Limnology and Freshwater Biology 2020 (5): 1055-1066 DOI:10.31951/2658-3518-2020-A-5-1055

Original Article

ISSN 2658-3518
LIMNOLOGY FRESHWATER BIOLOGY
www.limnolfwbiol.com

Pseudomonas bacteriophage AN14 – a Baikal-borne representative of *Yuavirus*

Evseev P.V.[1], Gorshkova A.S.[2], Sykilinda N.N.[1], Drucker V.V.[2], Miroshnikov K.A.[1]*

[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Miklukho-Maklaya Str., 16/10, Moscow, 117997, Russia*
[2] *Limnological Institute, Siberian Branch of the Russian Academy of Sciences, Ulan-Batorskaya Str., 3, Irkutsk, 664033, Russia*

ABSTRACT. Siphoviruses with a prolate capsid infecting *Pseudomonas* sp. were isolated from environmental sources through the world, and were recently defined as a separate taxonomic genus *Yuavirus*. Besides a distinguishable morphology, the hallmarks of the genus are heavily modified genomic DNA and a lytic infection cycle while putative lysogeny genes are proposed in the genomes. Bacteriophage AN14 was isolated from Lake Baikal ecosystem and was assigned as a member of *Yuavirus*. We report the biological and morphological features if this phage, as well as the comprehensive re-annotation of its genome. Modern bioinformatics and phylogenomic analysis resulted in the refinement of the taxonomic attribution of the YuA-like phages and highlighting the specific genomic and proteomic features typical for *Yuavirus* phages including AN14.

Keywords: Bacteriophage, *Pseudomon*as, genomics, phylogeny, taxonomy

1. Introduction

Bacteriophage AN14 infecting *Pseudomonas* was discovered in the aquatic ecosystem of Lake Baikal in 2010. According to the current nomenclature justified by the International Committee for Virus (ICTV) in 2020, phage AN14 belongs to the genus *Yuavirus* of family *Siphoviridae*. ICTV database lists six previously reported phages as members of *Yuavirus* — *Alphaproteobacteria* virus φJL001, and *Pseudomonas* viruses LKO4, M6, MP1412, PAE1 and YuA. Moreover, the NCBI database and taxonomic browser contains information on another 12 phages presumably belonging to the genus *Yuavirus*. Most of them infect *Pseudomonas* sp., however the set includes also a *Bordetella* phage LK3. Previously, other *Bordetella* phages (CN1, CN2, FP1 and MW2) were considered as members of *Yuavirus*. These viruses are currently attributed as a separate genus *Vojvodinavirus*. The representatives of *Yuavirus* phages were found in geographically distant environmental aqueous reservoirs. Extreme diversity of the bacterial hosts of this phage group comprising alpha-, beta- and gamma-proteobacteria raises questions on the correction and possible revision of the current taxonomic classification.

The presented work reports the parameters of the AN14 infection cycle and morphology. We have re-annotated the AN14 genome taking into account the recent data on structures and functions of phage proteins. Genomic and phylogenetic research has refined the taxonomic position of phages belonging to genus *Yuavirus*.

2. Materials and methods

2.1. Phage propagation and purification

Pseudomonas aeruginosa strain PAO1 (ATCC 15692), purchased from the American Type Culture Collection (ATCC), was used as a host for phage propagation. A water sample taken from Selenga river (52.090687 106.637384) (Fig. 1) (3 mL) was supplemented with 1 mL of 4× lysogeny broth (LB) and 40 μL overnight culture of PAO1, and incubated at 37°C for 18 h. Chloroform was added to a final concentration of 0.5% (v/v) for 4 h at 4°C. The suspension was then centrifuged at 7000× g for 20 min. The presence of bacteriophages and the titer in the supernatant was determined by the appearance of plaques on the bacterial lawn using the double-agar layer technique (Adams, 1959) with minor modifications. The phage from a single plaque was propagated using a liquid culture of the *P. aeruginosa* PAO1 host strain. The incubation was performed at 37°C until the lysis completion, and then chloroform was added. Bacterial debris was pelleted by centrifugation at 3000× g for 20 min. The phage lysate was precipitated with polyethylene glycol (PEG) 8000 (10%) – NaCl (0.6%) at 4°C overnight, centrifuged at 8000× g for 20 min at 4°C, resuspended in the SM buffer (50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 8 mM $MgSO_4$, 0.01% gelatin) and then 1 M KCl was added. The mixture was incubated on ice for 20 min and centrifuged (12000× g for 20 min at 4 °C) to precipitate PEG (Colombet et al., 2007). The phage preparation was purified by cesium chloride

*Corresponding author.
E-mail address: kmi@ibch.ru (K.A. Miroshnikov)

equilibrium gradient centrifugation at 22000× g (Beckman SW41 Ti rotor, Germany) for 2 h. The phage band was collected and dialyzed overnight against 0.01 M Tris–HCl (pH 7.5), 0.01 M $MgSO_4$, 0.15 M NaCl at 4°C. The titer of the resulting phage was about 10^{11} plaque forming units (PFU)/mL.

2.2. Phage one-step growth curve

Phage suspension was mixed with 1 mL of exponential phase culture of strain PAO1 at MOI = 0.1 and incubated at 37°C for 10 min for phage adsorption. The mixture was centrifuged at 10,000× g for 5 min to remove free phage particles. The pellet was resuspended in 50 mL of LB, and the culture was continuously incubated at 37°C. Samples were taken at 10 min intervals to 2 h, and phage titer was determined.

2.3. Electron microscopy

The morphology of phage AN14 was examined by negative strain transmission electron microscopy (TEM). Briefly, a single plaque was resuspended in 100 μL of saline (0.8% NaCl aqueous solution), centrifuged for 15 min at 4000× g, and the supernatant was filtered through a 0.22 μm filter. The prepared phages were placed on freshly-prepared formvar films with 2% (w/v) uranyl acetate. Morphological characteristics were observed under a transmission electron microscope (Zeiss LEO 906E, Germany).

2.4. Reannotation of phage genome

The previously deposited genomic sequence of phage AN14 was downloaded from GenBank (Accession number KX198613) and annotated by predicting and validating open reading frames (ORFs) using Prodigal 2.6.1 (Hyatt et al., 2010) and Prokka 1.17 (Seemann, 2014) pipelines. Identified ORFs were manually curated to ensure fidelity. Functions were assigned to ORFs using a BLAST search on a custom phage protein database compiled from annotated phage GenBank sequences, InterPro server (https://www.ebi.ac.uk/interpro/entry/InterPro) and HHpred server (https://toolkit.tuebingen.mpg.de) with Pfam-A_v32.0, NCBI_Consreved_Domain_v.3.16, SMART_v6.0, PRK_6.9, PDB, SCOPe70_2.07, ECOD_ECOD_F70_20190225 and COG_KOG_v1.0 databases. Custom BLAST databases were mounted with the BLAST tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). tRNA coding regions were searched with tRNAscan-SE (Schattner et al., 2005). The resulting genome map was visualized in Geneious Prime, version 2020.0.5 (https://www.geneious.com).

