

Translation of the original article

Kiryanova O.Yu., Kiselev S.S., Panyukov V.V., Nikulin N.A., Nazipova N.N., Chemeris A.V., Zimin A.A. *Mathematical Biology and Bioinformatics*. 2024;19(2):369–392.

doi: [10.17537/2024.19.369](https://doi.org/10.17537/2024.19.369)

===== TRANSLATIONS OF PUBLISHED ARTICLES =====

Using Multiplex Virtual PCR for Genetic Barcoding of *Tequatrovirus* Bacteriophages

Kiryanova O.Yu.¹, Kiselev S.S.², Panyukov V.V.³, Nikulin N.A.⁴,
Nazipova N.N.³, Chemeris A.V.⁵, Zimin A.A.⁴

¹*Institute of Petrochemistry and Catalysis RAS, Ufa Federal Research Center of the Russian Academy of Sciences Ufa, Russia*

²*Institute of Cell Biophysics, Pushchino Scientific Center for Biological Research, Russian Academy of Sciences, Pushchino, Russia*

³*Institute of Mathematical Problems of Biology RAS, Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russia*

⁴*G.K. Scryabin Institute of Biochemistry and Physiology of Microorganisms RAS, Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Pushchino, Russia*

⁵*Institute of Biochemistry and Genetics RAS, Ufa Federal Research Center of the Russian Academy of Sciences, Ufa, Russia*

Abstract. The paper considers the possibility of using multiplex virtual PCR to barcode the genomes of T4-type bacteriophages. The method of multiplex virtual PCR was compared to the phylogenomic analysis based on an alignment-free approach. The dataset analyzed was a set of genomes of phages of the genus *Tequatrovirus*, which were previously selected for pangenomic analysis. The genomes used were pre-checked for adequate annotation and, accordingly, sequencing accuracy. A total of 67 *Straboviridae* genomes were analyzed using multiplex virtual PCR; their barcodes were generated, and similarity parameters were determined. The results were represented in the form of two trees: a phylogenetic tree constructed on the basis of the neighbor-joining algorithm, and a phylogenomic tree based on the pairwise distance matrix between the sets of nucleotide 10-mers. The comparison of the two trees revealed some differences: in the branches of *Mosigvirus* and *Dhakavirus* (two distantly related genera taken as a reference) and the *Tequatrovirus* branch. A supposition was made that the differences pertained to the very algorithm of multiplex virtual PCR – which, in general, is more applicable for assessing similarity of closely related (with identity degree over 75 %), rather than distant genomes. Overall, the method of multiplex virtual PCR seems reliable enough to distinguish between taxonomic groups of large-genome (~170 kb) bacteriophages (including T4-type phages) at the level of genus. From a practical standpoint, the method is convenient and suitable for generating barcodes, which can then be included in the bacteriophage passport and used in the practice of phage therapy.

Key words: bacteriophage T4, genome barcoding, multiplex virtual PCR, bacteriophage phylogenomics, *Tequatrovirus*.

INTRODUCTION

A wide spread of antibiotic resistance genes among relevant human and veterinary bacterial pathogens has greatly limited the use of antimicrobial drugs in clinical practice, animal husbandry and aquaculture. The “era of antibiotics” approaches a crisis [1]. As a result, we see a revival of interest to bacteriophages as therapeutic agents, which started in Western medicine at the end of the 20th century [2–4]. In the USSR and the Russian Federation, phage preparations targeting various bacterial pathogens have been steadily produced since the 1930s and are freely available at the drugstores. Some of the most common bacteria targeted by such preparations are pathogenic strains of *E. coli*, which cause diarrhea [5].

Today, researchers are exploring the potential of phage therapy to suppress specific pathogenic bacteria in the microbiome. Bacteriophages are viruses that infect and kill bacteria and, strain-specific as they are, phages can selectively target harmful microorganisms while sparing beneficial ones. Correspondingly, phage therapy could be used to treat infectious diseases without disturbing the delicate balance of the microbiome. This makes phage preparations an attractive alternative to antibiotics [6, 7].

The viral nature of bacteriophages was discovered at the end of the 19th century. First, the British scientist Ernest H. Hankin discovered that waters of the Indian rivers Ganges and Yamuna possessed bactericidal properties. Investigating this intriguing phenomenon, Hankin passed the river water through a filter that retained bacteria and then heated the filtrate to 60°C. As it turned out, filtering did not reduce the bactericidal potency of water; heating, on the other hand, completely eliminated the antibacterial effect [8]. Next, the Russian microbiologist N. F. Gamaleya found out that lysates of the *Bacillus anthracis* culture contained some kind of a “lytic enzyme”, exhibiting properties similar to those of the Ganges water [9]. The one who actually discovered bacterial viruses was the British microbiologist Frederik W. Twort. He published a paper in *The Lancet* [10], putting forward a hypothesis about the viral nature of the “bacteria-dissolving enzyme”. The hypothesis was based on the observation that the efficiency of the “enzyme” was unusually high: “they (the colonies) dissolved even before they reached a size visible to the eye”, and “some of these cultures, slightly infected with the dissolving material, quickly became transparent and died”. In 1917, Félix d'Herelle introduced the term “bacteriophage” [11], which has become established as the name for bacterial viruses.

In his paper [10], F. W. Twort emphasized the importance of studying bacteriophages as a mean to combat infections of bacterial etiology. Elaborating this idea, he made two points of practical importance:

- 1) to study the toxicity of cocci and bacilli, one should separate them from bacteriophages;
- 2) it is possible and perspective to produce bacteriophage-based bacteriolytic drugs.

Phage therapy implies isolation and purification of specific phages – those that are known to target and kill the pathogenic bacteria causing the infection [12, 13]. These phages are then administered to the patient orally, topically or intravenously. The phages then selectively replicate within the target bacteria, causing their lysis and subsequent elimination from the body. This approach has shown promising results in the treatment of a variety of infectious diseases: first of all, intestinal and wound infections; some respiratory diseases; urinary tract infections; and skin pathologies of bacterial etiology [14].

Apart from isolating environmental phages, researchers design and test phage constructs with increased specificity and potency. The ability of such genetically modified phages to recognize and infect specific strains of bacteria is improved, which enhances the efficiency of phage therapy. In addition, engineered phages can be modified to carry “payloads” (e.g., antimicrobial peptides or genes encoding antimicrobial agents), which further enhances their ability to kill bacteria [12, 15]. Currently, however, the results of such promising studies do not go beyond the doors of research laboratories, since the use of genetically modified bacteriophages is prohibited by law in all countries around the world.

One of the first studies that revived interest to bacteriophages as therapeutic agents was Merrill's experiment of 1996 on spontaneous modification of a lambdoid phage upon its interaction with the murine immune system. The result was the production of a long-circulating phage λ [16]. As it turned out, this effect was achieved by a single amino acid substitution in the surface capsid protein D. There are almost no analogues of that study available in the literature. Another example of successful bacteriophage modification using the techniques of genetic engineering was the production of low-antigenic T4-type bacteriophages with a mutation in the *hoc* gene, encoding the main antigen protein of the phages [17].

The group of phages commonly used for antibacterial therapy is tailed bacteriophages. They are now combined into the order *Caudoviricetes*. Bacteriophages of this order can often be found in animals and humans, as well as soil and water samples. A prospective, in terms of therapy, subgroup of tailed bacteriophages is T4-type viruses of the genus *Tequatrovirus*, belonging to the family *Straboviridae*. Examples of the therapeutic use of bacteriophages of the genus *Tequatrovirus* bacteriophages and their cocktails, as well as advantages and disadvantages of these phages as antimicrobial agents are discussed here [18].

The objective of the present work was to assess suitability of multiplex virtual PCR as a method for generating barcodes, which could then be used for comparative phylogenetic analysis of complete bacteriophage genomes and their certification.

In this work, we have demonstrated how multiplex virtual PCR and genetic barcoding based on the detection of DNA polymorphism can be used for the creation of a bacteriophage passport. Genetic barcoding was previously used for identification and classification of animals and plants whose genomes were quite large (from several million to several billion base pairs) [19–21]. In this work, we have successfully applied this approach to analyze much smaller genomes (the characteristic size of the genomes of T4-related phages is 160–180 kb).

Traditionally, phage identification is based on genomic analysis, an integrated approach involving sequencing of the entire phage genome. Here, we have assessed the potential of genetic barcoding as a complementary tool to genomic analysis in the genetic characterization of phages. The procedure of barcoding has several advantages: it is simple, cost-effective and enables rapid processing of large sample arrays. Using it facilitates the pan-genomic analysis of relationships between bacteriophage genomes, as well as diagnostic testing of phage samples.

METHODS AND RESULTS

The dataset analyzed in the study contained 67 genomes of phages of the genus *Tequatrovirus* selected from a set of viral genomes previously used for pan-genomic analysis [22]. The set was downloaded from the NCBI Nucleotide Database and contained only properly annotated data, which were thoroughly checked for sequencing accuracy using cluster analysis. Initially, the annotations were checked for the presence of predicted ORFs and, in case of their absence, automatic annotation was carried out using Genome Annotation PATRIC with the parameters Domain: Viruses and Annotation recipe: Phage [23, 24]. Then, the amino acid sequences of ORFs were clustered using GET_HOMOLOGUES (with the parameters {minimum coverage, -C: 40%}; {minimum identity, -S: 50%}, {clustering algorithm: COGtriangle} and other parameters at their default values) [25, 26]. All the genomes were screened for the presence of core genes [27]. If no ORFs of core genes were found in the genome, its nucleotide sequence was automatically re-annotated using PATRIC and checked again. If there was a frameshift of a core gene, the genome was excluded from the analysis. The final set of sequences selected for the analysis contained genomes of phages of the genus *Tequatrovirus*, as well as several genomes of other *Straboviridae* viruses taken as outgroup references. The functions of the proteins encoded by pangenomic genes were verified using the genome annotation of *Escherichia* phage T4 (NC_000866.4) [28].

