Mathematical Biology and Bioinformatics
2023. V. 18. № 2. P. 323–346. doi: 10.17537/2023.18.323

========== **BIOINFORMATICS** ==========

Comparative analysis of *Actinobacteria* phage-plasmids and their transduction potential

Nikulin N.A.[*1], Kiselev S.S.[2], Panyukov V.V.[3], Lu Y.[4], Zimin A.A.[1]

[1]*G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms RAS, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Russia*
[2]*Institute of Cell Biophysics RAS, Federal Research Center "Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences", Pushchino, Russia*
[3]*Institute of Mathematical Problems of Biology RAS — the Branch of Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russia*
[4]*College of Life Sciences, Shanghai Normal University, Shanghai, China*

Abstract. The genus *Streptomyces* is the most extensively studied group within actinobacteria, which are widely used for the production of antibiotics and other metabolites. *Streptomyces* bacteriophages attract a lot of interest to study phage-host coevolution. One of the probable interactions is a bacteriophage-mediated horizontal transfer of genes encoding bacterial metabolites, which are important for humans. In this study, a search for *Streptomyces* bacteriophages having capacities necessary for efficient horizontal transfer was performed. Groups of bacteriophages of actinobacteria and their plasmids with relatively long GC-rich sequences containing both genes related to viral morphogenesis and plasmid-associated genes were revealed by comparative genome analysis. In one of these groups, homologs of proteins of DNA packaging into a capsid of *Punavirus* genus phages, that representatives are capable to perform a horizontal transfer of genes with high frequency were identified. Taking it into account, as well as the peculiarities of their genomic structure, horizontal gene transfer by this group of viruses and plasmids was assumed to occur. Considering the genome length and GC-content, it is thought that these viruses could be useful for producers construction. Moreover, some given examples showing the chimerical origin of phage-plasmids allow suggesting that recombination events between considered phages-plasmids occur at high frequency. New representatives of viruses with low similarity to known phage genomes were also detected among plasmids. A wide diversity of actinobacterial phage-plasmids that is related to their variability and a probable existence of a great number of sequences of this type of viruses in databases were found out.

Key words: *Actinobacteria phages, comparative genome analysis, transduction, Streptomyces bacteriophages, phage-plasmids.*

INTRODUCTION

The *Actinobacteria* belong to one of the largest bacterial phylum. They are distributed almost in all terrestrial, freshwater and marine biotopes [1, 2]. Most of the actinobacteria can form both mycelium inside substrate and aerial mycelium [3]. Currently, *Actinobacteria* are divided into 400 genera (NCBI taxonomy) by estimate [4]. The genus *Streptomyces* (phylum *Actinobacteria*) is the largest in terms of the number of species represented in it, the other large genera are *Actinomyces*, *Kitasatospora*, *Micromonospora*, *Nocardia*, *Micrococcus*, *Arthrobacter*, *Rhodococcus* [5].

[*]n.nikulin@ibpm.ru

Actinobacteria have capability of producing most of physiologically active bacterial metabolites that are used by humans. Majority of antibiotics applied for therapy in humans and for agriculture are produced by actinomycetes. *Actinobacteria* are used in pharmaceuticals, biotechnology, food industry, agriculture and as a source for industrial enzymes [5, 6]. Due to the capability of actinobacteria to utilize many pollutants contributing to their growth, they have the prospect of application in biological treatment of wastes generated because of human activity [7].

Actinobacteriophages are viruses that infect actinobacteria. Actinobacteriophages usually have mosaic genomes with high divergence among DNA sequences. Genomic diversity can be a result of the interactions between bacterial viruses and their hosts. Actinobacterial prophages can be involved in different systems of lysogenic conversion including defense systems protecting against infection of other phages.

When phages propagate, they can sometimes encapsidate host bacterial DNA to form transducing particles, which are inject bacterial DNA instead of a viral genome into the cell during infection of other bacteria. The DNA can then integrate into the chromosome by recombination or replicate as a plasmid in the new host cell.

This process of transferring bacterial DNA from one bacterium to another is known as genetic transduction [8]. The genetic information carried in transducing particles may have a strong influence on the bacterial recipients. Bacteria can acquire genes encoding antibiotic resistance or virulence factors by this route. Other genes acquired by bacteria can impart new capabilities and open new ecological niches for the bacterial recipients. This pathway promotes the formation of new bacterial strains, for examples, those that become more virulent and resistant to antibiotics.

It is suggested that through the process of transduction bacteria often acquire genes that allow them to quickly and efficiently adopt to changing environmental conditions. It was previously thought that the whole phage-mediated gene transfer occurs by one of two well-described mechanisms [9], generalized and specialized transduction, discovered in the 1950s [8]. With the recent discovery of lateral transduction [10] three mechanisms have been identified.

P1-mediated phage transduction is a standard approach for construction of new bacterial strains [11, 12]. One of the main aspects of this construction is the potential size of the phage head, which depends on the length of DNA packaged into capsid. Phage transduction potential is still not explored fully so far with regard to the genetics of actinobacteria.

In present research, the genomic features of bacteriophages of *Streptomyces* in the order *Caudovirales* with large genomes and related viruses of other bacteria were studied. The main goal of this work is to search for taxonomic groups of large bacteriophages capable of carrying out horizontal transfer of genetic information between genomes of *Streptomyces*. The large sizes of phage genomes can mediate horizontal transfer both of short DNA sequences and of long DNA sequences such as clusters of genes responsible for metabolism of any substance as a physiological compound, which is significant in medicine. These bacteriophages as well as methods of classical bacterial genetics and genome editing may become an essential tool for genetic manipulation with the genomes of *Streptomyces*.

MATERIALS AND METHODS

Genomes and annotation

Initially, genome sequences of *Streptomyces* bacteriophages from NCBI nr and Actinobacteriophage Databases [13] with sizes of more than 100 kbp were used.

Annotation or reannotation of nucleotide sequences of genomes and plasmids was performed with the Annotation service of the integrated PATRIC database [14, 15] including the following parameters: Domain [Bacteriophages], Annotation recipe [Bacteriophage]. Also, DFAST was used for the initial set of genomes [16].

Genome analysis

Clustering of homologous translated reading frames was carried out using GET_HOMOLOGUES [17] with parameters: COGtriangle (-G) algorithm [18], percentage of sequences identity 50 (-S), coverage 40 (-C), to include sequences with the same similarity into several clusters (-x), the remaining parameters without changes. The parameters were optimized manually, depending on the homologous clusters formed.

The pan-genome tree was obtained according to a matrix of the presence/absence of homologous clusters in GET_PHYLOMARKERS [19] using the estimate_pangenome_phylogenies script, the number of replications of the tree (-r) was 100. Visualization of the trees was performed with FigTree.

Phylogenomic trees were inferred based on pairwise distance matrices calculated using our own software IGPDE (InterGenomic Pairwise Distance Estimator), and were compared with trees obtained on the server VICTOR (comparison was carried out only for 71 *Streptomyces* phages with large genomes due to server limitation) [20]. The alignment-free method was applied, based on the *k*-mers (short oligonucleotides with a length of *k* bases). It should be noted that for double-stranded DNA, direct and inverted (reverse complement) sequences are considered as identical. To begin with, a set of oligonucleotides with a fixed length of *k* was generated for each genome or plasmid ($k = 10$). Moreover, *k*-mers were taken with a shift of one nucleotide. Then, for each pair of *k*-mer sets (*A* and *B*, for example), a similarity index *S* was calculated using the Sorensen (or Bray-Curtis) formula [21, 22], as in [23]:

$$S = \frac{2n(A \cap B)}{n(A) + n(B)}.$$

Here $n(A \cap B)$ stands for the number of elements in the intersection of two sets. The distance between two *k*-mer sets was estimated as: $D = 1 - S$. The obtained pairwise distances matrices were used to construct trees in MEGA X [24] with neighbor joining method [25].

To search for sequences homologous to Gilgamesh bacteriophage proteins, BLASTp [26] was used (matrix BLOSUM45 [27], word size: 2).

EasyFig 2.2.5 [28] with the use of BLASTn [26] was applied to compare genomic sequences. ORFs associated with virion morphogenesis, segregation, plasmid formation, etc. were extracted. The search for homologous sequences was carried out according to the detected homologous regions, homologous clusters and with using the BLASTp [26].

Network analysis

The bipartite network was constructed based on the presence/absence of homologous clusters, clusters consisting of sequences of only one genome were removed. For the network construction Cytoscape [29] was used with the following options: Prefuse Force Directed Layout algorithm, 10000 iterations, Default Spring length 100, Default Node Mass 10, the rest parameters remained unchanged.

For more effective visualization the nodes, whose genomes form known clusters/genus, as well as genomes located in common clades on the pan-genome tree and close to the network, were collapsed into common nodes. The size of such nodes reflects the number of nodes of the genomes included in it (each genome increases the size of the node by one unit of the standard size). The analysis of bipartite networks was carried out with the NetworkX Python package [30] using of Bipartite [31] and Distance Measures modules. To analyze the statistical significance of changes in network metrics, Wilcoxon test was used with the stat_compare_means() and compare_means() functions from the ggpubr package. *P*-values were adjusted with the Holm method [32]. The box plots were constructed using ggplot2.

Phylogenetic analysis

A tree based on the alignment of the amino acid sequences of large terminase subunit was inferred with maximum likelihood method [33] in IQ-TREE [34]. Alignment was obtained with MUSCLE [35]. Before phylogenetic reconstruction optimal substitution model was determined with ModelFinder [36], implemented in IQ-TREE (this test based on the bayesian information criterion [37]). Branch support levels were calculated with ultrafast bootstrap approximation [38].

RESULTS

Analysis of *Streptomyces* phages with large genomes

Initially, 71 genomes of *Streptomyces* bacteriophages, having the lengths of more than 100 kbp, were found. Almost all of them belonged to the *Siphoviridae* family; with the exception of Streptomyces phage BRock (refers to the *Myoviridae* family). When comparing the lengths of their genome sequences with GC-content (Figure 1), it was found that Streptomyces phage Gilgamesh and genomes of the BM cluster have the highest percentage of GC-pairs. Moreover, the Gilgamesh had the highest GC-content (71.3 %).

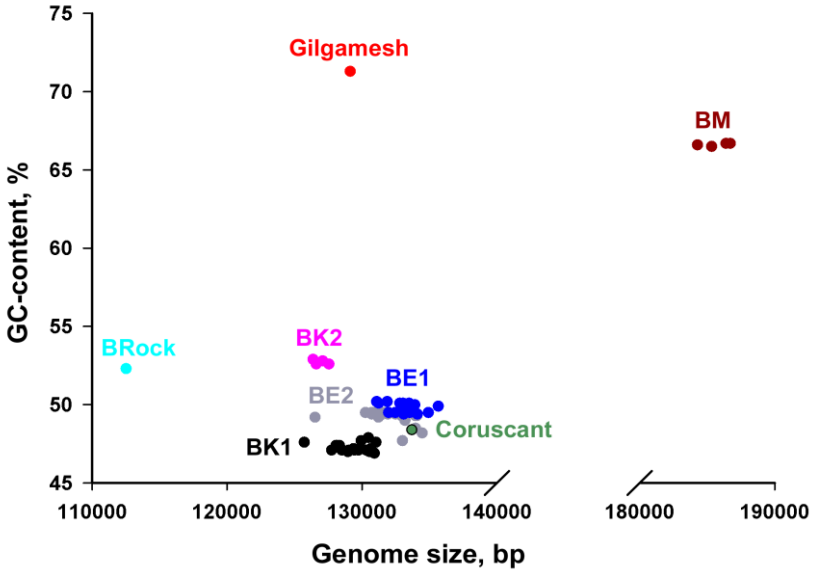

Fig. 1. Correspondence between the GC-content and genome size of *Streptomyces* phages with a genome length of more than 100 kbp.

Many of the *Streptomyces* genomes themselves have the same high level of GC-pairs [39–41], as a result of which it can be assumed that the pool of genes of Streptomyces phage Gilgamesh highly correlates with *Streptomyces*. Also, a phylogenomic tree for these phages was inferred (Figure 2).

According to the results, the branch of the Gilgamesh bacteriophage was located near the branch of the BM cluster, however the distance between them was large. So, it was decided to study the Streptomyces phage Gilgamesh and genomes containing homologous sequences of this virus.

