Biodiversity of new lytic bacteriophages infecting Shigella spp. in fresh water environment

Khashayar Shahin
Jiangsu Academy of Agricultural Sciences
Mohadeseh Barazandeh
Jiangsu Academy of Agricultural Sciences
Abolghasem Hedayatkhah
Curtin University
Mojtaba Mansoorianfar
Nanjing Forestry University
Ran Wang (ranwang@jaas.ac.cn)
Jiangsu Academy of Agricultural Sciences https://orcid.org/0000-0001-5994-1954

Research

Keywords: Water, Phage, Biodiversity, Bacteriophage, Shigella, Enterobacteriaceae

Posted Date: June 17th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-35862/v1

Abstract

Background Bacteriophages, viruses that infects and replicates within prokaryotic cells are the most abundant life forms in the environment, yet the vast majority of them have not been properly reported or even discovered. Almost all reported bacteriophages infecting the Enterobacteriaceae family, with E. coli being the major subject of the study, have been isolated from wastewater, sewage, and effluent resources. In the present study we focused on the distribution and biodiversity of Shigella phages in an aquatic ecosystem. Results While no Shigella bacteria was recovered from the Yangtze River, three lytic phages were isolated from this ecosystem and were subjected to biological, morphological, and genomic characteristics. Comparative genomics and phylogenetic analyses demonstrated that vB _SflM_004 isolate belongs to Myoviridae family, Felixounavirus genus of Ounavirinae subfamily, vB_SdyM_006 was classified under the same family, however, it is suggested to be in a new genus under Tevenvirinae subfamily with some other related bacteriophages. vB_SsoS_008 phage belongs to the Siphoviridae family, Tunavirus genus, Tunavirinae subfamily. The phages did not harbor any genes involved in the lysogenic cycles, and showed a high temperature and pH stability. Conclusions It can be concluded that isolation of bacteriophages could be independent of their bacterial host presence in the isolation environment.

Introduction

Bacteria-infecting viruses or bacteriophages (phages) are the most abundant biological entities on planet earth [1]. With an estimated minimum number of 10^{31} bacteriophages have the highest diversity with respect to genetics, morphology, host range, and infection cycles [2]. Our knowledge on bacteriophages is extremely limited for three main reasons. First, in theory, each and every bacterial species is a host for at least a phage. With our current knowledge on the field and the state of the art and technology, only a small percentage of bacteria can be grown in vitro thus the majority of the microorganisms are uncultivated (or yet-to-be-cultivated) [3]. Second, of the cultured one, only very few number have been used as hosts for phage isolation [3]. Finally, even with this limited number of host cells, almost all of the phages reported so far have been isolated from environments such as untreated sewage and hospital wastewater or wound since there is higher chance of isolating bacteriophages using such resources due to these resources being the most contaminated ecosystem containing a high number of pathogens. Considering the above facts, isolation of phages from other aquatic environments such as fresh water resources and reservoirs like rivers and lakes is of particular importance in view of their impact on both the microbial diversity and the ecological fate of photogenic bacteria.

Shigella is a gram-negative bacterial genus including four species: *S. boydii*, *S. dysenteriae*, *S. flexneri* and *S. sonnei* [4], all causing shigellosis with hundreds of millions of food/water born infections annually[5]. Although shigellosis can be usually treated with antibiotics, propagation of antibiotic-resistant strains has created many serious health problems in recent years [6], for example over the past few decades several shigellosis outbreaks have been reported into the geographical regions [7–10].

As an alternative to antibiotic, lytic bacteriophages can be used to control or to treat bacterial infections, a process known as phage therapy [11]. Despite the worldwide distribution of *Shigella* species, the high number of infected cases, and the great importance of food safety, only very few *Shigella*-infecting phages have been identified, studied, and reported so far. Almost all of these reported phages have been isolated from raw sewage samples. *Myoviridae* phages (pSs-1 and WZ1) [12, 13], *Siphoviridae* phages (vB_SflS-ISF001, vB_SsoS-ISF002, vB-SdyS-ISF003, SH6, Shfl1 and pSf-2) [14–19], and *Podoviridae* phages (pSb-1 and Sf6) [20, 21]are among the phages that infect different species of *Shigella.* Moreover, although phage therapy for controlling *Shigella* has a long history as the first phage research was done by Felix d'Herelle in the 1910s [22], nevertheless there is no comprehensive research on the abundance, distribution, and diversity of *Shigella* in natural aquatic environments yet.

Hence, in this study we focused on isolation and subsequent morphological, biological and genomic characterization of *Shigella*-infecting phages from Yangtze River as one of the biggest fresh water resources on the planet earth.

Materials And Methods

Bacterial stains and growth condition

All the *Shigella* bacteria used in the present study (Additional file 1, Table S1) have been previously isolated (unpublished data). They were all stored in tryptic soy broth (TSB) (Merck, Germany) containing 30% glycerol and kept at 70 ºC in the central bacterial strains collection of International Phage Research Center (IPRC) containing 30% glycerol at 70 ºC) [23]. In addition to these isolates, the type-stains of *S. sonnei* (ATCC 9290), *S. flexneri (ATCC 12022), Shigella dysenteriae* (PTCC 1188), *Shigella boydii* (ATCC 9207) and *Escherichia coli* (ATCC 25922) were used for determination of the bacteriophages host range. All the isolates were cultured routinely on brain heart infusion (BHI) agar (Merck, Germany) or in brain heart infusion broth (Merck, Germany) with constant shaking at 200 rpm and 37 ºC.

Bacteriophages isolation and morphology analysis

Four *Shigella* isolates including *S. flexneri* (w7) and *S. sonnei* (w44) which were recovered from water samples and also showed the highest antibiotic resistance, as well as *S. dysenteriae* (s.d.f1) and *Shigella boydii* (ATCC 9207) were used individually as the host bacteria for phages isolation following the previously described method with a slight modification [24]. One hundred milliliter of Yangtze River water samples (GPS coordinates of sampling locations include: 1- latitude: 32°06'37.8"N and longitude: 118°44'56.1"E; 2- latitude: 32°09'25.9"N and longitude: 118°50'48.8"E) (Additional file 1, Fig. S1) was centrifuged (10 min at 6 000 × g) and filtrated through 0.45 µm sterile syringe filters (JinTeng, China). The filtrate was used for phage isolation with (method I) or without (method II) pre-enrichment. In method I, 20 µl of the filtrate was dropped onto the surface of lawn cultures of the host bacteria and incubated at 37 ºC for at least 24 h. Following the emergence of clear plaques, one plaque was picked up for phage

purification procedure. In the pre-enrichment approach (method II) fifty milliliters of the filtrate water was added to 50 mL of the early- exponential culture of the host bacteria and incubated overnight with constant shaking (200 rpm). After centrifugation (10 min, 8 000 × g) and filtration (0.45-μm), 20 μl of the filtrate was spotted on BHI plates overlaid with the individual strain. Following clear phage plaque formation, the phage was purified following three repeats of single plaque isolation, elution, and re-plating. Phage propagation and phage titers determination were carried out regularly according to Clokie MR and Kropinski AM [25] protocols.

Phages transmission electron microscopy

The phage lysates (10^9 PFU mL^{-1}) were purified using centrifugation on cesium chloride gradient as described by Clokie MR and Kropinski AM [25], then stained with 2% phosphotungstic acid (PTA), and finally the grids were analyzed in a Hitachi HT7700 transmission electron microscope at an operating voltage of 100 kV at Nanjing Agricultural University (NAU), Nanjing, China.

Host range

The host ranges of the phages were determined using spot assays [26]. Ten microliters of the phage suspension (10^9 PFU mL^{-1}) was spotted onto double-layered BHI agar plates of each host strains (Additional file 1, Table S1). After an overnight incubation, the plates were checked for appearance of clear plaque (++), cloudy plaque (+) or no plaque (−).The efficiency of plating assay (EOP) was carried out against the *Shigella* isolates as described previously [27]. The EOP of the phage on their primary host strain (reference host) was considered as 1 and EOP of each strain were calculated as the ratio of phage titer on the tested bacterium to the phage titer on the reference host.

The phages biological characterizations

S. flexneri (w7), *S. dysenteriae* (s.d.f1) and *S. sonnei* (w44) were used as host bacteria in all the experiments. The thermostability and pH stability of the phages were evaluated by incubation of the phages lysate (10^9 PFU mL^{-1}) at wide range of temperatures (-20 to 80 ºC) and pH (2 to 12) as previously described [15]. The phage survival was measured using the overlay method and the titer is reported as a percentage of the control sample titers. The phages absorption rates to the surface of their bacterial hosts were determined as previously described [28]. Moreover, one-step growth experiments were done to determine phages burst size and latent periods [15]. The bacteriolytic potential of the phages were evaluated by monitoring the changes in OD_{600} absorbance of the phage/host mixture at different MOI as described previously [26]. The assays were performed in triplicate.

DNA extraction, DNA fingerprinting and whole genome sequencing

The stocks of purified phages were condensed using ultracentrifugation at 105 000 × g for 3 h at 4 ºC (Beckman Optima L-80 XP ultracentrifuge, TYPE 45 Ti rotor). The pellet was diluted in SM buffer and treated by 10 µg/mL DNase I and RNase I (Sigma, China) to digest any free DNA and RNA. The genomic DNA of phages was then extracted using phenol/chloroform/isoamyl alcohol protocol as described previously by Sambrook J and Russell DW [29]. Finally, the quality and quantity of the extracted DNA were examined using agarose gel electrophoresis and NanoDrop (Thermo Scientific, USA).

The digestion patterns produced by *EcoRI*, *EcoRV* and *HindIII* restriction enzymes (Thermo Fisher Scientific, US) were used for DNA fingerprinting analysis. The phage' DNA and the endonuclease were mixed individually according to manufacturer's protocol. After the incubation period, the DNA fragments were separated by 1% agarose gel electrophoresis at 90 V for 60 min. The DNA libraries and whole genome sequences were obtained using Illumina HiSeq NGS DNA sequencing system (TGS, Shenzhen, China). The raw sequencing data were assembled using SOAPdenovo (v2.04) at the default setting and the assembled sequences were deposited at DDBJ/EMBL/GenBank (Table 1).

