1.3–8.6% of the total number of reads in all the samples. This makes a striking contrast to the situation described for human viromes where ~2% of the contigs that com-

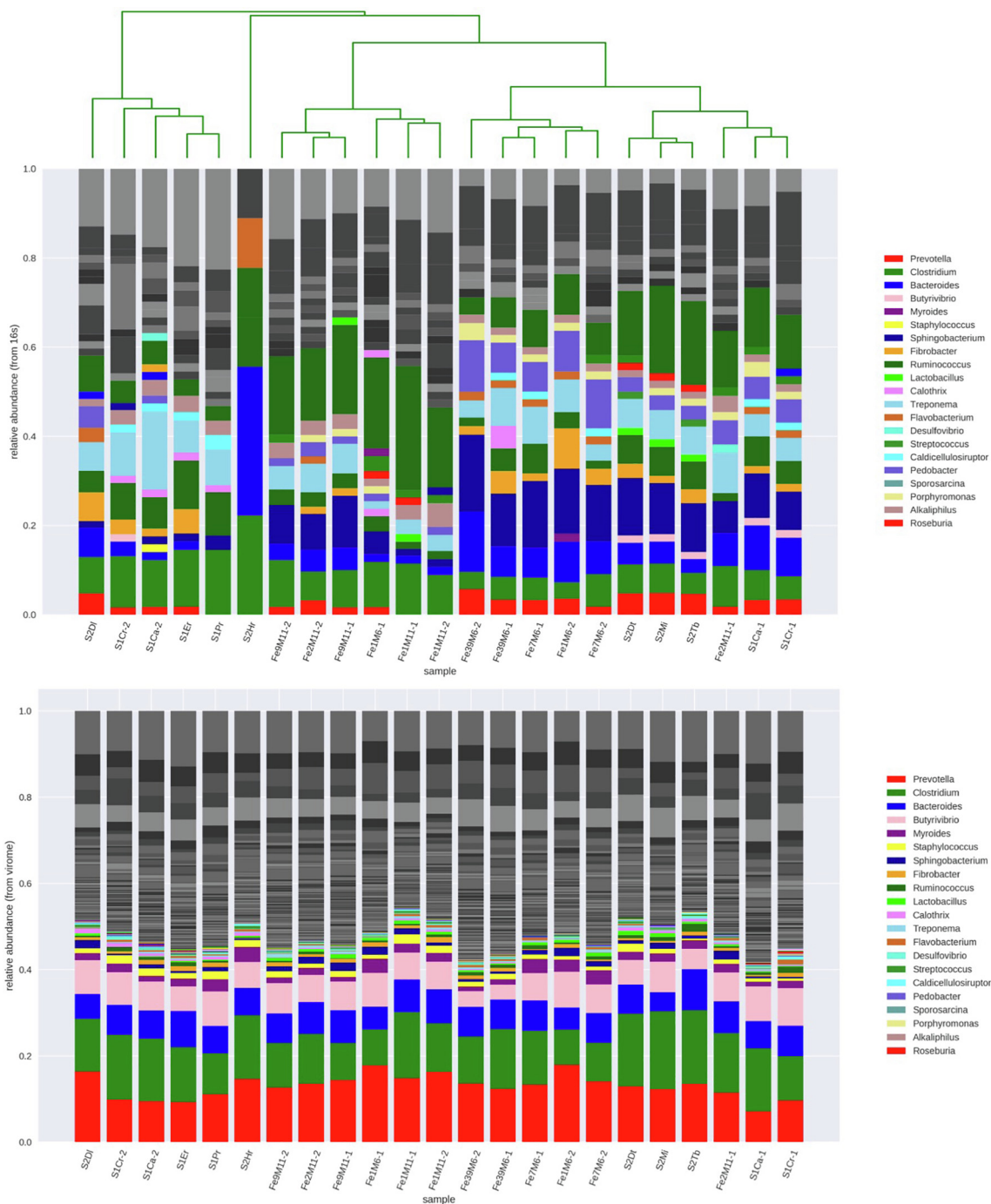


Fig. 6. Comparison of the samples by abundance of genus-level 16S bacterial OTUs (top) and by predicted host of phage contigs. Only the OTUs or genera overlapping between the 16S sequencing results and viral host prediction are shown in color. Other OTUs are shown in gray scale.

prised so called persistent personal viromes recruited 92.3% of VLP reads per sample [17].