Streptomyces phage Gilgamesh and far-related phages

Streptomyces phage Gilgamesh (MN234216.1) is a singleton (according to data from the Actinobacteriophage Database [13]), that is, a phage that does not have closely related organisms. It belongs to the *Siphoviridae* family. To characterize the genome composition of this phage, the homologs of its 156 proteins were searched among the translated ORFs of viral

genome sequences available in the nr database. It was found that 62 ORFs have homologs in other viruses. Then, genomes of bacteriophages containing at least 5 homologs of proteins of Gilgamesh were selected from the nr database. All together 546 genomes were found in the GenBank nucleotide collection database (accessed 1 October 2020). The supplementary file Table S1.xlsx includes accession numbers, belonging to a group of viruses, genome lengths, host bacterial strains for phages. For convenience, the genomes called uncultured Caudovirales phage were numbered.

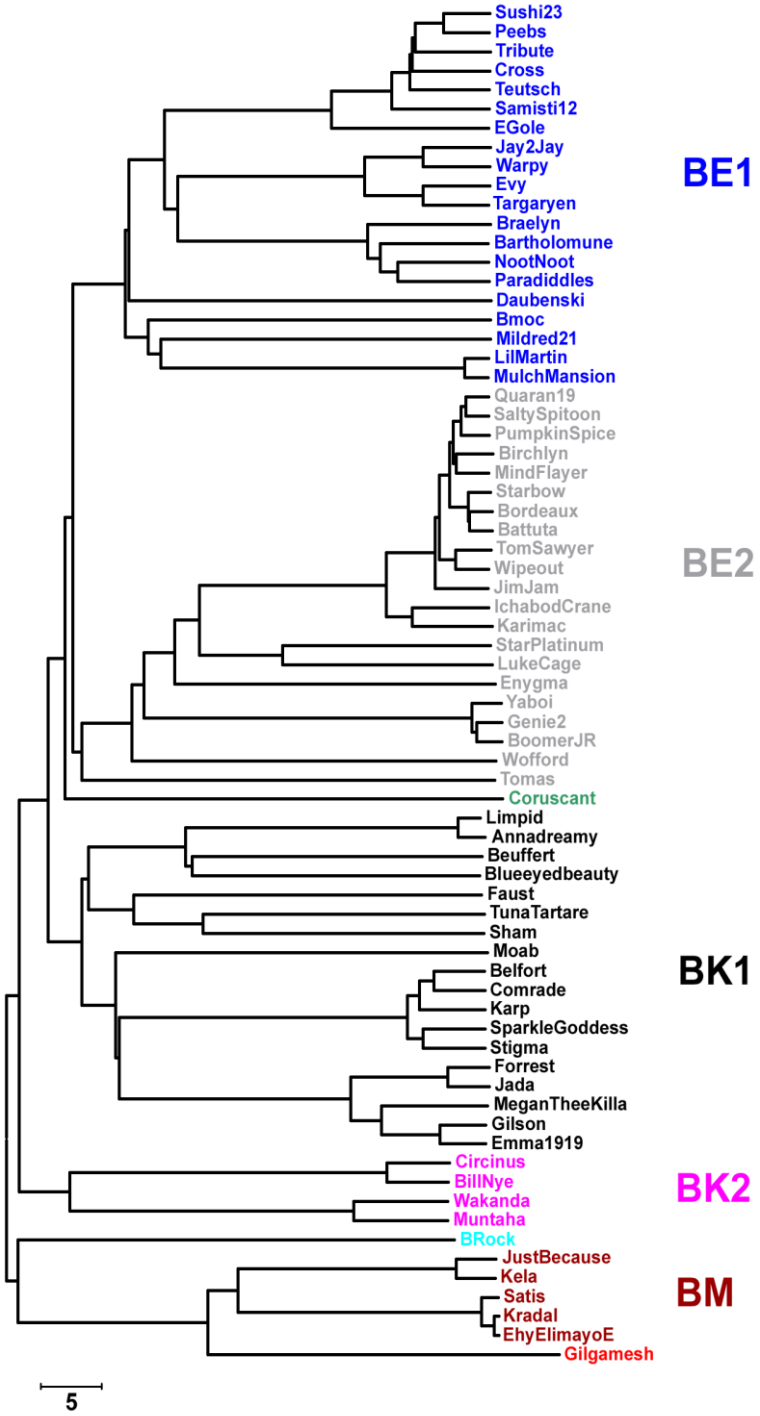

Fig. 2. A phylogenomic tree of *Streptomyces* phages, with a genome length of more than 100 kbp. The tree constructed with the neighbor joining method [25] from the pairwise distance matrix (see details in "Materials and Methods") calculated for 71 sets of 10-mers. The same color code as in Figure 1 corresponds to the clusters and subclusters from Actinobacteriophage Database [13]. The scale bar shows the Sorensen distance as a percentage.

Most viruses (439) with these genomes infected bacteria of the *Actinomycetia* class. The genomes of most of these viruses (about 430) are deposited in the Actinobacteriophage Database (Date of access: 1 October 2020). Most of the viruses (230) were phages infecting *Mycobacterium* and belonging to clusters E, J, L, N, Y, W, subcluster K1. The remaining genomes belong to phages of *Pseudomonas* (13), *Enterobacteriaceae* (13), *Aeromonas* (1), *Bacilli* (4), *Alteromonadales* (6), *Marinomonas* (1), *Streptococcus* (4), *Staphylococcus* (9), *Flavobacterium* (1), *Paracoccus* (1), *Vibrio* (1), *Pontimonas* (1). The rest 52 genomes were isolated from metagenomes.

Genome clustering was carried out. Altogether there were 9999 clusters. Core- and softcore-genomes were not formed. According to the results of clustering, there were 32 homologous clusters of translated ORFs (see supplementary Data S1.zip).

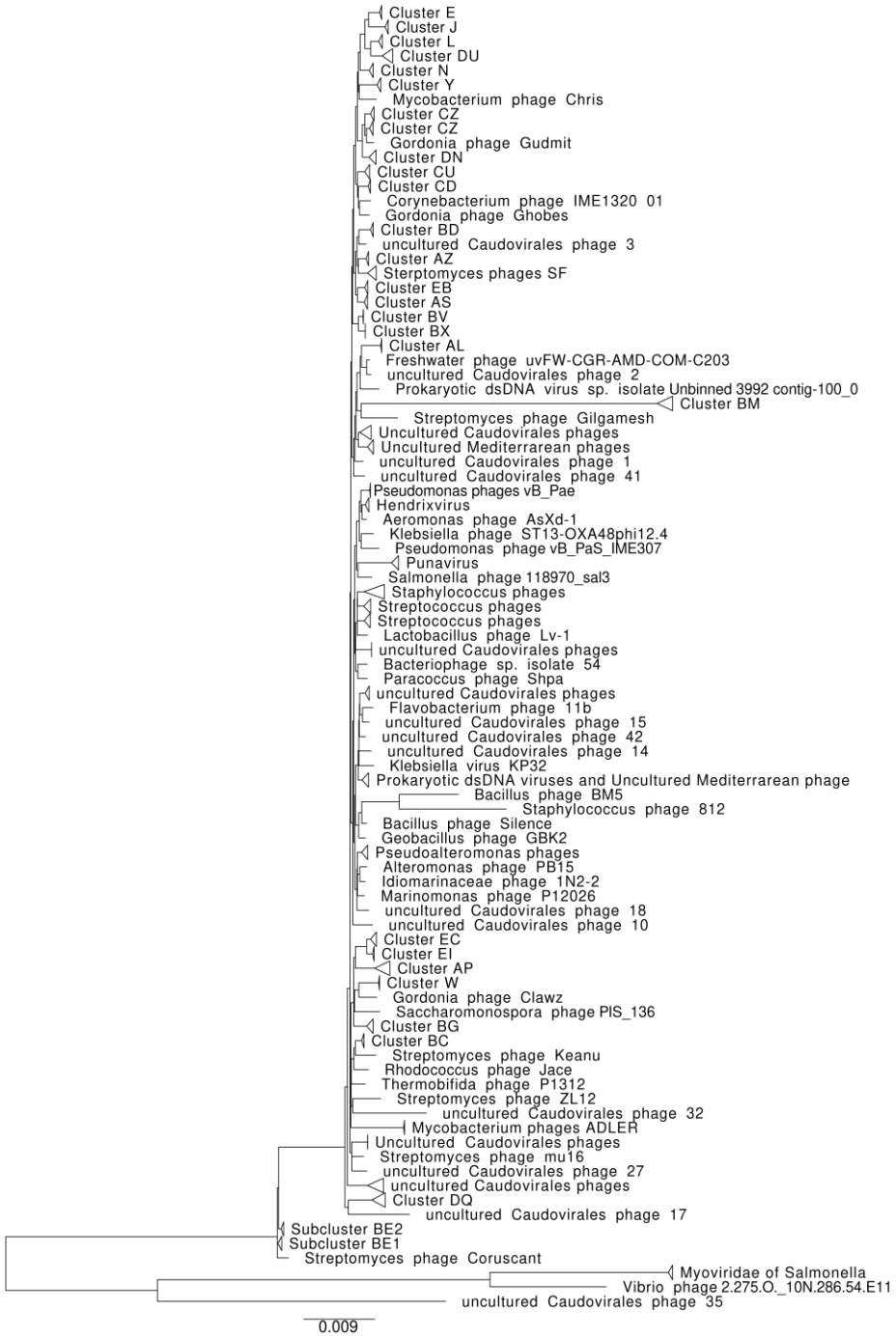

Fig. 3. A collapsed pan-genome tree for Streptomyces phage Gilgamesh genome and genomes of phages containing homologs of its proteins. The scale bar represents the number of expected substitutions per site under the binary GTR2+FO+ASC+R4 substitution model. The complete tree is given in Figure S1.

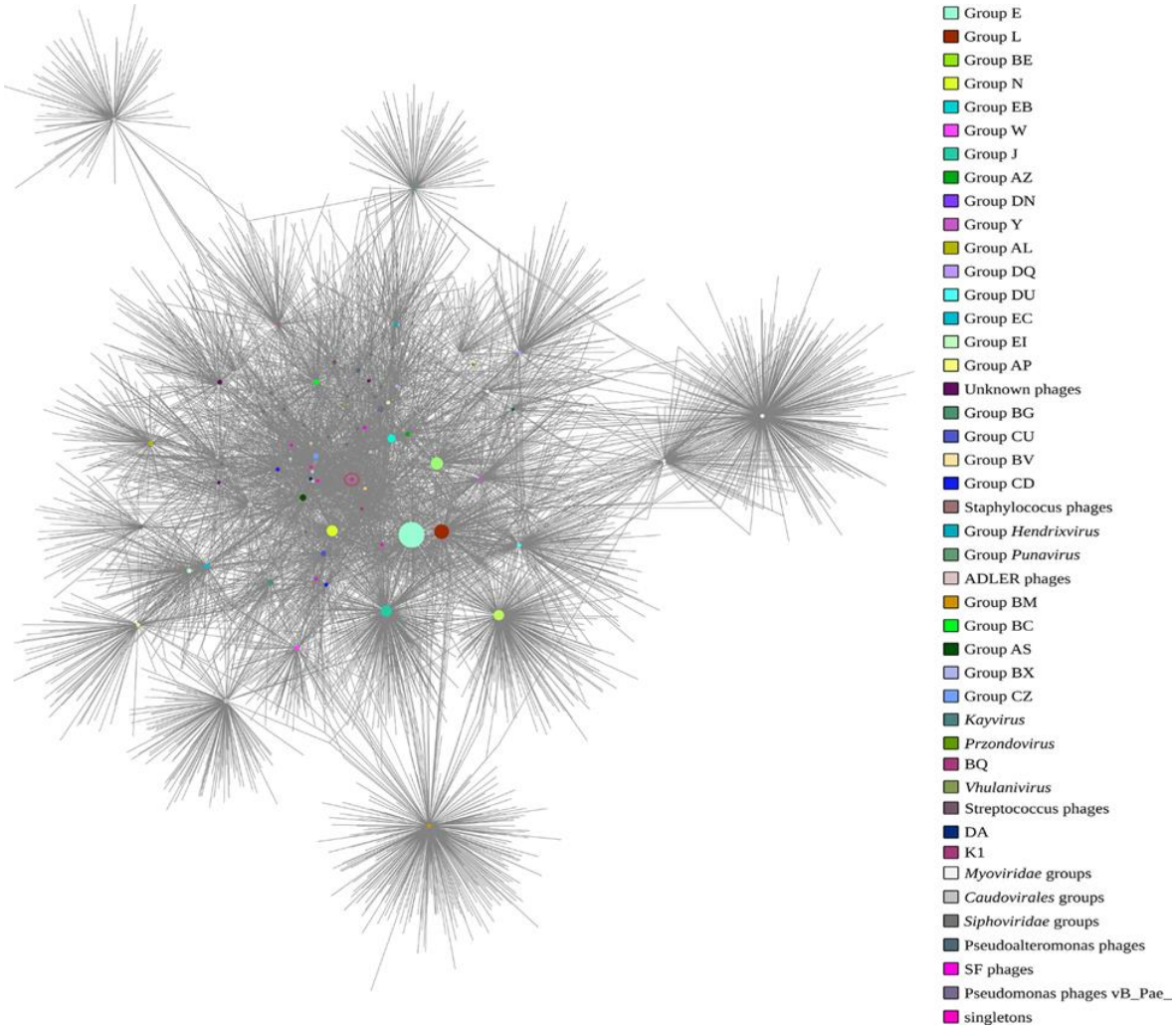

Fig. 4. Collapsed bipartite genome network of Streptomyces phage Gilgamesh and phages containing its homologs. The edges reflect the homologous protein clusters for Gilgamesh. Nodes of genomes and groups of viruses are marked with colored circles, nodes of homologous clusters are not marked. The node of phage Gilgamesh is marked with red ellipse. The complete network is shown in Figure S2.