The reference genome of bacteriophage T4 is 168903 bp long. It contains more than 400 protein-coding genes, 7 tRNA genes, and a number of short-RNA genes with unknown functions

linked to the tRNA genes. Some eukaryotic viruses (e.g., pox and herpes viruses) have genomes of a similar length. The files containing input data for modeling multiplex virtual PCR were in the FASTA format.

The phylogenomic tree was constructed in MEGA X [29] using the neighbor-joining method [30]. The tree was based on the pairwise distance matrix between the genomes of 67 bacteriophages; the distances were calculated with the IGPDE program, using an alignment-free approach [31]. The procedure of calculation was as follows. First, for each genome, a set of k -mers ($k = 10$) was generated with a 1-nucleotide step (direct and inverted sequences were considered as identical). Then, similarity indices (S) were determined between the pairs of sets of 10-mers using the Sorensen formula [32]. The distances were calculated using the formula $D = 1 - S$.

MEGA X [29] was also used to construct two phylogenetic trees; they were generated using the maximum-likelihood method [33], on the basis of aligned nucleotide/amino acid sequences for the major capsid protein. The sequences were aligned using the MUSCLE algorithm [34]. Models of optimal substitution were obtained using the built-in MEGA X test based on the Bayesian information criterion [35]. The levels of statistical support for branches were calculated after 2000 bootstrap replicates [36].

The method used in this work was based on a virtual multiplex RAPD analysis, allowing one to create barcodes of bacteriophage genomes. The computer modeling of multiplex virtual PCR revealed the primer annealing sites, as well as amplicon lengths (in the range from 51 to 500 bp). The range, in which amplicon lengths are identified, was based on the data of multiplex virtual PCR experiments, in which the length of reaction products was determined with a sequencing gel electrophoresis (accuracy, ± 1 bp). The method is described in more detail in [20]. More specifically, it implies a number of assumptions and limitations:

- *Complete annealing of primers.* The model assumes that the primer is fully complementary to the DNA fragment found. Incomplete annealing, hairpin formation and other factors that may affect PCR efficiency are not taken into account.
- *Creation of a barcode.* The results obtained are converted into a barcode format, where a strip corresponds to the detection of an amplicon of a certain length, and a space corresponds to its absence. The multiplicity of amplicon detection is not taken into account; the codified information only reflects the very fact of detection (presence/absence of an amplicon, in binary format).
- *The multiplex version of PCR* makes it possible to increase the amount of the reaction product and, hence, to create more detailed barcodes. It is important, since insufficient number of bands would complicate barcode comparison, making its results unclear and difficult to interpret.
- *Selection of primers.* Since the information content of the barcode depends on the primers, they should be selected carefully in regard to their nucleotide composition and length.

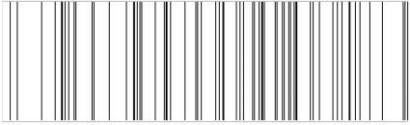
Advantages of the method:

- *Simplicity and affordability.* Virtual analysis does not require expensive equipment and reagents, which makes it affordable for many a research group.
- *Speed.* Computer modeling allows one to quickly obtain results without conducting time-consuming experiments.
- *Informativeness.* The barcode is a compact and informative “snapshot” of the genome, which can be used for various purposes.

The information about amplicon lengths was converted into the binary format according to the following procedure: the range of 51–500 bp was matched with a binary sequence of 450 positions. If an amplicon of a specific length was found during the analysis, it counted as 1 (or a stripe on the barcode); otherwise, as 0 (a space on the barcode). The resulting binary sequences were numerical snapshots of genomes, enabling their comparative analysis. Table 1 shows an example of the results obtained for the genome of the Enterobacteria phage JS10.

The results of multiplex virtual PCR were presented as a table in the .CSV format, which allowed one to determine the primers that would make the greatest contribution to the detected DNA polymorphism, as well as those whose contribution would be low or zero (i.e., unrecommended for the use in actual experiments). Figure 1 shows the results of such analysis obtained for the Enterobacteria phage Apltg8. One can see the annealing positions of 5'-ends of forward and reverse primers and the size of the resulting amplicon.

Table 1. A barcode generated for the Enterobacteria genome phage JS10

Genome	Amplicons b.p.	Binary sequence	Barcode
Enterobacteria phage JS10	059 067 094 109 116 117 120 124 130 132 150 152 169 188 196 197 200 203 205 220 230 232 233 237 253 265 267 271 287 291 295 296 310 314 318 328 330 335 338 339 341 353 355 361 363 368 370 374 376 377 392 395 402 417 422 427 435 436 442 447 458 472 495 497 500	00000000100000001000000000 00000000000000001000000000 000010000001100100010000010 10000000000000000010100000 00000000000010000000000000 000010000000110010010100000 000000000100000000001011000 10000000000000000100000000 000101000100000000000000001 000100011000000000000001000 100010000000001010000100110 100000000000101000001010000 101000101100000000000000100 100000010000000000000010000 100001000000011000001000010 00000000010000000000000100 00000000000000000000101001	

file	contig	direct primer	direct primer position	inverse primer	inverse primer position	amplicon size
Enterobacteria phage Apltg8.fasta	1	GGACTT	34087	TTGACC	34141	54
Enterobacteria phage Apltg8.fasta	1	GGTCAA	86181	AAGTCC	86238	57
Enterobacteria phage Apltg8.fasta	1	GCTGAT	38542	TACGAC	38600	58
Enterobacteria phage Apltg8.fasta	1	GGTCAA	57713	TTGACC	57775	62
Enterobacteria phage Apltg8.fasta	1	GTCGTA	149730	TTGACC	149796	66
Enterobacteria phage Apltg8.fasta	1	GCTGAT	14252	ATCAGC	14323	71
Enterobacteria phage Apltg8.fasta	1	GGACTT	97937	ATCAGC	98028	91
Enterobacteria phage Apltg8.fasta	1	GGACTT	119506	AAGTCC	119603	97
Enterobacteria phage Apltg8.fasta	1	GCTGAT	53492	ATCAGC	53602	110
Enterobacteria phage Apltg8.fasta	1	GTCGTA	144511	AAGTCC	144626	115
Enterobacteria phage Apltg8.fasta	1	GGTCAA	121427	TTGACC	121543	116
Enterobacteria phage Apltg8.fasta	1	GGTCAA	142471	ATCAGC	142597	126
Enterobacteria phage Apltg8.fasta	1	GGTCAA	24511	ATCAGC	24638	127
Enterobacteria phage Apltg8.fasta	1	GCTGAT	101376	ATCAGC	101513	137
Enterobacteria phage Apltg8.fasta	1	GGTCAA	138920	TTGACC	139070	150
Enterobacteria phage Apltg8.fasta	1	GGTCAA	34728	TTGACC	34880	152
Enterobacteria phage Apltg8.fasta	1	GGTCAA	88283	TACGAC	88447	164
Enterobacteria phage Apltg8.fasta	1	GCTGAT	62329	ATCAGC	62497	168
Enterobacteria phage Apltg8.fasta	1	GCTGAT	62320	ATCAGC	62497	177
Enterobacteria phage Apltg8.fasta	1	GTCGTA	38034	TTGACC	38216	182
Enterobacteria phage Apltg8.fasta	1	GGACTT	9456	ATCAGC	9650	194

Fig. 1. An example of file generated by the program BARCODE DNA ANALYZER (BDANALYZER) with the results of the analysis [37].

The distance between the binary sequences forming the barcodes of genomic sequences can be used to assess the degree of their similarity. To determine the degree of proximity of genomes, the Levenshtein distance or edit distance $d(S_1, S_2) = D_{12}$ was used, which was calculated using the formula, where two strings S_1 и S_2 with lengths M и N are compared:

$$D_{12} = \begin{cases} 0, & \text{if } i = 0, j = 0 \\ j, & \text{if } i > 0, j = 0 \\ i, & \text{if } i = 0, j > 0 \\ \min \begin{cases} D(i, j-1) + 1, \\ D(i-1, j) + 1, \\ D(i-1, j-1) + m(S_1(i), S_2(j)), \end{cases} & \text{if } i > 0, j > 0 \end{cases},$$

$$\text{here } m(S_1(i), S_2(j)) = \begin{cases} 0, & \text{if } S_1(i) = S_2(j) \\ 1, & \text{if } S_1(i) \neq S_2(j) \end{cases}, i = 1 \div N, j = 1 \div M.$$

As a result, we obtained a square matrix of pairwise distances between the sequences of the sample of length n . The distance matrix $\mathbf{S} = (D_{kp})$, $k = 1, \dots, n$, $p = 1, \dots, n$ was then used to construct a phylogenetic tree, which showed relationships between the genomes. In the tree constructed for our sample, two *Mosigvirus* phages (RB69 and vB_EcoM_JS09) and two *Dhakavirus* phages (JS10 and JS98), all from the family *Straboviridae*, were used as reference genomes. Their position on the tree allowed us to determine the foundation of the branch that goes to the genus *Tequatrovirus*.