Most of the bacteriophage genotypes comprising the bulk of equine intestinal dsDNA viromes are unrelated or very distantly related to known phage genotypes. Only 31 out of 1152 identified VCs simultaneously contained horse virome contigs and known bacteriophage genomes. Moreover, manual analysis of the 10 largest contigs, did not allow (with a single exception) to assign these sequences to any known tailed phage group. Interestingly, we estimated that only 10–30% of this immense phage diversity are represented by temperate bacteriophages. These values are in marked contrast to human fecal viromes that are dominated by

temperate bacteriophages [27,64]. Noteworthy, the high prevalence of virulent bacteriophages in horses is in agreement with previous data on the diversity of coliphages isolated from horse feces [42–13]. The coliphages isolated from the human feces were reported to be mainly temperate [13,64]. At the same time, high richness and high evenness of the viral community as well as the lack of the correlation between the abundance of the host 16S-based OTUs and abundance of the predicted phages to these hosts (Fig. 6) indicate that the community most probably contains multiple viruses for many (if not for all) the bacterial species present in the samples. This pattern may support the elevated diversity at the strain level, might be maintained by a kill-the-winner type

mechanism [65]. Our metagenomic data does not provide any direct estimates of bacterial diversity and/or phage-host relationships at the strain level. However, high intraspecies diversity of *E. coli* within horse feces associated with high diversity of co-occurring coliphages, having relatively narrow host ranges, was previously demonstrated using culture-based approaches [42–43]. Additional culture-based evaluation of the strain-level diversity of a more prevalent species than *E. coli* combined with characterization of its co-occurring phages, may help to shed more light over the pattern of the phage-host relationships in the horse gut ecosystem.

Despite the observed high virome diversity, our data suggest that a healthy horse (in the feral population the animals without visible abnormalities, wounds and marked anomalies were considered as healthy) intestinal virome includes a certain number of conserved components. The human core-virome was defined by [24] as a set of viral genotypes present in more than 50% of human fecal viromes. However, the limited amount of data (22 samples from 14 animals) makes this criterion less useful for evaluation of our data. At the same time, we may benefit from the known history of strict isolation of the population of the feral horses preserved on an island in Rostovski national reserve (population F) from any contact with other horses. The factor of geographical isolation makes direct transfer of the viral genotypes even between the ancestors of these animals over the last 80–100 years unlikely. Nevertheless, we could find a significant number of bacteriophage genotypes present in all three populations. Approximately 3% (192 out of 6438) of contigs detected in the samples were universally present in all three locations. At the higher taxonomic level 262 of 1130 VCs (approximately corresponding to the genus or subfamily level of relatedness) detected in the individual samples were present in all populations (Fig. 5). Moreover, despite the long history of isolation, populations S2 and F shared 552 of 875 VCs. Thus, a significant fraction of bacteriophage OTUs of species or genus level are widely present in the horse intestinal viromes, but the fractions of these common taxa may vary significantly. Higher similarity of the fecal viromes compositions of the populations F and S2 compared to their distance to the population S1 (Figs. 4 and 5) may be explained by similar diets (grass only or grass and forages compared to grass and grain diets). Given the fact that the distribution of abundances of the viral genotypes in the individual viromes is even (Fig. 2 and Table S2), such variations may obscure the commonality of the viromes composition because many shared components are present below current limits of detection. At the same time, increasing the detection sensitivity using less stringent criteria may lead to a high frequency of false detection of viral genotypes. The deep sequencing of several viromes from different location using, for example, high output Illumina sequencing and combined with long-reads single-molecule based sequencing (e.g. Oxford nanopore) may allow characterization of the repertoire of the core components of equine virome. It is logical to expect that some endemic bacteriophage genotypes may also exist in certain populations, especially in the isolated animal groups, such as the population F. However, high diversity of the viromes does not allow the identification of them at the metagenome sequence coverage levels achieved in this work.

Another remarkable feature of the horse viromes is revealed by clustering of the individual viromes compositions according to the sampling site (Fig. 4). The tightest cluster was formed by the samples from the population F. The clustering of these samples did not reflect the social structure of the herd. Only the samples taken from the same animals always clustered together.