On the pan-genome tree (Figure 3 and Figure S1) (the best model: GTR2+FO+ASC+R4, log-likelihood of the consensus tree: -79394.082) it is seen that the closest clade to the branch of the phage Gilgamesh is a BM cluster, as well as on the phylogenomic tree (Figure 2). Unique features of the viruses in this cluster are a relatively large size of genomes for bacteriophages (more than 180 kbp), and a long capsid length (about 285 nm) with a small capsid width (about 47 nm). Only the lytic cycle of reproduction was found in these phages. They are bacteriophages infecting *Streptomyces*. However, the distance between the branch of Gilgamesh and the BM clade is very large, which also indicates that this virus is a singleton.

In addition, a visual inspection of the bipartite network (Figure 4 and Figure S2) reveals that the node representing Gilgamesh is located separately from the node of the BM cluster. The BM node itself is located remotely from other nodes of the network.

Also, the Gilgamesh's node does not form a separate dense subnetwork with any genome nodes, which is less connected to the rest of the network, which is typical of cluster genome nodes. Visually, this can be seen in Figure 5 and Figure S3, where only the edges are left, part of which is the Gilgamesh's node and/or nodes of clusters of homologous proteins with which it is connected.

In total, a complete bipartite network was formed by 5642 nodes and 58993 edges. The diameter of the network is 6, the radius is 3. The remaining network characteristics are presented in Table S2.xlsx.

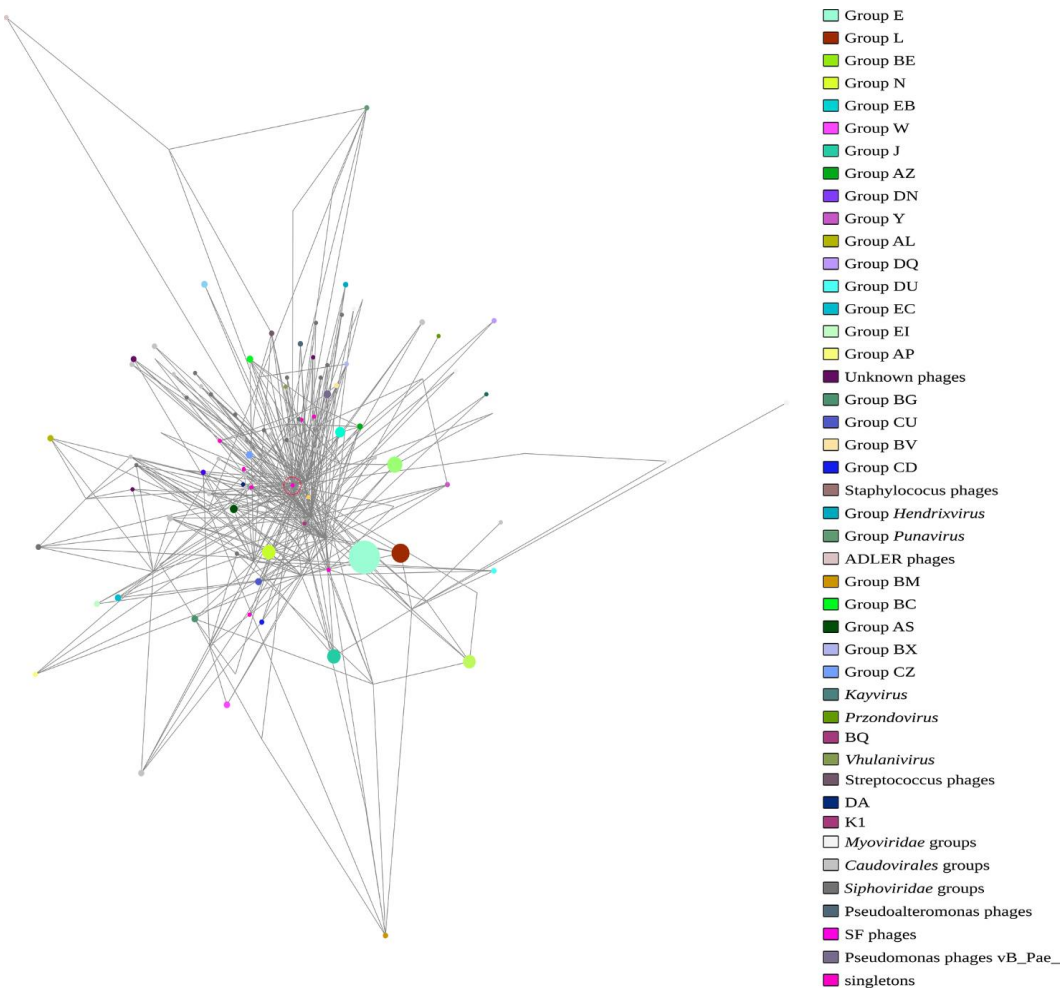

Fig. 5. Collapsed bipartite network reflecting genome nodes and edges of homologous clusters of Gilgamesh and the genome node of Gilgamesh. Genome nodes are marked with colored circles, cluster nodes are not marked. The node of Streptomyces phage Gilgamesh is marked with red ellipse. A whole network is represented in Figure S3.

The central nodes (whose eccentricity is equal to the radius of the network) are two genome nodes, namely Streptomyces phage Gilgamesh and Gordonia phage Schmidt. The cluster of homologs of the large terminase subunit (a protein of phage Gilgamesh was not included into this cluster) had the largest number of edges (358) among the nodes of clusters of homologous translated ORFs. Portal protein (a protein of phage Gilgamesh was included into this cluster) was next according to the number of nodes (356).

While considering the homologs of the translated ORFs, as well as the annotation of the genome of bacteriophage Gilgamesh, it was found that proteins associated with the formation of the plasmid form are present. Taking into account that some of the homologs of these proteins were found in bacteriophages, for which the existence of the plasmid form was determined (e.g, Streptomyces phage ZL12 [42]), a search for the homologous sequences of the translated ORFs of Gilgamesh genome among the plasmids from the nr database was performed. We focused on plasmids characterized as complete genome, which contained at least 5 homologs of Gilgamesh proteins. As a result, 25 plasmids were found (see Table S1.xlsx) satisfying the requirements. Their sequences were annotated or reannotated. Table S1 demonstrates the lengths and access numbers for this plasmids.

Streptomyces phage Gilgamesh and related plasmids

Repeated clustering with previously used genomes of phages and plasmids was carried out. There were 15114 clusters. As before, core- and softcore-genome were not formed.

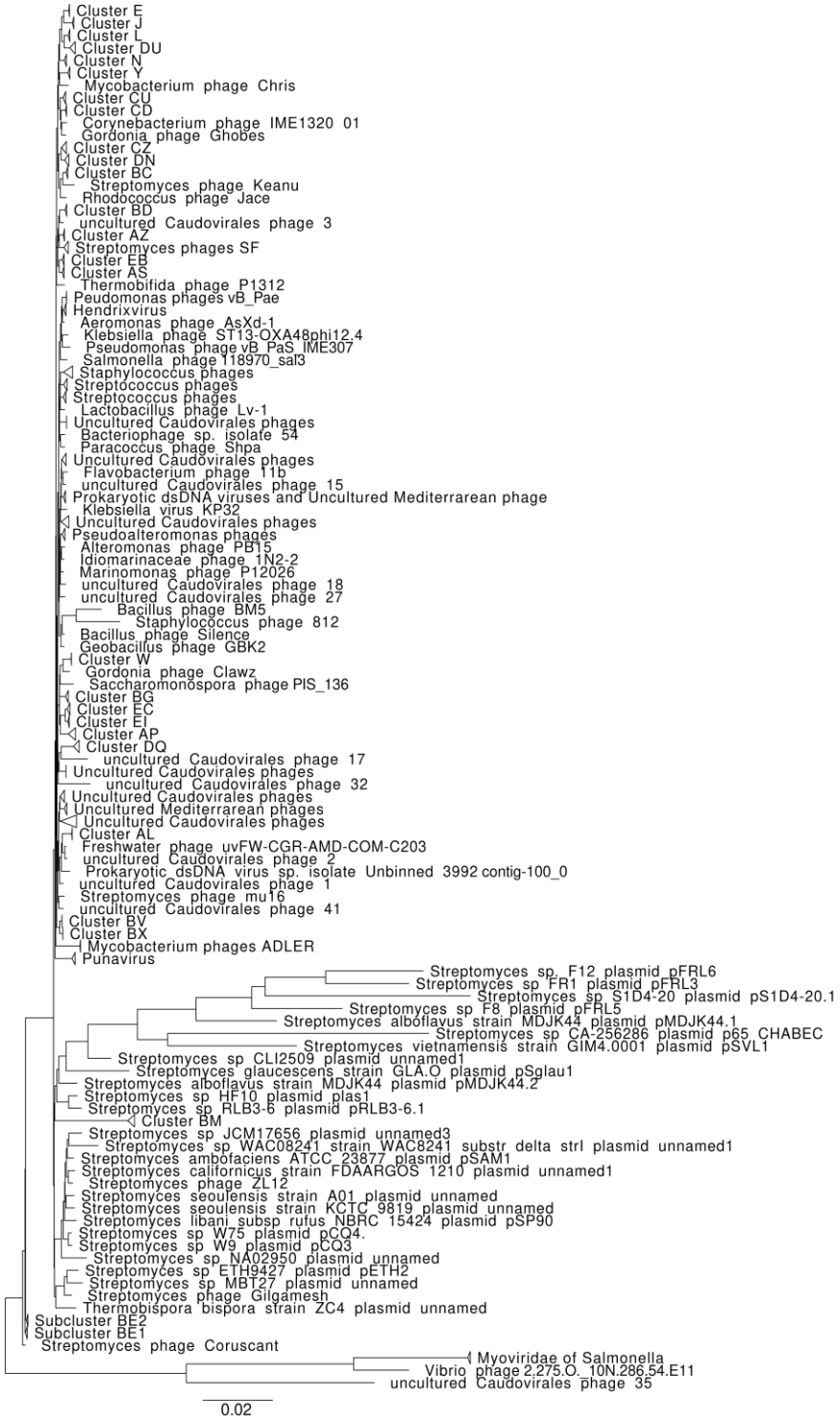

Fig. 6. A collapsed pan-genome tree for Streptomyces phage Gilgamesh and phages and plasmids containing homologs of its proteins. The scale bar represents the number of expected substitutions per site under the binary GTR2+FO+ASC+R5 substitution model. The complete tree is given in Figure S4.

According to the clustering results, 81 translated ORFs of the Gilgamesh genome had homologs among the translated ORFs of plasmids. Totally, 85 translated ORFs of the Gilgamesh genome had homologs among the translated ORFs of all genomes used (see Data S2.zip). It is worth noting that some of the homologs of one protein did not form a common

cluster, which is most likely due to the features of the COGtriangle algorithm [18], which requires at least 3 sequences to be similar to each other in terms of established parameters in order to form a cluster.

Then a common pan-genome tree has been obtained (Figure 6 and Figure S4) (best model: GTR2+FO+ASC+R5, best score: -113605.489).

As a result, the genome of bacteriophage Gilgamesh was turned out into a clade formed by 3 sequences, among which there were 2 plasmids: *Streptomyces* sp. ETH9427 plasmid pETH2, *Streptomyces* sp. MBT27 plasmid unnamed. If we take a common clade of the higher level, it consists of 13 plasmids and 2 phages. Gilgamesh was one of these phages and Streptomyces phage ZL12 was second. It has been already mentioned previously [42] that ZL12 is a phage-plasmid. Phage-plasmids are extrachromosomal elements that possess the genes of both phages and plasmids and are capable of forming a virion and existing in the host cell in plasmid form [43].

