Data in brief 27 (2019) 104606

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Draft genome sequence data of T-5 like *Salmonella* bacteriophage ΦSP3 with demonstrated therapeutic potential

K.S. Sritha, Jeena Augustine, Sarita G. Bhat*

Department of Biotechnology, Cochin University of Science and Technology, Cochin, 682022, Kerala, India

ARTICLE INFO

Article history:
Received 17 December 2018
Received in revised form 16 September 2019
Accepted 26 September 2019
Available online 4 October 2019

Keywords:
Genome
Bacteriophage
Biocontrol agent
Salmonella

ABSTRACT

This data article describes the complete draft genome sequence of *Salmonella* specific bacteriophage ΦSP3 isolated from chicken intestinal contents. The ΦSP3 genome was sequenced by paired end runs using Illumina HiSeq 2500 with 100X coverage. Phylogenetic analysis using major capsid gene and genome wide comparison were performed to understand bacteriophage evolutionary relationship. Genome sequence of bacteriophage ΦSP3 was deposited in GenBank under the accession number MG387042.

1. Data

Bacteriophage ΦSP3 was isolated from chicken intestinal contents obtained from retail market in Cochin (*10.060256 N; 76.321881E*) as a therapeutic agent against *Salmonella* [1]. The physicochemical characteristics of the bacteriophage were studied in detail [1]. The phage significantly reduced bacteria applied on chicken cuts, especially at refrigerated conditions, making it an ideal candidate for storage applications [2]. A consortium of phages including ΦSP3 also increased the longevity of *C. elegans* infected with *Salmonella* [3], indicating its ability to control infections.

* Corresponding author.
E-mail address: sgbhat@cusat.ac.in (S.G. Bhat).

https://doi.org/10.1016/j.dib.2019.104606

Specifications Table

Subject area	*Biology*
More specific subject area	*Genomics*
Type of data	*Genome sequence data*
How data was acquired	*shot gun method using Illumina HiSeq 2500 with paired end runs*
Data format	*Raw and analyzed*
Experimental factors	*chicken intestinal content obtained from retail market*
Experimental features	*Draft genome sequence of Salmonella bacteriophage ΦSP3*
Data source location	*Cochin, India (10.060256 N; 76.321881E)*
Data accessibility	*Genome sequence was deposited in GenBank under the accession number* MG387042

Value of the Data

- The sequence data confirms the lytic nature and absence of toxic genes of ΦSP3 bacteriophage, fulfilling with the requirements for future commercialization.
- Data contributes to phage genomics which requires many more phages to be sequenced and analyzed, to fill the gaps of unidentified proteins as well as evolutionary relationships.
- Phage genome data can also be used for screening and Identification of novel antimicrobial proteins.

The transmission electron micrograph of ΦSP3 showed a bacteriophage with a hexagonal head (53.77 ± 0.38 nm) and a distinguishable long non-contractile tail (123.66 ± 0.32 nm), which are typical morphological features of family *Siphoviridae*, Fig. 1a [4]. The draft genome sequence of ΦSP3 genome had a size of 109,106 bp with 39.5% GC content. Many short overlapping regions between adjoining genes were frequently detected. There were 166 predicted ORFs in the genome of which 57 ORFs were with assigned functions (S1). 55 genes were transcribed in rightward direction (strand +) while 111 genes on the leftward direction (strand -). 16 tRNA encoding genes were identified. No gene related to phage lysogeny was detected, confirming the lytic nature of the phage. Moreover, the absence of genes encoding virulence and allergy inducing genes makes ΦSP3 highly innocuous for application. Single gene analysis using major capsid gene (ORF 149) was used to determine bacteriophage evolutionary relationship. It was observed that ΦSP3 clustered together with the T5 like phages (Fig. 1b). Another approach to understand the relationship is via whole genome comparison with related phages. This approach revealed that ΦSP3 was 97% similar to *Salmonella* phage SPC35. Aligning ΦSP3 with well-studied T5 genome [10] showed 95% similarity, but did not show any similarity to T4, T7 and T3 phages.

The genome map of ΦSP3 is displayed in Fig. 2. Genome annotation analysis showed that ΦSP3 genome is functionally organized into modules containing gene clusters involved in different functions viz genes required for inactivation of host genome as well as transfer of DNA, genes involved in nucleotide metabolism, lytic processes, packing and morphogenesis cluster. The DNA end structures can be predicted from the terminase amino acid sequence as the enzymes that generate the virion DNA ends are quite diverse. These different types of ends reflect differing DNA replication. Accordingly, the amino acid sequence of large terminase genes can be used to predict the packing strategy of phages [5,6]. The large terminase gene with 1316 bp (ORF 156) of phage ΦSP3 clustered with phages having known DNA termini and packing mechanisms. Phage ΦSP3 terminase gene clustered with that of T5 which shows that they have complex concatemeric packing mechanisms (Fig. 3) Moreover, the blast alignment with *Salmonella* phage SPC35 showed presence of long terminal repeats with a size of around 9000 bp.