Significant fractions of contigs and VCs were found in at least 50% of the samples in each population, however the percentage of the shared contigs and VCs was lower in the group of the horses stabled in the city equestrian centre (S1). These findings may be

explained by the significant exchange of viral genotypes between the animals. Horses in stable 1 (S1) are held in individual boxes, which is much more restrictive of behaviour facilitating virus exchange through a fecal-oral route (see [39]). In the population from stable 2 (S2) during the summer season, horses spend most of their time at the pasture, and in the feral population (F) they have no human-imposed restrictions at all. During our field work we regularly observed behaviour that may allow viral exchange (for example, during the spring time large groups of horses take the mud baths in the freshwater pools, were some animals may defecate and from which they also may drink), though detailed recording of the behavior falls out the scope of this work. In such conditions, the level of all-to-all exposure may erase the signal from tighter contacts within a harem group.

The transfer of phage genotypes between horses was earlier observed by the detection of the highly related coliphage isolates that could be obtained from multiple animals held in the same stable but could not be discovered in other locations [45,66–67]. These observations are in good agreement with the metagenomic data indicating that the transfer of the phages between the individual viromes is not limited to the minor viromes fraction(s) such as coliphages. So, the individuality and stability of the intestinal viromes are less pronounced in horses compared to humans [16–17,27].

Summarizing all the data, we conclude that horse intestinal viromes appear to be a more open ecological system than has been inferred from human viromes. The bacteriophages of equine intestinal viromes represent a large pool of novel viral groups including the high level taxa such as families or subfamilies (we mean here new contemporary understanding of phage families, not the old *Siphoviridae* – *Myoviridae* – *Podoviridae* grouping within *Caudovirales*). This work provides an essential starting point from which the full genetic diversity for phages can be explored using long-read sequencing and culture based methods. Additional work is also required to analyze temporal stability of the horse viromes. It also has to be mentioned that because of the limited size of our sample series in this study the observed features of horse viromes should not be interpreted as eco-physiological norm for the horses in general. However, our data suggest that each of the three populations were adequately sampled, and therefore wider biogeographical study with many samples from different location analyzed may better allow for more reliable generalization of the conclusions than the increase in the number of the samples from the same populations.

4. Material and methods

4.1. The horse populations and sampling

The sampling was performed in three geographically separated horse populations. The stable 1 population (S1) was located in a children's equestrian club in Detski park Fili, Moscow, Russia, and represents typical stabled horses. These animals are kept in the boxes and taken outdoors for a limited time to be exercised (1–4 h per day) and to have a rest (ca. 2 h). These animals diet is typical for sportive horses and is composed of forages, supplemented with, oats, offal and carrots. The animals have an *ad libitum* access to water. The population from stable 2 (S2) was a group of horses living in a stable located in the village of Tretyakovo, Klinski district of Moskovskaya oblast, Russia. Horses are stabled in boxes and fed by forages (carrots or apples are occasionally given to them), but they spend 8–16 h (depending on the season) per day in the field where they can graze. Access to water is not limited for this population, with water provided twice a day during dry weather, where the horses can drink *ad libitum*. In the population S2 four animals were sampled only once (in October 2018). Thus,

the living conditions and diet of these three populations represent almost the whole spectrum of the living conditions of the horses as a species.

The herd of feral horses (F population) inhabiting the island Vodny in the salty lake Manych-Gudilo in Rostovskaya oblast, Russia. This island belongs to the core part national Reserve “Rostovski” and therefore no business activity is allowed there, and the visits to the island are restricted. The horses on the island Vodny do not receive any feeding from humans and their diet is only the grass they can forage, their access to drinking water is variable. From late autumn to spring, the animals can drink from the pools or consume snow; the grass is also frequently wet due to rain and/or to dew-fall. In summer, water is piped from the terrestrial beach once per day, and limited (ca. 3 L per min) output to an old water hole that exists in the middle of the island [68]. The water from the holes is slightly saline. The access to these water sources differs significantly for different animals depending on their ranks in the social groups (harems) and on the rank of the harem stallion of their group among the harem stallions of the herd (at the moment of sampling in May and October 2017 there were 17–18 harem groups and a group of bachelor stallions). Samples of freshly voided feces were collected immediately after natural defecation and placed in sterile plastic containers. The containers were placed on ice and transported to the laboratory. The samples from the populations S1 and S2 were processed within 24 h, the samples from the population F – within 72 h.