Table 1
Morphologic and genomic characteristics of the isolated phages in current study.

	vB_SflM_004	vB_SdyM_006	vB_SsoS_008
Host	*S. flexneri*	*S. dysenteriae*	*S. sonnei*
Isolation method	II	I	I
Plaque size	1.5-2 mm	2 mm	2.5–2.7 mm
Head diameter	99.6 ± 8 nm	92.7 ± 2 nm	59.2 ± 2 nm
Tail length (relaxed form)	107.2 ± 6 nm	106.2 ± 4	171.9 ± 5
Width (relaxed form)	15.5 ± 1 nm	15.5 nm	6.9 ± 1 nm
Tail length (contracted form)	50.2 ± 2 nm	54.1 ± 1 nm	-
Width (contracted form)	15.8 ± 1 nm	15 ± 1 nm	
Family	Myoviridae	Myoviridae	Siphoviridae
Sequencing platform	Illumina HiSeq	Illumina HiSeq	Illumina HiSeq
Fold coverage	4,115×	3,247×	1,405×
Genome length	85,887 bp	166,138 bp	50,414 bp
G + C%	38.6	31.5	45.6
No. of coding sequences	135	252	83
tRNA	0	9	0
GenBank ID	MK295205	MK295204	MK335533

Bioinformatic analysis

Open reading frames (ORFs) were detected using Prokaryotic GeneMark.hmm version 3.25 (http://opal.biology.gatech.edu/genemark/gmhmmp.cgi) [30] and NCBI ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) and translated to protein sequences using ExPASy translate tool (http://web.expasy.org/translate/). Molecular weight and isoelectric pH of the predicted ORF were estimated using ExPASy compute pI/Mw tool (http://web.expasy.org/compute_pi/) [31]. tRNAscan-SE was used to find any tRNA sequence [32]. Functional and conserved domains of the predicted ORFs proteins were analyzed using a couple of software and online tools including Basic Local Alignment Search Tool (BLASTp), (https://blast.ncbi.nlm.nih.gov/Blast.cgi), HHpred (https://toolkit.tuebingen.mpg.de/#/tools/hhpred) [33], Pfam (http://pfam.xfam.org/search#tabview=tab1) [34] and InterProScan (http://www.ebi.ac.uk/interpro/search/sequence-search) [35]. The promoter sequences were identified

using BPROM program of the Softberry website with the maximum allowable distance from the starting codon of a gene at 100 bp [36].

Comparative genomic analysis and phylogeny

The whole genome sequences of the taxonomically close and related phages were obtained from NCBI database (https://ncbi.nlm.nih.gov) and were used for comparison of the sequences at both genome and proteome levels using EasyFig v 2.2.3 [37]. In addition, CoreGenes 3.5 was used for core gene analysis [38]. The phylogenetic tree was constructed using Mega 7.0 based on UPGMA (unweighted pair group method with arithmetic mean) with 2 000 bootstrap replication [39].

Results

Phages isolation and morphology

Phages were isolated from Yangtze River according to their ability to lyse and generate clear plaques with or without hallow zones using several *Shigella* species as the host cells. The transition electron microscopy micrographs shows that the isolated phages for *S. flexneri* and *S. dysenteriae* have icosahedral head, contractile tail, collar, and base plate, the typical properties *Myoviridae* family of bacteriophages (Fig. 1-A to D). Moreover, the TEM micrograph of the isolated phage for *S. sonnei* shows that it has an icosahedral head and a non-contractile tail, a similar structure to that of the *Siphoviridae* phages (Fig. 1E and F). The head diameter, tail length and width of the phages are summarized in Table 1. The phages were designated as vB_SflM_004, vB_SdyM_006 and vB_SsoS_008 according to their host species and phage family.

Bacteriophages host ranges

The host range of the isolated phages were tested on a wide range of bacteria including *Shigella* isolates as well as standard strains of gram negative and gram positive bacteria. The vB_SdyM_006 phage was capable of producing clear plaque only on *S. dysenteriae* isolates (3/3 isolates), while vB_SflM_004 and vB_SsoS_008 produced either clear or cloudy plaques on most of the tested *S. flexneri* and *S. sonnei* isolates (Additional file 1, Table S1). A relatively wide range of EOP (0.12 ± 0.07 ~ 1) of the phages was observed against different isolates of *Shigella* spp. (Additional file 1, Table S2).

DNA fingerprinting

The DNA fingerprinting of the isolated phages were obtain using restriction endonucleases *EcoRI*, *EcoRV* and *HindIII*. The obtained restriction pattern revealed that the genome of the phages vB _SflM_004 and vB_SsoS_008 were digested only with *EcoRV* and the genome of phage vB_SdyM_006 was digested with

EcoRV and *HindIII*. The observed differences in the DNA fingerprints in terms of size and pattern (Fig. 1) imply that the genome size and sequence of the isolated phages were different from each other.

Basic biological characteristics

The thermo- and pH stability of the phages were tested at a wide range of temperatures and pH values (Fig. 2). Titer of all the 3 phages were stable (> 90%) at -20 to 40 ºC, but it started to decrease when incubated at 50 ºC for 1 h. While by further increase in the temperature to 80 ºC, vB_SdyM_006 and vB_SsoS_008 could not be recovered, the vB _SflM_004 phage was still recovered at this temperature, but lost its activity when incubated at 90 ºC (Fig. 2- A). In the case of pH stability, the highest activity was observed at pHs ranging from 6 to 8. Incubation at basic pH of 12 (for all phages) and acidic pHs of 4 (for vB_SsoS_008) and 3 (for vB_SdyM_006 and vB _SflM_004) led to deactivation of the phages (Fig. 2-B). The one-step growth curves demonstrated that the phages vB _SflM_004, vB_SdyM_006 and vB_SsoS_008 were started to release from their host cells after 30, 50 and 15 min, respectively. Moreover, the burst sizes were estimated to be about 139 ± 29, 93 ± 15 and 94 ± 9 virions per single bacterium for vB _SflM_004, vB_SdyM_006 and vB_SsoS_008, respectively (Fig. 2-C). Additionally, as shown in Fig. 2-D the phages particles were absorbed immediately after incubation where vB _SflM_004, vB_SdyM_006 and vB_SsoS_008 phages were fully absorbed on their host cell after 12, 6 and 10 min, respectively.

Genome Analysis

The fold coverage, genome size, G + C contents and other general genome features of the phages vB_SflM_004, vB_SdyM_006 and vB_SsoS_008 is presented in Table 1.

Genome Analysis of vB_SflM_004

ATG was detected as the start codon in all of the ORFs. The Opal (TGA), Ochre (TAA) and Amber (TAG) stop codons were presented in 59, 49 and 20 ORFs, respectively. The BPROM search detected 16 promoters (Additional file 2) with consensus sequences at -10 (tttTAtaaT) and − 35 (TTcAca) (the capital letters indicated conserved nucleotides). Nine Rho-factor independent termination sites were also detected in the vB_SflM_004 genome with FindTerm online software (Fig. 3).

In general, the genomic organization and genetic analysis of the phage vB_SflM_004 demonstrated that the genome contained 135 possible open reading frames (ORFs) including 20 ORFs encoding structural proteins, 25 ORFs for metabolism-related proteins, 4 ORFs associated with bacterial lysis-like proteins, 83 ORFs encoding hypothetical proteins which showed relatively high similarity to the previously described phage hypothetical proteins with no clear understanding of their functions yet, and the 3 remaining ORFs which showed no similarity to any known proteins in the databases (Fig. 3). The list of 135 ORFs as well as their details and annotation is provided in the Additional file 3. The detected genes involved in the

bacterial cell lysis were lysozyme (*gp111*), holin (*gp79*) and two spanin (o-spanin, *gp28* and i-spanin, *gp29*) which are similar to the previously reported genes in the *Ounavirinae* Subfamily including *Felixounavirus* (*gp28* and *gp29*), *Mooglevirus* (*gp111*) and *Suspvirus* (*gp79*). In addition, two pairs of rIIA/rIIB proteins were also detected at the semi-beginning (*ORF32* and *ORF33*) and the end (*ORF134* and *ORF135*) of vB_SflM_004 genome which could play role in regulation of bacterial lysis (Additional file 3). The gene products involved in the metabolism/regulation pathways of vB_SflM_004 were identified as different types of DNA polymerases, kinases, reductases, protease, nucleases, hydrolysis, and regulatory proteins with relatively high similarity to those of *Ounavirinae* Subfamily (check Additional file 3 for more detail). The structural and assembly genes were encoding the tail fiber proteins, tail sheath, tail protein, tail tube protein, minor tail protein, tail assembly protein, Major capsid, pro-head assembly scaffold protein and a head maturation protease. Some of these proteins were similar to those available in the GenBank database. For instance, tail tube, major capsid, and tail protein were almost identical (≥ 97%) to the respective predicted gene products of phages vB_EcoM_Alf5, SF19, Meda and SF13. On the other hand, the tail proteins and the major capsid protein represented a low identity (≤ 55%) to the previously reported phage proteins. The gene distribution pattern (Fig. 3 and Additional file 3) shows that about half of the gene products of ORF81 to ORF110 were identified as structural proteins. Same as other viruses, bacteriophages tend to have the genes with similar function close to each other in a compact arrangement [40]. Thus, it is possible that the remaining ORFs (which have been considered as hypothetical proteins) in this region of the genome may have structural function.

The highest similarity of the hypothetical proteins was to those of phage SF13 (13 out of 83 ORF) with a clear concordant relation in their gene products function and their respective identified conserved domains. However, in the case of conserved domains of DUF3277 and DUF3383 no clear relations were found (Additional file 3).