2. Experimental design, materials, and methods

The phage morphology was determined by transmission electron microscopy (Model JOEL JEM-100X).

Phage DNA was extracted as per Sambrook *et al.* [7], and purity was determined in terms of 260/280 and 260/230 ratios. Phage genome was sequenced by shot gun method using Illumina HiSeq 2500 with

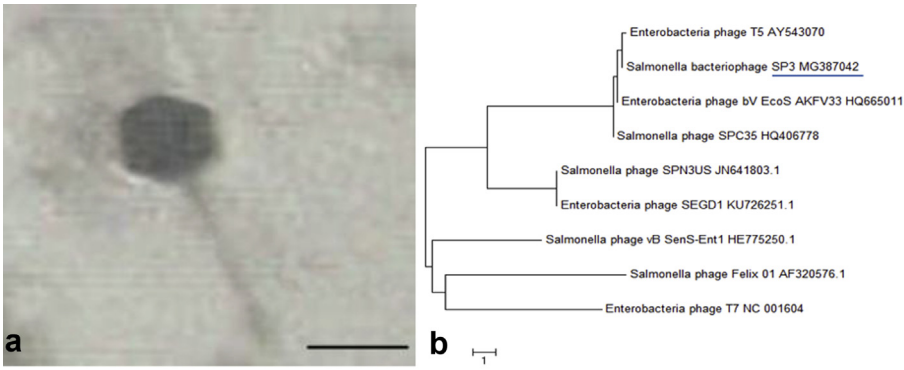

Fig. 1. a) Transmission electron micrograph image of phage ΦSP3 stained with 1% uranyl acetate (bar represents 60nm). b) Phylogenetic tree based on major capsid gene of selected bacteriophages. The gene sequences are compared using the ClustalW program, and the phylogenetic tree was generated using the neighbour-joining method and 1000 bootstrap replicates.

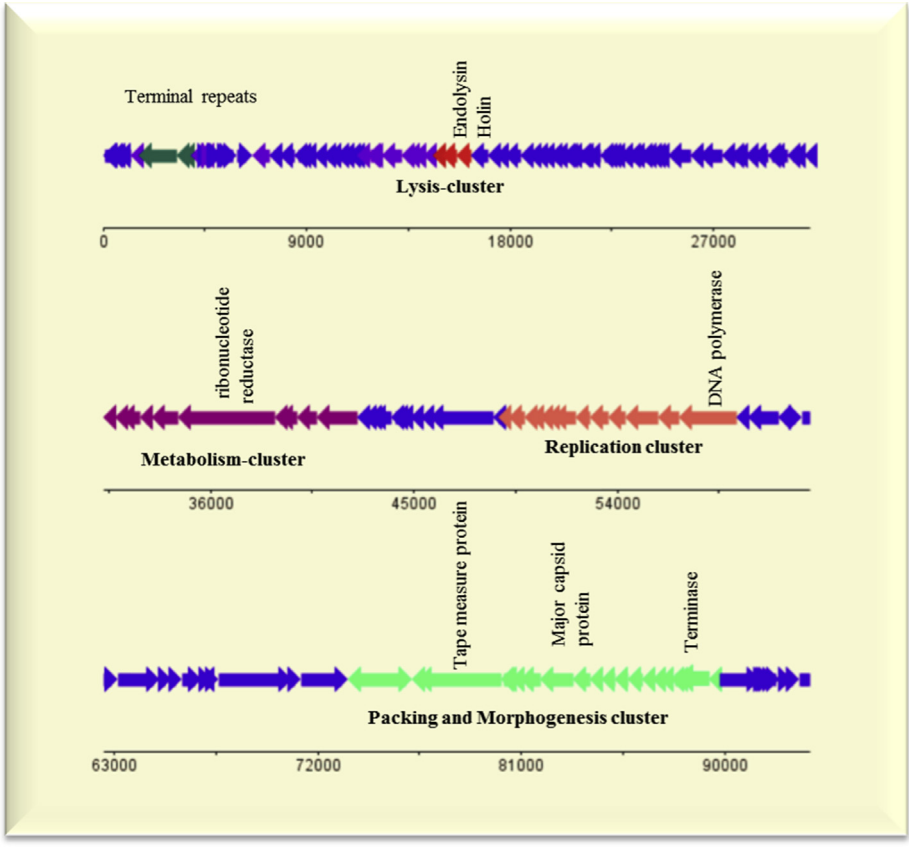

Fig. 2. Genome map of bacteriophage ΦSP3 (DNA plotter).

