![American Society for Microbiology | Microbiology Resource Announcements]

Metagenomic Exploration of Koumiss from Kazakhstan

Andrey Bogoyavlenskiy,[a,b] Madina Alexyuk,[b] Pavel Alexyuk,[b] Makhabbat Amanbayeva,[a] Elmira Anarkulova,[b] Anar Imangazy,[a] Aliya Bektuganova,[a] Vladimir Berezin[b]

[a]Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[b]Research and Production Center for Microbiology and Virology, Almaty, Kazakhstan

ABSTRACT   Here, we report a metagenomic analysis of koumiss from Kazakhstan. In this study, shotgun metagenomic sequencing of the RNA and DNA viral community was performed.

Bacterial infections of the lung and gastrointestinal tract remain one of the main problems of medicine in a number of countries (1–3). The emergence of antibiotic-resistant variants of microorganisms has led to the need to search for new nonstandard methods of treating bacterial infections. One of these methods is koumiss treatment of pulmonary and gastrointestinal infections in regions with developed horse breeding. It was shown that the composition of koumiss is of great importance for the additional treatment of microbial diseases (4, 5). In our research, the virome of koumiss (fermented dairy product made from mare's [*Equus caballus*] milk) was studied to diagnose the presence of phages in this dairy product that can lyse pathogenic microorganisms, creating a therapeutic effect. Koumiss prepared on meat and dairy cattle farms in Talgar near Almaty (43°18′N, 77°14′E) was used for the study.

A shotgun metagenomic study of the DNA and RNA viral community, including sample preparation, sequencing, and bioinformatics analysis, was mainly performed as described in our previous studies (6–8). Briefly, 2 L koumiss was centrifuged at 15,000 rpm for 15 min at 4°C, filtered using a 0.45-mm membrane to remove most of the bacteria, and concentrated by ultracentrifugation using a Beckman Coulter Avanti J30I ultracentrifuge at 29,000 rpm for 2 h at 4°C. Total DNA was isolated using a PureLink genomic DNA extraction kit (Thermo Fisher Scientific, USA) and stored at 2 to 8°C. Viral RNA was extracted using a QIAamp viral RNA minikit (Qiagen). The sample extracts were pretreated with RNase-free DNase (Promega). Double-stranded cDNA was obtained using a SuperScript double-stranded cDNA synthesis kit (Invitrogen) according to the manufacturer's instructions. Genomic DNA and synthesized double-stranded cDNA were pooled. The paired-end library was prepared using a Nextera XT DNA sample preparation kit (Illumina, USA) according to the manufacturer's protocol. Sequencing of the library was conducted on the Illumina MiSeq platform using the MiSeq reagent kit v3 (2 × 300-bp cycles; Illumina). The obtained sequences (2,719,939 reads) were tested for quality using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) and Trimmomatic v0.36 (9) within the Genome Detective tool (10). Reads with average quality of less than 20 and length less than 200 bp were removed. Taxonomic identification of viral sequences was performed by Kaiju using the BLASTn v2.5.0 algorithm (10) against the NCBI RefSeq viral complete genome database (26 February 2021; 20 GB) (11). Default parameters were used for all software. The sequence reads were considered "identified" if they had a relative in the reference database with an E value of $<10^{-5}$ and a bit score of >50.

The raw data contained 2,719,939 paired sequence reads. After quality processing, we obtained 2,033,388 single reads; of these, 573,066 sequences (28.2% of the data set) were identified as viral, belonging to double-stranded DNA (dsDNA) viruses with no RNA stage (84%), retro-transcribing viruses (2%), single-stranded DNA (ssDNA) viruses (2%),
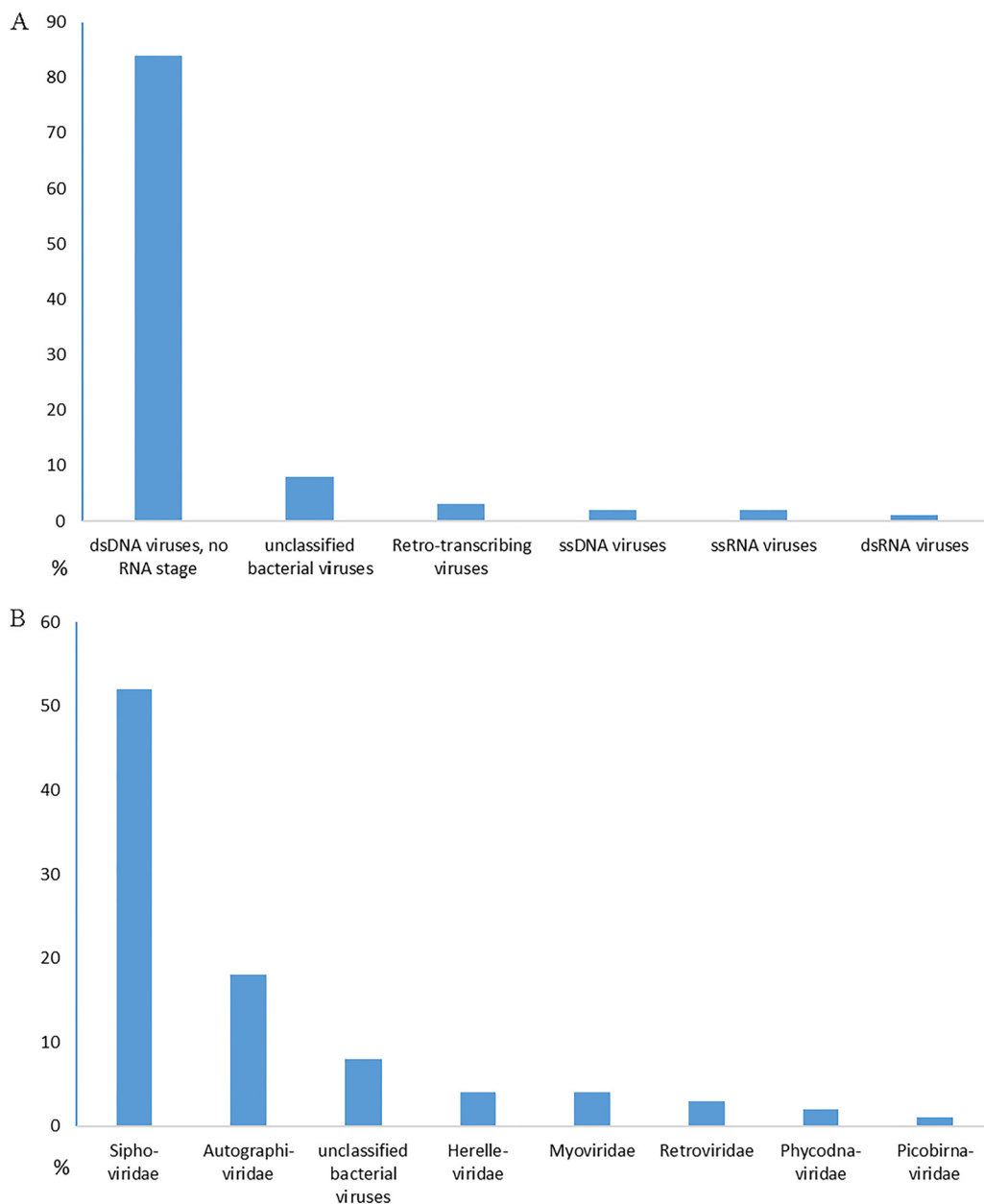
A