2.5. Phylogenetic analysis

Phage reference genomes were downloaded from NCBI GenBank (ftp://ftp.ncbi.nlm.nih.gov/genbank). Whether necessary, the genomes were annotated using Prokka (Seemann, 2014), with a custom phage protein database compiled from annotated phage GenBank sequences. A search for homologous sequences was conducted using a BLAST search and found sequences were checked for the presence of annotated homologous genes in NCBI genomes. Genes were extracted from GenBank annotations. For some unannotated sequences, ORFs were found by Glimmer3 (Delcher et al., 1999). ORFs were validated and corrected by comparison with known homologous genes. Protein alignments were made with MAFFT (Katoh et al., 2002) (L-INS-i algorithm, BLOSUM62 scoring matrix, 1.53 gap open penalty, 0.123 offset value). The alignments were trimmed manually and with trimAL (Capella-Gutiérrez et al., 2009) with gappyout settings. Best protein models were found with MEGAX 10.0.5 (Kumar et al., 1994). Phylograms were generated based on the amino acid sequences of proteins and their concatenated alignments, using Geneious Prime and MAFFT for sequence alignment. Trees were constructed using the maximum likelihood (ML) method with an RAxML program (Stamatakis, 2014) with a WAG + G protein model and the robustness of the trees was assessed by bootstrapping (1000).

Fig.1. Map of Lake Baikal. The sampling station is indicated by an asterisk.

2.6. Whole-genome and proteome analysis

Average nucleotide identity (ANI) was computed using the OrthoANIu tool (Lee et al., 2016), employing USEARCH (http://www.drive5.com/usearch/) over BLAST (https://www.ezbiocloud.net/tools/orthoaniu) with default settings and with an EzBioCloud server (https://www.ezbiocloud.net/tools/ani). Genome comparison was made with BRIG (Alikhan et al., 2011). Proteome analysis was performed with BPGA software (Chaudhari et al., 2016).

2.7. 3D homology modeling, alignment and visualization

Protein remote homology detection, 3D structure prediction and template-based homology prediction were made by Phyre2 protein fold recognition server (Kelley et al., 2015) (http://www.sbg.bio.ic.ac.

uk/~phyre2), HHpred (https://toolkit.tuebingen.mpg.de/tools/hhpred). The obtained structures were aligned and visualized with UCSF Chimera (Goddard et al., 2018).

3. Results and discussion

3.1. General biological characteristics of phage AN14

Bacteriophage AN14 was isolated and propagated using the type strain PAO1 of *Pseudomonas aeruginosa*. It produces small clear plagues (~1 mm diameter). The one-step growth curve shows a latent period of 40 ± 5 min, slow lysis, and a burst size about 110 PFU/cell at 37°C (Fig. 2). The infection range of phage AN14 is considered as relatively narrow. Only two of 53 assayed environmental and clinical isolates of *P. aeruginosa* were susceptible to AN14.

Electron microscopic imaging reveals AN14 as a typical member of the *Siphoviridae* family (Fig. 3), having a flexible, non-contractile tail (~150 nm) and an elongated capsid (~70 x 50 nm) (B2 morphotype). These dimensions are typical for all representatives of *Yuavirus* where the morphology was recorded.

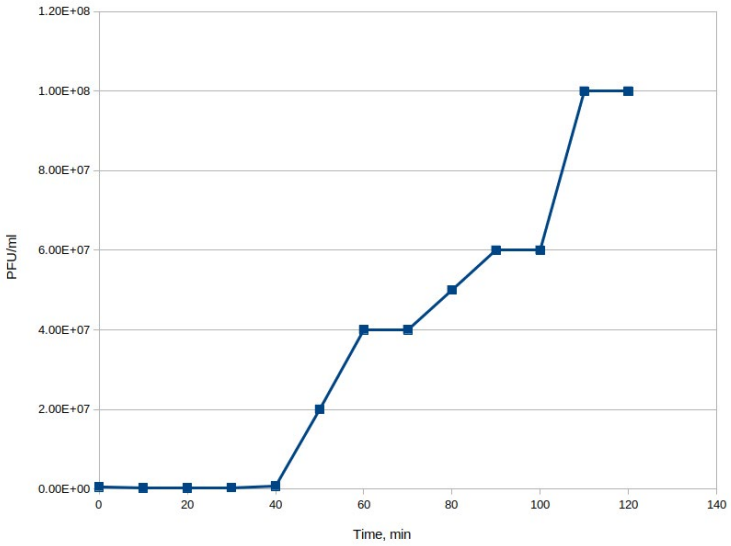


Fig.2. One step growth curve of AN14 using *P. aeruginosa* PAO1 as a host.

3.2. AN14 genome — general features

The number of entries in Genbank phage database grows rapidly (Sayers et al., 2019), as well as the number of resolved protein structures in the corresponding databases (https://www.rcsb.org/stats/growth/growth-released-structures). Exploiting the recent available data enables us to improve the quality of existing annotations.

The length of phage AN14 dsDNA genome is 60,973 bp (accession number KX198613), and it exists in a circular permuted form. Therefore, the selection of the "zero point" is arbitrary, and a number of alternative ways of gene numeration are offered in GenBank annotations. We have chosen to align the AN14 genome with the genome of a type phage YuA, since it logically reflects the employment of genes in the infection cycle. The genome of AN14 (Fig. 4, Table 1) contains 86 putative ORFs, putative functions of 58 proteins can be predicted, and 28 ORFs are assigned as hypothetical proteins. There are no tRNA genes found in the genome. All genes are oriented in the same direction. G + C content of the AN14 genome is 64.5%, evenly distributed through the genome (Fig. 4). *Yuavirus* phages infecting *Pseudomonas* have the G + C value most close to the average G + C (65%) of the bacterial host. This may be the trait of a long-term adaptation of the phage to its host (Ceyssens et al., 2008). Notably, the G + C content of the genomes of *Pseudomonas* YuA-like phages (64.3%-64.7%) is higher than in the genomes of *Alphaprotebacteria* phage φJL001 (62%) and *Bordetella* phage LK3 (64.0%).

The genes of AN14 genome are clustered in three blocks – the first one encodes structural and lysis proteins, the second one is responsible for nucleotide metabolism and modification proteins, and the third block encodes replication, transcription and other proteins. The genes of structural proteins comprise about one third of the genome and are well conserved in

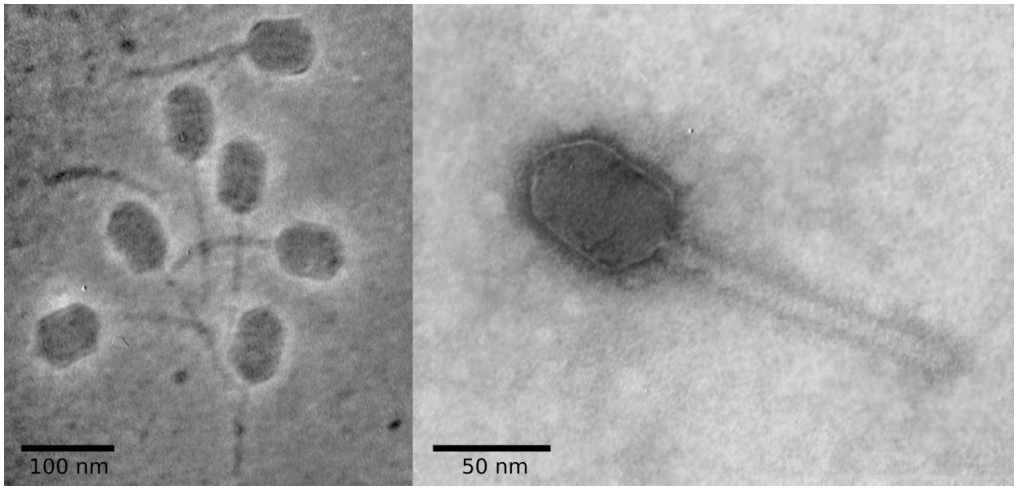


Fig.3. Transmission electron microscopy image of phage AN14. Staining with 2% uranyl acetate. Scale bars – 50 nm and 100 nm.