BLASTN analysis of the phage vB_SflM_004 genome revealed that the genome of the phage was highly similar (~ 94% similarity with > 75% query coverage) to *Escherichia coli* phage 11, phage 12, *Enterobacteria* phage WV8 and *Salmonella* phage BPS15Q2. As shown in (Fig. 4) the dotplot analysis of these bacteriophages using Gepard demonstrated a considerable sequence similarity between vB_SflM_004 and the other related phages with a few remarkable differences such as deletion of an approximately 10 kb region at position around 18 000. Comparison of the genome with other close phages using CoreGenes showed that 68% of the proteins were shared with *Ounavirinae* subfamily in which the entire lysis group proteins, some genes with structural or regulatory functions, as well as some of the hypothetical proteins were conserved (Fig. 4 and Additional file 4). Fig S2 depicts the relatedness of vB_SflM_004 and other phages with high homology using Easyfig software.

Genome Analysis of vB_SdyM_006

The genome of vB_SdyM_006 contains 252 ORFs (Additional file 3) and 9 tRNA coding regions (Table S3). The only identified start codon was ATG. Ocher, Amber and Opal stop codons were identify in 91, 50

and 111 ORFs, respectively. A BPROM search identified 31 promoters, with the consensus sequences of ATG**TAT**AA**T** and **TT**TA**A**T at the −10 and −35 positions, respectively (the conserved bases were presented in bold) (Additional file 2). In addition, the only identified potential Rho-factor independent termination site were located after the gene encoding the inhibitor of the prohead protease (*gp187*) (Fig. 5).

With regard to the comprehensive genetic analysis of vB_SdyM_006 and the homology-based search of its 252 ORFs, the predicted ORFs could be clustered into five groups. Forty-five ORFs were predicted as structure proteins, nearly from *ORF133* to *ORF182* (Additional file 3). Tail completion and sheath stabilizer protein, head completion protein, baseplate wedge subunit, baseplate wedge tail fiber connector, baseplate wedge subunit and tail pin, short tail fibers, fibritin, neck protein, tail sheath stabilization protein, tail sheath protein, tail tube protein, portal vertex of head, prohead core protein, prohead core scaffold protein, major capsid protein, capsid vertex protein, Membrane protein, baseplate tail tube initiator, baseplate tail tube cap, baseplate hub subunit, and baseplate hub distal subunit all were detected in this region and have a high similarity rates to the respective predicted gene products of phages vB_PmiM_Pm5461, PM2, phiP4-3 and vB_MmoM_MP1 [41–44]. Moreover, a small group of five genes was detected close to end of the genome (from *ORF233* to *ORF237*) encoding different parts of the tail structure including long tail fiber proximal subunit, long tail fiber proximal connector, long tail fiber distal connector, long tail fiber distal subunit and distal long tail fiber assembly catalyst which had 100% similarity to vB_PmiM_Pm5461 and phiP4-3 (Additional file 3) [42, 44].

Within the lysis functions, the *ORF98* encodes an endolysin with peptidase activity (conserved domain pfam13539). The *gp238* was identified as a Holin lysis mediator due its high similarity to the respective predicted gene product of phage vB_PmiM_Pm5461. The *gp85* (lysis inhibition regulator) and *gp200* (rIII lysis inhibition accessory protein) are predicted to have the regulatory roles in the lysis pathway. It is worth mentioning that *ORF137* (baseplate hub + tail lysozyme) and *ORF157* (head core scaffold protein + protease) encodes bifunctional proteins whose contain either C-terminal or N-terminal sequences and showed a relatively high similarity with those of the cell wall lysozymes.

Terminases, the proteins responsible for packaging the phage genome were detected almost in the middle of the genome (*ORF150* and *ORF151*), showing a ≥ 94% similarity to the small and large subunits of phage vB_MmoM_MP1 terminase (Fig. 5 and additional file 3). Furthermore, the conserved domains of DNA_Packaging (pfam11053) and Terminase_6 (pfam03237) were identified in the small and large subunits of the terminase, respectively.

The predicted genes involved in the metabolic and regulatory functions were including several DNA-associated genes (DNA polymerase, helicase, primase, ligase, topoisomerase and endonuclease proteins), RNA-assosiated genes (RNA polymerase, tRNA synthetase, ligase, endonuclease and RNaseH proteins), different types of exonuclease, recombinase, anti-sigma factors, sigma factors, anaerobic NTP reductase, thioredoxin, kinase, host translation inhibitors, and several other genes. Most of the predicted proteins showed a high identity (≥ 90%) with the counterpart proteins of *Tevenvirinae* Subfamily of

phages while some others had no similarity (*gp71*, *gp74*, *gp100*, *gp129*, *gp191* and *gp212*) (additional file 3).

Based on the BLASTN analysis, the genome sequence of vB_SdyM_006 had 98% (97% query coverage) and 97% (73% query coverage) similarity to the genome sequences of *Proteus* phages, phiP4-3 and vB_PmiM_Pm5461, respectively. Moreover, the sequences alignment of these three phages using Gepard software showed a higher similarity between vB_SdyM_006 and phiP4-3 than vB_SdyM_006 and vB_PmiM_Pm5461 (Fig. 4). Furthermore, the relatedness of vB_SdyM_006 and other phages with high degree of homology was determined using Easyfig software (Fig S3)

The CoreGene analysis showed that vB_SdyM_006 shared ~ 84% similarity with that of the encoded proteins of the mentioned phages above (score > 70), including 111 hypothetical proteins and 101 known proteins with different functions. These protein coding genes were spread out all along the genome and were not restricted to any particular region (Fig. 4 and additional file 4).

Genome analysis of vB_SsoS_008

The genome of vB_SsoS_008 contained 83 putative ORFs, of which the function of 33 ORFs was predicted (additional file 3), and the other 50 ORFs were assigned as hypothetical proteins in which 47 ORFs had similarities with the hypothetical proteins of bacteriophages vB_EcoS_SH2, Sfin-1, T1, SH6 and phi2457T while the other 3 ORFs were evidently unique to UAB_Phi87 and showed no similarity with the already deposited sequences. Twelve sequences with conserved consensus sequences of gTtTAatAT (−10) and TTgCaA (−35) were identified as promoter and were distributed throughout the phage genome (the conserved bases were presented in capital letter) (additional file 2). All of the ORFs started with an ATG codon, with Opal (36 ORFs), Ochre (30 ORFs) and Amber (17 ORFs) stop codons. Only one Rho-independent terminator was identified by FindTerm (Fig. 6). The genome of vB_SsoS_008 contained no tRNA or pseudo-tRNA genes.

The vB_SsoS_008 ORFs was encoding known protein that can be classified into 5 functional groups. The structural group contained 21 proteins including portal protein (gp26), capsid proteins (gp28-31, tail proteins (gp41, 43–55, 61 and 62). All of the structural proteins showed a relatively high to high similarity (85–100%) with the respective predicted gene products of phages B_EcoS_SH2, Sfin-1, T1, SH6 and phi2457T, except *gp46* in which only 58% similarity was observed to the tail fibers protein of *Shigella* phage Sfin-1. Detection of pfam05939 conserved domain (Phage_min_tail) in this gene approved the function of *gp46* as the tail fibers. The second group includes 8 proteins predicted to be associated with nucleotide metabolism and its regulation. The product of these genes facilitate genome replication, transcription and translation. These proteins are DNA methylase (gp3), kinase (gp17), nuclease (gp58), recombination protein (gp59), DNA primase (gp63) and primase (gp64), helicase (gp66) and methyltransferase (gp68) which showed ≥ 80% similarity to the counterpart proteins of phages Sfin-1, T1 and phi2457T (additional file 3). The third group includes the necessary protein involved in the bacterial cell lysis process. The two genes, 76 and 77, are predicted to encode endolysin and spanin, respectively,

and had 90% (query coverage of 65%) and 84% (query coverage of 79%) identity with their counterpart proteins of *Shigella* phage Sfin-1and *Shigella* phage SH6, respectively. Interestingly, holin gene was found neither close nor far from the lysine gene. The DNA packaging complex consisted of large (gp25) and small (gp24) subunits of terminase was categorized as the fourth group. The large subunit of this complex had a high identity (96%, query coverage of 100) while the small subunit had only 73% similarity (query coverage of 90%) with the counterpart proteins of the related phages (additional file 3).

BLASTN analysis of the phage vB_SsoS_008 genome showed approximately 91.2% (query coverage 90%), 91.7% (query coverage 84) and 90.5% (query coverage 96%) similarity with *Shigella* phage SH6, *Enterobacteria* phage T1, and *Shigella* phage Sfin-1, respectively. The CoreGenes analysis (score > 70) revealed that vB_SsoS_008, *Shigella* phage SH6, *Enterobacteria* phage T1, and *Shigella* phage phi2457T had ~ 60% proteins in common including the structural, DNA packaging, metabolic, endolysin and hypothetical proteins (Fig. 4 and additional file 4). In addition, the alignment of nucleotide sequences using Gepard software showed a high similarity between vB_SsoS_008 with *Shigella* phage SH6, *Enterobacteria* phage T1, and *Shigella* phage phi2457T (Fig. 4). Furthermore, the relatedness of the vB_SsoS_008 and other phages with high homology was determined using the Easyfig (Fig S4).

Phylogenetic analysis

The phylogenetic relationship between the isolated phages and other similar phages available in online databases was studied using construction of the phylogenic tree based on major capsid sequences that were identified in all of these phages (Fig. 7). Both vB _SflM_004 and vB_SdyM_006 were clustered as a member of the *Myoviridae* family. However, their major capsid sequences were different enough to classify them into lower taxa levels in which vB _SflM_004 was clustered into *Felixounavirus* genus of *Ounavirinae* and vB_SdyM_006 was only classifiable to a subfamily level and as a member of the *Tevenvirinae*. The constructed phylogenic tree based on the major capsid sequences suggests that vB_SdyM_006 along with phiP4-3, PM2 and vB_PmiM_Pm5461 phages could be considered as a new genus in *Tevenvirinae* subfamily due to the considerable phylogenic distance with other related members such as those of *Tequatrovirus* genus. Moreover, the phylogenic analysis indicated that phage vB_SsoS_008 should be added to *Tunavirus* genus, *Tunavirinae* subfamily of *Siphoviridae* family.