4.2. Extraction of the viromes, DNA isolation and sequencing

The viromes were extracted as it described in [69] with minor modifications. Briefly, 10 g of the fecal sample was suspended in 100 ml of the extraction buffer (0.2 M NaCl, 0.1 mg/ml Na₃, 1 mg/ml Tween 20 (Sigma-Aldrich, USA)) and extracted on the planetary shaker at 200 rpm, room temperature for 4 h. The coarse material was then filtered out using meltblown tissue (Miracle cloth meltblown fabric), then pelleted by centrifugation at 10,000 g for 15 min. The supernatant was carefully separated. The samples were filtered through a combined filter composed of Whatman GF-F glass fiber paper and a layer of diatomite (Hyflo Super-Cel). The DNase was added to the filtrate up to 0.01 mg ml⁻¹ and the filtrate was incubated for 1 h at room temp. The virus-like particles were then PEG-precipitated by adding dry NaCl to 0.6 M and dry PEG 6000 (Panreac), dissolving both on orbital shaker (45 min) and allowing precipitate to form in a refrigerator (+4°C, 5–6 days). Brown-greenish precipitates containing VLPs were directly extracted with CTAB using a protocol described in [70].

Ion proton shotgun sequencing metavirome DNA (approx. 1000 ng) was fragmented to a mean size of 200–300 bp using the Covaris S220 System (Covaris, Woburn, Massachusetts, USA). Then, an Ion XpressPlus Fragment Library Kit (Life Technologies) was employed to prepare a barcoded shotgun library. Emulsion PCR was performed using the OneTouch system (Life Technologies). Beads were prepared using the One Touch 2 and Template Kit v2, and sequencing was performed using Ion Proton 200 Sequencing Kit v2 and the P1 Ion chip. The reads were deposited to Sequence read archive (SRA) database, the accession numbers are given in the Table S1.

4.2.1. 16S rRNA genes libraries sequencing

For 16S sequencing the total DNA was extracted from the feces (the samples saved at -70 °C were used). A 16S rRNA sequencing library was constructed according to the 16S metagenomics sequencing library preparation protocol (Illumina, San Diego, CA, USA) targeting the V3 and V4 hypervariable regions of the 16S rRNA gene using the Nextera XT Index Kit (Illumina) with a dual indexing strategy. Samples were sequenced on the MiSeq sequenc-

ing platform (Illumina) using a 2 × 250 cycle V2 kit, following standard Illumina sequencing protocols. For 16S analysis we used greengenes (version gg_13_5) database and Megan5 software.

4.3. Bioinformatic analysis

Prior to assembly reads were quality controlled by trimming with Sickle [71] with default settings. Contaminating by horse DNA was removed by mapping all reads to EquCab3.0 (GCA_002863925.1) as reference genome, using bbmap with the following settings ‘minid = 0.95’ with any reads that mapped removed prior to assembly [72]. Mapping suggested minimal contamination of horse DNA with the highest percentage of reads that mapped from any library at 0.07%. Metagenomes were assembled with MEGAHITv1.1.2, using the following parameters ‘-k-min 21 --k-max 255 --k-step 10 -t 30’. Reads were mapped back against resultant contigs using BBmap ‘minid = 0.95 covstats rpkms’ [72]. Resultant bam and sam files were processed using Samtools v1.6 [73]. To assess the level of bacterial DNA contamination all samples were processed with SortMeRNA v2.1 to check for contaminating rRNA reads ‘sortmerna --ref /usr/local/bioinf/sortmerna-2.1/rRNA_databases/silva-bac-16_s-id90.fasta,/usr/local/bioinf/sortmerna-2.1/index/silva_b90:/usr/local/bioinf/s\$’ [74] and also using viromeQC. For read based assessment of viral diversity centrifuge was used with default settings and database of known phage genomes. For contig based assessment of viral diversity, contigs were first filtered with DeepVirFinder to remove any contigs that are of likely bacterial origin, only contigs with a p value < 0.05 and were greater than 5 kb in length were considered for further analysis [75]. Contigs were considered to be present within a sample if the average coverage of mapped reads was ≥1X over ≥70% of the sample, as recommended in other studies [4]. The relative abundance of contigs within each sample was determined by counting the number of reads mapped to each contig, divided by the length of the contig (Kbp) to give RPK. The sum of all RPK values per sample was divided 1 000 000, with each RPK divided by the per million scaling factor. Processing of data was carried out in R, using the PhyloSeq [47] library to calculate diversity statistics. To identify circular contigs LAST (lastal -s 1 -x 300 -f 0 -T) was used to identify the ends of contigs that overlapped [76].