Table 1. Functional assignments of *Pseudomonas* phage AN14 genes.

Name	From	To	Length	Direction	locus tag
helicase	379	2073	1695	forward	gp01
hypothetical protein	2081	2434	354	forward	gp02
exonuclease inhibitor protein	2494	3039	546	forward	gp03
ATP-binding protein	3182	4021	840	forward	gp04
exodeoxyribonuclease V subunit RecD	4123	5316	1194	forward	gp05
ribonucleotide reductase	5377	7200	1824	forward	gp06
uracil-DNA glycosylase	7342	7848	507	forward	gp07
hypothetical protein	7948	8631	684	forward	gp08
hypothetical protein	8642	8905	264	forward	gp09
alpha-glutamyl/putrescinyl thymine pyrophosphorylase clade 3	8915	9790	876	forward	gp10
1-aminocyclopropane-1-carboxylate deaminase	9792	10628	837	forward	gp11
radical SAM protein	10700	12142	1443	forward	gp12
thymidylate kinase	12197	12760	564	forward	gp13
alpha-glutamyl/putrescinyl thymine pyrophosphorylase clade 1	12772	13743	972	forward	gp14
hypothetical protein	13799	14359	561	forward	gp15
hypothetical protein	14372	14485	114	forward	gp16
thymidylate synthase/dCMP hydroxymethylase	14482	15537	1056	forward	gp17
putative phosphohydrolase	15550	16185	636	forward	gp18
hypothetical protein	16160	16387	228	forward	gp19
hypothetical protein	16356	16742	387	forward	gp20
DNA polymerase A	16742	18736	1995	forward	gp21
hypothetical protein	18733	19047	315	forward	gp22
phage_holin_1-like protein	19044	19238	195	forward	gp23
deoxycytidylate deaminase	19235	19696	462	forward	gp24
restriction endonuclease	19689	20087	399	forward	gp25
lambda-repressor-like protein	20072	20398	327	forward	gp26
DNA-primase	20411	22822	2412	forward	gp27
dsRBD-like protein	22833	23048	216	forward	gp28
membrane protein	23160	23525	366	forward	gp29
hydrolase, metallopeptidase	23564	24742	1179	forward	gp30
hypothetical protein	24739	25080	342	forward	gp31
hypothetical protein	25092	25361	270	forward	gp32
nucleotide pyrophosphohydrolase	25358	25969	612	forward	gp33
hypothetical protein	25966	26442	477	forward	gp34
hypothetical protein	26439	26768	330	forward	gp35
hypothetical protein	26787	27083	297	forward	gp36
hypothetical protein	27086	27373	288	forward	gp37
hypothetical protein	27378	27647	270	forward	gp38
hypothetical protein	27644	27808	165	forward	gp39
serine/threonine-protein kinase	27787	28404	618	forward	gp40
all-alpha NTP pyrophosphatase-like protein	28431	29114	684	forward	gp41
hypothetical protein	29114	29395	282	forward	gp42

Name	From	To	Length	Direction	locus tag
3'-5' exonuclease	29397	29969	573	forward	gp43
DNA-directed RNA polymerase subunit alpha	29962	30153	192	forward	gp44
hypothetical protein	30254	30862	609	forward	gp45
hypothetical protein	30991	31254	264	forward	gp46
DNA binding protein	31306	32193	888	forward	gp47
diguanylate-cyclase GGDEF-like protein	32198	32668	471	forward	gp48
antirestriction protein	32763	33176	414	forward	gp49
hypothetical protein	33234	33563	330	forward	gp50
hypothetical protein	33586	33798	213	forward	gp51
hypothetical protein	33795	34190	396	forward	gp52
hypothetical protein	34187	34489	303	forward	gp53
hypothetical protein	34486	34815	330	forward	gp54
hypothetical protein	34812	35000	189	forward	gp55
hypothetical protein	35004	35255	252	forward	gp56
terminase small subunit	35924	36451	528	forward	gp57
terminase large subunit	36451	37998	1548	forward	gp58
portal protein	38056	39594	1539	forward	gp59
membrane protein	39591	39704	114	forward	gp60
minor capsid protein	39706	40914	1209	forward	gp61
minor structural protein	40983	41765	783	forward	gp62
major capsid protein	41961	42893	933	forward	gp63
hypothetical protein	43000	43605	606	forward	gp64
hypothetical protein	43633	43803	171	forward	gp65
Rz-like lysis protein	43807	44376	570	forward	gp66
Rz1-like protein	44039	44305	267	forward	gp67
holin	44373	44708	336	forward	gp68
lysozyme, lytic murein transglycosylase F	44693	45418	726	forward	gp69
capsid decoration protein	45418	45936	519	forward	gp70
head-tail attachment protein	45940	46308	369	forward	gp71
tail protein	46305	46757	453	forward	gp72
major tail tube protein	46791	48323	1533	forward	gp73
tail assembly chaperone protein	48386	48844	459	forward	gp74
tail assembly chaperone protein	48868	49134	267	forward	gp75
putative tail competion protein	49118	49564	447	forward	gp76
tail tape measure protein	49567	52392	2826	forward	gp77
putative tail fiber protein	52402	53931	1530	forward	gp78
structural protein	53946	54944	999	forward	gp79
tail assembly structural protein	54946	56616	1671	forward	gp80
tail assembly structural protein	56613	57422	810	forward	gp81
membrane protein	57436	57669	234	forward	gp82
structural protein	57669	57872	204	forward	gp83
tail component	57856	60027	2172	forward	gp84
structural protein	60027	60512	486	forward	gp85
membrane protein	60580	60924	345	forward	gp86

