

## Research Article

# Metagenomic Profiles of the Intestinal Virome of Long-Livers

Elizaveta V. Starikova<sup>1\*</sup>, Ksenia M. Klimina<sup>1</sup>, Anastasia O. Eudokimova<sup>2</sup>, Ksenia A. Yeruslanova<sup>3</sup>, Denis A. Gudkov<sup>1</sup>, Daria A. Kashtanova<sup>3</sup>, Yulia V. Kotovskaya<sup>3</sup> and Olga N. Tkacheva<sup>3</sup>

<sup>1</sup>Federal Research and Clinical Centre of Physical-Chemical Medicine, Moscow, Russia

<sup>2</sup>Moscow Institute of Physics and Technology, Moscow, Russia

<sup>3</sup>Russian Clinical and Research Center Of Gerontology, Moscow, Russia

**ORCID**

Elizaveta V. Starikova, 0000-0001-6582-210X

**Abstract.** The microbial community of the human intestine is important for maintaining human health. It has been reported that the gut microbiome changes with age, and it can be enriched with certain beneficial bacteria while also losing certain commensal bacteria. Little is known about the gut virome of long-livers. Our research aimed to extract, sequence and analyze the viral fraction of long-livers' gut microbiota in comparison with those of young adults and the elderly. We were thereby able to characterize the gut virome profiles and viral diversity of three age groups.

**Keywords:** aging, gut microbiome, viral metagenomics, bacteriophages

Corresponding Author:

aaaaaaa; email:

estarikova@rcpcm.org

**Dates**

Published 13 January 2022

Publishing services provided by  
Knowledge E

© Elizaveta V. Starikova et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

Selection and Peer-review under the responsibility of the 8th Scientific and Practical Conference Conference Committee.

## 1. Introduction

Human intestinal virome is an important part of the microbial community of the human gut and is now actively studied in relation to human health and disease. The first studies on human gut virome have discovered crAssphage [1], a highly abundant and widespread bacteriophage that wasn't possible to identify before metagenomic era. Many of the virome studies are performed on WGS metagenomic data via mining viral sequences from metagenomic assemblies. Using metagenomic sequencing in combination with virome extraction methods allows us to analyze so-called "viral dark matter" - virus-like sequences found in metagenomic data that do not match any of the known viruses. In our study, we extract and analyze gut viromes in relation to age, using faecal samples of the three age groups.

## 2. Materials and Methods

 OPEN ACCESS

## 2.1. Sample collection

Fecal samples were collected individually from three groups of healthy female subjects living in Moscow: homestay long-livers (95-101 years, n=8), postmenopausal women (57-65 years, n=7) and young adults (19-33 years, n=8). Samples were transported to the laboratory within a few hours after collection, and then stored at -20 °C. In total, 23 samples were collected.

## 2.2. Virome extraction and sequencing

Viral particles were extracted from faecal suspension using a modified protocol for virome extraction that involved ultracentrifugation in CsCl. Extracted fractions were stained with SYBR green to evaluate the presence of viral particles. Samples containing the sufficient amount of viral particles were treated with DNase in order to avoid contamination and subsequently purified. Total viral DNA was extracted from purified viral particles, and DNA libraries were prepared using NEBNext Ultra II DNA Library Prep Kit. Sequencing was performed using Illumina HiSeq 2500 Sequencing System.

## 2.3. Read quality control

Sequencing read quality was evaluated using FastQC. Reads were filtered by quality and trimmed using trimmomatic [2]. Human DNA was removed by mapping the reads on human genome sequence (hg19).

## 2.4. Viral profiles

In order to obtain each sample's reference-based viral profiles, we have mapped the filtered viral reads to a set of 2681 bacteriophage genomes downloaded from RefSeq in June 2020. The mapping was performed using bowtie2 read alignment tool (v2.2.6) with default parameters. For each of the bacteriophage genomes, mean coverage depth was calculated using bedtoolsgenomecov (v2.17.0) instrument.

## 2.5. Viral alpha-diversity

We have evaluated viral alpha-diversity (a measure of species diversity within a habitat) for each of the samples in two ways, using reference and non-reference approaches. For

the reference approach, we have calculated the number of bacteriophages identified in a sample via mapping viral reads on bacteriophage genomes from RefSeq. We have used a threshold of 80% read coverage by width.

For the non-reference approach, we have calculated the number of viral operational taxonomic units (OTUs). In order to do that, we have performed metagenomic assembly using MEGAHIT assembler [3]. For each of the resulting contigs, proteins were predicted using Prodigal gene recognition and translation tool [4]. Each of the predicted proteins was compared to a set of phage-specific hidden markov model (HMM) profiles corresponding to capsid protein families from prokaryotic Virus Orthologous Groups (pVOG) [5] database (<http://dmk-brain.ecn.uiowa.edu/pVOGs/>) using hmmscan tool from HMMER3 package (<https://www.ebi.ac.uk/Tools/hmmer/>). We have calculated a number of viral OTUs for each of the samples as a number of assembly contigs encoding at least one protein that hits any of pVOG capsid profile HMMs with E-value < 0.001.