Contigs were annotated automatically using Prokka v1.12 using the following settings ‘--meta’, using a custom database phage proteins [77]. This database was constructed by extracting all proteins from publically available phage genomes within the European Nucleotide Archive, which were downloaded as previously described [78] and the database created with scripts included with Prokka.

Further annotation was provided by the use of hmmpfiles using hmmscan with the prokaryotic Viral Orthologous Groups (pVOG) collection of hmm profiles using a cutoff value of 1E⁻⁵ [7–8]. To identify putative temperate phages a method akin to Shkoporov & Hill was used. We utilised a set of PFAM hmms (PF07508, PF00589, PF01609, PF03184, PF02914, PF01797, PF04986, PF00665, PF07825, PF00239, PF13009, PF16795, PF01526, PF03400, PF01610, PF03050, PF04693, PF07592, PF12762, PF13359, PF13586, PF13610, PF13612, PF13701, PF13737, PF13751, PF13808, PF13843, and PF13358) that are specific to bacteriophage transposase, integrase, recombinase, resolvase and excisionases.

Putative hosts were predicted using WIsH [57]. A database of 9075 complete bacterial genomes was downloaded from Genbank (Jan 2018) and models were constructed for each bacterial genome within WIsH. Null parameters were calculated for each bacterial model using 7000 bacteriophage genomes. Hosts were predicted for each phage contig in the virome, with only predictions that had a pvalue of <0.05 considered for further analysis.

4.4. Closest relatives

A custom database of all known phages genomes was produced by extraction of ~10,000 complete phage genomes from genbank as previously described. A MASH database was produced for using sketch -s 10000. Each contig was queried against this database using mash dist function, with the top hit that had a distance of <0.05 assigned as it closest known relative. To cluster contigs at the genus level, vContact2 was used with the following settings "--rel-mode 'Diamond' --db 'ProkaryoticViralRefSeq85-Merged' --pcs-mode MCL --vcs-mode ClusterONE". The network graphs was visualized in Cytoscape and using Python package graphviz_layout.

4.5. Nucleotide sequences accession numbers.

The metagenomic reads, assembled contigs and 16S rRNA genes sequences were deposited in GenBank. The bioproject accession numbers for each sample are listed in the Table S1.

Diversity indices were produced by use the R Phyloseq package [47].

CRedit authorship contribution statement

V.V. Babenko: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - review & editing. **A. Millard:** Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Validation, Visualization, Writing - original draft, Writing - review & editing. **E.E. Kulikov:** Investigation, Writing - review & editing. **N.N. Spasskaya:** Investigation, Writing - review & editing. **M.A. Letarova:** Investigation, Writing - review & editing. **D.N. Konanov:** Formal analysis, Validation, Writing - review & editing. **I. Sh. Belalov:** Formal analysis, Investigation, Visualization, Writing - review & editing. **A.V. Letarov:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge T. Redgwell for help with proofreading and V. N. Filatova for help with the samples collection. The work at RC Biotechnology RAS was partially supported by Russian Foundation for Basic Research, Russia, grant #18-29-13029 (3 part of the work was performed before the grant acquisition under the state assignment from the Ministry of Education and Science of Russian Federation). AM's work at UoL was funded by MRC grants MR/L015080/1 & MR/T030062/1. Bioinformatic analysis was in part carried out on infrastructure provided by MRC-CLIMB.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.10.036>.