Discussion

Independent distribution of Shigella **phages from** Shigella **bacterial cells**

There are many reports concerning isolation of *Shigella* phages from sources such as municipal wastewater, drainage ponds, and surface runoff after regional occurrence of shigellosis, which could be due to the contact of *Shigella* -contaminated resource (such as the municipal wastewaters) with other water resources that usually are *Shigella* free (such as the runoff waters) [4, 9, 45–48]. In the present study, a number of *Shigella*-infecting phages were isolated from a large freshwater resource, the Yangtze River, and in an area with no prevalence of *Shigella*. This is especially interesting and important because

no *Shigella* bacteria were recovered during the entire sampling period. Therefore, the relatively high abundance and diversity of *Shigella* phages could be an indication of their omnipresence in environments, especially in aquatic environments.

In general, the lipopolysaccharides of Gram-negative bacteria includes the O-antigen as the source of the bacterial diversity at the serotypes level, and the outer core as the source of diversity at the species level [49]. In *Shigella* spp. the inner core is conserved [50]. Since both vB _SflM_004 and vB_SsoS_008 phages infected all of the tested *Shigella sonnei* and *Shigella flexneri* isolates (with the exception of 3 isolates) in this study, it can be said that their function was independent of the O-antigen. The vB_SdyM_006 phage, however, only lysed *Shigella dysenteriae* isolates and had no effect on the isolates of other genera. Based on these observations, it can be inferred that the vB _SflM_004 and vB_SdyM_006 phages act as O-antigen-independent phages and are capable of lysing various species and serotypes, while vB_SdyM_006 is more specific and may only work on the serotypes of a certain species.

The presence of *Shigella* bacteria in various water resources has been recorded in different parts of the world. Connor TR, Barker CR, Baker KS, Weill F-X, Talukder KA, Smith AM, Baker S, Gouali M, Thanh DP and Azmi IJ [51] and Shahin K, Bouzari M, Wang R and Khorasgani MR [4] acknowledged that neither *Shigella boydii* nor *Shigella dysenteriae* was the dominant endemic species in any country. However, *Shigella sonnei* is the dominant species in industrialized countries [5]. *Shigella flexneri*, which has had a longer and more frequent history of presence, is also distributed worldwide [4]. We expected in our observations to see coordination between the distribution and presence of the bacteria and of their infecting phages. Therefore, at first glance, it was expected that the frequency of *Shigella sonnei* and *Shigella flexneri* would be high in the studied aqueous environment (at least at the time of phage isolation). Nevertheless, despite using an enrichment step no *Shigella bacteria* were recovered during this research. Considering the proven specific nature of the phage-host interaction in many cases, a plausible hypothesis is that these phages have high stability in such environmental conditions and therefore remained dormant in the environment until the proper host reappears. In other words, *Shigella* and their infecting phages entered the environment and after a while, with the destruction of the bacterial host, the phages remained in the environment. However, given the very high abundance of the phages (estimated at 10^{31}) on our planet and the bacteria in natural environment [2], as well as the culturability of only a very small percentage of bacterial species under laboratory conditions [52], another improbable but possible hypothesis is that there are other host(s) for vB _SflM_004, vB_SdyM_006, and vB_SsoS_008 phages. In other words, these phages may have infected some of the yet-to-be-cultivated bacteria.

Doore SM, Schrad JR, Perrett HR, Schrad KP, Dean WF and Parent KN [53] hypothesized that *Shigella flexneri* might be the dominant species in the studied environment (water resources of Michigan, USA) in view of a high abundance of *Shigella flexneri* phages in those ecosystems. In our study, isolation of vB _SflM_004 without pre-enrichment (method II) indicates its higher frequency than that of the other two phage. The higher structural, host, and biological diversity in the isolated phages of the present study than the *Shigella* phages isolated from the Michigan and Nebraska aquatic environments could be justify

due to the size of the studied water resource (the Yangtze River) and the repeated sampling at different times in the present investigation

Suitable biological properties for stability in the environmental

The thermos- and pH- stability tests demonstrated that for all of three phases (vB _SflM_004, vB_SdyM_006, and vB_SsoS_008) more than 50% of the infectivity was preserved at temperatures of -20 to 60 °C and pHs of 6 to 10. Similar to our observation, it was reported frequently that *Shigella* phages have higher stability under alkaline conditions compared to acidic conditions [15, 17, 19, 26, 53]. The absence of any significant changes in the phage titers at -20 to 40 °C and in neutral to somewhat alkaline pH is an indication of the high stability of the phages in normal environments (neutral or slightly acidic/alkaline pH and ambient temperatures of $\leq$ 40 °C).

One-step growth curve analysis and phage adsorption rate provide a comprehensive perspective of phage interactions with their bacterial host. In general, short adsorption time indicates a more powerful initial attachment behavior of a phage to its host surface, while shorter latent period and larger burst size are indications of a higher of lytic potency of a phage [15]. Therefore, in the case of the isolated *Shigella* phages in this study, it can be assumed that almost immediately after the introduction of an appropriate host cell to the environment, these phages identify the host cells, get attached, then replicate very fast and keep their titer high in the environment.

The isolated Shigella phages have unique DNA fingerprints

The use of restriction enzymes to determine the genome size of bacteriophages and to differentiate them from each other has a long history [54]. The obtained DNA fingerprint profiles indicated that vB _SflM_004 and vB_SdyM_006 phages were different from each other and belong to two different groups despite both of them being a member of *Myoviridae* phages as it proved through the whole genome sequencing. Moreover, comparison of these profiles with the already reported *Shigella* phages such as pSf-1, pSf-2, Shfl1, vB_SflS-ISF001, vB_SsoS-ISF002, and vB-SdyS-ISF003 shows a clear difference. Therefore, it can be concluded that the DNA fingerprint profiles can be used as a genome-based typing method to identify similar phages. Our results demonstrate that rapidness, cost-effectiveness, proper degree of sensitivity, in addition to not using any sophisticated laboratory equipment are the advantages of this test and makes it an appropriate method for preliminary analysis of phage diversity.

High diversity of Shigella phages

Although vB _SflM_004 and vB_SdyM_006 phages were both of the *Myoviridae* family, there were substantial phylogenetic differences to placed vB _SflM_004 in the *Felixounavirus* genus, *Ounavirinae* subfamily of *Myoviridae* family. Moreover, although vB_SdyM_006 together with Phi4-3, PM2, Pm5461, CGG4-1, PEi20, and PST, and T4 phages were phylogenetically placed in the *Tevenvirinae* subfamily, these phages were sufficiently different from Pss1, Sf 22, PST14 phages (of the *Tequatrovirus* genus) to be clustered in a new genus together with phages closely related to it such as Phi4-3, PM2, and PM5461

[41–44]. Several phage families are capable of infecting *Shigella* bacteria. According to the latest online version of the ICTV (2018), totally there are 24 *Shigella*-infecting phages in the *Siphoviridae* (5 phages), *Podoviridae* (5 phages), *Myoviridae* (13 phages), and *Ackermannviridae* (1 phage) families. Through searching the public databases, a total 35 phages were found as the *Shigella*-infecting phages between 2016 and 2018. However since then until January 2020, another 73 complete sequences of *Shigella* phages (39 *Myoviridae*, 24 *Siphoviridae*, 8 *Podoviridae*, and 2 *Ackermannviridae*) were registered at NCBI (https://www.ncbi.nlm.nih.gov/nuccore). In the present study, isolation of more phages from *Myoviridae* family and even without using pre-enrichment in the case of vB_SflM_004, is in agreement with the general pattern mentioned above, that is a higher frequency of bacteriophages from *Myoviridae* family in the previous studies.

Conclusion

Isolation of the three types of lytic phages while no *Shigella* bacteria were recovered implies that the presence of phages in an environment could be completely independent of the presence of their target host bacteria. Such observation is highly likely due to a high level of stability of these phages in the environment, or in a less-likely hypothesis, due to the phages having other host (s) among the yet-to-be-cultivated bacteria. Furthermore, *Shigella* phage isolation using methods with or without enrichment also demonstrated coordination in distribution of these phages in freshwater and other environments. The, genomic and bioinformatics analyses did not identify any genes involved in the lysogenic cycles in the genome sequences of the three isolated phages, and no plaques suspected of harboring lysogenic phages were detected during the tests. Both of these observations indicate the absolute linearity of these phages and, given the desirability of their other biological properties, they can be introduced as suitable antibacterial candidates for bio-control of *Shigella* in different conditions.

Declarations

Ethics approval and consent to participate

Not applicable

Availability of data and materials

All data generated or analysed during this study are included in this published article. In addition the whole genome sequences are available under accession numbers of MK295205, MK295204 and MK335533 in GenBank.

Acknowledgements

The authors would like to thank all colleagues who made this research successful.

Funding

This study was supported by funds from the National Natural Science Foundation of China (NSFC No. 31950410562, 31602078 and 31702297) and Natural Science Foundation of Jiangsu Province (BK20180054).

Author contributions

KS and RW designed the research, KS, MB, and AH contributed to data analysis and drafted the manuscript. MB, LZ, HB, MM, MP and TH performed the laboratory phage works, DNA extraction and sequencing. All authors contributed to manuscript finalization.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Suttle CA: Viruses in the sea. Nature 2005, 437:356-361.
2. Hendrix RW, Hatfull GF, Ford ME, Smith MC, Burns RN: Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage**.** In Horizontal Gene Transfer. Elsevier; 2002: 133-VI
3. Hatfull GF, Hendrix RW: Bacteriophages and their genomes. Current Opinion in Virology 2011, 1:298-303.
4. Shahin K, Bouzari M, Wang R, Khorasgani MR: Distribution of antimicrobial resistance genes and integrons among *Shigella* spp. isolated from water sources. Journal of Global Antimicrobial Resistance 2019, 19:122-128.
5. WHO: Shigellosis: disease burden, epidemiology and case management. Weekly Epidemiological Record 2005, 80:94-99.
6. Baker KS, Dallman TJ, Field N, Childs T, Mitchell H, Day M, Weill F-X, Lefèvre S, Tourdjman M, Hughes G: Horizontal antimicrobial resistance transfer drives epidemics of multiple *Shigella* species. Nat Commun 2018, 9:1462.
7. Puzari M, Sharma M, Chetia P: Emergence of antibiotic resistant *Shigella* species: A matter of concern. J Infect Public Health 2018, 11:451-454.
8. Löfdahl M, Ivarsson S, Andersson S, Långmark J, Plym-Forshell L: An outbreak of *Shigella dysenteriae* in Sweden, May-June 2009, with sugar snaps as the suspected source. Eurosurveillance 2009, 14:29-35.