The phylogenetic tree was constructed with the program MEGA X [29] using the neighbor-joining method [30]. The input data was a matrix of pairwise distances between the binary sequences of the sample. Multiplex virtual PCR was performed with a set of 4 primers: GATGAT, AGACTT, AGTCAA and ATCGTA. The primers were generated in the program GATCGGenerator [37], which yielded nucleotide sequences of a given structure: these were sequences without homopolymer regions, with a given number of dinucleotides. The required number of sequences, their length, GC-content, and the percentage of dinucleotides was specified as input data in the program. In addition, it was possible to generate sequences with the complete exclusion of homopolymer regions. The generator did not allow the repetition of the same motifs (2–5 nucleotides long) more than four times in a row. After generation of the sample, nucleotide sequences were manually selected out of it for further analysis.

The number of primers was determined on the basis of the results of barcode construction tests. A set of 6 primers yielded a large number of amplicons, which did not allow an unambiguous comparative analysis. The clearest picture (about 50% of the barcode bands) was obtained using four primers.

In the first simulation run, a set of primers with 50% GC-content was selected: GCTGAT, GGACTT, GGTCAA and GTCGTA. However, the phylogenetic tree constructed on the basis of the results obtained (Fig. 2) had some discrepancies with the previous data and pan-genomic analysis of the genomes. Therefore, taking into account that genomes of phages of the genus *Tequatrovirus* are AT-rich, we decided to adjust the parameters of search for annealing sites, specifying the adenine and thymine content of the primers as $\geq 50\%$. The results of the second simulation run are shown in Figure 3.

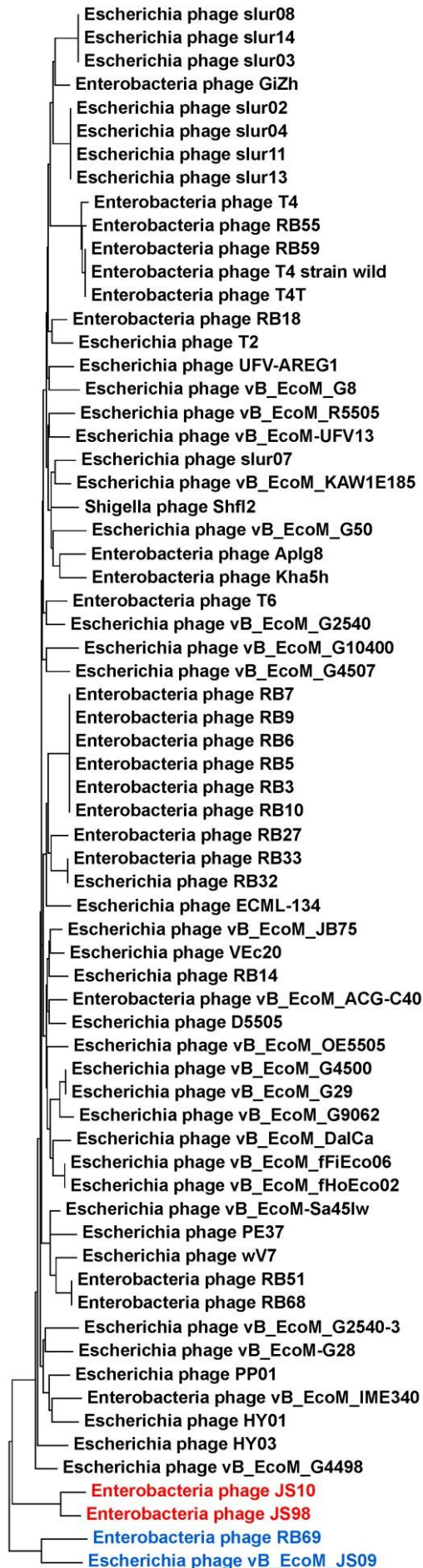


Fig. 2. A phylogenetic tree based on the results of multiplex virtual PCR using primers with 50% GC-pairs. The phylogenetic tree was constructed in the program MEGA X [29] with the neighbor-joining method [30]. The phages of the genera *Tequatrovirus*, *Mosigvirus* and *Dhakavirus* are shown in black, blue and red color, respectively.

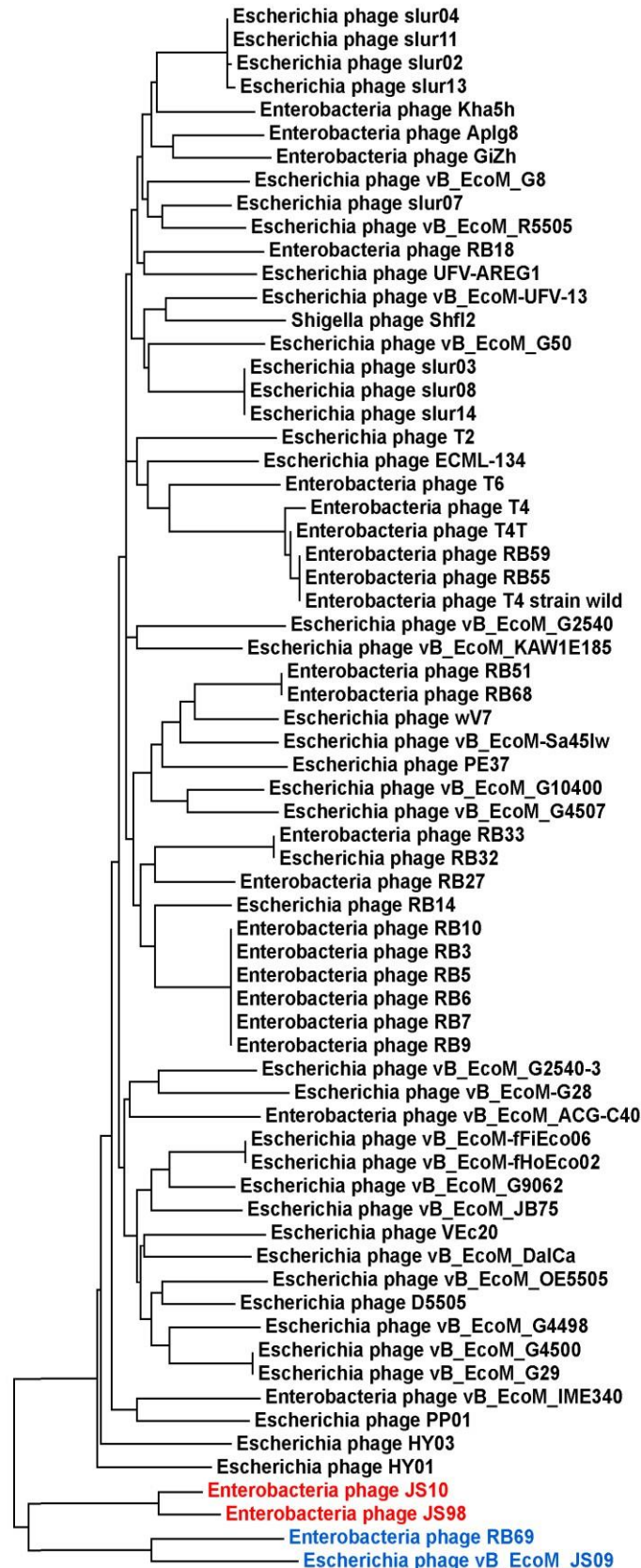


Fig. 3. A phylogenetic tree based on the results of multiplex virtual PCR with AT-rich primers. The phylogenetic tree was constructed in the program MEGA X [29] using the neighbor-joining method [30]. The phages of the genera *Tequatrovirus*, *Mosigvirus* and *Dhakavirus* are shown in black, blue and red color, respectively.

The phylogenetic tree shown in the Figure 3 has two main branches, with one of them containing all the genomes of bacteriophages of the genus *Tequatrovirus*. The genomes of Escherichia phage slur04, Escherichia phage slur11, Escherichia phage slur02, Escherichia phage slur13, Enterobacteria phage Kha5h, Enterobacteria phage Aplg8, Enterobacteria phage GiZh, Escherichia phage vB_EcoM_G8, Escherichia phage slur07, Escherichia phage vB_EcoM_R5505, Enterobacteria phage RB18, Escherichia phage UFVAREG1, Escherichia phage vB_EcoM-UFV13, Shigella phage Shfl2, Escherichia phage vB_EcoM_G50, Escherichia phage slur03, Escherichia phage slur08, Escherichia phage slur14 form one common branch. It should be noted that this branch mainly contains phages from two samples (slur- and vB_EcoM-phages).

Samples, which bacteriophages are isolated from, often contain closely related phage strains. This is reflected by the content of this tree branch. Another example is the branch containing the genomes of Enterobacteria phages phage Enterobacteria phage RB51, Enterobacteria phage RB68, Escherichia phage wV7, Escherichia phage vB_EcoMSa45lw, Escherichia phage PE37, Escherichia phage vB_EcoM_G10400, Escherichia phage vB_EcoM_G4507, Enterobacteria phage RB33, Escherichia phage RB32, Enterobacteria phage RB27, Escherichia phage RB14, Enterobacteria phage RB10, Enterobacteria phage RB3, Enterobacteria phage RB5, Enterobacteria phage RB6, Enterobacteria phage RB7, Enterobacteria phage RB9. As one can see, this branch contains a group of RB phages [28]. Although the RB-phage group is highly diverse in composition and contains representatives of, at least, two genera [28], many of these phages, which were isolated from a lake near the Cold Spring Harbor village, are closely related. The branches of this tree are almost identical to the branches on the phylogenomic tree constructed using the alignment-free method [31] (Fig. 4) and on the trees obtained by maximum likelihood after alignment of genes or amino acid sequences for the major capsid proteins of the phages (Fig. 5 and 6, respectively).