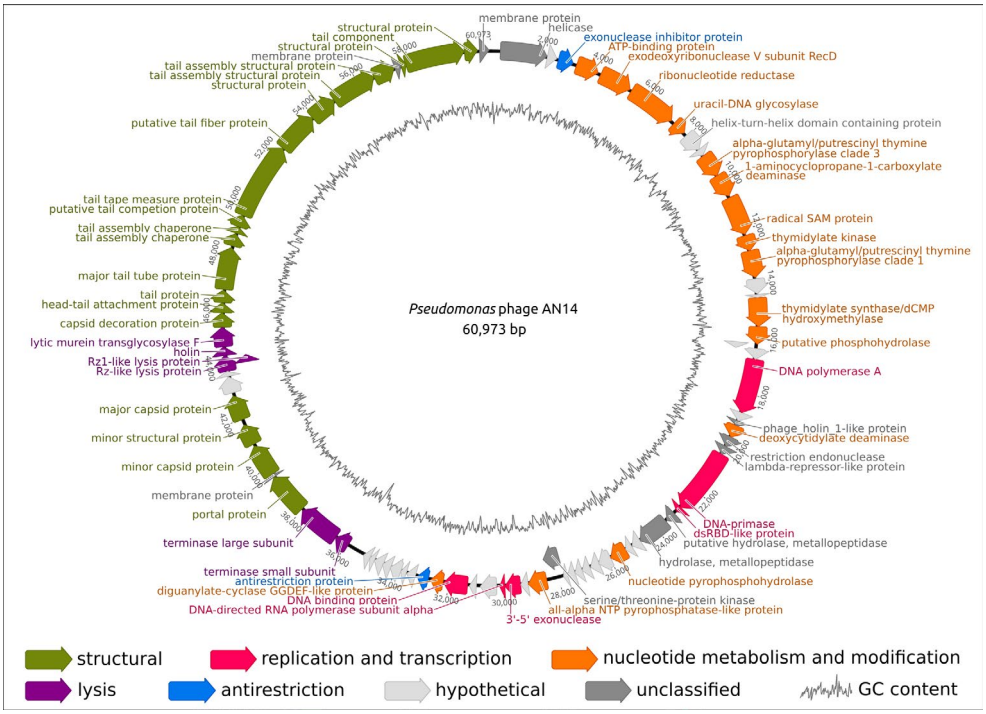


Fig.4. Circular diagram of functional assignments of the *Pseudomonas* phage AN14 genome. In total, 86 protein-coding genes are shown as colored blocks. The direction of transcription is shown by arrows. The GC content of the genome sequence is indicated by the internal grey curve.

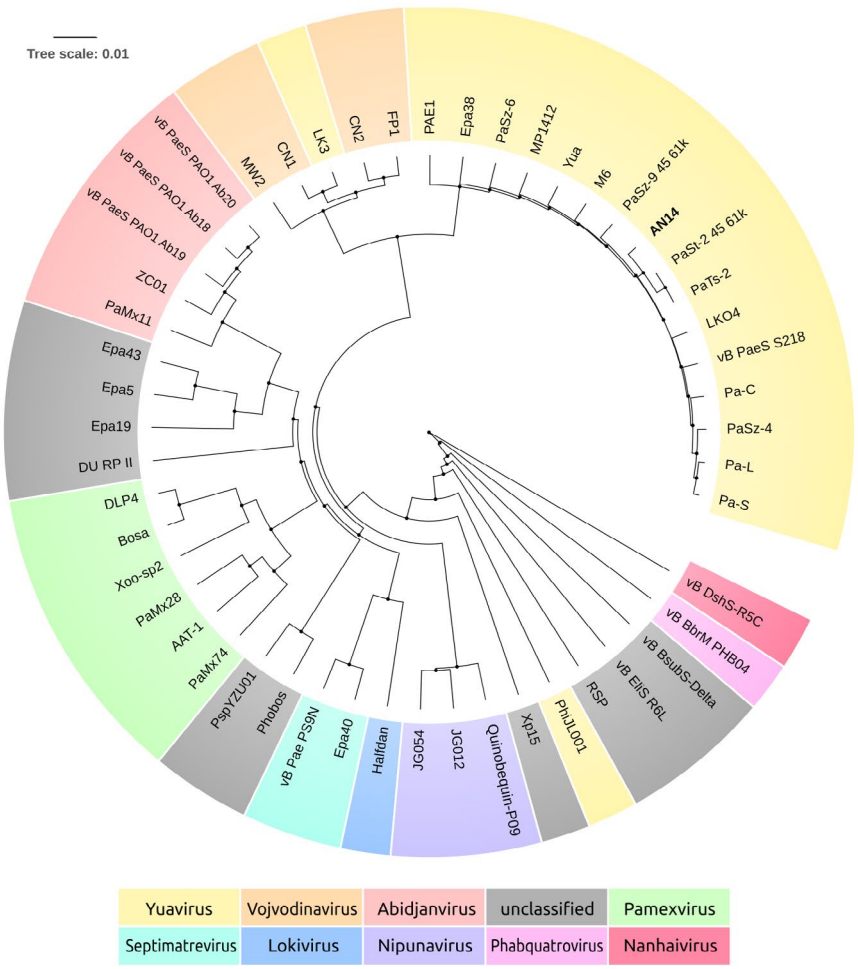
the representatives of the genus. The functions of most of these genes can be predicted using BLAST and HMM-HMM pipelines. A noticeable feature of phage YuA is the helical-shaped decoration of the prolate capsid with gene product (gp) 63 (Sokolova et al., 2010).The gene encoding the decoration protein is conservative among *Yuavirus* genomes and corresponds to *g70* in AN14.

The block for terminal lysis of the host cell in *Pseudomonas Yuavirus* phages contains genes encoding peptidoglycan-lysing enzyme, which has been proposed to function as murein transglycosylase F, and the holin located immediately downstream. The HMM-HMM and BLAST search also revealed the presence of nested genes encoding Rz-like and Rz1-like proteins adjacent to holin. Such proteins have been shown to form a complex spanning the periplasmic space, providing more efficient lysis of the host cell (Berry et al., 2008; Kongari et al., 2018). All these lysis proteins are conserved in the genomes of *Yuavirus* phages infecting *Pseudomonas* and *Vojvodinavirus* phages infecting *Bordetella*, but have no close homologs in *Alphaprotebacteria* phage φJL001, which is misclassified, as we suggest. Therefore, we propose that the *Yuavirus* phages encode the combined lysis module involving endolysin, holin, and Rz-like/ Rz1-like proteins.

One of the noticeable features of AN14 predicted proteome is a significant number of proteins involved in nucleotide modification and repair. At least 8 genes for such proteins can be found, and their homologs are present in other *Yuavirus* genomes. A large set of potential nucleotide modification proteins correlates with the experimentally found high content of modified bases in the genomes of phages YuA (Ceyssens et al., 2008) and M6 (Lee et al., 2018). Another interesting feature of AN14 genome is the presence an anti-restriction system (Spoerel et al., 1979), including antirestriction protein (gp49) and predicted exonuclease inhibitor protein (gp03). The AN14 genome contains diguanylate-cyclase GGDEF domain protein (gp48). This protein can function as diguanylate cyclase (DGC) to form cyclic di-GMP, a global second messenger controlling bacterial motility and sessility. Higher activity of DGC results to an elevated cyclic di-GMP level plays an important role in the early stage of biofilm formation in *P. aeruginosa* (Simm et al., 2004; Bae et al., 2012). We have also identified genes presumably participating in amino acid metabolism such as serine/threonine-protein kinase (gp13). The previous annotation of AN14 genome and the annotations of the genomes of YuA-like phages genomes have stated a presence of an integrase-like protein (gp27), but we propose that this protein is most likely a DNA primase. Thus, we can reinterpret important features of the lifecycle of *Yuavirus* phages (see below).

3.3. Taxonomy

Calculations of average nucleotide identity (ANI) between AN14 and all 21067 phage sequences deposited in the NCBI GenBank (up to July 2020) with OrthoANIu have revealed 50 phages with genomic ANI 53.8%–97.8% compared to AN14 (Fig. 5), which can be considered as relatives to phage AN14. The proteomic tree clusters *Yuavirus* phages infecting *Pseudomonas* (including AN14) as a distinct group (Fig. 5). According to the formal 95% ANI cutoff for the species definition among prokaryotes (Konstantinidis and Tiedje, 2005; Varghese et al., 2015) that is usually applied to the phages, most representatives of *Yuavirus* are the strains (phage races) within the same species (Fig. 5).