Thereafter, a bipartite network with all genomes, including plasmids has been created (Figure 7 and Figure S5).

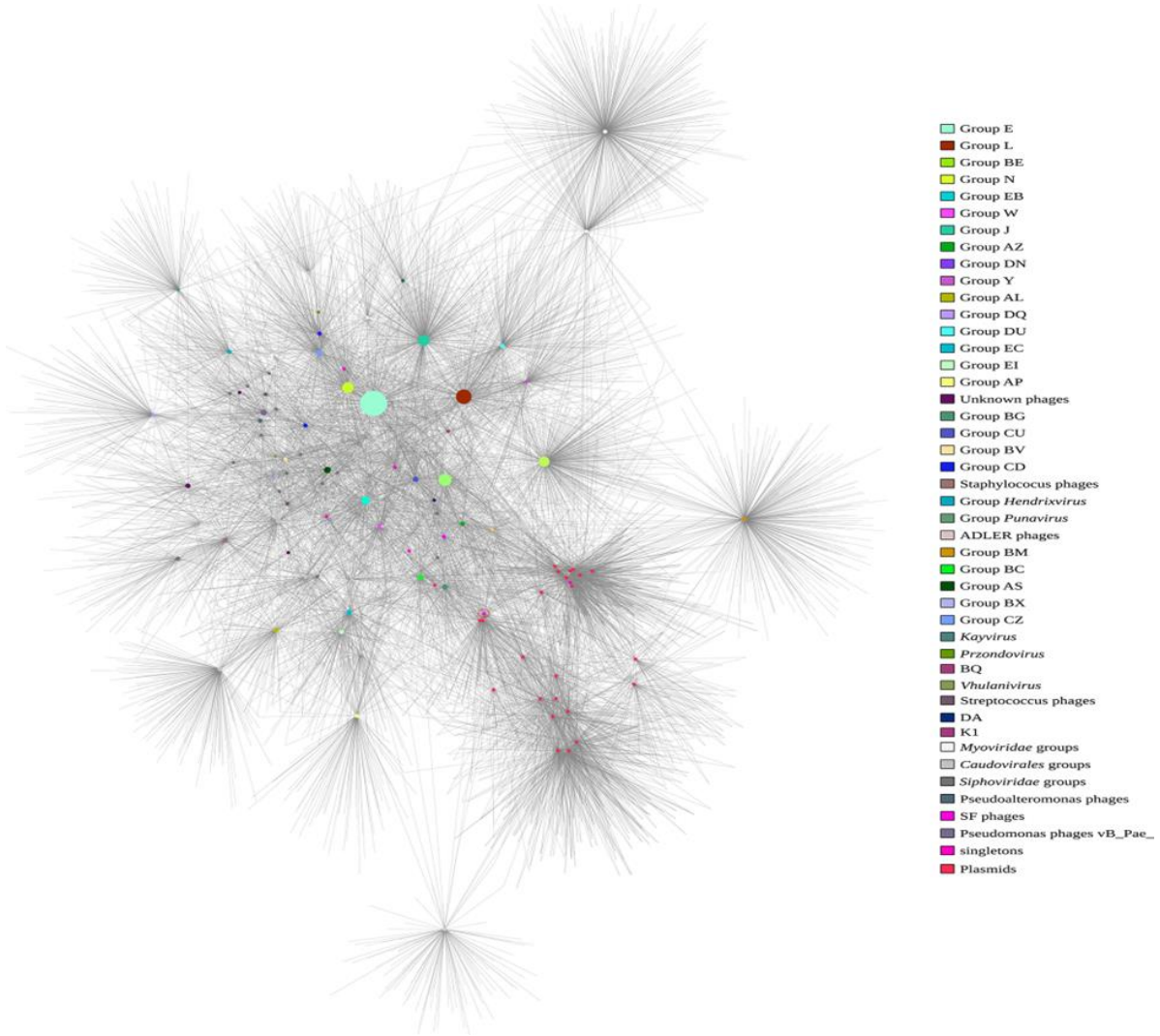

Fig. 7. A collapsed bipartite genome network of Streptomyces phage Gilgamesh and phages and plasmids containing its homologs. The nodes of separate genomes and groups of viruses are marked with colored circles; the nodes of homologous clusters are not marked. The node of phage Gilgamesh is marked with the red ellipse. The whole network is shown in Figure S5.

The number of nodes was equal to 6482, the number of edges was 61614, and the network diameter was 6, the radius was 3. Table S3.xlsx represents other characteristics for the

network and separate nodes. When looking at the subnetwork generated by genome nodes of Gilgamesh, homologous clusters of its proteins, and the nodes of other genomes (Figure 8 and Figure S6), a network densification formed by genome nodes of Gilgamesh, *Streptomyces* sp. MBT27 plasmid unnamed, *Streptomyces* sp. ETH9427 plasmid pETH2 and their edges is viewed.

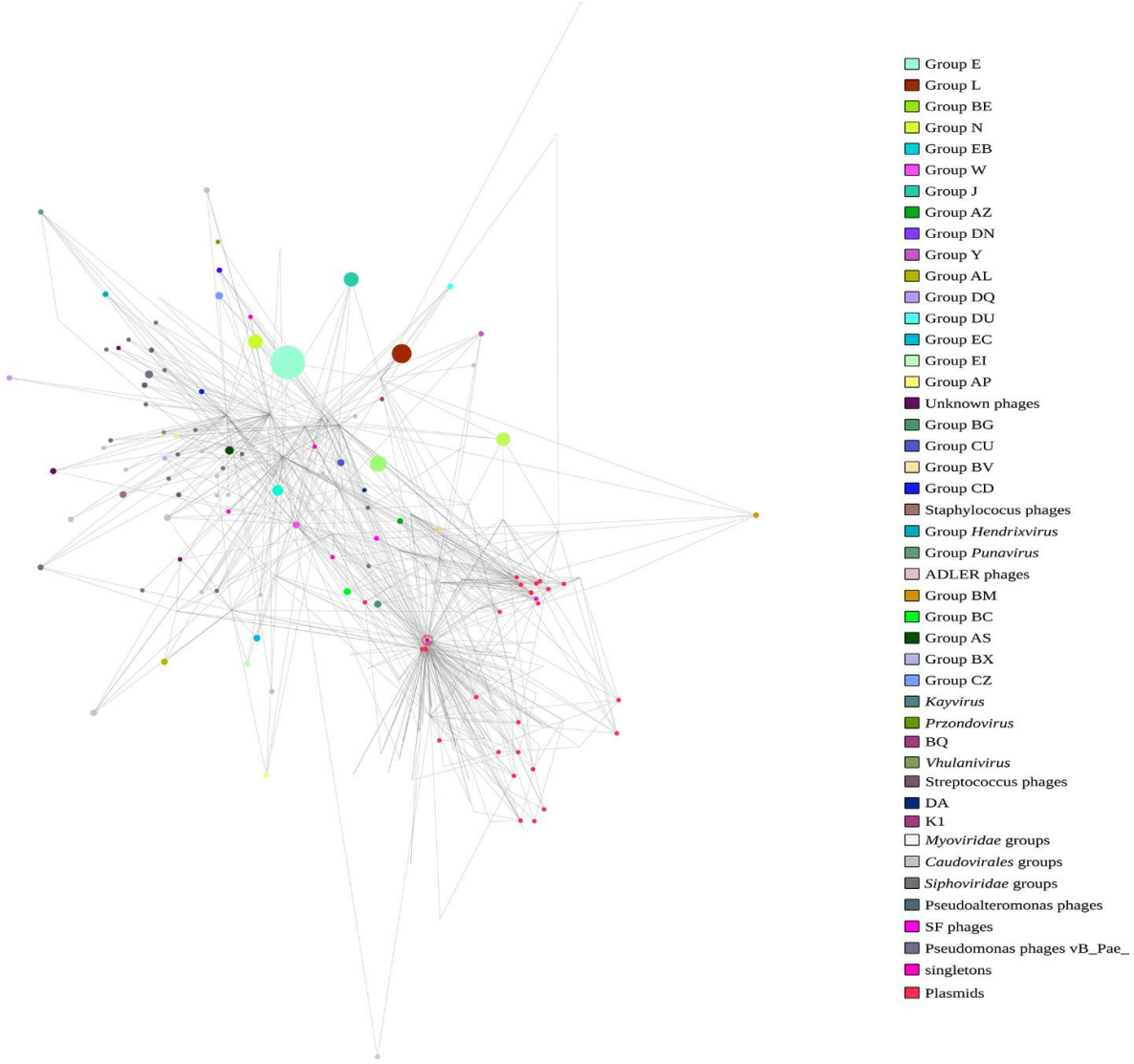

Fig. 8. A collapsed bipartite network reflecting nodes of genomes and edges of homologous Gilgamesh clusters and the node of the Gilgamesh genome. The nodes of phage genomes are marked with colored circles; the nodes of clusters are not marked. The node of Streptomyces phage Gilgamesh is marked with the red ellipse. The whole network is given in Figure S6.

The genome node of bacteriophage Gilgamesh with eccentricity equaled to the network radius was the central node. The cluster of homologs of the large terminase subunit (Gilgamesh protein was not included in this cluster) had maximum number of edges (388) among cluster nodes of homologous translated ORFs. In second place in the number of nodes (359) there was the portal protein (Gilgamesh protein was not included in this cluster).

Effect of plasmids addition on bipartite network characteristics

A comparison of metrics for two networks (without and with plasmids) led to the following results. Total average clustering coefficient for the network without plasmids was approximately 0.43, for the network with plasmids it was 0.4. The average clustering coefficient of genome nodes for the network without plasmid genomes was approximately 0.077, for the network with plasmids it was 0.072. The average clustering coefficient of

homologous cluster nodes for the network without plasmids was approximately 0.464, for the network with plasmids it was 0.436. The diameters and radiuses of the networks were the same and equal to 6 and 3, respectively. The central nodes of the networks were described above, in both cases one of or the only central node was the node of the genome of Streptomyces phage Gilgamesh.

Table 1. Increase or decrease in metrics for groups and singletons of phage depends on the number of added plasmids. ↓ - value decrease, ↑ - value increase, ↑↓ - some of values decreased and some of values increased. Singletons are represented separately from each other

Group	Betweenness centrality	Closeness centrality	Degree centrality	Clustering coefficient
Arquatrovirinae	↓	↓	↓	↓
Cluster AL	↓	↑↓	↓	↓
Cluster AP	↓	↓	↓	↑
Cluster AS	↓	↓	↓	↓
Cluster AZ	↓	↓	↓	↓
Cluster BC	↓	↑	↓	↑
Cluster BD	↓	↓	↓	↓ (BD3↑)
Cluster BE	↓	↓	↓	↓
Cluster BG	↓	↑	↓	↓
Cluster BM	↓	↓	↓	↑↓
Cluster BQ	↓	↑	↓	↓
Cluster BV	↓	↓	↓	↓
Cluster BX	↓	↑	↓	↑
Cluster CD	↓	↓	↓	↓
Cluster CU	↓	↓	↓	↑↓
Cluster CZ	↓	↓	↓	↓
Cluster DA	↓	↓	↓	↓
Cluster DN	↓	↑	↓	↑↓
Cluster DQ	↓	↓	↓	↓
Cluster DU	↓	↓	↓	↓
Cluster E	↓	↓	↓	↓
Cluster EB	↓	↓	↓	↓
Cluster EC	↓	↓	↓	↓
Cluster EI	↓	↓	↓	↓
Cluster J	↓	↓	↓	↓
Cluster L	↓	↓	↓	↓
Cluster N	↑↓	↑↓	↓	↓
Cluster W	↓	↑	↓	↓
Cluster Y	↓	↓	↓	↓
Hendrixvirus	↓	↓	↓	↑↓
Kayvirus	↓	↓	↓	↑
Phietavirus	↓	↓	↓	↑↓
Przondovirus	↓	↓	↓	↓
Punavirus	↓	↑↓	↓	↓
Subcluster K1	↓	↑	↓	↓
Triavirus	↓	↓	↓	↓
Vhulanivirus	↓	↓	↓	↓
Gordonia phage Clawz	↓	↓	↓	↓
Mycobacterium phage Adler	↓	↑	↓	↓
Rhodococcus phage Jace	↓	↓	↓	↓
Streptomyces phage Gilgamesh	↑	↓	↑	↓
Streptomyces phage SF1	↓	↓	↓	↓
Streptomyces phage SF3	↓	↓	↓	↓
Streptomyces phage ZL12	↑	↑	↑	↓
Streptomyces phage mu16	↓	↓	↓	↑
Gordonia phage Gudmit	↓	↑	↓	↑

Table 1 presents changes in metrics of the known groups (families, genera, clusters and singletons). Table S4.xlsx presents metrics of each node of phage genomes before and after addition of plasmids.