9. He F, Han K, Liu L, Sun W, Zhang L, Zhu B, Ma H: Shigellosis outbreak associated with contaminated well water in a rural elementary school: Sichuan Province, China, June 7–16, 2009. PLoS ONE 2012, 7:e47239.
10. Abaidani I, Raju PA, Al-Shualli I, Al-Sa'di K, Al-Shaqsi N, Al-Khatri A: Shigellosis outbreak in Al Batinah south governorate, Oman: case-control study. Sultan Qaboos Univ Med J 2015, 15:382–389.
11. Shahin K, Bouzari M, Wang R, Yazdi M: Prevalence and molecular characterization of multidrug-resistant *Shigella* species of food origins and their inactivation by specific lytic bacteriophages. International Journal of Food Microbiology 2019, 305:108252.
12. Jamal M, Chaudhry WN, Hussain T, Das CR, Andleeb S: Characterization of new *Myoviridae* bacteriophage WZ1 against multi-drug resistant (MDR) *Shigella dysenteriae*. Journal of Basic Microbiology 2015, 55:420-431.
13. Jun JW, Giri SS, Kim HJ, Yun SK, Chi C, Chai JY, Lee BC, Park SC: Bacteriophage application to control the contaminated water with *Shigella*. Scientific Reports 2016, 6:22636.
14. Shahin K, Bouzari M, Wang R: Complete genome sequence analysis of a lytic *Shigella flexneri* vB_SflS-ISF001 bacteriophage. Turkish Journal of Biology 2019, 43:99-112.
15. Shahin K, Bouzari M, Wang R: Isolation, characterization and genomic analysis of a novel lytic bacteriophage vB_SsoS-ISF002 infecting *Shigella sonnei* and *Shigella flexneri*. Journal of Medical Microbiology 2018, 67:376-386.
16. Shahin K, Bao H, Komijani M, Barazandeh M, Bouzari M, Hedayatkhah A, Zhang L, Zhao H, He T, Pang M, Wang R: Isolation, characterization, and PCR-based molecular identification of a *siphoviridae* phage infecting *Shigella dysenteriae*. Microbial Pathogenesis 2019, 131:175-180.
17. Jun JW, Kim JH, Shin SP, Han JE, Chai JY, Park SC: Characterization and complete genome sequence of the *Shigella* bacteriophage pSf-1. Research in microbiology 2013, 164:979-986.
18. Jun JW, Kim HJ, Yun SK, Chai JY, Lee BC, Park SC: Isolation and comparative genomic analysis of T1-Like *Shigella* bacteriophage pSf-2. Current microbiology 2016, 72:235-241.
19. Hamdi S, Rousseau GM, Labrie SJ, Tremblay DM, Kourda RS, Slama KB, Moineau S: Characterization of two polyvalent phages infecting *Enterobacteriaceae*. Scientific reports 2017, 7:40349.
20. Jun JW, Yun SK, Kim HJ, Chai JY, Park SC: Characterization and complete genome sequence of a novel N4-like bacteriophage, pSb-1 infecting *Shigella boydii*. Research in microbiology 2014, 165:671-678.
21. Casjens S, Winn-Stapley DA, Gilcrease EB, Morona R, Kühlewein C, Chua JE, Manning PA, Inwood W, Clark AJ: The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. J Mol Biol 2004, 339:379-394.
22. Duckworth DH: " Who discovered bacteriophage?". Bacteriol Rev 1976, 40:793.
23. Mikoleit M: Laboratory protocol: Biochemical identification of *Salmonella* and *Shigella* using an abbreviated panel of tests. WHO Global Foodborne Infections Network, Geneva, Switzerland 2010.

24. Yazdi M, Bouzari M, Ghaemi EA, Shahin K: Isolation, Characterization and Genomic Analysis of a Novel Bacteriophage VB_EcoS-Golestan Infecting Multidrug-Resistant *Escherichia coli* Isolated from Urinary Tract Infection. Scientific Reports 2020, 10:1-13.
25. Clokie MR, Kropinski AM: Bacteriophages. Springer; 2009.
26. Shahin K, Bouzari M: Bacteriophage application for biocontrolling *Shigella flexneri* in contaminated foods. Journal of Food Science and Technology 2018, 55:550-559.
27. Shahin K, Bouzari M, Komijani M, Wang R: A New Phage Cocktail Against Multidrug, ESBL-Producer Isolates of *Shigella sonnei* and *Shigella flexneri* with Highly Efficient Bacteriolytic Activity. Microbial Drug Resistance 2020.
28. Kropinski AM, Mazzocco A, Waddell TE, Lingohr E, Johnson RP: Enumeration of bacteriophages by double agar overlay plaque assay. Methods Mol Biol 2009, 501:69-76.
29. Sambrook J, Russell DW: Molecular cloning: a laboratory manual. 2 edn. NY: Cold Spring Harbor Laboratory; 2001.
30. Besemer J, Lomsadze A, Borodovsky M: GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic acids research 2001, 29:2607-2618.
31. Gasteiger E, Hoogland C, Gattiker A, Duvaud Se, Wilkins MR, Appel RD, Bairoch A: Protein identification and analysis tools on the ExPASy server. Springer; 2005.
32. Schattner P, Brooks AN, Lowe TM: The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic acids research 2005, 33:W686-W689.
33. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V: A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 2018, 430:2237-2243.
34. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, Bateman A, Eddy SR: HMMER web server: 2015 update. Nucleic Acids Research 2015, 43:W30-W38.
35. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 1997, 25:3389-3402.
36. Salamov VSA, Solovyevand A: Automatic annotation of microbial genomes and metagenomic sequences. Metagenomics and its applications in agriculture Nova Science Publishers, Hauppauge 2011:61-78.
37. Sullivan MJ, Petty NK, Beatson SA: Easyfig: a genome comparison visualizer. Bioinformatics 2011, 27:1009-1010.
38. Zafar N, Mazumder R, Seto D: CoreGenes: a computational tool for identifying and cataloging" core" genes in a set of small genomes. BMC Bioinformatics 2002, 3:12.
39. Kumar S, Stecher G, Tamura K: MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution 2016, 33:1870-1874.

40. Bardina C, Colom J, Spricigo DA, Otero J, Sánchez-Osuna M, Cortés P, Llagostera M: Genomics of three new bacteriophages useful in the biocontrol of *Salmonella*. Frontiers in Microbiology 2016, 7:545.
41. Oliveira H, Pinto G, Oliveira A, Noben J-P, Hendrix H, Lavigne R, Łobocka M, Kropinski AM, Azeredo J: Characterization and genomic analyses of two newly isolated *Morganella* phages define distant members among *Tevenvirinae* and *Autographivirinae* subfamilies. Scientific Reports 2017, 7:1-14.
42. Morozova V, Kozlova Y, Shedko E, Babkin I, Kurilshikov A, Bokovaya O, Bardashova A, Yunusova A, Tikunov A, Tupikin A: Isolation and characterization of a group of new *Proteus* bacteriophages. Archives of Virology 2018, 163:2189-2197.
43. Männistö RH, Kivelä HM, Paulin L, Bamford DH, Bamford JK: The complete genome sequence of PM2, the first lipid-containing bacterial virus to be isolated. Virology 1999, 262:355-363.
44. Gkoutzourelas A, Barmakoudi M, Bogdanos DP: A Bioinformatics Analysis Reveals Novel Pathogens as Molecular Mimicry Triggers of Systemic Sclerosis. Mediterranean Journal of Rheumatology 2020, 31:50.
45. Garner E, Benitez R, von Wagoner E, Sawyer R, Schaberg E, Hession WC, Krometis L-AH, Badgley BD, Pruden A: Stormwater loadings of antibiotic resistance genes in an urban stream. Water Research 2017, 123:144-152.
46. Egoz N, Shmilovitz M, Kretzer B, Lucian M, Porat V, Raz R: An outbreak of *Shigella sonnei* infection due to contamination of a municipal water supply in northern Israel. Journal of Infection 1991, 22:89-93.
47. Barnell A, Bennet J, Chehey R, Greenblatt J: *Shigella sonnei* outbreak associated with contaminated drinking water--Island Park, Idaho, August 1995. MMWR Morb Mortal Wkly Rep 1996, 45:229-231.
48. Arias C, Sala M, Domínguez A, Bartolomé R, Benavente A, Veciana P, Pedrol A, Hoyo G: Waterborne epidemic outbreak of *Shigella sonnei* gastroenteritis in Santa Maria de Palautordera, Catalonia, Spain. Epidemiology & Infection 2006, 134:598-604.
49. Kropinski AM, Sulakvelidze A, Konczy P, Poppe C: *Salmonella* phages and prophages—genomics and practical aspects**.** In Salmonella. Springer; 2007: 133-175
50. Brooke JS, Valvano MA: Biosynthesis of inner core lipopolysaccharide in enteric bacteria identification and characterization of a conserved phosphoheptose isomerase. J Biol Chem 1996, 271:3608-3614.
51. Connor TR, Barker CR, Baker KS, Weill F-X, Talukder KA, Smith AM, Baker S, Gouali M, Thanh DP, Azmi IJ: Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. Elife 2015, 4:e07335.
52. Buchanan RE, Gibbons NE: Bergey's manual of determinative bacteriology*.* 8 edn. Baltimore: Williams and Wilkins; 1974.
53. Doore SM, Schrad JR, Perrett HR, Schrad KP, Dean WF, Parent KN: A cornucopia of *Shigella* phages from the Cornhusker State. Virology 2019, 538:45-52.