Phage	ANI
Pseudomonas phage AN14	100.0
Pseudomonas virus M6	97.8
Pseudomonas phage PaSz-4	97.8
Pseudomonas phage Pa-L	97.6
Pseudomonas phage Pa-S	97.6
Pseudomonas phage PAE1	97.4
Pseudomonas phage PaSt-2_45_61k	97.4
Pseudomonas phage PaTs-2	97.4
Pseudomonas virus PaSz-6	97.3
Pseudomonas phage LKO4	97.0
Pseudomonas phage vB_PaeS_S218	97.0
Pseudomonas phage MP1412	96.3
Pseudomonas phage Pa-C	96.2
Pseudomonas phage Epa38	95.5
Pseudomonas phage PaSz-9_45_61k	95.4
Pseudomonas virus Yua	93.9
Bordetella virus LK3	78.9
Bordetella phage MW2	78.4
Bordetella phage CN1	78.2
Bordetella phage CN2	77.3
Bordetella phage FP1	76.1
Pseudomonas phage Epa19	71.2
Pseudomonas phage PaMx74	70.2
Pseudomonas phage AAT-1	69.2
Pseudomonas phage vB_PaeS_PAO1_Ab18	68.1
Pseudomonas phage vB_PaeS_PAO1_Ab19	67.5
Pseudomonas phage JG054	67.3
Pseudomonas phage Epa43	67.0
Stenotrophomonas phage DLP4	66.9
Xanthomonas phage Xp15	66.7
Pseudomonas phage vB_PaeS_PAO1_Ab20	66.4
Pseudomonas phage PaMx28	65.6
Pseudomonas phage ZC01	65.3
Xanthomonas phage Bosa	65.3
Pseudomonas phage PaMx11	65.2
Pseudomonas phage Epa5	65.0
Pseudomonas phage Quinobequin-P09	64.7
Xanthomonas phage Xoo-sp2	63.5
Pseudomonas phage JG012	62.7
Brevundimonas phage vB_BsubS-Delta	61.8
Bordetella phage vB_BbrM_PHB04	60.8
Alphaproteobacteria phage phiJL001	60.8
Erythrobacter phage vB_EliS_R6L	59.8
Escherichia phage Halfdan	59.8
Dinoroseobacter phage vB_DshS-R5C	58.2
Pseudomonas phage vB_Pae_PS9N	57.9
Pseudomonas phage RSP	56.7
Ralstonia phage DU_RP_II	56.7
Pseudomonas phage Epa40	56.4
Bacteriophage Phobos	54.1
Pseudomonas phage PspYZU01	53.8

Fig.5. Proteome UPGMA tree constructed with BPGA pipeline (left diagram) and table of ANI values compared to AN14, calculated with orthoANIu (right). The tree scale bars show estimated substitutions per site.

Terminase is one of the most conserved protein complexes encoded in bacteriophage genomes, and it has been frequently used for taxonomic grouping of phages (Smith et al., 2013). In order to construct a consistent taxonomy and phylogenetic positioning of phage AN14, we performed a BLAST search using the terminase large subunit protein sequences and GenBank phage database, and constructed a list of phages belonging to the taxa, representatives of which were found with E -value < 10^{-8}. The maximum likelihood (ML) phylogeny of terminase large subunit protein sequence (Fig. 6) clearly showed the affiliation of AN14 with the *Yuavirus* clade. The tree groups 16 *Pseudomonas* phages into a distinct clade and points to *Vojvodinavirus* as the genus closest evolutionally. These phages are listed in Table 2. Contrary to the current ICTV classification, Alphaprotebacteria phage φJL001 (AY576273) (Lohr et al., 2005) seems to belong to a group comparatively distant from *Yuavirus*, and *Bordetella* virus LK3 (KX961385) belongs to the genus *Vojvodinavirus*. These conclusions are in the agreement with proteome clustering, ANI calculations, genome sequence comparison among 10 AN14-related phage genomes, and ML phylogeny based on 39 core proteins concatenated sequences (Fig. 5, Fig. 6).

The terminase phylogeny points to *Pseudomonas* phages of the *Abijanvirus* genus as the next closest group to *Vojvodinavirus*. More distant relatives also represent the phages of gram-negative bacteria, mostly *Pseudomonas* spp. (Fig. 6). Phylogenetic analysis suggests that *Alphaprotebacteria* phage φJL001 terminase have diverged from the common clade comprising *Yuavirus* and related groups even earlier than the divergence to genera *Stenhofvirus*, *Pamexvirus*, *Abijanvirus* and *Vojvodinavirus* occurred.

3.4. *Yuavirus* pan-genome and pan-proteome features

The pan-genome analysis as well as ANI data demonstrated the high similarity of *Pseudomonas Yuavirus* phages. 55 core protein clusters (cut-off 0.5) have representatives in genomes of all *Yuavirus* phages infecting *Pseudomonas*. According to COG classification, about 37% of *Yuavirus* core proteome can be associates with the replication, recombination and repair protein category (L), 25% falls into the nucleotide transport and metabolism (F) category, 25% of core proteome belongs to the transcription (K) category, and about 91% of accessory proteome falls into the signal transduction mechanisms (T) category (Fig. 7). The analysis of *Yuavirus* proteins revealed the presence of homing endonucleases in some genomes (Bae et al., 2012; Dyson et al., 2016).