Figure 9 shows the box plots of metric values for all nodes of bacteriophage genomes before and after addition of plasmids. Figure S7 shows changes in the metric values of all nodes of bacteriophage genomes before and after addition of plasmids.

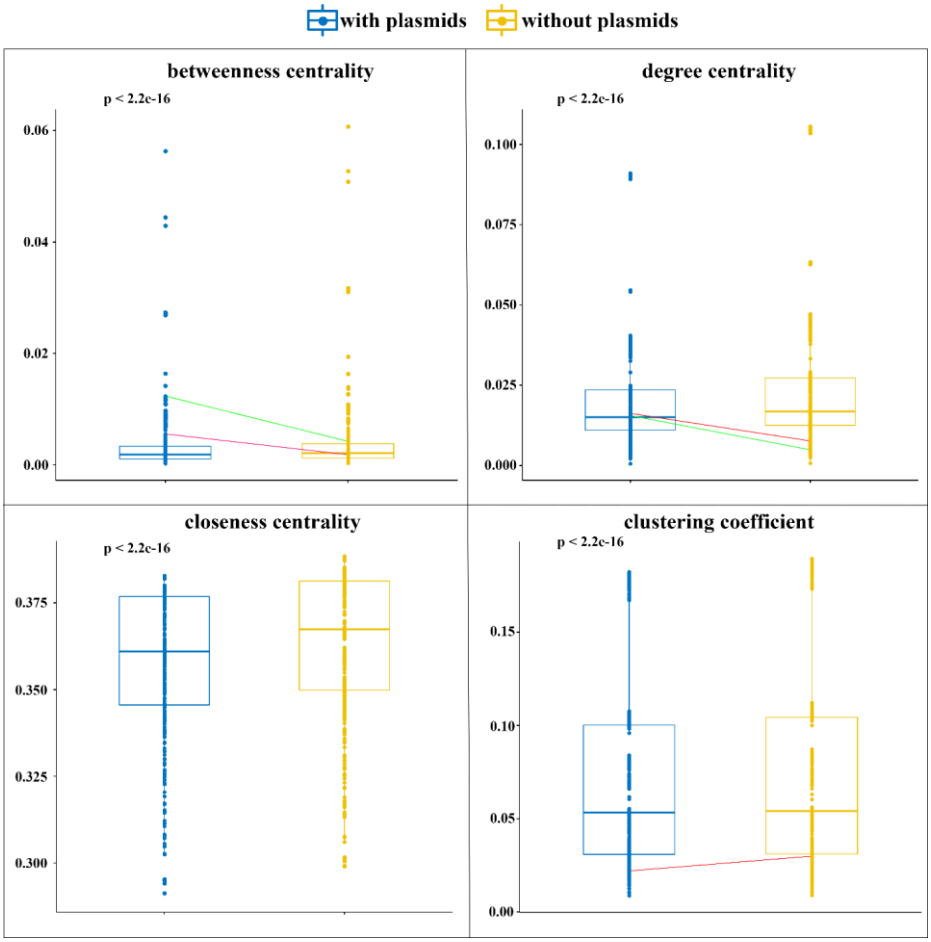

Fig. 9. The box plots of metrics for genome nodes of bacteriophages depends on whether plasmids are added or not. The colored lines denote the most substantial changes. The red line indicates a change in the index for a node of Streptomyces phage Gilgamesh, the green line depicts Streptomyces phage ZL12.

The Wilcoxon test revealed that addition of plasmids had a statistically significant effect on all measured network index values ($p < 2.2e-16$).

Having studied the genome node metrics, it was found that the greatest changes occurred in two nodes such as Streptomyces phage Gilgamesh and Streptomyces phage ZL12. Probably, it is so because the proteins of the Gilgamesh were used as reference while searching for plasmids and the ZL12 is phage-plasmid.

Phages-plasmids

A phylogenomic tree with plasmids and phages was constructed (Figure 10 and Figure S8).

According to the results, the clade mentioned earlier contains not only plasmids and two genomes of the phages but also the genome of Saccharomonospora phage PIS 136. The difference between the pan-genome tree and phylogenomic tree in this case could be because PIS 136 has many genes that show no similarity to other genomes. As a result, this virus is located separately on the pan-genome tree. Moreover, with an increase in the values of two

network metrics regarding the node of the genome of PIS 136, a betweenness centrality value, and a degree centrality value (in this case, it reflects the number of homologous clusters), we can assume that the phylogenomic tree in this case better matches the evolutionary relation of PIS 136 to the genomes of phage-plasmids mentioned above.

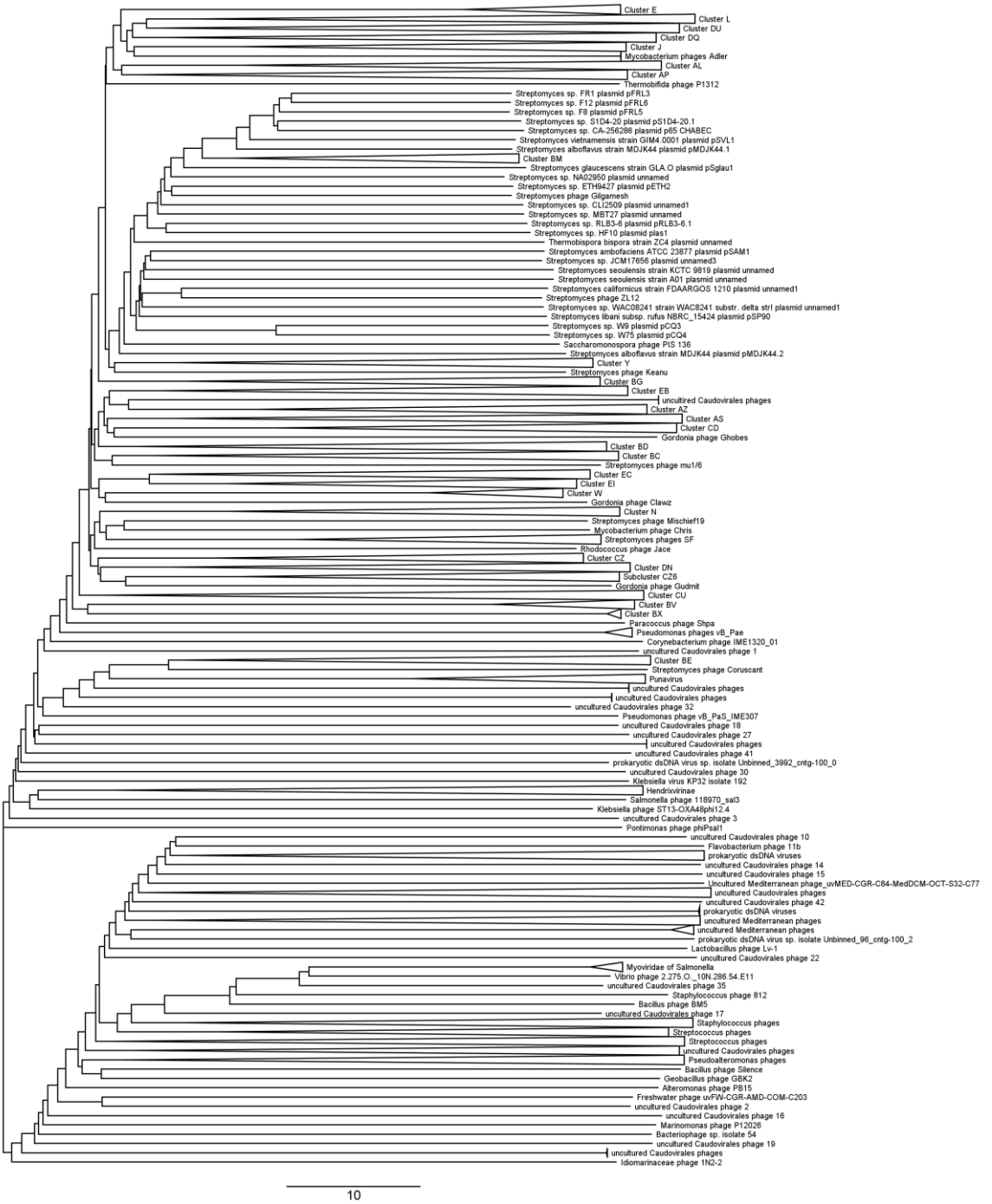

Fig. 10. A collapsed phylogenomic tree for Streptomyces phage Gilgamesh and phages and plasmids containing its homologs. The tree constructed with the neighbor joining method [25] from the pairwise distance matrix calculated for 572 sets of 10-mers. The scale bar shows the Sorensen distance as a percentage. A complete tree is shown in Figure S8.

Further, the search for homologous proteins between the phages and plasmids which were included in the common clade was carried out to determine whether they belonged to phage-plasmids or not.

Having explored a similarity of plasmid proteins to the Gilgamesh proteins, we have found that *Streptomyces* sp. MBT27 plasmid unnamed, *Streptomyces* sp. ETH9427 plasmid

pETH2 have homologs of all known Gilgamesh morphogenesis proteins. Figure 11 demonstrates that these plasmids exhibit high similarity of the sequences to Gilgamesh; moreover, they have a similar system which is necessary for the formation of plasmid, its segregation etc. (DNA translocase, RusA-like resolvase, ParB, ParA, Cre) but is partly different in homology of the proteins and location of its genes. Additionally, these plasmids, as well as Gilgamesh, contains a large ORF (near the gene of portal protein) with unknown functions and low homology between the plasmids. Only pETH2, according to annotation, has adenine-specific DNA methylase domain, N12 class, helicase superfamily C-terminal domain, prolipoprotein diacylglyceryl transferase, Superfamily I DNA or RNA helicase, DNA polymerase III subunits gamma and tau (in this case the size of ORF is approximately 23 kbp).

Gilgamesh has the following homologs with the other plasmids, which may point to its possible ability to exist in plasmid form: DNA translocase, RusA-like resolvase, Cre recombinase, ParB, ParA, tyrosine integrase.

Phage ZL12, as well as Gilgamesh (Figure 11), has genomic regions that are similar to plasmids located in the same dense subnetwork. A comparison between genome of phage-plasmid ZL12 and 9 *Streptomyces* plasmids is shown in Figure S9. Also, these plasmids have all virion morphogenesis proteins, and typical plasmid-associated proteins (relating to the formation, segregation, replication, etc.).

ZL12 has the homologs with the other plasmids, that point to its ability to exist in plasmid form. They are resolvase, Cre recombinase, tyrosine integrase, ParB, ParA.