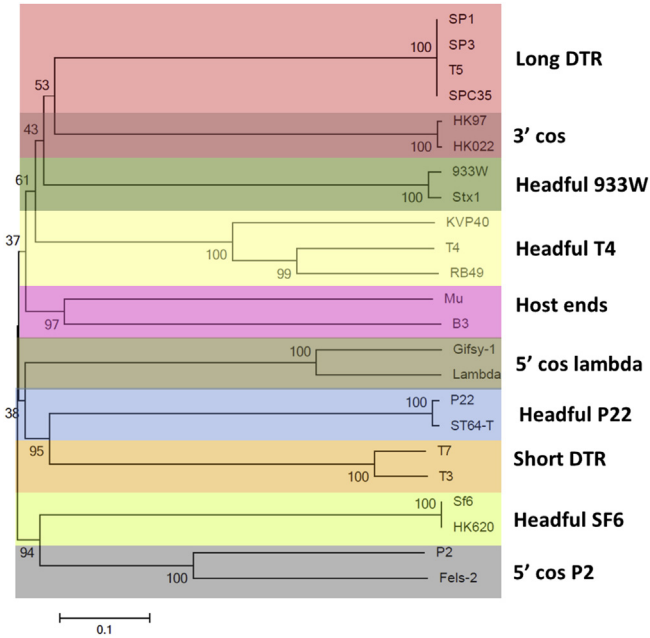

Fig. 3. a) Phylogenetic tree on based amino acid sequences of terminase gene. ΦSP3 were clustered with phages of known termini and with experimentally determined packaging mechanisms. The DNA termini structures of phages that have been experimentally determined are indicated on right side on each cluster [5] structures have been experimentally determined are indicated.

paired end runs with 100X Coverage. The whole genome sequences were assembled using IVA [8] and SEQuel for correcting errors [9]. Genes were predicted using GeneMarkS [10]. Predicted ORFs were annotated with BLASTX, Uniprot, NCBI Conserved Domain Database (CDD). DNA Plotter was used to construct phage genome map [11] and tRNA gene prediction by tRNAscan-SE [12]. Phylogenetic tree depicting the evolutionary relationship of *Salmonella* bacteriophage was generated based on terminase gene and major capsid gene by neighbour-joining method [13] using MEGA 7.0 software [14].

Acknowledgments

The authors acknowledge the financial support of KSCSTE through the project F.No. 009/SRSHS/2012/CSTE to Sarita G Bhat.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.104606.

References

[1] J. Augustine, S.G. Bhat, Physicochemical characterization of a T5-like *Salmonella* phage [Phi] SP-3, J. Microbiol. Biotechnol. Food Sci. 4 (2014) pp102–pp107, https://doi.org/10.15414/jmbfs.2014.4.2.102-107.
[2] J. Augustine, S.G. Bhat, Bio control of *Salmonella* Enteritidis in spiked chicken cuts by lytic bacteriophages ΦSP-1 and ΦSP-3, J. Basic Microbiol. 55 (2015) 500–503.

[3] J. Augustine, M.V. Gopalakrishnan, S.G. Bhat, Application of ΦSP-1 and ΦSP-3 as a therapeutic strategy against *Salmonella* Enteritidis infection using *Caenorhabditis elegans* as model organism, FEMS Microbiol. Lett. 356 (2014) pp113–pp117.
[4] M. Krupovic, B.E. Dutilh, E.M. Adriaenssens, J. Wittmann, F.K. Vogensen, M.B. Sullivan, J. Rumnieks, D. Prangishvili, R. Lavigne, A.M. Kropinski, J. Klumpp, Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee, Arch. Virol. 161 (2016) pp1095–pp1099.
[5] S.R. Casjens, E.B. Gilcrease, D.A. Winn-Stapley, P. Schicklmaier, H. Schmieger, M.L. Pedulla, R.W. Hendrix, The generalized transducing *Salmonella* bacteriophage ES18: complete genome sequence and DNA packaging strategy, J. Bacteriol. 187 (2005) 1091–1104.
[6] S.R. Casjens, E.B. Gilcrease, Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions, Bacteriophages Methods Protocols Vol. 2 Molec. Appl. Aspect. (2009) 91–111.
[7] J. Sambrook, E. Fritsch, I. Maniatis, Molecular Cloning – A Laboratory Manual, second ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, 2000.
[8] M. Hunt, A. Gall, S.H. Ong, J. Brener, B. Ferns, P. Goulder, E. Nastouli, J.A. Keane, P. Kellam, T.D. Otto, IVA: accurate de novo assembly of RNA virus genomes, Bioinformatics 31 (2015) 2374–2376.
[9] R. Ronen, C. Boucher, H. Chitsaz, P. Pevzner, SEQuel: improving the accuracy of genome assemblies, Bioinformatics 12 (2012) i188–i196.
[10] J. Besemer, A. Lomsadze, M. Borodovsky, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions, Nucleic Acids Res. 29 (2001) 2607–2618.
[11] T. Carver, N. Thomson, A. Bleasby, M. Berriman, J. Parkhill, DNAPlotter: circular and linear interactive genome visualization, Bioinformatics 25 (2009) 119–120.
[12] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, Nucleic Acids Res. 25 (1997) 955–964.
[13] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[14] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874.