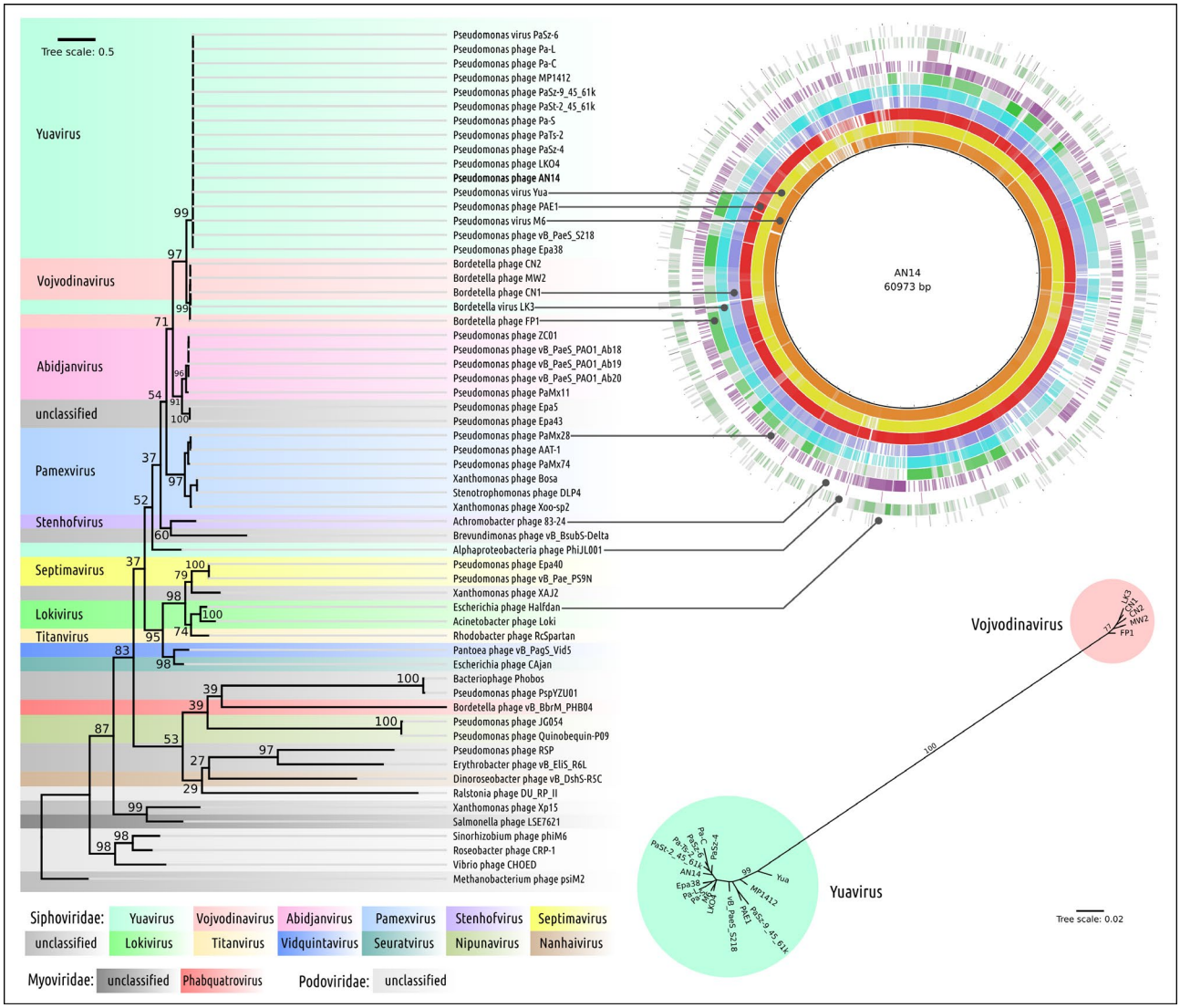


Fig.6. Best-scoring tree found by ML search with RAxML based on the terminase large subunit protein sequences (left figure), genome sequence comparison among 10 AN14-related phage genomes exhibiting co-linearity detected by TBLASTX (upper-right circular diagram), and best-scoring tree found by ML search with RAxML based on 39 core proteins concatenated sequences (lower-right figure). Bootstrap support values are shown near their branching as a percentage of 1000 replicates. The tree scale bars show estimated substitutions per site.

The 39 concatenated protein phylogeny (Fig. 6) testifies that *Pseudomonas* phage YuA have diverged first among 16 *Pseudomonas Yuavirus* phages. Pan-genome analysis demonstrates that phage YuA shares the least number of common proteins with other 15 phages of this group. Meanwhile, *Pseudomonas* phage AN14 seems to be evolutionary closer to phages M6, PaTs-2_45_61k and PaTs-2 (Fig. 7, Table 2) and contains no genes absent in the genomes of *Yuavirus* phages.

3.5. Infection cycle of *Yuavirus* phages

Previously, it was shown experimentally that *Pseudomonas* phage YuA (Ceyssens et al., 2008) and *Bordetella* phages CN1, FP2, LK3 and MW2 (Petrovic et al., 2017), closely related to AN14, infect their hosts through a lytic cycle only. At least, the lysogenic behavior has not been observed, and we could not find any prophage sequences resembling the AN14 genome within the published *Pseudomonas* genomes with the BLAST search. More distantly related phage φJL001 that infects an uncharacterized marine alphaproteobacterium, JL001, was reported to be a temperate phage, but any stable lysogens were not detected as well. While induction of φJL001 and the homoimmunity characteristics of φJL001 resemble true lysogeny, the phage φJL001 genome does not integrate into host cellular replicons, possibly demonstrating a case of pseudolysogeny (Lohr et al., 2005).

The main rationale to state a possible temperate lifestyle of *Yuavirus* phages is the presence of the putative conserved integrase gene (*g27* in case of AN14) and the repressor (*g26*) in the genome. However, the homologs of AN14 gp27, annotated in φJL001, YuA, M6 and other related phages as integrases, may perform a different function not related to lysogeny. Alternatively, it may be attributed as DNA primase/polymerase belonging to AEP-primases. This proposition arises from HMM-HMM search and 3D homologous modelling (Fig. 8). The comparison with the closest structural homologs reveals that the simulated structure of AN14 gp27 contains two domains. N-terminal domain may possess the priming

activity, and the larger C-terminal domain possesses the helicase activity, supposedly belonging to the helicase Superfamily 3. According to HMM-HMM searches conducted with HHpred and Pyre2 server, the RepB' protein of the broad-host-range plasmid RSF1010 (Geibel et al., 2009) is the closest structural homolog of the AN14 DNA primase priming domain (>99% probability), and the helicase domain of the deep-sea vent phage NrS-1 polymerase (Chen et al., 2020) seems to be the closest known structural homolog of the AN14 helicase domain. Enzymes of this group may be involved to DNA reparation processes (Guilliam et al., 2015). Putative λ-like repressor encoded in AN14, as well as its homolog in φJL001, might hypothetically participate in the homoimmunity formation observed in the course at φJL001 infection (Lohr et al., 2005). Repressor in phage λ was shown to prevent the superinfection (Fogg et al., 2010).

Undoubtedly, this hypothesis requires a future experimental proof. However, it gives a reasonable explanation for the long-disputed contradiction between the presence of predicted lysogeny genetic apparatus, and lysogeny never observed experimentally for *Yuavirus* phages.

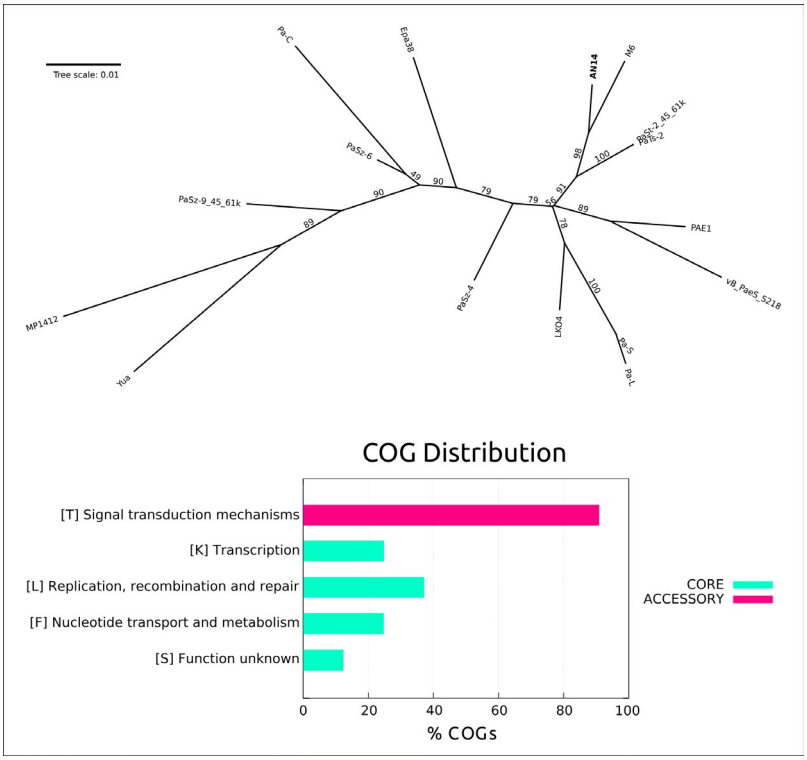


Fig.7. Best-scoring tree found by maximum likelihood (ML) search with RAxML based on 55 core proteins concatenated sequences and COG distribution diagram of *Yuavirus* pan-proteome constructed with BPGA pipeline. Bootstrap support values are shown near their tree branching as a percentage of 1000 replicates. The tree scale bars show estimated substitutions per site.