B

FIG 1 (A) Viral groups of the koumiss virome. (B) The main viral families of the koumiss virome. Microsoft Excel 2016 was used to generate the plots.

ssRNA viruses (2%), dsRNA viruses (1%), and unclassified bacterial viruses (8%) (Fig. 1A). Despite the fact that representatives of 70 families of viruses were detected, 7 of them comprised 92% of all identified viruses (Fig. 1B). In total, we revealed a high viral diversity in the samples of koumiss (721 virotypes from 70 families). Among these viral groups, viruses were found that are able to lyse the main groups of microorganisms that cause diseases of the lungs and gastrointestinal tract. Among these microorganisms are those belonging to the genera *Streptococcus*, *Mycobacterium*, *Campylobacter*, *Pseudomonas*, *Acinetobacter*, etc.

Our metagenomic research allowed us to start exploring the diversity of viral communities in koumiss, which will make it possible to assess the role of koumiss in traditional medicine.

Data availability. The raw sequence reads are available under BioProject accession number PRJNA775816.

REFERENCES

1. Budden KF, Gellatly SL, Wood DLA, Cooper MA, Morrison M, Hugenholtz P, Hansbro PM. 2017. Emerging pathogenic links between microbiota and the gut-lung axis. Nat Rev Microbiol 15:55–63. https://doi.org/10.1038/nrmicro.2016.142.
2. Feng Q, Chen WD, Wang YD. 2018. Gut microbiota: an integral moderator in health and disease. Front Microbiol 9:151. https://doi.org/10.3389/fmicb.2018.00151.
3. Sencio V, Machado MG, Trottein F. 2021. The lung–gut axis during viral respiratory infections: the impact of gut dysbiosis on secondary disease outcomes. Mucosal Immunol 14:296–304. https://doi.org/10.1038/s41385-020-00361-8.
4. Tang H, Ma H, Hou Q, Li W, Xu H, Liu W, Sun Z, Haobisi H, Menghe B. 2020. Profiling of koumiss microbiota and organic acids and their effects on koumiss taste. BMC Microbiol 20:85. https://doi.org/10.1186/s12866-020-01773-z.
5. Zhang M, Dang N, Ren D, Zhao F, Lv R, Ma T, Bao Q, Menghe B, Liu W. 2020. Comparison of bacterial microbiota in raw mare's milk and koumiss using PacBio single molecule real-time sequencing technology. Front Microbiol 11:581610. https://doi.org/10.3389/fmicb.2020.581610.
6. Amanbayeva M, Anarkulova E, Bogoyavlenskiy A, Alexyuk M, Imangazy A, Berezin V. 2021. Metagenomic exploration of *Atelerix albiventris* gut microbiome. Microbiol Resour Announc 10:e01342-20. https://doi.org/10.1128/MRA.01342-20.
7. Alexyuk MS, Bogoyavlenskiy AP, Alexyuk PG, Moldakhanov YS, Berezin VE. 2020. Metagenome analysis of surface waters of the Shardara Reservoir, the largest artificial reservoir in southern Kazakhstan. Microbiol Resour Announc 9:e00053-20. https://doi.org/10.1128/MRA.00053-20.
8. Alexyuk M, Bogoyavlenskiy A, Amanbayeva M, Alexyuk P, Moldakhanov Y, Anarkulova E, Imangazy A, Berezin V. 2020. Virome structure of the Small Aral Sea. Microbiol Resour Announc 9:e01023-20. https://doi.org/10.1128/MRA.01023-20.
9. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
10. Vilsker M, Moosa Y, Nooij S, Fonseca V, Ghysens Y, Dumon K, Pauwels R, Alcantara LC, Vanden Eynden E, Vandamme A-M, Deforche K, de Oliveira T. 2019. Genome Detective: an automated system for virus identification from high-throughput sequencing data. Bioinformatics 35:871–873. https://doi.org/10.1093/bioinformatics/bty695.
11. Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 33:D501–D504. https://doi.org/10.1093/nar/gki025.