Another branch of the tree is almost entirely composed of vB_EcoM phages: Escherichia phage vB_EcoM_G2540-3, Escherichia phage vB_EcoM-G28, Enterobacteria phage vB_EcoM_ACG-C40, Escherichia phage vB_EcoM-fFiEco06, Escherichia phage vB_EcoM-fHoEco02, Escherichia phage vB_EcoM_G9062, Escherichia phage vB_EcoM_JB75, Escherichia phage VEc20, Escherichia phage vB_EcoM_DalCa, Escherichia phage vB_EcoM_OE5505, Escherichia phage D5505, Escherichia phage vB_EcoM_G4498, Escherichia phage vB_EcoM_G4500 and Escherichia phage vB_EcoM_G29. This topology is another illustration of the point made above, and the mutual arrangement of these branches agrees well with the topology of trees obtained by alternative methods. The branches containing genomes of Enterobacteria phage vB_EcoM_IME340, Escherichia phage PP01, Escherichia phage HY03, Escherichia phage HY01, are located close to the branching base of phages of the genus *Tequatrovirus* phages, which is also similar to the topology of this branch obtained on the basis of phylogenomic analysis and multiple superposition of protein and gene 23.

As can be seen, topological similarity within the branch of T4 phages is quite high. The similarity of genome sequences within this genus varies from 90 to 99% and, in addition, T4 phages are known to have a very high frequency of recombination. For these reasons, comparison of their individual genes – even conservative ones, such as gene 23 – is more suitable to determine the place of the genus within *Straboviridae* than to analyze intragenomic relationships. A phylogenomic analysis (e.g., comparison of all their oligonucleotide 10-mers) seems to be more appropriate in this kind of situation complicated by the enhanced intragroup homologous recombination. The topological similarity of trees constructed on the basis of multiplex virtual PCR and phylogenomic analysis suggests that a limited information about the number of oligonucleotides and distances between them can be sufficient for an adequate comparative description of these genomes.

The barcode based on the data of multiplex virtual PCR contains information sufficient for the precise identification of a phage and can, therefore, be written down in the phage passport as a code for its evolutionary history and taxonomic position. The modern ICTV taxonomy is based

on the evolutionary relationships between bacteriophages and is rooted in their biology, matching the data of pan-genome analysis [22].

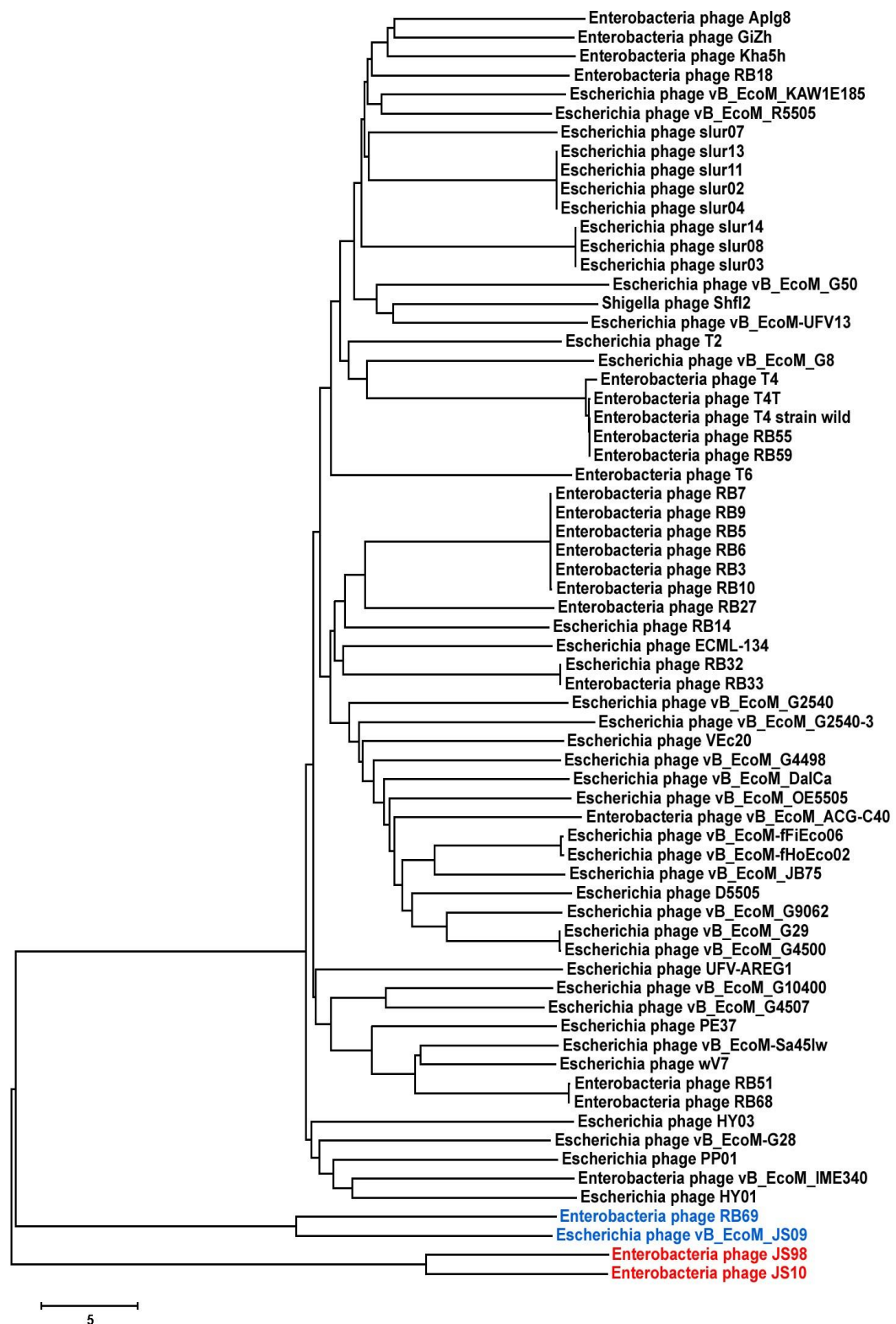


Fig. 4. A phylogenomic tree based on the pairwise intergenomic distance matrix calculated from the analysis of nucleotide 10-mers. The tree was constructed with the program MEGA X [29] using the neighbor-joining method [30]. The phages of the genera *Tequatrovirus*, *Mosigvirus* and *Dhakavirus* are shown in black, blue and red color, respectively. The scale bar indicates the Sørensen distance (%).