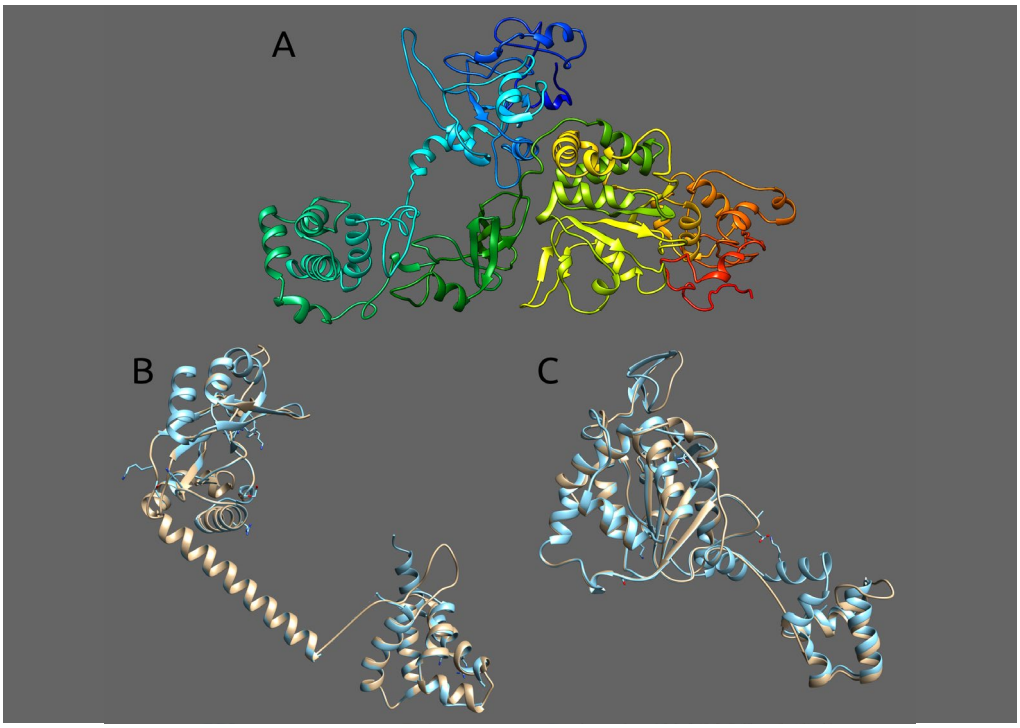


Fig.8. 3D structure homology modelling of the AN14 DNA primase (gp27). (A) Final structure obtained by Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2). The model is colored based on a rainbow gradient scheme, where the N-terminus of the polypeptide chain is colored blue and the C-terminus is colored red. (B) 3D-alignment of obtained AN14 DNA primase protein N-terminal domain protein model (colored pale brown) with best-fitting model (PDB entry 3H20 of crystal structure of primase RepB' from plasmid RSF1010, colored blue) made with UCSF Chimera. (C) 3D-alignment of obtained AN14 DNA primase protein C-terminal domain protein model (colored pale brown) with best-fitting model (PDB entry 6K9C of crystal structure of NrS-1 C terminal region from *Nitratiruptor* phage NrS-1, colored blue) made with UCSF Chimera.

Table 2. Genomic features of *Yuavirus* phages

Phage	NCBI accession #	Year	Location	Source	Host strain	Genome, bp	ORF #	Reference
AN14	KX198613	2010	Russia	River water	PAO1	60973	86	This work
YuA	AM749441	2008	Russia	Pond water	PAO1	58663	77	(Ceyssens et al., 2008)
M6	DQ163916	2006	Canada	N/A	N/A	59446	N/A	(Kwan et al., 2006)
LKO4	KC758116	2008	Greece	N/A	N/A	61818	77	Direct submission
MP1412	JX131330	2011	Korea	N/A	PAO1	61167	77	(Bae et al., 2012)
PAE1	KT734862	2015	Australia	Sewage	PAO9505	62181	88	(Dyson et al., 2016)
vB_PaeS_S218	MF490239	2015	Italy	N/A	N/A	61680	86	Direct submission
PaTs-2	MH791403	2017	China	Freshwater	PAO1	61820	90*	Direct submission
PaSz-4	MH791406	2017	China	River water	PAO1	61942	90*	Direct submission
Pa-C	MN871456	2017	China	Sea water	PAO1	60633	87*	Direct submission
Pa-L	MN871463	2017	China	Sea water	PAO1	61641	87*	Direct submission
Pa-S	MN871469	2017	China	Sea water	PAO1	61639	91*	Direct submission
PaSz-6	MN871482	2017	China	Sea water	PAO1	54656	77*	Direct submission
PaSt-2_45_61k	MN871478	2017	China	Sea water	PAO1	61865	90*	Direct submission
PaSz-9_45_61k	MN871486	2017	China	Sea water	PAO1	61157	84*	Direct submission
Epa38	MT118302	2020	USA	N/A	N/A	61775	89	Direct submission
Bordetella virus LK3	KX961385	2013	Serbia	N/A	ATCC 10580	59831	79	Direct submission
Alpha-proteobacteria phage PhiJL001	AY576273	2005	USA	N/A	N/A	63649	91	(Lohr et al., 2005)

N/A – data not available

* Genomic annotations were performed in the course of this work. Annotation tables are available upon request from the Authors

4. Discussion

A group of *Pseudomonas*-infecting siphoviruses with an elongated capsid was discovered in mid-2000's. The first sequenced representative of this group was phage M6 (Kwan et al., 2006). However the phage YuA with comprehensively studied genomics and proteomics (Ceyssens et al., 2008) was chosen as a type phage for the recently established genus *Yuavirus*. In the past decade numerous *Yuavirus* phages were isolated from different habitats (sewage, fresh and sea water reservoirs) in Eurasia, America and Australia (Table 2). An isolation of AN14 from the ecosystem of Lake Baikal further spreads the known geography of *Yuavirus* phages. Genomics of most described phages of this group are very similar (Fig. 5, Fig. 6), and phage AN14 shares all morphological, biological and genomic features of *Yuavirus*. The representatives of this genus are perfectly accommodated to their bacterial host and possess a complicated apparatus of nucleotide modification and reparation. This may be a possible reason for slow evolution and low genomic diversity within studied YuA-like viruses.

Reannotation of AN14 genome using modern bioinformatic approaches yielded the attribution of putative functions of most phage proteins. Application of phylogenetic and pan-genomic analysis enabled us to refine composition of the genus *Yuavirus*. We suggest that the host range of *Yuavirus* is limited to *Pseudomonas* sp., similar to the related genus *Vojvodinavirus* including *Bordetella* phages only. The phages previously attributed to this group are distant evolutionally enough to be considered as members of other genera.

Finally, we hypothesize the alternative function of the protein conserved among all *Yuavirus* phages including AN14. Predicted integrase suggesting the possibility of lysogenic behavior may be the primase/helicase performing a different metabolic role, and, therefore, YuA-like phages possess strictly lytic infection cycle. After an experimental proof this suggestion may be important for the choice of *Yuavirus* phages for phage therapy purposes.