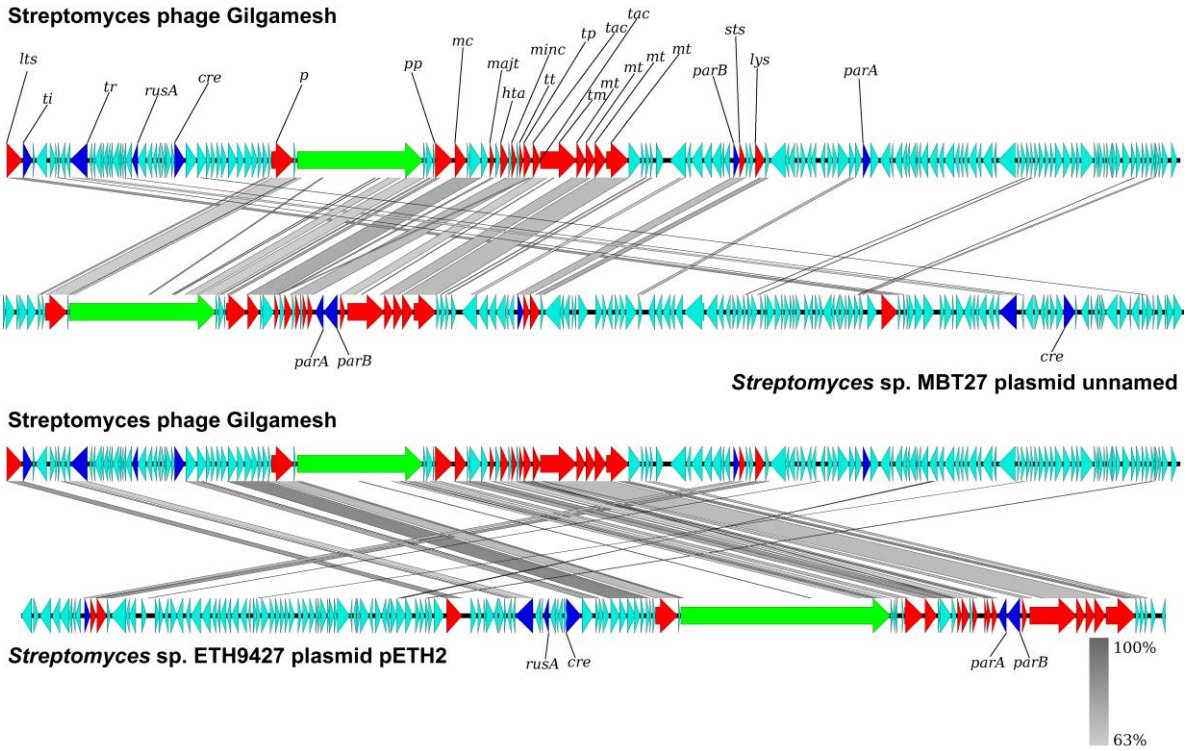

Fig. 11. A comparison between genome of phage Gilgamesh and 2 *Streptomyces* plasmids. Red color denotes the known ORFs of virion morphogenesis; blue – plasmid formation, segregation etc.; green – giant ORFs; *lts* – large terminase subunit, *ti* – tyrosine integrase, *tr* – DNA translocase, *rusA* – RusA-like resolvase, *cre* – Cre protein, *p* – portal protein, *pp* – prohead protease, *mc* – major capsid protein, *majt* – major tail protein, *hta* – head-to-tail adaptor, *minc* – minor capsid protein, *tt* – tail terminator, *tp* – tail protein, *tac* – tail assembly chaperone, *tm* – tape measure protein, *mt* – minor tail protein, *parB* – ParB-like dsDNA partitioning protein, *sts* – small terminase subunit, *lys* – endolysin, *parA* – ParA-like dsDNA partitioning protein.

Phage ZL12 and FDAARGOS 1210 plasmid unnamed1, WAC08241 strain WAC8241 substr. delta strI plasmid unnamed1, KCTC 9819 plasmid unnamed, JCM17656 plasmid

unnamed3, pSAM1, pCQ4, pCQ3, pSP90, A01 plasmid unnamed form a common branch on a phylogenomic tree, a pan-genome tree, and the nodes of these genomes form dense subnetwork on bipartite network.

Saccharomonospora phage PIS 136 has 1 common homolog with the known proteins of ZL12 morphogenesis and plasmids – the large terminase subunit.

PIS 136 has homolog with other plasmids which may point to the ability of this phage to exist in plasmid form. It is XerC integrase, however, no homologs of proteins that could point directly to this ability were found.

Interestingly, PIS 136 has DNA methylase, which is similar to cytosine-specific methyltransferase in the phages of the genus *Pseudotevenvirsus* (it is specific for the representatives of this genus). In addition, the gene of this methyltransferase is located between the genes of DNA polymerase and helicase. It has been shown in our previous study [44] that this region usually includes genes for the synthesis of some non-canonical DNA bases and their modifications. Taking into account that in *Pseudotevenvirsus* this gene is located in the other region, further study of the homologs of the genes of this methyltransferase and the regions of bases synthesis may extend our knowledge about diversity of phages due to non-canonical bases.

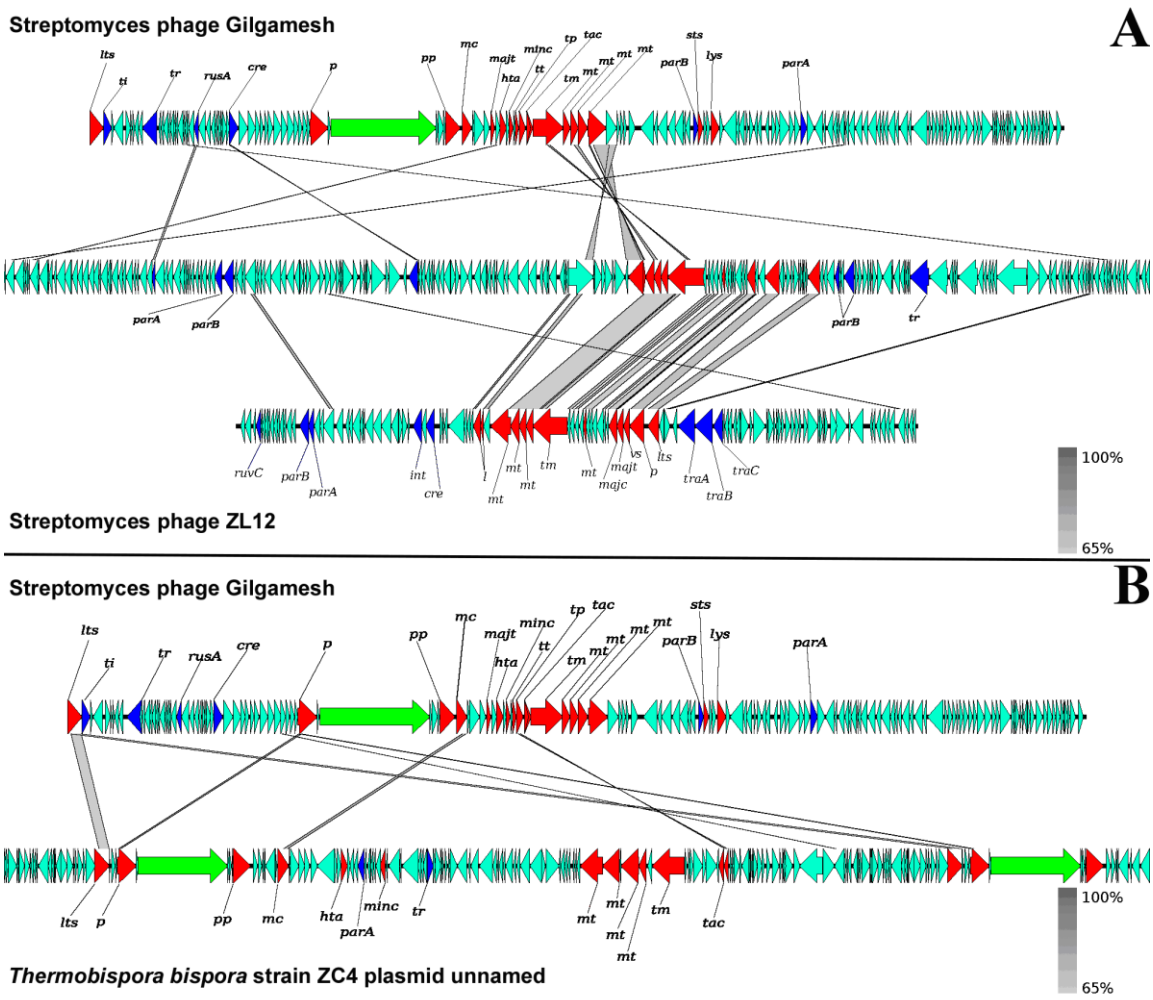

Fig. 12. A comparison between genomes of Streptomyces phage ZL12 and/or Streptomyces phage Gilgamesh and plasmids. The abbreviations are given in the legend to Figure 11. Red color marks the known (found) genes of virion morphogenesis, blue depicts genes associated with the plasmid part, green denotes large ORFs. A – *Streptomyces* sp. NA02950 plasmid unnamed (in the center). B – *Thermobispora bispora* strain ZC4 plasmid unnamed, *ruvC* – Holliday junction resolvase RuvC, *l* – endolysin, *vs* – virion structure protein, *traA*, *traB*, *traC* – conjugal transfer protein.

Phages from the BM cluster have large terminase subunits which are similar to ones of Gilgamesh and plasmids, as well as endolysins, which are similar to ZL12 and plasmid endolysins.

The genomes of *Streptomyces* sp. NA02950 plasmid unnamed and *Thermobispora bispora* strain ZC4 plasmid unnamed are interesting cases (Figure 12).

Streptomyces sp. NA02950 plasmid unnamed on the bipartite network (Figure 7) is located between two dense subnetworks formed by the genome nodes of phage Gilgamesh, phage ZL12 and plasmids. This is also seen in the homology of virus morphogenesis genes and plasmid part (Figure 12,A). Some genes have a high homology to the ones typical of phage ZL12 and related plasmids and other genes were similar to Gilgamesh and related plasmids. This fact may indicate a chimerical origin of NA02950 plasmid.

On the network, *Thermobispora bispora* strain ZC4 plasmid unnamed was located remotely from plasmids and phages-plasmids, closer to the network of other phages. ZC4 showed the highest similarity to the genes of morphogenesis and plasmid part of phage Gilgamesh, but there were not so many homologs (Figure 12,B). Most of the morphogenesis genes exhibited similarity to other phages. It was found that this plasmid can be repeated, probably because of incorrect annotation, due to a suggested mechanism of the DNA packaging into a capsid (headful packaging). This plasmid as well as Gilgamesh has a long ORF next to the portal protein gene.

The presence of proteins relating to morphogenesis of bacteriophages may indicate that the studied plasmids are also phages or defective phages, taking into account that most of them are isolated from the bacteria of the genus *Streptomyces*, have other homologs with the bacteriophages (totally 345 homologous clusters with all the phages). Especially in this case it is worthy to mention plasmids *Streptomyces* sp. MBT27 plasmid unnamed, *Streptomyces* sp. ETH9427 plasmid pETH2, which have the greatest amount of protein homologs with Gilgamesh and 9 plasmids forming the common branch with ZL12 described earlier.

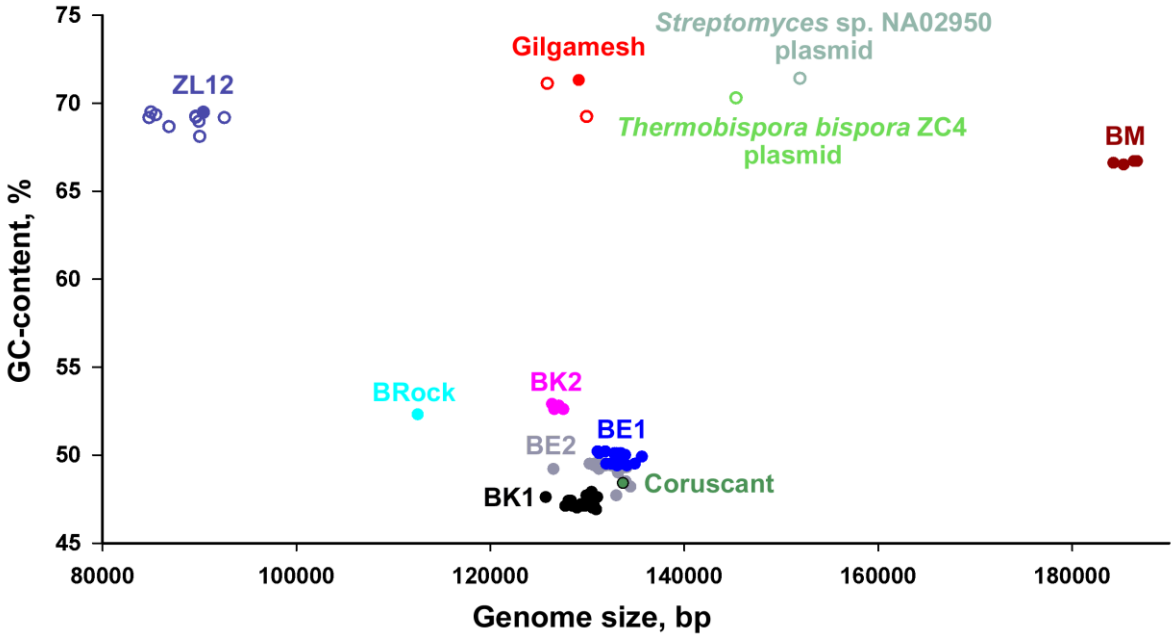

Fig. 13. The size and GC-content of plasmids and related phage genomes. The related sequences are marked with the same color. The plasmids are indicated by empty circles.

It is also notable that we failed to find homologs of the proteins associated with morphogenesis of phages in plasmids that form a separate clade above the clade of BM cluster on the pan-genome tree. Nevertheless, on the phylogenomic tree these plasmids form a common clade with genomes of Gilgamesh and the related plasmids (Figure 10). On the

network they are located separately from plasmids without forming a dense subnetwork (Figure 7 and Figure S5).