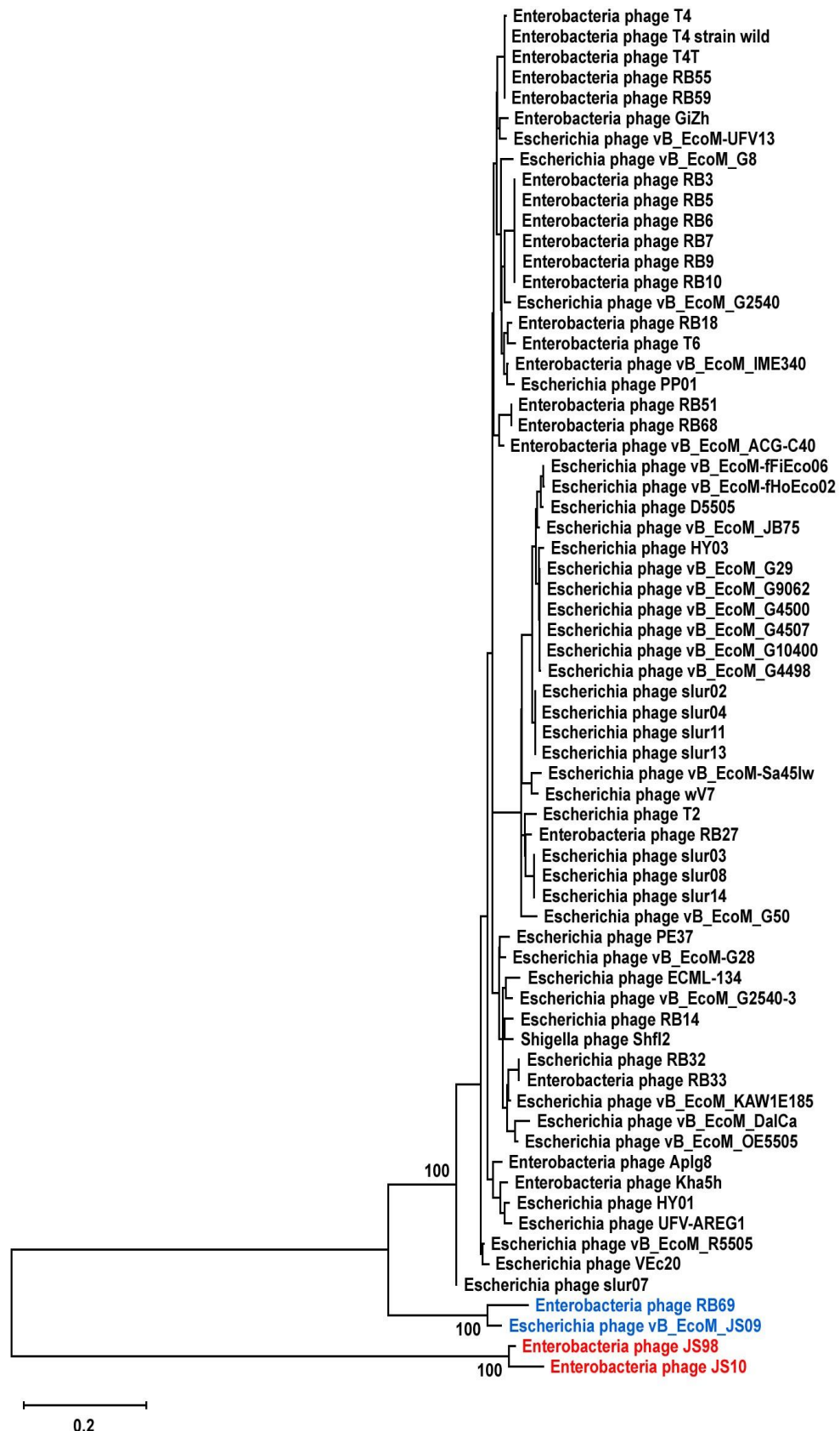


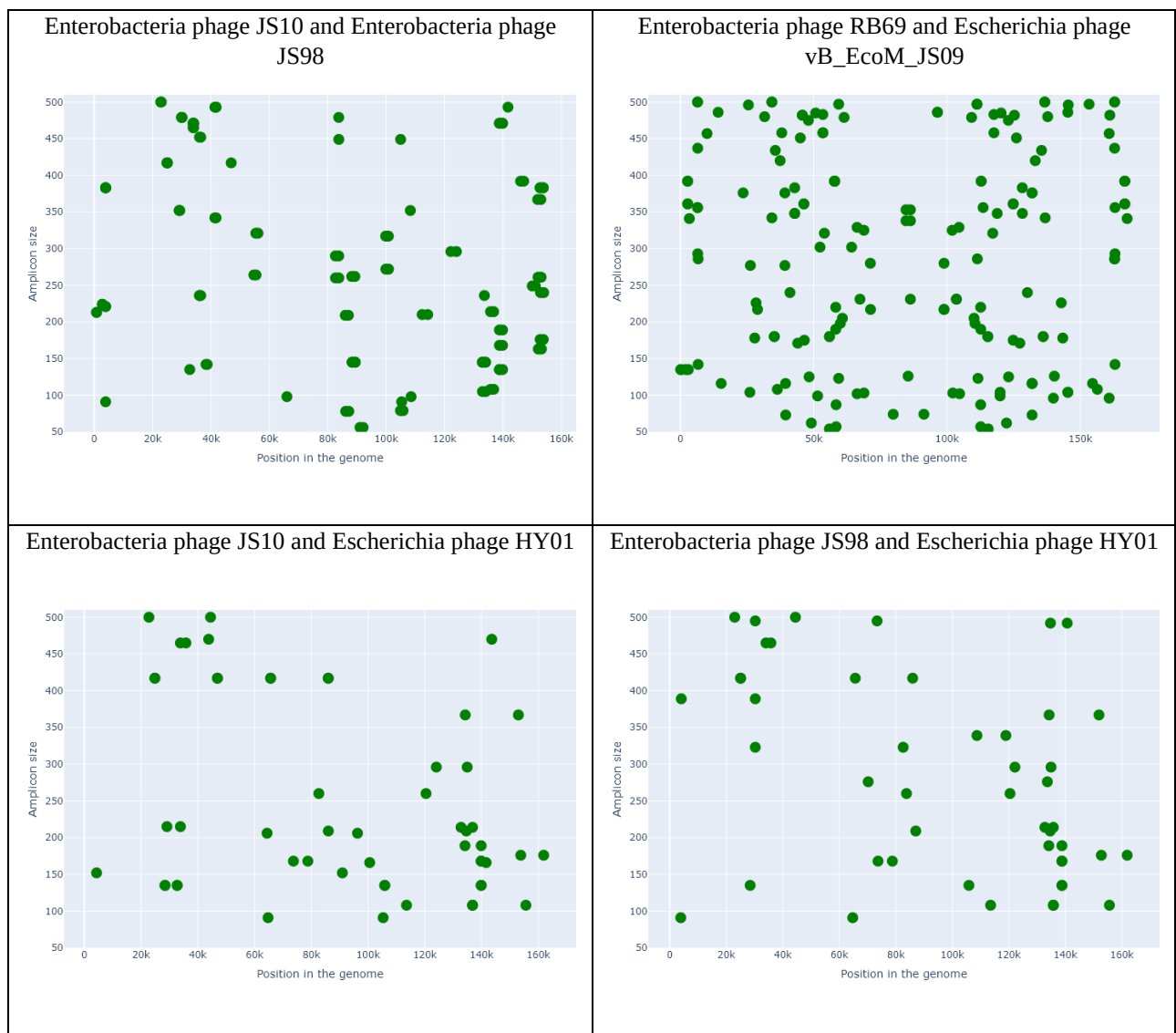
Fig. 5. A phylogenetic tree inferred from the alignment of nucleotide sequences for the gene of major capsid protein. The tree was constructed with the program MEGA X [29] using the maximum likelihood method. The optimal substitution model is GTR + G + I [38, 39]. The levels of bootstrap support for branches are given as percentages of 2000 iterations. The phages of the genera *Tequatrovirus*, *Mosigvirus* and *Dhakavirus* are shown in black, blue and red color, respectively. The scale bar indicates the number of substitutions per site.

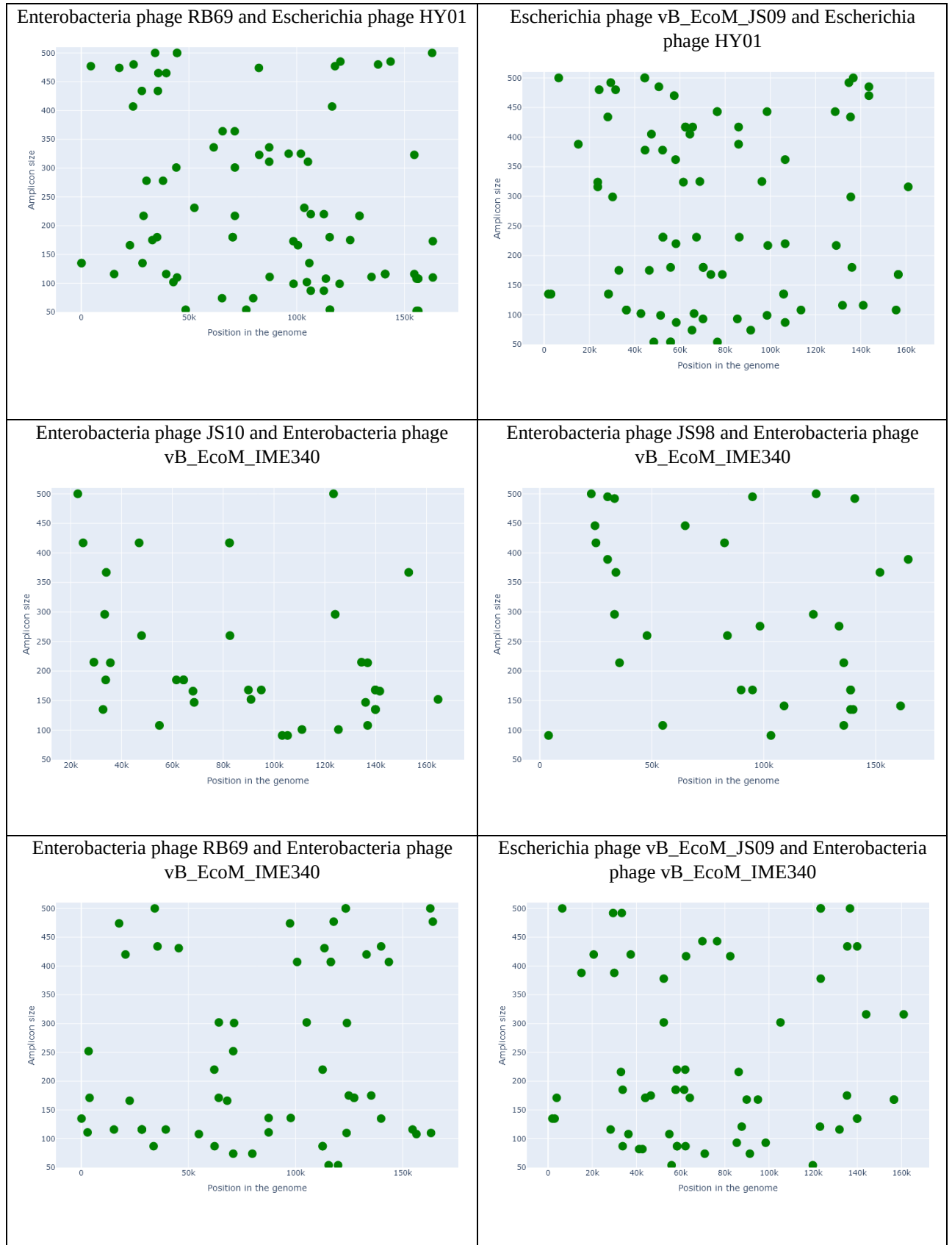


Fig. 6. A phylogenetic tree inferred from the alignment of amino acid sequences of the major capsid protein. The tree was constructed with the program MEGA X [29] using the maximum likelihood method. The optimal substitution model is WAG + G + I [39, 40]. The levels of bootstrap support for branches are given as percentages of 2000 iterations. The phages of the genera *Tequatrovirus*, *Mosigvirus* and *Dhakavirus* are shown in black, blue and red color, respectively. The scale bar indicates the number of substitutions per site.

The trees presented in Figures 2–6 show that the three groups of bacteriophages considered here form compact subbranches, with the representatives of different genera not mixing with each other. This confirms adequacy of all the methods used. In respect to the tree topology, their outcome only differs in minor details. For example, at the reference groups of *Dhakavirus* and *Mosigvirus* phages. On the tree built with AT-rich primers, they are on the same branch separated from other branches (Fig. 3). On the tree inferred from multiplex virtual PCR using primers with 50% GC-pairs, *Dhakavirus* is on the same branch with *Tequatrovirus*, whereas *Mosigvirus* is on a separate branch (Fig. 2). The trees based on phylogenomic analysis (Fig. 4), and on alignment of nucleotide (Fig. 5) or amino acid sequences (Fig. 6) of the major capsid protein have a different arrangement of these three groups.