Acknowledgments

This study was carried out under the State Task No. 0345-2019-0003 (AAAA-A16-116122110061-6) "Microbial and Viral Communities in the Biofilms of Freshwater Ecosystems: Taxonomic Diversity, Functional Characteristics, and Biotechnological Potential". The microscopic studies were performed in the Instrument Center «Electron Microscopy» affiliated in the Shared Facilities for Physical and Chemical Research «Ultramicroanalysis» based in Limnological Institute of the Siberian Branch of the RAS.

References

Adams M. 1959. Bacteriophages. New York: Interscience Publishers.

Alikhan N.F., Petty N.K., Ben Zakour N.L. et al. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12. DOI: 10.1186/1471-2164-12-402

Bae H.-W., Chung I.-Y., Sim N. et al. 2012. Complete genome sequence of *Pseudomonas aeruginosa* siphophage MP1412. Journal of Virology 86. DOI: 10.1128/JVI.01403-12

Berry J., Summer E.J., Struck D.K. et al. 2008. The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. Molecular Microbiology 70: 341-351. DOI: 10.1111/j.1365-2958.2008.06408.x

Capella-Gutiérrez S., Silla-Martínez J.M., Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972-1973. DOI: 10.1093/bioinformatics/btp348

Ceyssens P.J., Mesyanzhinov V., Sykilinda N. et al. 2008. The genome and structural proteome of YuA, a new *Pseudomonas aeruginosa* phage resembling M6. Journal of Bacteriology 190: 1429-1435. DOI: 10.1128/JB.01441-07

Chaudhari N.M., Gupta V.K., Dutta C. 2016. BPGA-an ultra-fast pan-genome analysis pipeline. Scientific Reports 6. DOI: 10.1038/srep24373

Chen X., Su S., Chen Y. et al. 2020. Structural studies reveal a ring-shaped architecture of deep-sea vent phage NrS-1 polymerase. Nucleic Acids Research 48: 3343-3355. DOI: 10.1093/nar/gkaa071

Colombet J., Robin A., Lavie L. et al. 2007. Virioplankton "pegylation": use of PEG (polyethylene glycol) to concentrate and purify viruses in pelagic ecosystems. Journal of Microbiology Methods 71: 212-219. DOI: 10.1016/j.mimet.2007.08.012

Delcher A.L., Harmon D., Kasif S. et al. 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Research 27: 4636-4641. DOI: 10.1093/nar/27.23.4636

Dyson Z.A., Seviour R.J., Tucci J. et al. 2016. Genome sequences of *Pseudomonas oryzihabitans* phage POR1 and *Pseudomonas aeruginosa* phage PAE1. Genome Announcement 4. DOI: 10.1128/genomeA.01515-15

Fogg P.C.M., Allison H.E., Saunders J.R. et al. 2010. Bacteriophage Lambda: a paradigm revisited. Journal of Virology 84: 6876-6879. DOI: 10.1128/jvi.02177-09

Geibel S., Banchenko S., Engel M. et al. 2009. Structure and function of primase RepB′ encoded by broad-host-range plasmid RSF1010 that replicates exclusively in leading-strand mode. Proceedings of the National Academy of Sciences of the United States of America 106: 7810-7815. DOI: 10.1073/pnas.0902910106

Goddard T.D., Huang C.C., Meng E.C. et al. 2018. UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Science 27: 14-25. DOI: 10.1002/pro.3235

Guilliam T.A., Keen B.A., Brissett N.C. et al. 2015. Primase-polymerases are a functionally diverse superfamily of replication and repair enzymes. Nucleic Acids Research 43: 6651-6664. DOI: 10.1093/nar/gkv625

Hyatt D., Chen G.-L., Locascio P.F. et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11. DOI: 10.1186/1471-2105-11-119

Katoh K., Misawa K., Kuma K. et al. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30: 3059-3066. DOI: 10.1093/nar/gkf436

Kelley L.A., Mezulis S., Yates C.M. et al. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols 10: 845-858. DOI: 10.1038/nprot.2015.053

Kongari R., Rajaure M., Cahill J. et al. 2018. Phage spanins: diversity, topological dynamics and gene convergence. BMC Bioinformatics 19. DOI: 10.1186/s12859-018-2342-8

Konstantinidis K.T., Tiedje J.M. 2005. Genomic insights that advance the species definition for prokaryotes. Proceedings of the National Academy of Sciences of the United States of America 102: 2567-2572. DOI: 10.1073/pnas.0409727102

Kumar S., Tamura K., Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Bioinformatics 10: 189-191. DOI: 10.1093/bioinformatics/10.2.189

Kwan T., Liu J., Dubow M. et al. 2006. Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. Journal of Bacteriology 188: 1184-1187. DOI: 10.1128/JB.188.3.1184-1187.2006

Lee I., Kim Y.O., Park S.C. et al. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. International Journal of Systematic and Evolutionary Microbiology 66: 1100-1103. DOI: 10.1099/ijsem.0.000760

Lee Y.-J., Dai N., Walsh S.E. et al. 2018. Identification and biosynthesis of thymidine hypermodifications in the genomic DNA of widespread bacterial viruses. Proceedings of the National Academy of Sciences of the United States of America 15: 3116-3125. DOI: 10.1073/pnas.1714812115

Lohr J.E., Chen F., Hill R.T. 2005. Genomic analysis of bacteriophage ΦJL001: insights into its interaction with a sponge-associated alpha-proteobacterium. Applied and Environmental Microbiology 71: 1598-1609. DOI: 10.1128/AEM.71.3.1598-1609.2005

Petrovic A., Kostanjsek R., Rakhely G. et al. 2017. The first Siphoviridae family bacteriophages infecting *Bordetella bronchiseptica* isolated from environment. Microbial Ecology 73: 368-377. DOI: 10.1007/s00248-016-0847-0

Sayers E.W., Cavanaugh M., Clark K. et al. 2019. GenBank. Nucleic Acids Research 47: 94-99. DOI: 10.1093/nar/gky989

Schattner P., Brooks A.N., Lowe T.M. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research 33: W686-W689. DOI: 10.1093/nar/gki366

Seemann T. 2014. Prokka: rapid prokaryotic genome

annotation. Bioinformatics 30: 2068-2069. DOI: 10.1093/bioinformatics/btu153

Simm R., Morr M., Kader A. et al. 2004. GGDEF and EAL domains inversely regulate cyclic di-GMP levels and transition from sessility to motility. Molecular Microbiology 53: 1123-1134. DOI: 10.1111/j.1365-2958.2004.04206.x

Smith K.C., Castro-Nallar E., Fisher J.N.B. et al. 2013. Phage cluster relationships identified through single gene analysis. BMC Genomics 14. DOI: 10.1186/1471-2164-14-410

Sokolova O.S., Filchikov M.V., Pechnikova E.V. et al. 2010. Structure and virulence of *Pseudomonas aeruginosa* Bacteriophages. Microscopy and Microanalysis 16: 1090-1091. DOI: 10.1017/S1431927610059994

Spoerel N., Herrlich P., Bickle T.A. 1979. A novel bacteriophage defence mechanism: the anti-restriction protein. Nature 278: 30-34. DOI: 10.1038/278030a0

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313. DOI: 10.1093/bioinformatics/btu033

Varghese N.J., Mukherjee S., Ivanova N. et al. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Research 43: 6761-6771. DOI: 0.1093/nar/gkv657