We compared GC-content and size of the plasmids, which are related to phage ZL12 and phage Gilgamesh, as well as *Streptomyces* sp. NA02950 plasmid unnamed, *Thermobispora bispora* strain ZC4 plasmid unnamed, with phage genomes used previously (Figure 13).

It turned out that the related ZL12 plasmids and the phage itself had a very high percent similarity to GC-content and similar size. The related plasmids of Gilgamesh also had similar GC-content and size. In the case of *Streptomyces* sp. NA02950 plasmid unnamed, *Thermobispora bispora* strain ZC4 plasmid unnamed, as well as phages-plasmids and plasmids described above, had a high GC-content and sufficiently large sizes, which were different from the other phages and plasmids.

Another interesting point which is worth to mention is bacteriophage Gilgamesh has circularly permuted ends (this characteristic was taken from Actinobacteriophage Database). Moreover, a large terminase subunit of this phage is similar to terminases of phages of the genus *Punavirus* and they can form a single clade on a phylogenetic tree (see Figure S10) together with the phages of AL cluster (infecting *Arthrobacter*) which also have circularly permuted ends, closely related plasmids and two phages from metagenomes. Escherichia virus P1 is a type phage of this genus. Thus, we can assume that bacteriophage Gilgamesh packages its DNA by a headful mechanism. Taking into account the characteristic features of the ends, the packaging, and the ability to exist in plasmid form, it is possible to assume that Gilgamesh is capable of transduction, which is at least general one, typical for *Punavirus*. In addition, given that closely related plasmids *Streptomyces* sp. MBT27 plasmid unnamed, *Streptomyces* sp. ETH9427 plasmid pETH2 show similarity to Gilgamesh morphogenesis proteins, we may suggest these phage-plasmids also have the ability to transduce. Moreover, genomes sizes of this phage and plasmids are similar (approximately 125–130 kbp, see Figure 13).

DISCUSSION

For this study, we initially considered phage genomes with the largest length which were amongst phages infecting *Streptomyces*. It was found that Streptomyces phage Gilgamesh having a high GC-content but there were no closely related viruses observed within *Streptomyces* phages.

Interestingly, the sequences of plasmids related to Gilgamesh and included in this study as well as, according to known data, the bacteria, from which one of the plasmids has been isolated, have similar GC-content. Moreover, analysis of *Streptomyces* genomes revealed that representatives of this genus display high GC-content. The size of Gilgamesh and related plasmids is over 100 kbp, it is similar to one of two patterns of sizes of the phage-plasmids reported by [43]. This size, as it is predicted, is due to the presence of genetic modules both phages and plasmids.

For the first time, the existence of the plasmid form was observed for the lambda bacteriophage [45, 46]. Later, the plasmid prophage was detected and studied for bacteriophage P1 and related viruses [47–49], as well as for satellite bacteriophage P4 [50, 51]. In addition, linear lambdoid phage-plasmid N15 was discovered and investigated [52–54]. A number of artificial phage-plasmids were constructed for practical and research purposes [55–57]. Recently, the global spread of the natural phage-plasmids in different groups of bacterial plasmids and bacteriophages was shown [43].

When analyzing Gilgamesh, it was found that it has homologs of genes which are necessary for the formation of a plasmid form, segregation, etc., moreover, the plasmids most related to it have similar plasmid genes and genes of phage morphogenesis (Figure 11). As a result, we suppose that Gilgamesh and 2 related plasmids (*Streptomyces* sp. MBT27 plasmid unnamed, *Streptomyces* sp. ETH9427 plasmid pETH2) are the phage-plasmids. It was also

reported that *Streptomyces* sp. MBT27 plasmid unnamed may belongs to phage-plasmids [43].

The point of interest was the presence of giant ORFs in Gilgamesh and related plasmids. According to the annotation of this ORF in pETH2 it contains domains of proteins associated with replication, however, little homology to other giant ORFs located in the same region does not make it possible to talk about the presence of the same domains in them. Moreover, the network analysis showed that the addition of plasmids, that have the homologs of Gilgamesh proteins, into the pan-genome network of phages, had the greatest impact on the nodes of the Gilgamesh and Streptomyces phage ZL12 genomes. According to the data, Streptomyces phage ZL12 is a phage-plasmid [42]. The 9 plasmids have the most related sequences to it. After analysis of these plasmids, we found that they all have phage morphogenesis genes, as well as plasmid genes, similar to the known ZL12 genes (Figure S9). It was shown that one of the plasmids related to ZL12, namely KCTC 9819 plasmid unnamed, belongs to phage-plasmids [43]. The possibility of existence in the form of a lytic phage was detected for the pCQ4 plasmid [58]. It is also worth noting that all these plasmids have a high GC-content.

As a result, we can assume that the sequences forming on the pan-genome tree a separate clade, including the genomes of Gilgamesh, ZL12 and related plasmids, are the sequences of phage-plasmids or descendants of phage-plasmids. In addition, the possibility of existence of these phages and plasmids in the form of phage-plasmids is indicated by the following.

For genomes related to Gilgamesh, this may be indicated by a close homology of the large terminase subunits with proteins of representatives of the genus *Punavirus* which are phage-plasmids. Besides, Gilgamesh, as well as these representatives, has circular permutations, which altogether points to a similar mechanism of DNA packaging into a capsid. Also, Gilgamesh and related plasmids have Cre protein which is required to form a circular shape; it is also typical for *Punavirus*. In addtion, they have ParA and ParB homologs necessary for the plasmid segregation. However, it is interesting that, genes of these proteins are located in a different genomic regions, which is more typical for ZL12 and its related plasmids. This difference between related genomes was demonstrated in [59], using examples of related actinobacteriophages with missing/present ParAB cassettes. It was suggested that differences in these systems in closely related viruses may be necessary to diversify segregation systems in order to avoid incompatibility [59].

The ZL12 phage has a locus for conjugal transfer, which includes *traA*, *traB*, *traC* genes [42]. We demonstrated that most of its related plasmids have a TraA homolog (Figure S9), however, TraB and TraC homologs were observed only in the most related plasmids such as *Streptomyces seoulensis* strain A01 plasmid and *Streptomyces libani* subsp. *rufus* NBRC 15424 plasmid pSP90. In other plasmids, it was also possible to detect the presence of ORFs with similar location and the same length of TraB and TraC, but we failed to identify homology. This situation may be explained by the high frequency of recombination of these genes. Moreover, related plasmids had most of the genes homologous to ZL12 genes of virion morphogenesis, as reported earlier.

Having analyzed data on genes, phylogenomic and pan-genome network analyses, it can be concluded with great probability that Gilgamesh, ZL12 and their most related plasmids represent a group of phage-plasmids. In addition, the recombination processes between the Gilgamesh and ZL12 groups can be indicated by the cases with the genes of the ParA, ParB proteins in plasmids most related to Gilgamesh, which were mentioned above.

The genome nodes of the AL cluster showed a partial increase in closeness centrality. The representatives of this cluster have Par-like genes which are necessary for segregation of plasmids [60]. Moreover, they also have circular permutations of the ends and, according to the results of phylogenetic analysis of the large subunit of terminase, they have similar sequences of this protein with the proteins of *Punavirus* and Gilgamesh. Altogether, this may indicate that they belong to phage-plasmids.

The nodes of the AP cluster exhibited an increase in clustering coefficient. Previously was reported that these phages possess ParB-like proteins [60].

In the nodes of the BC cluster clustering coefficient and betweenness centrality became greater. The cluster representatives have circularly permuted ends (according to Actinobacteriophage Database [13]. Interestingly, one of the representatives of this cluster (phage SV1) was the object of the study of transduction [61]. In the nodes of the W cluster closeness centrality increased. The representatives of this cluster contain Par-like proteins, the ends with circular permutations.

The BM cluster and *Punavirus* are described above. It should be noted that clustering coefficient of the BM cluster increased slightly and *Punavirus* exhibited a negligible change in closeness centrality. Saccharomonospora phage PIS 136 was also described above, betweenness centrality and degree centrality of this node increased. It was showed that this phage has a circular genome [62], that, taking into account previously reported data, could give evidence in support of supposition that this phage is a phage-plasmid.

Betweenness centrality of the node of Thermobifida phage P1312 became higher. According to [63] this phage has circular permutations of the ends, a headful packaging strategy, a chromosome segregation gene, and circular DNA.

In the nodes of the related phages Mycobacterium phage Adler, Mycobacterium phage ADLER F1725 closeness centrality increased. ParA-like protein is detected in these phages. Thus, in some phages with increased metrics, we managed to identify similar DNA packaging systems, genes associated with plasmid modules. Especially, it is worth noting, the AL cluster and the Saccharomonospora phage PIS 136, which according to the results of other analyses (phylogenetic and phylogenomic, respectively), have formed common clades with the phage-plasmids analyzed earlier. These viruses can also be phage-plasmids.

The mechanism of generalized transduction is primarily related to the way DNA is packaged into a capsid. Classical examples: generalized transduction by T4, P1 mutants etc. [11, 64]. These phages have the headful packaging mechanism of the DNA packaging into a capsid. It was shown [65] that with the phylogenetic analysis of the terminase, it is possible to predict the mechanism of DNA packaging into a capsid. As it was demonstrated (Figure S10), Gilgamesh and related plasmids, as well as phages of the AL cluster have the greatest homology to *Punavirus* phages. Thus, taking into account their circular permutations of the ends and the presence of genes similar to Cre, these phages share a similar packaging mechanism with *Punavirus*. In this context, it is most likely that these phages are transducers.

In addition, given the size of the Gilgamesh genome and related plasmids, the heads of these viruses can accommodate sufficiently large DNA fragments. As a result, these phage-plasmids are very promising from the point of view of genetic manipulation with *Streptomyces*. Moreover, taking into account their GC-content, evolutionary relations with other viruses and plasmids, the presence of genes which are necessary for Cre-Lox recombination, that can be used as a genetic tool, these phages are on the whole of interest to genetic engineering.

In sum, the research has revealed a group of related phage-plasmids (or defective phages) with various gene transfer mechanisms; genes share large homology with *Streptomyces* and similar GC-content with ones. These phages were helpful to find out new opportunities in the coevolutionary relationships between phages and bacteria.

The study was supported by the Russian Science Foundation grant No. 22-25-00669, https://rscf.ru/project/22-25-00669/

REFERENCES

1. Subramani R., Aalbersberg W. Culturable rare *Actinomycetes*: diversity, isolation and marine natural product discovery. *Appl. Microbiol. Biotechnol.* 2013. V. 97. P. 9291–9321. doi: 10.1007/s00253-013-5229-7
2. Mohammadipanah F., Wink J. *Actinobacteria* from arid and desert habitats: diversity and biological activity. *Front. Microbiol.* 2016. V. 6. Article № 1541. doi: 10.3389/fmicb.2015.01541
3. Azman A.S., Othman I., Velu S.S., Chan K.G., Lee L.H. Mangrove rare actinobacteria: taxonomy, natural compound, and discovery of bioactivity. *Front. Microbiol.* 2015. V. 6. Article № 856. doi: 10.3389/fmicb.2015.00856
4. Nouioui I., Carro L., Garcia-Lopez M., Meier-Kolthoff J.P., Woyke T., Kyrpides N.C., Pukall R., Klenk H.P., Goodfellow M., Goker M. Genome-based taxonomic classification of the phylum *Actinobacteria*. *Front. Microbiol.* 2018. V. 9. Article № 2007. doi: 10.3389/fmicb.2018.02007
5. Amin D.H., Abdallah N.A., Abolmaaty A., Tolba S., Wellington E.M.H. Microbiological and molecular insights on rare actinobacteria harboring bioactive prospective. *Bull. Natl. Res. Cent.* 2020. V. 44. Article № 5. doi: 10.1186/s42269-019-0266-8
6. Subramani R., Sipkema D. Marine rare actinomycetes: a promising source of structurally diverse and unique novel natural products. *Mar. Drugs.* 2019. V. 17. Article № 249. doi: 10.3390/md17050249
7. Chaturvedi S., Khurana S.M.P. In: *Plant Biotechnology: Progress in Genomic Era.* Ed. Khurana S., Gaur R. Singapore: Springer, 2019. P. 277–307.
8. Zinder N.D., Lederberg J. Genetic exchange in *Salmonella. J. Bacteriol.* 1952. V. 64. P. 679–699. doi: 10.1128/jb.64.5.679-699.1952
9. Thierauf A., Perez G., Maloy A.S. Generalized transduction. *Methods Mol. Biol.* 2009. V. 501. P. 267–286. doi: 10.1007/978-1-60327-164-6_23
10. Chen J., Quiles-Puchalt N., Chiang Y.N., Bacigalupe R., Fillol-Salom A., Chee M., Fitzgerald J.R., Penades J.R. Genome hypermobility by lateral transduction. *Science.* 2018. V. 362. P. 207–212. doi: 10.1126/science.aat5867
11. Moore S.D. Assembling new *Escherichia coli* strains by transduction using phage P1. *Methods Mol. Biol.* 2011. V. 765. P. 155–169. doi: 10.1007/978-1-61779-197-0_10
12. Thomason L.C., Costantino N., Court D.L. *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* 2007. Chapter 1. P. 1.17.1–1.17.8. doi: 10.1002/0471142727.mb0117s79
13. Russell D.A., Hatfull G.F. PhagesDB: the actinobacteriophage database. *Bioinformatics.* 2017. V. 33. P. 784–786. doi: 10.1093/bioinformatics/btw71
14. Brettin T., Davis J. J., Disz T., Edwards R.A., Gerdes S., Olsen G.J., Olson R., Overbeek R., Parrello B., Pusch G.D. et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 2015. V. 5. Article № 8365. doi: 10.1038/srep08365
15. Davis J.J., Wattam A.R., Aziz R.K., Brettin T., Butler R., Butler R.M., Chlenski P., Conrad N., Dickerman A., Dietrich E.M. et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res.* 2020. V. 48. P. D606–D612. doi: 10.1093/nar/gkz943
16. Tanizawa Y., Fujisawa T., Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics.* 2018. V. 34. P. 1037–1039. doi: 10.1093/bioinformatics/btx713