Table 2. Two-dimensional maps of primer annealing sites with matching amplicon sizes obtained in virtual multiplex PCR simulations





To obtain additional information, we conducted a pairwise comparison of genomes of the following viruses: Escherichia phage HY01, Enterobacteria phage JS10, Enterobacteria phage JS98, Enterobacteria phage RB69, Escherichia phage vB_EcoM_JS09, Enterobacteria phage vB_EcoM_IME340. The results plotted as amplicon lengths versus their location in the genome are given in Table 2.

The percentage of amplicons that match in size and are close in location from the total number of amplicons of the same size is shown in Table 3. The data clearly indicate the presence of homologous DNA fragments in the genomes of Enterobacteria phages JS10 and JS98. It is noteworthy that on the phylogenetic tree, Enterobacteria phage RB69 and Escherichia phage vB_EcoM_JS09, belonging to the genus *Mosigvirus*, are in the same clade – despite the fact that amplicon matching in this pair is less than 13%. Perhaps, in the future, we should pay closer attention to amplicons that are located in identical positions in the genomes, but differ in length by 1-2 nucleotides. Such a difference may indicate minor nucleotide substitutions in these loci. Comparing reference genomes with that of Escherichia phage HY01 demonstrates that *Dhakavirus* phages (JS10 and JS98) are genetically closer to HY01 than *Mosigvirus* phages (RB69 and vB_EcoM_JS09).

Table 3. Percentage of matching amplicons in coinciding genome pairs

Pairs of genomes	Match percentage
Enterobacteria phage JS10 and Enterobacteria phage JS98	88.3%
Enterobacteria phage RB69 and Escherichia phage vB_EcoM_JS09	12.9%
Enterobacteria phage JS10 and Escherichia phage HY01	33.3%
Enterobacteria phage JS98 and Escherichia phage HY01	40.0%
Enterobacteria phage RB69 and Escherichia phage HY01	33.3%
Escherichia phage vB_EcoM_JS09 and Escherichia phage HY01	25.8%
Enterobacteria phage JS10 and Enterobacteria phage vB_EcoM_IME340	23.5%
Enterobacteria phage JS98 and Enterobacteria phage vB_EcoM_IME340	6.3%
Enterobacteria phage RB69 and Enterobacteria phage vB_EcoM_IME340	13.6%
Escherichia phage vB_EcoM_JS09 and Enterobacteria phage vB_EcoM_IME340	32.0%

Thus, the use of multiplex virtual PCR and genetic barcoding opens up new possibilities for the study of bacteriophage genomes. The approach investigated in this work provides a rapid and accurate test for identification of bacteriophages of the genus *Tequatrovirus*. After collecting more data on the viruses of the *Straboviridae* family, genetic barcoding can facilitate studies of their diversity, ecology and evolution. It may also serve a simple, yet important practical purpose of labeling bacteriophage samples used in medicine, veterinary science, sanitary treatment of environmental objects etc.

DISCUSSION

At the time of its emergence, the idea of a “genetic barcode”, corresponding to specific regions of the genome and serving as a marker for the biological species, looked promising. It seemed possible to assign unique barcodes to all the living organisms and then use them to

identify these species. The information included in the barcode might not be sufficient to resolve unsettled issues of phylogeny, but it should be enough to quickly determine the approximate place of a species within the well-studied phylogenetic domain.

In bacteria, for example, the region of the genome that can be barcoded is the region around the gene of 16S rRNA of the small ribosomal subunit [41, 42]. Bacterial species are distinguished by the sequences of four genes, including the 16S rRNA gene and the internal transcribed spacer (ITS; see Table 1 in [43]). This genetic marker is used for the so-called 16S profiling (also known as 16S metagenomic research), which is currently available in many a medical laboratory around the world. It should be noted that, in most cases, 16S metagenomics is only genus-specific, not allowing one to accurately identify the species of bacteria [44, 45]. To determine the bacterial species, it is necessary to use additional genetic markers or conduct a whole-genome sequencing. Barcoding of eukaryotic microbiota is also based on rRNA of the small ribosomal subunit; however, it is about eukaryotic ribosomes, so the approach is called 18S metagenomics [46–48]. 18S metagenomics is used to examine the fungal and protozoan biodiversity of medical and veterinary samples.

In 2003, there emerged a new concept of DNA barcoding, which was based on nucleotide divergence and aimed at precise identification of organisms at the level of species [49]. To assess global biodiversity using the new taxonomic tool, international initiatives were launched in hundreds of countries. To date, more than 320,000 species: animals, plants, fungi, bacteria (archaea included) – have been barcoded [43]. Eight mitochondrial and five nuclear genes are used to distinguish animal species; two nuclear and 15 plastid genes, for plants; four mitochondrial and eight nuclear genes, for fungi. Ribosomal RNA and ITS genes are utilized in all of these classification systems.

Other prospective approaches to the barcoding of organisms are based on the assessment of their conservative high-order characteristics (as has been done in this work) or whole-genome textual characteristics, as was done in [22].

However, it quickly became clear that there is no single short sequence or single genetic marker, which can predict the location and length of the sequence that needs to be analyzed to determine the virus, since the diversity of viruses in nature is many orders of magnitude greater than that of cellular life forms. Viruses and their groups are so diverse in various segments of their genomes that it is simply impossible to find common genetic markers which would be shared within viral taxa. Therefore, to identify a virus, a full-length genome sequencing is now used. *Straboviridae* are currently represented by 782 genomes (accessed: 18.10.2024) in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), with at least 50 new genomes appearing per year. The discovery of new *Straboviridae* viruses is mainly driven by the prospects of their use in the treatment of the diseases caused by gram-negative bacteria (primarily gamma-proteobacteria and, in particular, pathogenic *Enterobacteria*). Among *Straboviridae*, the genus with the highest practical potential is *Tequatrovirus*, which is represented by 280 (date of access: 18.10.2024) full genomes (<https://www.ncbi.nlm.nih.gov/taxonomy>). Obviously, the analysis of this rapidly growing body of genomic information requires a variety of bioinformatics approaches.

In this paper, we have used barcoding to identify bacteriophages whose genomes are much smaller than those of plants and animals – by six orders of magnitude [20, 21]. The results obtained demonstrate that the approach is universal and can be applied to genomes of any size.

The method of multiplex virtual PCR used here is not tied to any specific marker sequence in the genome. Instead, it defines a quantitative characteristic reflecting the size variation of replicas that can theoretically be generated with the help of special primers during multiplex virtual PCR.

To assess the efficiency of the proposed virtual multiplex PCR method, we have compared its results with the results of other approaches used for comparative analysis of bacteriophage genomes. The first approach, which we developed earlier [31], is based on estimating the evolutionary distance between organisms as a difference between their *k*-mer dictionaries. Correspondingly, to evaluate evolutionary distances, phylogenomic trees have been constructed for the same sample of 67 genomes, including bacteriophages from the genus *Tequatrovirus*, as

well reference genera *Mosigvirus* and *Dhakavirus*. The second approach is based on the alignment of nucleotide/amino acid sequences for the major capsid protein, the most frequently used genetic marker in T4 phages. It utilizes the maximum likelihood algorithm. Comparison of the trees have shown that the results produced by different methods are in good agreement with each other, thereby validating the methodology proposed.

CONCLUSIONS

An important purpose of future research in the field of therapy of infectious diseases is the development of specific pathogens-targeting strategies. The traditional broad-spectrum antibiotics can lead to the appearance of drug-resistant strains and disrupt the commensal microbiota. By identifying specific microbial species or virulence factors responsible for the infection, one can develop treatment courses which will selectively target these pathogens while sparing beneficial members of the microbiome.

The use of bacteriophages as an alternative or complementary medicine for the struggle with infectious diseases is another area of future research. Bacteriophages have advantages over antibiotics, including high specificity to their bacterial hosts and a lower likelihood of developing resistance. Studies aimed at the isolation of new bacteriophages from natural sources, as well as genetical modification of the known viruses are currently underway.

The great variety of bacteriophages and multiple areas of their use (phage processing, phage therapy, phage display, etc.) demand for standardization of terms, protocols – and also methods of bacteriophage certification. The Subcommittee on Bacterial and Archaeal Viruses of the International Committee on Taxonomy of Viruses (ICTV) established a systematic nomenclature of phage genomes and consistent annotation of phage sequences [50–54]. This service is supposed to form a “biospecimen navigation platform”. However, for the service to function, it is necessary to develop simple and effective tools (first of all, a “phage passport”) [55]. The passport might need to be defined within each national jurisdiction, but ideally it should be globally competent, covering the requirements for all the jurisdictions and using a common language/protocols for the genomic research. The passport system should serve as a basis for the phage-therapy regulatory framework, veterinary phage processing and public health needs requiring equal access. The developers of phage drugs are discussing the organizational and legal issues of such a “bacteriophage passport” based on the biological characteristics of the phages [56–58].

Bacteriophage T4 and its closest relatives (T-even bacteriophages) are classical objects of biological research. All the isolates of T4-type bacteriophages tested to date have a conservative genetic module that encodes components of the virion, including gene 23 (*g23*). This is the gene for the major capsid protein. The analysis of its sequence reveals a striking level of diversity among T4-type bacteriophages.