References

- [1] Fred EB. Antony van Leeuwenhoek: on the three-hundredth anniversary of his birth. *J Bacteriol* 1933;25(iv):2–18.
- [2] Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. *Genome Med* 2016;8:51.
- [3] Barko PC, McMichael MA, Swanson KS, Williams DA. The gastrointestinal microbiome: a review. *J Vet Intern Med* 2018;32:9–25.
- [4] Heintz-Buschart A, Wilmes P. Human gut microbiome: function matters. *Trends Microbiol* 2018;26:563–74.
- [5] Torres J, Mehndru S, Colombel J-F, Peyrin-Biroulet LJTL. *Crohn's Disease* 2017;389:1741–55.
- [6] Khanna S, Raffals LE. The microbiome in Crohn's disease: role in pathogenesis and role of microbiome replacement therapies. *Gastroenterol Clin North Am* 2017;46:481–92.
- [7] Chen X, Devaraj S. Gut microbiome in obesity, metabolic syndrome, and diabetes. *Curr Diab Rep* 2018;18:129.
- [8] Maruvada P, Leone V, Kaplan LM, Chang EB. The human microbiome and obesity: moving beyond associations. *Cell Host Microbe* 2017;22:589–99.
- [9] Picardo SL, Coburn B, Hansen AR. The microbiome and cancer for clinicians. *Crit Rev Oncol Hematol* 2019;141:1–12.
- [10] Nitschke A, Deonandan R, Konkle AT. The link between autism spectrum disorder and gut microbiota: a scoping review. *Autism* 2020. <https://doi.org/10.1177/1362361320913364>.
- [11] Vuong HE, Yano JM, Fung TC, Hsiao EY. The microbiome and host behavior. *Annu Rev Neurosci* 2017;40:21–49.
- [12] Warner BB. The contribution of the gut microbiome to neurodevelopment and neuropsychiatric disorders. *Pediatr Res* 2019;85:216–24.
- [13] Letarov A, Kulikov E. The bacteriophages in human- and animal body-associated microbial communities. *J Appl Microbiol* 2009;107:1–13.
- [14] Seo SJ, Kweon MN. Virome-host interactions in intestinal health and disease. *Curr Opin Virol* 2019;37:63–71.
- [15] Wang H, Ling Y, Shan T, Yang S, Xu H, Deng X, Delwart E, Zhang WJVe. 2019. Gut virome of mammals and birds reveals high genetic diversity of the family Microviridae. *5:vez013*.
- [16] Moreno-Gallego JL, Chou SP, Di Rienzi SC, Goodrich JK, Spector TD, Bell JT, et al. Virome diversity correlates with intestinal microbiome diversity in adult monozygotic twins. *Cell Host Microbe* 2019;25(261–272):e5.
- [17] Shkoporov AN, Hill C. Bacteriophages of the Human Gut: The "Known Unknown" of the Microbiome. *Cell Host Microbe* 2019;25:195–209.
- [18] Neil JA, Cadwell KJ. 2018. The intestinal virome and immunity. *201:1615–1624*.
- [19] Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, et al. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host Microbe* 2019;26(764–778):e5.
- [20] Emet C, Ruffin M, Lamendella R. Enteric virome and carcinogenesis in the gut. *Dig Dis Sci* 2020;65:852–64.
- [21] Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 2011;21:1616–25.
- [22] Koonin EV, Yutin N. The crass-like phage group: how metagenomics reshaped the human virome. *Trends Microbiol* 2020;28:349–59.
- [23] Garmeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A. Studying the gut virome in the metagenomic era: challenges and perspectives. *BMC Biol* 2019;17:84.
- [24] Manrique P, Bolduc B, Walk ST, van der Oost J, de Vos WM, Young MJ. Healthy human gut phageome. *Proc Natl Acad Sci U S A* 2016;113:10400–5.
- [25] Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Bolding L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* 2014;5:4498.
- [26] Shkoporov AN, Khokhlova EV, Fitzgerald CB, Stockdale SR, Draper LA, Ross RP, et al. PhiCrAss001 represents the most abundant bacteriophage family in the human gut and infects *Bacteroides* intestinalis. *Nat Commun* 2018;9:4781.
- [27] Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 2010;466:334–8.
- [28] Shkoporov AN, Clooney AG, Sutton TDS, Ryan FJ, Daly KM, Nolan JA, et al. The human gut virome is highly diverse, stable, and individual specific. *Cell Host Microbe* 2019;26(527–541):e5.
- [29] Duarte AM, Silva JM, Brito C, Teixeira D, Melo F, Ribeiro B, et al. Faecal virome analysis of wild animals from Brazil. *Viruses* 2019;11.
- [30] Bergner LM, Orton RJ, Benavides JA, Becker DJ, Tello C, Biek R, et al. Demographic and environmental drivers of metagenomic viral diversity in vampire bats. *Mol Ecol* 2019. <https://doi.org/10.1111/mec.15250>.
- [31] Chen L, Gu W, Liu C, Wang W, Li N, Chen Y, et al. Characteristics of the tree shrew gut virome. *PLoS ONE* 2019;14:e0212774.
- [32] Lima DA, Cibulski SP, Tochetto C, Varela APM, Finkler F, Teixeira TF, et al. The intestinal virome of malabsorption syndrome-affected and unaffected broilers through shotgun metagenomics. *Virus Res* 2019;261:9–20.
- [33] Li H, Li H, Wang J, Guo L, Fan H, Zheng H, et al. The altered gut virome community in rhesus monkeys is correlated with the gut bacterial microbiome and associated metabolites. *Virol J* 2019;16:105.
- [34] Zhao G, Droit L, Gilbert MH, Schiro FR, Didier PJ, Si X, et al. Virome biogeography in the lower gastrointestinal tract of rhesus macaques with chronic diarrhea. *Virology* 2019;527:77–88.
- [35] Wang H, Ling Y, Shan T, Yang S, Xu H, Deng X, et al. Gut virome of mammals and birds reveals high genetic diversity of the family Microviridae. *Virus Evol* 2019;5(vez013).
- [36] Garber A, Hastie P, Murray JA. Factors influencing equine gut microbiota: Current knowledge. *J Equine Vet Sci* 2020;88:102943.
- [37] Hintz H, Cymbaluk N. Nutrition of the horse. *Annu Rev Nutr* 1994;14:243–67.
- [38] Зелевский НВ. 2007. *Анатомия лошади*. ИКЛ.