17. Contreras-Moreira B., Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 2013. V. 79. P. 7696–7701. doi: 10.1128/AEM.02411-13
18. Kristensen D.M., Kannan L., Coleman M.K., Wolf Y.I., Sorokin A., Koonin E.V., Mushegian A. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics.* 2010. V. 26. P. 1481–1487. doi: 10.1093/bioinformatics/btq229
19. Vinuesa P., Ochoa-Sanchez L.E., Contreras-Moreira B. GET_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus *Stenotrophomonas. Front. Microbiol.* 2018. V. 9. Article № 771. doi: 10.3389/fmicb.2018.00771
20. Meier-Kolthoff J.P., Goker M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics.* 2017. V. 33. P. 3396–3404. doi: 10.1093/bioinformatics/btx440
21. Sorensen T. A method of establishing groups of equal amplitudes in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter.* 1948. V. 5. P. 1–34.
22. Bray J.R., Curtis J.T. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs. 1957.* V. 27. P. 325–349. doi: 10.2307/1942268
23. Panyukov V.V., Kiselev S.S., Ozoline O.N. Unique *k*-mers as strain-specific barcodes for phylogenetic analysis and natural microbiome profiling. *Int. J. Mol. Sci.* 2020. V. 21. Article № 944. doi: 10.3390/ijms21030944
24. Kumar S., Stecher G., Li M., Knyaz C., Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 2018. V. 35. P. 1547–1549. doi: 10.1093/molbev/msy096
25. Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987. V. 4. P. 406–425. doi: 10.1093/oxfordjournals.molbev.a040454
26. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. Basic local alignment search tool. *J. Mol. Biol.* 1990. V. 215. P. 403–410. doi: 10.1016/S0022-2836(05)80360-2
27. Henikoff S., Henikoff J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA.* 1992. V. 89. P. 10915–10919. doi: 10.1073/pnas.89.22.10915
28. Sullivan M.J., Petty N.K., Beatson S.A., Easyfig: a genome comparison visualizer. *Bioinformatics.* 2011. V. 27. P. 1009–1010. doi: 10.1093/bioinformatics/btr039
29. Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D., Amin N., Schwikowski B., Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003. V. 13. P. 2498–2504. doi: 10.1101/gr.1239303
30. Hagberg A.A., Schult D.A., Swart P.J. In: *Proceedings of the 7th Python in Science Conference.* Ed. Varoquaux G., Vaught T., Millman J. Pasadena, 2008. P. 11–15.
31. Latapy M., Magnien C., Del Vecchio N. Basic notions for the analysis of large two-mode networks. *Social Netw.* 2008. V. 30. P. 31–48. doi: 10.1016/j.socnet.2007.04.006
32. Holm S. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* 1979. V. 6. P. 65–70.
33. Felsenstein J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 1981. V. 17. P. 368–376. doi: 10.1007/BF01734359
34. Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 2015. V. 32. P. 268–274. doi: 10.1093/molbev/msu300

35. Edgar R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004. V. 32. P. 1792–1797. doi: 10.1093/nar/gkh340
36. Kalyaanamoorthy S., Minh B.Q., Wong T.K.F., von Haeseler A, Jermiin L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods.* 2017. V. 14. P. 587–589. doi: 10.1038/nmeth.4285
37. Schwarz G. Estimating the dimension of a model. *Ann. Stat.* 1978. V. 6. P. 461–464. doi: 10.1214/aos/1176344136
38. Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 2018. V. 35. P. 518–522. doi: 10.1093/molbev/msx281
39. Ventura M., Canchaya C., Tauch A., Chandra G., Fitzgerald G.F., Chater K.F., van Sinderen D. Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.* 2007. V. 71. P. 495–548. doi: 10.1128/MMBR.00005-07
40. Heinsch S.C., Hsu S.Y., Otto-Hanson L., Kinkel L., Smanski M.J. Complete genome sequences of *Streptomyces* spp. isolated from disease-suppressive soils. *BMC Genomics.* 2019. V. 20. Article № 994. doi: 10.1186/s12864-019-6279-8
41. Subramaniam G., Thakur V., Saxena R.K., Vadlamudi S., Purohit S., Kumar V., Rathore A., Chitikineni A., Varshney R.K. Complete genome sequence of sixteen plant growth promoting *Streptomyces* strains. *Sci. Rep.* 2020. V. 10. Article № 10294. doi: 10.1038/s41598-020-67153-9
42. Zhong L., Cheng Q., Tian X., Zhao L., Qin Z. Characterization of the replication, transfer, and plasmid/lytic phage cycle of the Streptomyces plasmid-phage pZL12. *J. Bacteriol.* 2010. V. 192. P. 3747–3754. doi: 10.1128/JB.00123-10
43. Pfeifer E., Moura de Sousa J.A., Touchon M., Rocha E.P.C. Bacteria have numerous distinctive groups of phage-plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic Acids Res.* 2021. V. 49. P. 2655–2673. doi: 10.1093/nar/gkab064
44. Nikulin N.A., Zimin A.A. Influence of non-canonical DNA bases on the genomic diversity of *Tevenvirinae*. *Front. Microbiol.* 2021. V. 12. Article № 632686. doi: 10.3389/fmicb.2021.632686
45. Falkow S., Baron L.S. Plasmid formation after lambda bacteriophage infection of *Escherichia coli-Salmonella typhosa* hybrids. *J. Bacteriol.* 1970. V. 102. P. 288–290. doi: 10.1128/jb.102.1.288-290.1970
46. Lieb M. λ mutants which persist as plasmids. *J. Virol.* 1970. V. 6. P. 218–225. doi: 10.1128/JVI.6.2.218-225.1970
47. Jaffe-Brachet A., D'Ari R. Maintenance of bacteriophage P1 plasmid. *J. Virol.* 1977. V. 23. P. 476–482. doi: 10.1128/JVI.23.3.476-482.1977
48. Sternberg N., Hoess R. The molecular genetics of bacteriophage P1. *Annu. Rev. Genet.* 1983. V. 17. P. 123–154. doi: 10.1146/annurev.ge.17.120183.001011
49. Jacob A.E., Barth P.T., Grinter N.J. Compatibility properties of P1 and φAMP prophages. *Mol. Gen. Genet.* 1975. V. 141. P. 263–267. doi: 10.1007/BF00341804
50. Goldstein R., Sedivy J., Ljungquist E. Propagation of satellite phage P4 as a plasmid. *Proc. Natl. Acad. Sci. USA.* 1982. V. 75. P. 515–519. doi: 10.1073/pnas.79.2.515
51. Ghisotti D., Briani F., Forti F., Piazza F., Polo S., Sabbattini P., Sturniolo T., Terzano S., Zangrossi S., Zappone M., Sironi G., Deho G. Multiple regulatory mechanisms controlling phage-plasmid P4 propagation. *FEMS Microbiol. Rev.* 1995. V. 17. P. 127–134. doi: 10.1111/j.1574-6976.1995.tb00194.x
52. Ravin V.K., Shulga M.G. Evidence for extrachromosomal location of prophage N15. *Virology*. 1970. V. 40. P. 800–807. doi: 10.1016/0042-6822(70)90125-X
53. Rybchin V.N., Svarchevsky A.N. The plasmid prophage N15: a linear DNA with covalently closed ends. *Mol. Microbiol.* 1999. V. 33. P. 895–903. doi: 10.1046/j.1365-2958.1999.01533.x

54. Ravin N.V. N15: the linear phage-plasmid. *Plasmid*. 2011. V. 65. P. 102–109. doi: 10.1016/j.plasmid.2010.12.004
55. Strizhov N., Tikhomirova L. Construction of recombinant plasmid carrying the λ DNA fragment responsible for prophage integration. *Nucleic Acids Res.* 1978. V. 5. P. 1767–1777. doi: 10.1093/nar/5.6.1767
56. Donoghue D.J., Sharp P.A. Construction of a hybrid bacteriophage-plasmid recombinant DNA vector. *J. Bacteriol.* 1978. V. 136. P. 1192–1196. doi: 10.1128/jb.136.3.1192-1196.1978
57. Mel'nikov A.A., Chernov A.P., Fodor I.I. Lambda plasmidophages and their properties. *Mol. Biol. (Moscow).* 1985. V. 19. P. 610–616. Article in Russian.
58. Cheng Q., Zhong L., Qin Z. Plasmid pCQ4 and its phage ΦCQ4 of endophytic *Streptomyces* sp. from *Artemisia annua* L. *Acta Microbiologica Sinica (Wei Sheng Wu Xue Bao).* 2012. V. 52. P. 825–831.
59. Dedrick R.M., Mavrich T.N., Ng W.L., Cervantes Reyes J.C., Olm M.R., Rush R.E., Jacobs-Sera D., Russell D.A., Hatfull G.F. Function, expression, specificity, diversity and incompatibility of actinobacteriophage *parABS* systems. *Mol. Microbiol.* 2016. V. 101. P. 625–644. doi: 10.1111/mmi.13414
60. Klyczek K.K., Bonilla J.A., Jacobs-Sera D., Adair T.L., Afram P., Allen K.G., Archambault M.L., Aziz R.M., Bagnasco F.G., Ball S.L. et al. Tales of diversity: genomic and morphological characteristics of forty-six *Arthrobacter* phages. *PLoS ONE*. 2017. V. 12. Article № e0180517. doi: 10.1371/journal.pone.0180517
61. Smith M.C., Hendrix R.W., Dedrick R., Mitchell K., Ko C.C., Russell D., Bell E., Gregory M., Bibb M.J., Pethick F. et al. Evolutionary relationships among actinophages and a putative adaptation for growth in *Streptomyces* spp. *J. Bacteriol.* 2013. V. 195. P. 4924–4935. doi: 10.1128/JB.00618-13
62. Bajpai R., Soni V., Khandrika L., Jangir P.K., Sharma R., Agrawal P. Genome sequence of a novel actinophage PIS136 isolated from a strain of *Saccharomonospora* sp. *J. Virol.* 2012. V. 86. Article № 9552. doi: 10.1128/JVI.01529-12
63. Cheepudom J., Lee C.C., Cai B., Meng M. Isolation, characterization, and complete genome analysis of P1312, a thermostable bacteriophage that infects *Thermobifida fusca*. *Front. Microbiol.* 2015. V. 6. Article № 959. doi: 10.3389/fmicb.2015.00959
64. Wilson G.G., Young K.Y., Edlin G.J., Konigsberg W. High-frequency generalized transduction by bacteriophage T4. *Nature.* 1979. V. 280. P. 80–82. doi: 10.1038/280080a0
65. Casjens S.R., Gilcrease E.B., Winn-Stapley D.A., Schicklmaier P., Schmieger H., Pedulla M.L., Ford M.E., Houtz J.M., Hatfull G.F., Hendrix R.W. The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J. Bacteriol.* 2005. V. 187. P. 1091–1104. doi: 10.1128/JB.187.3.1091-1104.2005

Received 22.08.2023.
Revised 06.09.2023.
Published 18.10.2023.