Analysis of molecular events in T4-infected *Escherichia coli* revealed some of the most important principles of biology, including the relationship between gene structures and their products, virus-induced acquisition of metabolic functions, and morphogenesis of complex structures through sequential interactions of gene products rather than sequential activation of genes. T4-type bacteriophages and related strains were involved in the first formulations of many fundamental biological concepts: unambiguous recognition of nucleic acids as a genetic material, definition of the notion of “gene” by studying mutations in its fine structure, recombinational and functional assays, demonstration that the genetic code is triplet, discovery of mRNA, the importance of DNA recombination and replication, light-dependent and light-independent mechanisms of DNA repair, DNA restriction and modification, intron-exon self-splicing in prokaryotes, translational bypass etc. All these discoveries indicate how attractive this group of phages is when it comes to extending new approaches to such a large and diverse group of species as viruses. Bacteriophage T4 and related phages are among the most studied biological objects, which can be of help in the development of new bioinformatic approaches. In this study,

we have investigated applicability of multiplex virtual PCR for identifying bacteriophage genomes, but this approach can also be applied to viruses of eukaryotes comparable in length with bacteriophages. The validity of such an extrapolation is based on the similarity of medium-length viral genomes, as well as principles of evolution of viruses.

This paper proposes a method for barcoding bacteriophage genomes, which makes it possible to unambiguously identify the strain of every phage species. This method of certification is universal: its applicability for identification of animals and plants has already been demonstrated. In the study, we have obtained barcodes (binary sequences) for a sample of the most conservative bacteriophages, and the phylogenetic trees constructed on the basis of these data are shown to be consistent with the previously obtained trees based on other algorithms for calculating distances between genomes.

The study was supported by a grant from the Russian Science Foundation No. 24-64-00017, <https://rscf.ru/project/24-64-00017/>

REFERENCES

1. Ventola C.L. The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics*. 2015. V. 40. No. 4. P. 277–283.
2. Ventola C.L. The antibiotic resistance crisis: part 2: management strategies and new agents. *Pharmacy and Therapeutics*. 2015. V. 40. No. 5. P. 344–352.
3. Górski A., Międzybrodzki R., Weber-Dąbrowska B., Fortuna W., Letkiewicz S., Rogóż P., Jończyk-Matysiak E., Dąbrowska K., Majewska J., Borysowski J. Phage therapy: combating infections with potential for evolving from merely a treatment for complications to targeting diseases. *Front. Microbiol.* 2016. V. 7. Article No. 1515. doi: [10.3389/fmicb.2016.01515](https://doi.org/10.3389/fmicb.2016.01515)
4. Golkar Z., Bagasra O., Pace D.G. Bacteriophage therapy: a potential solution for the antibiotic resistance crisis. *J. Infect. Dev. Ctries.* 2014. V. 8. No. 2. P. 129–136. doi: [10.3855/jidc.3573](https://doi.org/10.3855/jidc.3573)
5. Nikulin N., Nikulina A., Zimin A., Aminov R. Phages for treatment of *Escherichia coli* infections. *Prog. Mol. Biol. Transl. Sci.* 2023. V. 200. P. 171–206. doi: [10.1016/bs.pmbts.2023.03.011](https://doi.org/10.1016/bs.pmbts.2023.03.011)
6. Łusiak-Szelachowska M., Międzybrodzki R., Drulis-Kawa Z., Cater K., Knežević P., Winogradow C., Amaro K., Jończyk-Matysiak E., Weber-Dąbrowska B., Rękas J. et al. Bacteriophages and antibiotic interactions in clinical practice: What we have learned so far. *J. Biomed. Sci.* 2022. V. 29. Article No. 23. doi: [10.1186/s12929-022-00806-1](https://doi.org/10.1186/s12929-022-00806-1)
7. Romero-Calle D., Guimarães Benevides R., Góes-Neto A., Billington C. Bacteriophages as alternatives to antibiotics in clinical care. *Antibiotics*. 2019. V. 8. Article No. 138. doi: [10.3390/antibiotics8030138](https://doi.org/10.3390/antibiotics8030138)
8. Hankin E.H. L'action bactericide des eaux de la Jumna et du Gange sur le vibron du cholera. *Annales de l'Institut Pasteur*. 1896. V. 10. P. 511–523. (In French).
9. Gamaleya N.F. Bacteriolysins – enzymes that destroy bacteria. *Russian archive of pathology*. 1898. No. 6. P. 607–613. (In Russ.)
10. Twort F.W. An investigation on the nature of ultra-microscopic viruses. *Lancet*. 1915. V. 186. P. 1241–1243. doi: [10.1016/S0140-6736\(01\)20383-3](https://doi.org/10.1016/S0140-6736(01)20383-3)
11. d'Herelle F. Sur un microbe invisible antagoniste des bacilles dysentériques. *C. R. Hebd. Seances Acad. Sci.* 1917. V. 165. No. 11. P. 373–375. (In French).
12. Moelling K., Broecker F., Willy C. A wake-up call: we need phage therapy now. *Viruses*. 2018. V. 10. Article No. 688. doi: [10.3390/v10120688](https://doi.org/10.3390/v10120688)
13. Abedon S., Danis-Wlodarczyk K., Alves D. Phage therapy in the 21st century: is there modern, clinical evidence of phage-mediated efficacy? *Pharmaceuticals*. 2021. V. 14. Article No. 1157. doi: [10.3390/ph14111157](https://doi.org/10.3390/ph14111157)

14. Kamel M., Aleya S., Alsubih M., Aleya L. Microbiome dynamics: a paradigm shift in combatting infectious diseases. *J. Pers. Med.* 2024. V. 14. No. 2. Article No. 217. doi: [10.3390/jpm14020217](https://doi.org/10.3390/jpm14020217)
15. Chang C., Yu X., Guo W., Guo C., Guo X., Li Q., Zhu Y. Bacteriophage-mediated control of biofilm: a promising new dawn for the future. *Front. Microbiol.* 2022. V. 13. Article No. 825828. doi: [10.3389/fmicb.2022.825828](https://doi.org/10.3389/fmicb.2022.825828)
16. Merrill C.R., Biswas B., Carlton R., Jensen N.C., Creed G.J., Zullo S., Adhya S. Long-circulating bacteriophage as antibacterial agents. *Proc. Natl Acad. Sci. USA.* 1996. V. 93. P. 3188–3192. doi: [10.1073/pnas.93.8.3188](https://doi.org/10.1073/pnas.93.8.3188)
17. Zimin A.A., Butanaev A.M., Suzina N.E., Skoblikov N.E., Sakhabutdinova L.R., Prisyazhnaya N.V., Kononenko S.I., Kashchaev A.G. Construction of bacteriophage T4 mutants with reduced antigenicity. *Scientific journal of KubSAU.* 2017. No. 134(10). Article No. 34. doi: [10.21515/1990-4665-134-034](https://doi.org/10.21515/1990-4665-134-034) (In Russ.)
18. Nikulin N.A., Kononenko S.I., Koshchaev A.G., Zimin A.A. Construction of therapeutic phage cocktails based on bacteriophages: advantages and disadvantages. *Scientific journal of KubSAU.* 2017. No. 133(9). Article No. 63. doi: [10.21515/1990-4665-133-063](https://doi.org/10.21515/1990-4665-133-063) (In Russ.)
19. Kiryanova O.Yu., Kiryanov I.I., Kuluev B.R., Garafutdinov R.R., Chemeris A.V., Gubaidullin I.M. Multiplexed *in silico* RAPD analysis for genome barcoding. *Mathematical Biology and Bioinformatics.* 2022. T. 17. No. 2. P. 208–229. doi: [10.17537/2022.17.208](https://doi.org/10.17537/2022.17.208) (In Russ.)
20. Kiryanova O.Yu., Garafutdinov R.R., Chemeris D.A., Gerashchenkov G.A., Giniyatov Yu.R., Gubaidullin I.M., Chemeris A.V. Canine DNA polymorphism (*Canis familiaris* L.). VI. Genomic *in silico* barcoding of canine genomes and genomes of their wild relatives. *Biomika.* 2022. Vol. 14. No. 1. P. 59–67. doi: [10.31301/2221-6197.bmcs.2022-4](https://doi.org/10.31301/2221-6197.bmcs.2022-4) (In Russ.)
21. Kiryanova O.Yu., Kuluev B.R., Kuluev A.R., Mardanshin I.S., Gubaidullin I.M., Chemeris A.V. Multiplex *in silico* RAPD analysis of a number of related plants with different genome sizes and prospects for such an approach for DNA certification of agricultural plant varieties. *Biomika.* 2020. Vol. 12. No. 2. P. 194–210. doi: [10.31301/2221-6197.bmcs.2020-10](https://doi.org/10.31301/2221-6197.bmcs.2020-10) (In Russ.)
22. Nikulin N.A., Zimin A.A. Influence of non-canonical DNA bases on the genomic diversity of *Tevenvirinae*. *Front. Microbiol.* 2021. V. 12. Article No. 632686. doi: [10.3389/fmicb.2021.632686](https://doi.org/10.3389/fmicb.2021.632686)
23. Brettin T., Davis J., Disz T., Edwards R.A., Gerdes S., Olsen G.J., Olson R., Overbeek R., Parrello B., Pusch G.D., Shukla M. 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 No. 8365. doi: [10.1038/srep08365](https://doi.org/10.1038/srep08365)
24. Wattam A.R., Davis J.J., Assaf R., Boisvert S., Brettin T., Bun C., Conrad N., Dietrich E.M., Disz T., Gabbard J.L. et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* 2017. V. 45. No. D1. P. D535–D542. doi: [10.1093/nar/gkw1017](https://doi.org/10.1093/nar/gkw1017)
25. 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. No. 12. P. 1481–1487. doi: [10.1093/bioinformatics/btq229](https://doi.org/10.1093/bioinformatics/btq229)
26. Contreras-Moreira B., Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* 2013. V. 79. No. 24. P. 7696–7701. doi: [10.1128/AEM.02411-13](https://doi.org/10.1128/AEM.02411-13)