54. Braun Jr V, Hertwig S, Neve H, Geis A, Teuber M: Taxonomic differentiation of bacteriophages of *Lactococcus lactis* by electron microscopy, DNA-DNA hybridization, and protein profiles. Microbiology 1989, 135:2551-2560.
55. Bao H, Shahin K, Zhang Q, Zhang H, Wang Z, Zhou Y, Zhang X, Zhu S, Stefan S, Wang R: Morphologic and genomic characterization of a broad host range *Salmonella enterica* serovar Pullorum lytic phage vB_SPuM_SP116. Microbial Pathogenesis 2019, 136:103659.

Figures

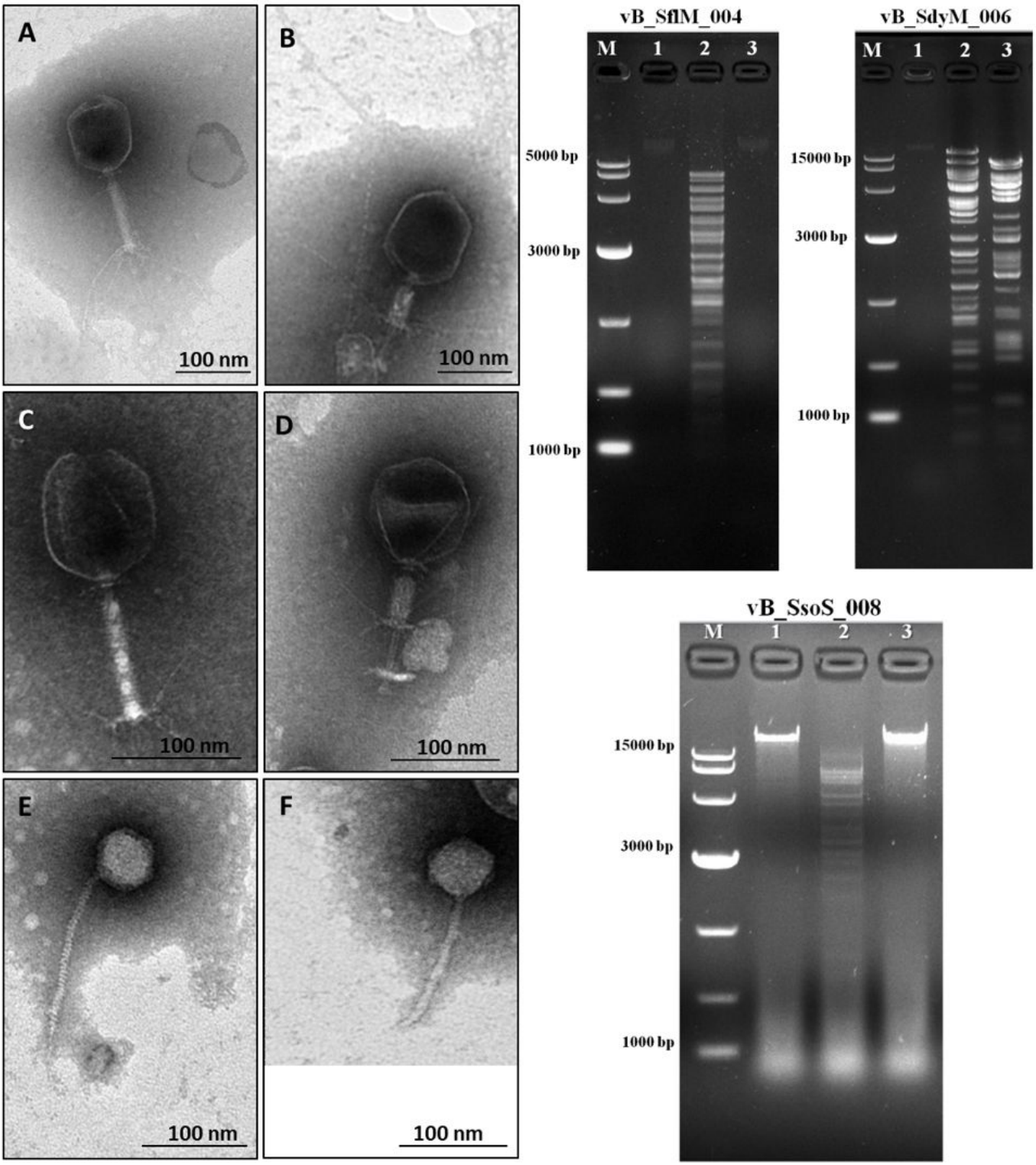

Figure 1

Left side, the electron micrograph of phages vB _SflM_004 (A and B), vB_SdyM_006 (C and D) and vB_SsoS_008 (E and F). The samples were negatively stained with 2 % phosphotungstic acid (PTA). Scale bars 100 nm. Right side, the DNA fingerprinting analysis of the genomic DNA of phage vB _SflM_004, vB_SdyM_006 and vB_SsoS_008. The genome was digested with EcoRI (line1), EcoRV (line2) and HindIII (line3). M line represents the DNA marker.

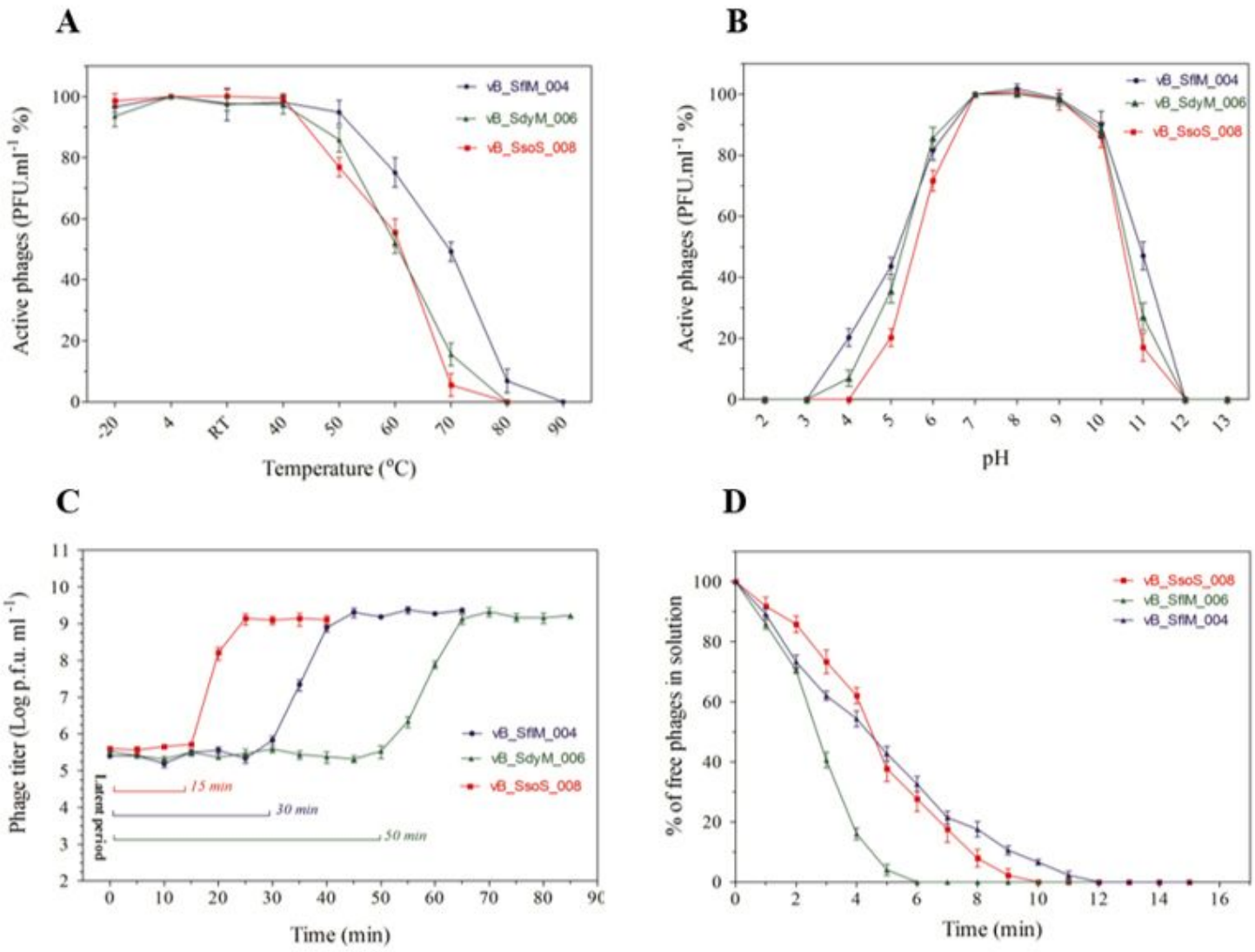

Figure 2

The thermostability (A), pH stability (B), one-step growth curve (C) and adsorption rate (D) of vB _SflM_004, vB_SdyM_006 and vB_SsoS_008. The error bars indicate standard deviations (SD).