- [39] Waring GH. Horse behavior. second ed. Norwich, N.Y: Noyes Publishing; 2003.
- [40] Cann AJ, Fandrich SE, Heaphy S. Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes. *Virus Genes* 2005;30:151–6.
- [41] Park J, Kim EB. Differences in microbiome and virome between cattle and horses in the same farm. *Asian-Australas J Anim Sci* 2020;33:1042–55.
- [42] Golomidova A, Kulikov E, Isaeva A, Manykin A, Letarov A. The diversity of coliphages and coliforms in horse feces reveals a complex pattern of ecological interactions. *Appl Environ Microbiol* 2007;73:5975–81.
- [43] Isaeva AS, Kulikov EE, Tarasyan KK, Letarov AV. A novel high-resolving method for genomic PCR-fingerprinting of Enterobacteria. *Acta Naturae* 2010;2:82–8.
- [44] Golomidova AK, Kulikov EE, Prokhorov NS, Guerrero-Ferreira RC, Ksenzenko VN, Tarasyan KK, et al. Complete genome sequences of T5-related *Escherichia coli* bacteriophages DT57C and DT571/2 isolated from horse feces. *Arch Virol* 2015;160:3133–7.
- [45] Babenko VV, Golomidova AK, Ivanov PA, Letarova MA, Kulikov EE, Manolov et al. Phages associated with horses provide new insights into the dominance of lateral gene transfer in virulent bacteriophages evolution in natural systems; 2019. doi: 10.1101/542787 bioRxiv:542787.
- [46] Kulikov EE, Isaeva AS, Rotkina AS, Manykin AA, Letarov AV. Diversity and dynamics of bacteriophages in horse feces. *Mikrobiologiya* 2007;76:271–8.
- [47] McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013;8.
- [48] Kim BR, Shin J, Guevarra R, Lee JH, Kim DW, Seol KH, et al. Deciphering Diversity Indices for a Better Understanding of Microbial Communities. *J Microbiol Biotechnol* 2017;27:2089–93.
- [49] Torsvik V, Goksoyr J, Daae FL. High diversity in DNA of soil bacteria. *Appl Environ Microbiol* 1990;56:782–7.
- [50] Kim D, Song L, Breitwieser FP, Salzberg SL. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res* 2016;26:1721–9.
- [51] Graziotin AL, Koonin EV, Kristensen DM. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res* 2017;45:D491–8.
- [52] Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008;36:W5–9.
- [53] Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* 2018;430:2237–43.
- [54] Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;17:132.
- [55] StatQuest. July 22, 2015 2015. RPKM, FPKM and TPM, clearly explained. <https://www.rna-seqblog.com/rpkm-fpkm-and-tpm-clearly-explained/>. Accessed
- [56] Edwards RA, Vega AA, Norman HM, Ohaeri M, Levi K, Dinsdale EA, et al. Global phylogeography and ancient evolution of the widespread human gut virus crAssphage. *Nat Microbiol* 2019;4:1727–36.
- [57] Galiez C, Siebert M, Enault F, Vincent J, Soding J. WISH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* 2017;33:3113–4.
- [58] Kelekna P. The horse in human history. New York: Cambridge University Press, Cambridge; 2009.