27. Petrov V.M., Ratnayaka S., Nolan J.M., Miller E.S., Karam J.D. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Viol. J.* 2010. V. 7. Article No. 292. doi: [10.1186/1743-422X-7-292](https://doi.org/10.1186/1743-422X-7-292)
28. Miller E.S., Kutter E., Mosig G., Arisaka F., Kunisawa T., Rüger W. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* 2003. V. 67. No. 1. P. 86–156. doi: [10.1128/MMBR.67.1.86-156.2003](https://doi.org/10.1128/MMBR.67.1.86-156.2003)
29. 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. No. 6. P. 1547–1549. doi: [10.1093/molbev/msy096](https://doi.org/10.1093/molbev/msy096)
30. Saitou N., Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 1987. V. 4. No. 4. P. 406–425. doi: [10.1093/oxfordjournals.molbev.a040454](https://doi.org/10.1093/oxfordjournals.molbev.a040454)
31. Nikulin N.A., Kiselev S.S., Panyukov V.V., Lu Y., Zimin A.A. Comparative analysis of *Actinobacteria* phage-plasmids and their transduction potential. *Mathematical Biology and Bioinformatics*. 2023. V. 18. P. 323–346. doi: [10.17537/2023.18.323](https://doi.org/10.17537/2023.18.323)
32. 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.
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](https://doi.org/10.1007/BF01734359)
34. 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](https://doi.org/10.1093/nar/gkh340)
35. Schwarz G. Estimating the dimension of a model. *Ann. Stat.* 1978. V. 6. P. 461–464. doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)
36. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 1985. V. 39. P. 783–791. doi: [10.1111/j.1558-5646.1985.tb00420.x](https://doi.org/10.1111/j.1558-5646.1985.tb00420.x)
37. Kiryanova O.Yu., Kiryanov I.I., Garafutdinov R.R., Chemeris A.V., Gubaidullin I.M. Computer program No. 2021667097 GATCGgenerator. URL: <https://onlinepatent.ru/software/2021667097/> (registration date 10/25/2021). (In Russ.)
38. Tavaré S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. Mathem. Life Sci.* 1986. V. 17. P. 57–86.
39. Gu X., Fu Y.X., Li W.H. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol. Biol. Evol.* 1995. V. 12. P. 546–557. doi: [10.1093/oxfordjournals.molbev.a040235](https://doi.org/10.1093/oxfordjournals.molbev.a040235)
40. Whelan S., Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 2001. V. 18. P. 691–699. doi: [10.1093/oxfordjournals.molbev.a003851](https://doi.org/10.1093/oxfordjournals.molbev.a003851)
41. Church D.L., Cerutti L., Gürtler A., Griener T., Zelazny A., Emler S. Performance and application of 16S rRNA gene cycle sequencing for routine identification of bacteria in the clinical microbiology laboratory. *Clin. Microbiol. Rev.* 2020. V. 33. Article No. e00053-19. doi: [10.1128/CMR.00053-19](https://doi.org/10.1128/CMR.00053-19)
42. Ludwig W., Schleifer K.H. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiol. Rev.* 1994. V. 15. No. 2-3. P. 155–173. doi: [10.1111/j.1574-6976.1994.tb00132.x](https://doi.org/10.1111/j.1574-6976.1994.tb00132.x)
43. Guo M., Yuan C., Tao L., Cai Y., Zhang W. Life barcoded by DNA barcodes. *Conserv. Genet. Resour.* 2022. V. 14. No. 4. P. 351–365. doi: [10.1007/s12686-022-01291-2](https://doi.org/10.1007/s12686-022-01291-2)
44. Regueira-Iglesias A., Balsa-Castro C., Blanco-Pintos T., Tomás I. Critical review of 16S rRNA gene sequencing workflow in microbiome studies: From primer selection to advanced data analysis. *Mol. Oral Microbiol.* 2023. V. 38. No. 5. P. 347–399. doi: [10.1111/omi.12434](https://doi.org/10.1111/omi.12434)

45. Morsli M., Salipante F., Magnan C., Dunyach-Remy C., Sotto A., Lavigne J.P. Direct metagenomics investigation of non-surgical hard-to-heal wounds: a review. *Ann. Clin. Microbiol. Antimicrob.* 2024. V. 23. No. 1. Article No. 39. doi: [10.1186/s12941-024-00698-z](https://doi.org/10.1186/s12941-024-00698-z)
46. Kuo M.T., Chen J.L., Hsu S.L., Chen A., You H.L. An omics approach to diagnosing or investigating fungal keratitis. *Int. J. Mol. Sci.* 2019. V. 20. No. 15. Article No. 3631. doi: [10.3390/ijms20153631](https://doi.org/10.3390/ijms20153631)
47. Drevinek P., Hollweck R., Lorenz M.G., Lustig M., Bjarnsholt T. Direct 16S/18S rRNA gene PCR followed by Sanger sequencing as a clinical diagnostic tool for detection of bacterial and fungal infections: a systematic review and meta-analysis. *J. Clin. Microbiol.* 2023. V. 61. No. 9. Article No. e0033823. doi: [10.1128/jcm.00338-23](https://doi.org/10.1128/jcm.00338-23)
48. Poulsen S.H., Søgaard K.K., Fuursted K., Nielsen H.L. Evaluating the diagnostic accuracy and clinical utility of 16S and 18S rRNA gene targeted next-generation sequencing based on five years of clinical experience. *Infect. Dis. (Lond.)*. 2023. V. 55. No. 11. P. 767–775. doi: [10.1080/23744235.2023.2241550](https://doi.org/10.1080/23744235.2023.2241550)
49. Hebert P.D., Cywinska A., Ball S.L., deWaard J.R. Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 2003. V. 270. No. 1512. P. 313–321. doi: [10.1098/rspb.2002.2218](https://doi.org/10.1098/rspb.2002.2218)
50. Zhu Y., Shang J., Peng C., Sun Y. Phage family classification under *Caudoviricetes*: A review of current tools using the latest ICTV classification framework. *Front. Microbiol.* 2022. V. 13. Article No. 1032186. doi: [10.3389/fmicb.2022.1032186](https://doi.org/10.3389/fmicb.2022.1032186)
51. Turner D., Adriaenssens E.M., Lehman S.M., Moraru C., Kropinski A.M. Bacteriophage taxonomy: a continually evolving discipline. *Methods Mol. Biol.* 2024. V. 2734. P. 27-45. doi: [10.1007/978-1-0716-3523-0_3](https://doi.org/10.1007/978-1-0716-3523-0_3)
52. Krupovic M., Turner D., Morozova V., Dyal-Smith M., Oksanen H.M., Edwards R., Dutilh B.E., Lehman S.M., Reyes A., Baquero D.P. et al. Bacterial Viruses Subcommittee and Archaeal Viruses Subcommittee of the ICTV: update of taxonomy changes in 2021. *Arch. Virol.* 2021. V. 166. No. 11. P. 3239–3244. doi: [10.1007/s00705-021-05205-9](https://doi.org/10.1007/s00705-021-05205-9)
53. Zerbini F.M., Siddell S.G., Lefkowitz E.J., Mushegian A.R., Adriaenssens E.M., Alfenas-Zerbini P., Dempsey D.M., Dutilh B.E., García M.L., Hendrickson R.C. et al. Changes to virus taxonomy and the ICTV Statutes ratified by the International Committee on Taxonomy of Viruses (2023). *Arch. Virol.* 2023. V. 168. No. 7. Article No. 175. doi: [10.1007/s00705-023-05797-4](https://doi.org/10.1007/s00705-023-05797-4)
54. Adriaenssens E., Brister J.R. How to name and classify your phage: an informal guide. *Viruses*. 2017. V. 9. No. 4. Article No. 70. doi: [10.3390/v9040070](https://doi.org/10.3390/v9040070)
55. Lin R.C.Y., Iredell J., Khatami A. After the paper: Australian Phage Network. URL: <https://communities.springernature.com/posts/after-the-paper-australian-phage-network> (accessed 18.10.2024)/
56. Onsea J., Soentjens P., Djebara S., Merabishvili M., Depypere M., Spriet I., De Munter P., Debaveye Y., Nijs S., Vanderschot P. et al. Bacteriophage application for difficult-to-treat musculoskeletal infections: development of a standardized multidisciplinary treatment protocol. *Viruses*. 2019. V. 11. No.10. Article No. 891. doi: [10.3390/v11100891](https://doi.org/10.3390/v11100891)
57. Pirnay J.P., Verbeken G., Ceyssens P.J., Huys I., De Vos D., Ameloot C., Fauconnier A. The magistral phage. *Viruses*. 2018. V. 10. No. 2. Article No. 64. doi: [10.3390/v10020064](https://doi.org/10.3390/v10020064)
58. Gibson S.B., Green S.I., Liu C.G., Salazar K.C., Clark J.R., Terwilliger A.L., Kaplan H.B., Maresso A.W., Trautner B.W., Ramig R.F. Constructing and characterizing bacteriophage libraries for phage therapy of human infections. *Frontiers in Microbiology*. 2019. V. 10. Article No. 2537. doi: [10.3389/fmicb.2019.02537](https://doi.org/10.3389/fmicb.2019.02537)

Received 19.11.2024.

Published 05.12.2024.