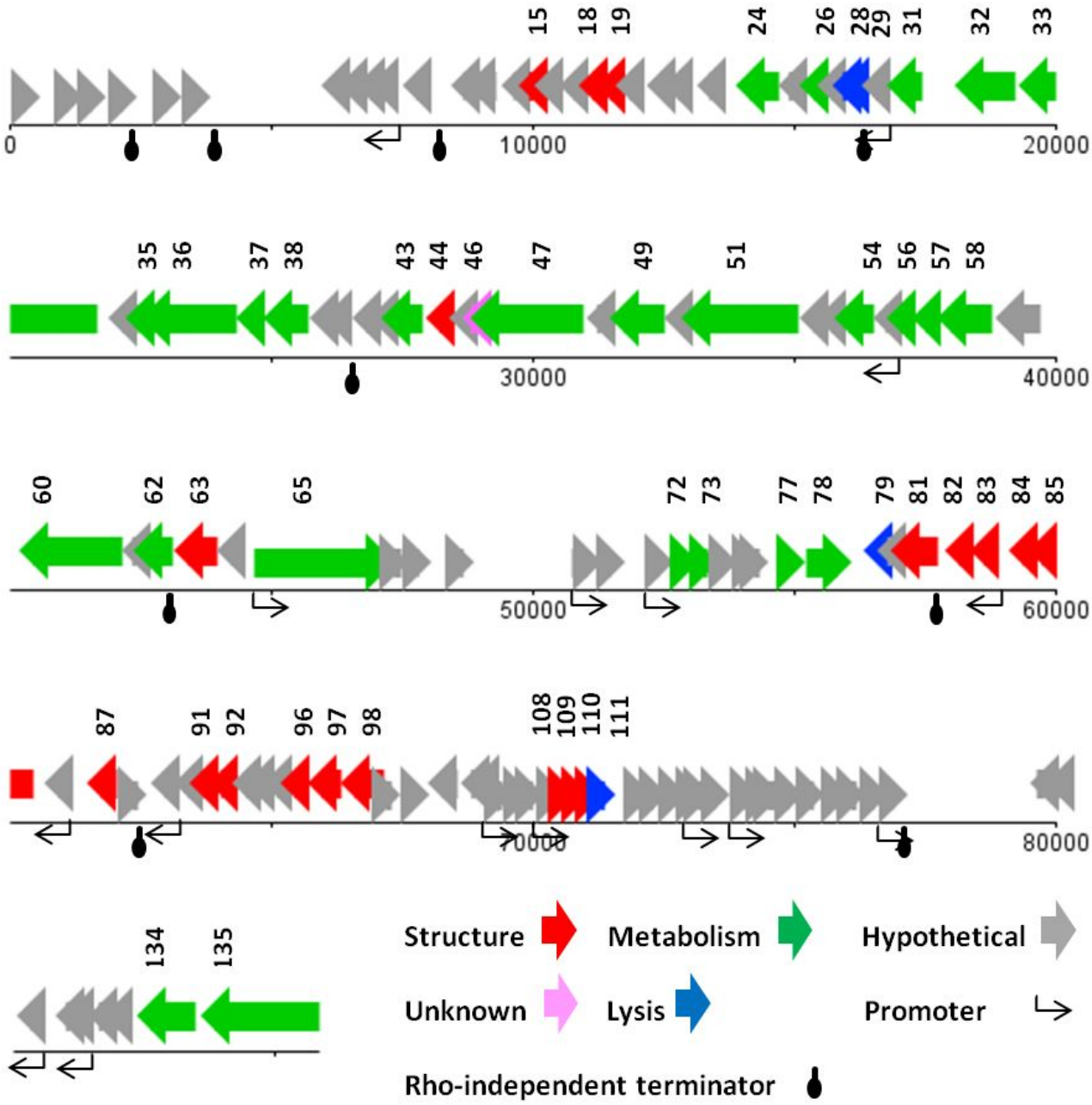

Figure 3

The linear genome map of phage vB_SflM_004.

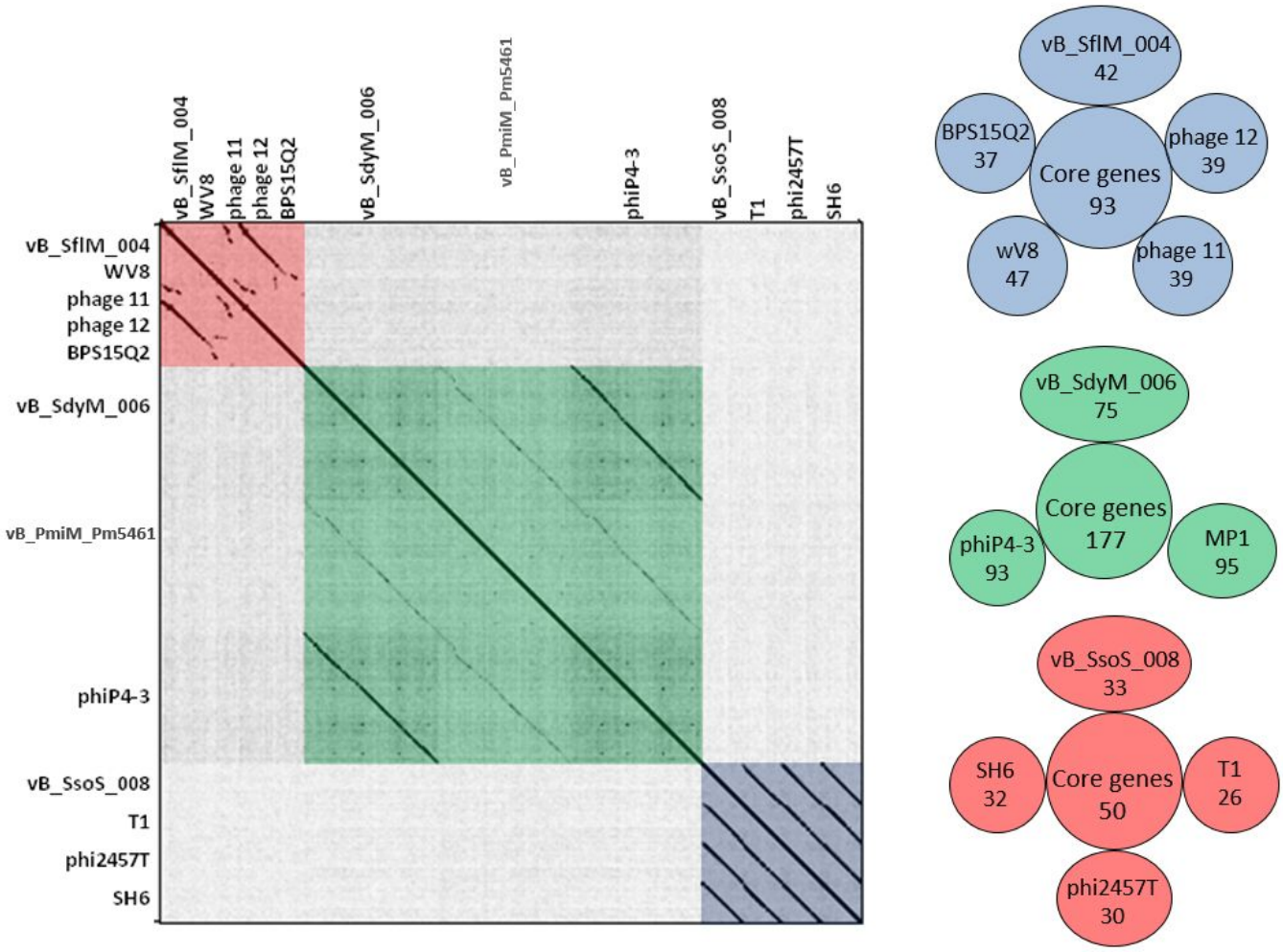

Figure 4

Left side, Dot plot alignment of the nucleotide sequences of the isolated phages vB _SflM_004, vB_SdyM_006 and vB_SsoS_008 with that of the close related phages. The FASTA file of 1300800 bp file was compared to itself using GEPARD. Dark diagonal lines parallel to the main diagonal shows strong and continuous sequence similarity, while pale lines indicate weaker sequence relationships due to interruptions that result in a discontinuous line. The red, green and blue boxes were added to illustrate the phage sequences related to vB _SflM_004, vB_SdyM_006 and vB_SsoS_008. Right side, the shared conserved proteins of vB _SflM_004, vB_SdyM_006 and vB_SsoS_008 phages with their close related phages (SH6, Shfl1, ADB-2, JMPW2). Only the gene products with >75 score in CoreGenes analysis were considered as homolog.