- [59] Murcia PR. Clinical insights: The equine microbiome. *Equine Vet J* 2019;51:714.
- [60] Langner K, Blaue D, Schedlbauer C, Starzonek J, Julliard V, Vervuert I. Changes in the faecal microbiota of horses and ponies during a two-year body weight gain programme. *PLoS ONE* 2020;15:e0230015.
- [61] Lindenbergh F, Krych L, Kot W, Fielden J, Frokiaer H, van Galen G, et al. Development of the equine gut microbiota. *Sci Rep* 2019;9:14427.
- [62] Patterson-Kane JC, Karikoski NP, McGowan CM. Paradigm shifts in understanding equine laminitis. *Vet J* 2018;231:33–40.
- [63] Bulmer LS, Murray JA, Burns NM, Garber A, Wemelsfelder F, McEwan NR, et al. High-starch diets alter equine faecal microbiota and increase behavioural reactivity. *Sci Rep* 2019;9:18621.
- [64] Mathieu A, Dion M, Deng L, Tremblay D, Moncaut E, Shah SA, et al. Virulent coliphages in 1-year-old children fecal samples are fewer, but more infectious than temperate coliphages. 2020;11:1–12.
- [65] Weinbauer MG. Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 2004;28:127–81.
- [66] Golomidova AK, Kulikov EE, Babenko VV, Ivanov PA, Prokhorov NS, Letarov AV. *Escherichia coli* bacteriophage Gostya9, representing a new species within the genus T5virus. *Arch Virol* 2019;164:879–84.
- [67] Golomidova AK, Kulikov EE, Prokhorov NS, Guerrero-Ferreira Rcapital Es C, Knirel YA, Kostryukova ES, et al. Branched lateral tail fiber organization in T5-like bacteriophages DT57C and DT571/2 is revealed by genetic and functional analysis. *Viruses* 2016;8.
- [68] Minoransky VA, Uzdenov VM. Feral horses of Vodny island (Manych-Gudilo lake, Rostov region, Russia). *Vesti biosfernogo zapovednika "Askania-nova"* 2011;13:135–45.
- [69] Kulikov EE, Golomidova AK, Prokhorov NS, Ivanov PA, Letarov AV. High-throughput LPS profiling as a tool for revealing of bacteriophage infection strategies. *Sci Rep* 2019;9:2958.
- [70] Thomas JC, Khoury R, Neeley CK, Akroush AM, Davies ECJBE. 1997. A fast CTAB method of human DNA isolation for polymerase chain reaction applications. 25:233–5.
- [71] Joshi N, Fass J. 2011. Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.21)[Software].
- [72] Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).
- [73] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- [74] Kopylova E, Noé L, Pericard P, Salson M, Touzet H. Sortmerna 2: ribosomal rna classification for taxonomic assignment, p. In (ed),
- [75] Ren J, Song K, Deng C, Ahlgren NA, Fuhrman JA, Li Y, Xie X, Sun F. 2018. Identifying viruses from metagenomic data by deep learning. arXiv preprint arXiv:180607810.
- [76] Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res* 2011;21:487–93.
- [77] Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–9.
- [78] Michniewski S, Redgwell T, Grigonyte A, Rihtman B, Aguilo-Ferretjans M, Christie-Oleza J, et al. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. *Environ Microbiol* 2019;21:2112–28.