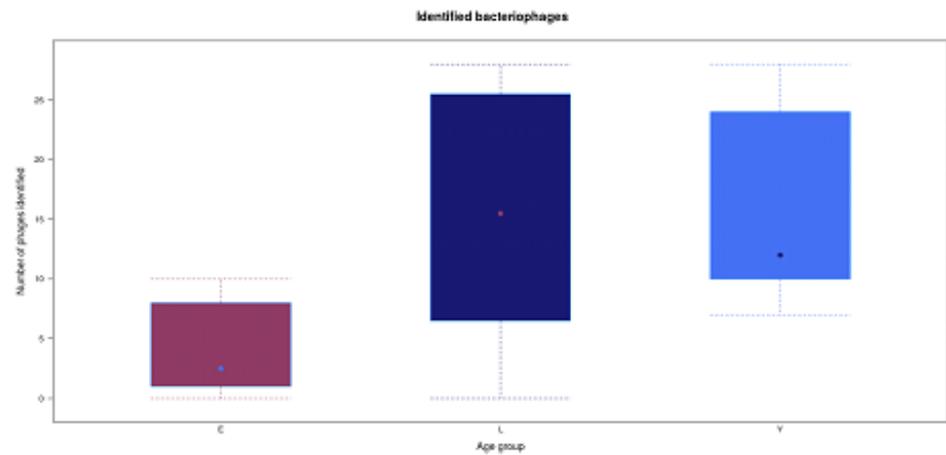
## 3. Results and Discussion

### 3.1. Viral profiles

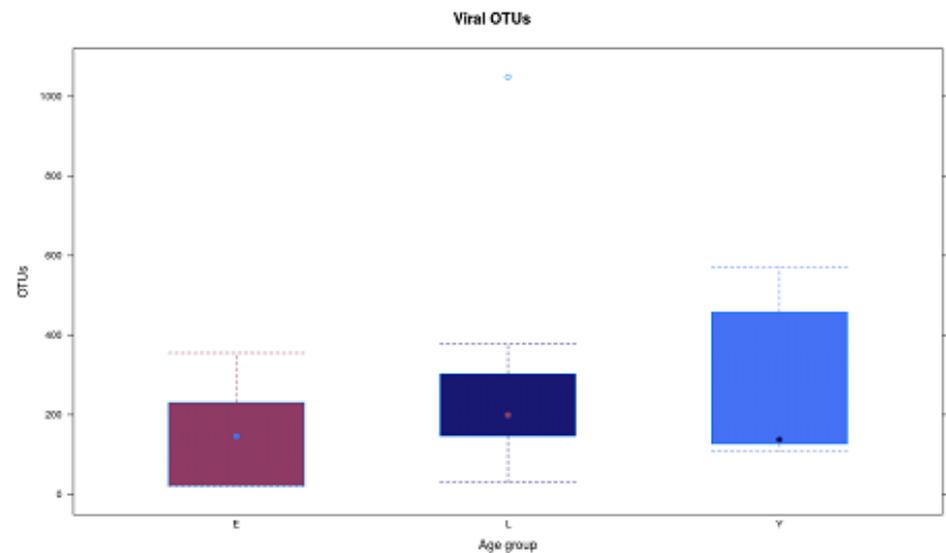
In total, we have identified 78 bacteriophages that were covered by sequencing reads by at least 80% of their length. The most abundant bacteriophage identified in the analyzed samples was CrAssphage (NC\_024711.1), a highly abundant human stool-associated bacteriophage recently discovered in metagenomes. It was identified in 7 of the 23 analyzed samples with mean read coverage depth from 16 up to 13040. Two other abundant bacteriophages having mean read coverage more than 1500 were Lactococcus phage 340 (NC\_021853.1) and Lactococcus phage 63301 (NC\_031017.1). We have compared all the bacteriophages' abundances using ANOVA test and have found no significant differences between age groups. Then, we have used a non-reference approach.

### 3.2. Viral alpha-diversity

Using a reference-based approach, we have shown that the mean number of bacteriophage species identified was higher in the long-livers' group as compared to elderly group and comparable to those of the young adults group. Non-reference approach have shown similar results: the mean number of distinct contigs containing putative capsid genes was higher in long-livers' group as compared to elderly group and was



**Figure 1:** Viral alpha-diversity calculated as a number of identified bacteriophages. “E” stands for the “elderly” group, “L” stands for the “long-livers” group, and “Y” stands for the “young adults” group.



**Figure 2:** Viral alpha-diversity calculated as a number of viral OTUs identified. “E” stands for the “elderly” group, “L” stands for the “long-livers” group, and “Y” stands for the “young adults” group.

comparable to those of the young adults group. The results can be seen on Figures 1 and 2.

## 4. Conclusion

Thus, we have evaluated long-livers gut virome for the first time. We haven't found significant differences in known bacteriophages' abundances between groups. However, we have shown using two different approaches that viral diversity in long-livers' group analyzed in this study is higher than those in the elderly group. Further characterization of the gut virome of different age groups will hopefully result in a better understanding of complex processes that a human organism undergoes during aging.

## 5. Acknowledgements and Funding

### References

- [1] Dutilh BE, Cassman N, McNair K et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications*. 2014;5(1):1-11.
- [2] Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.
- [3] Li D, Liu CM, LuoR, Sadakane K, Lam TW. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674-1676.
- [4] Hyatt D, ChenGL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11(1):119-125.
- [5] Graziotin AL, Koonin EV, Kristensen DM. Prokaryotic virus orthologous groups (pVOGs): A resource for comparative genomics and protein family annotation. *Nucleic Acids Research*. 2016. 45(D1):D491-D498.