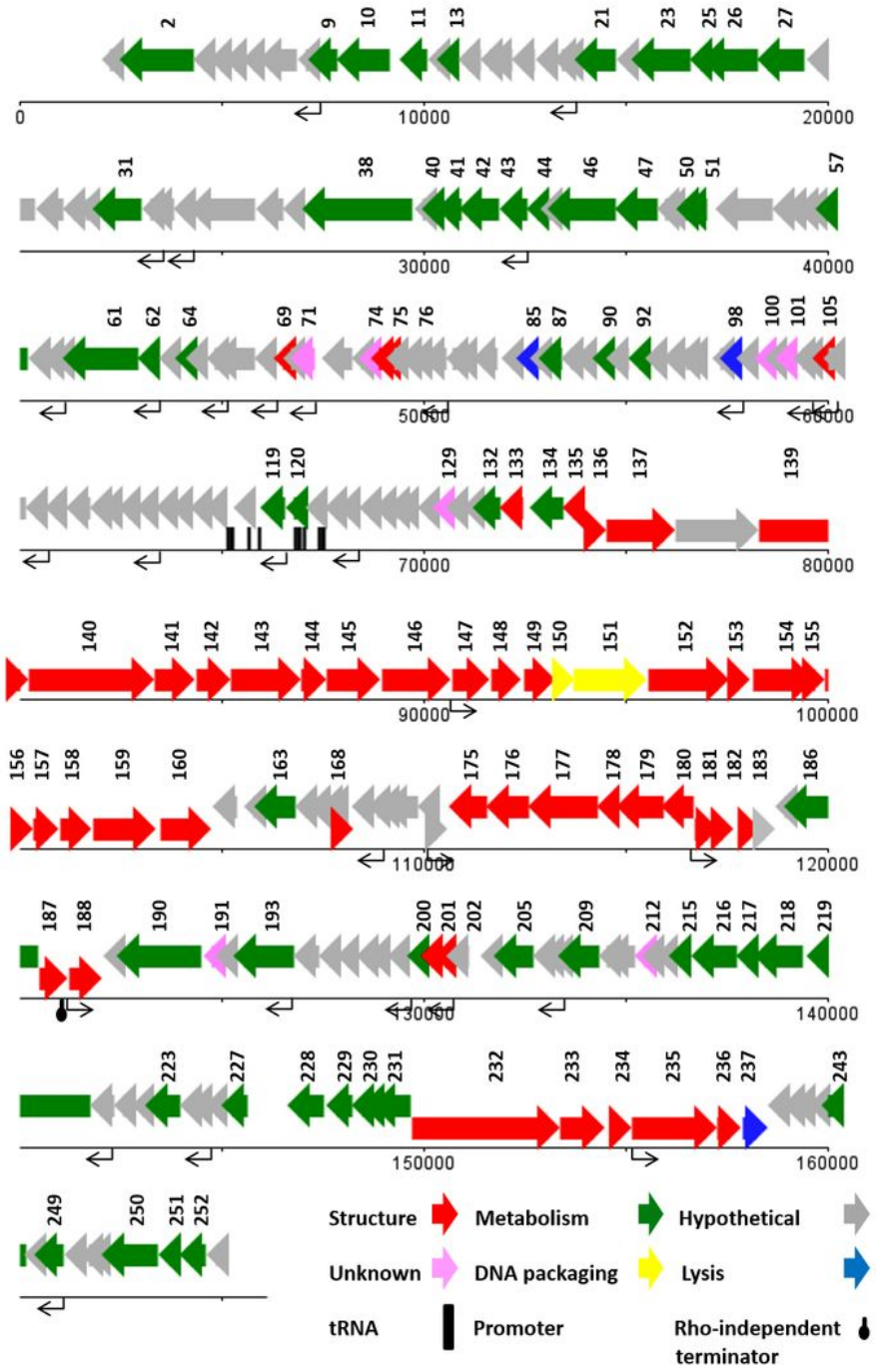

Figure 5

The linear genome map of phage vB_SdyM_006.

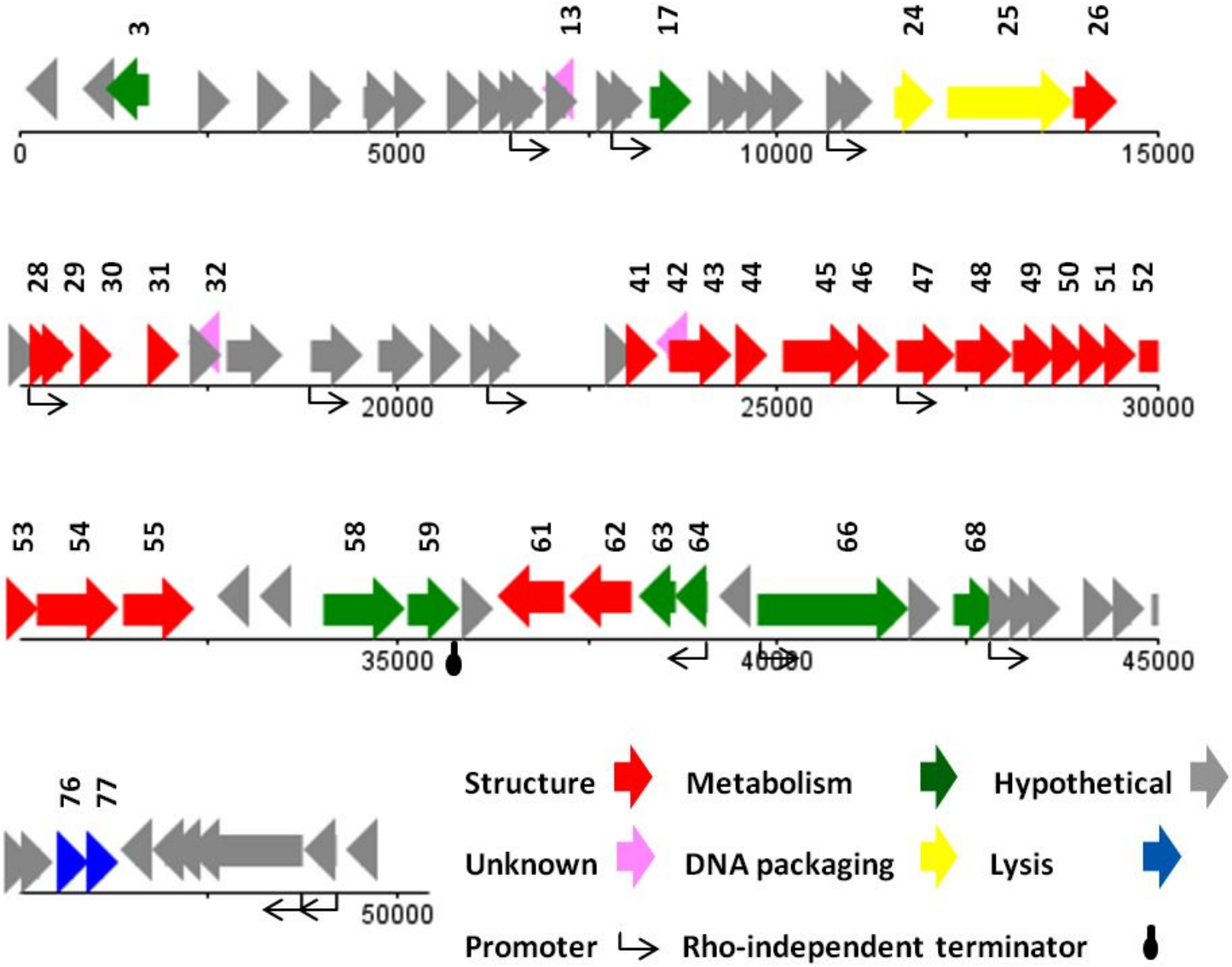

Figure 6

The linear genome map of phage vB_SsoS_008.

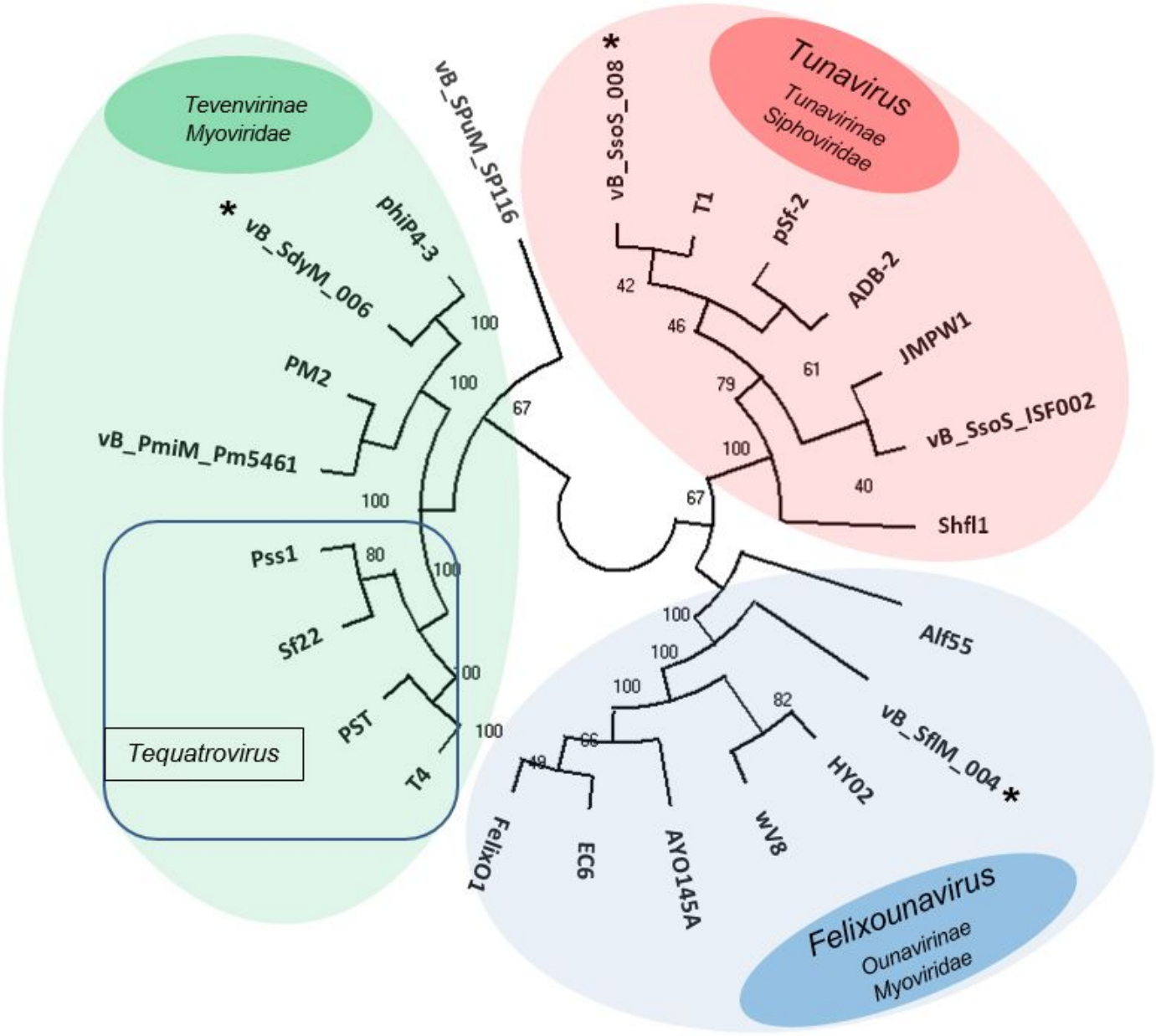

Figure 7

UPGMA tree phylogenic analysis of the major capsids of phages vB _SflM_004, vB_SdyM_006 and vB_SsoS_008. The tree was rooted using the major capsid of Salmonella enterica serovar Pullorum lytic phage vB_SPuM_SP116 [55] as the out group. The bootstrap percentages are shown next to each node. Phage groups as defined in the ICTV virus taxonomy (https://talk.ictvonline.org/taxonomy/) release are represented by different colors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile4putativepromotors.pdf
- Additionalfile3coregene.pdf
- Additionalfile2Annotation.pdf
- Additionalfile1.pdf