

# Genome sequences of bacteriophages that infect *Salmonella* Typhi from Bangladesh

Shuborno Islam,<sup>1</sup> Rathindranath Kabiraj,<sup>1</sup> Himadree Sarkar,<sup>1</sup> Preonath Chondrow Dev,<sup>1</sup> Afroza Akter Tanni,<sup>1</sup> Deb Purna Keya,<sup>1</sup> Apurba Rajib Malakar,<sup>1</sup> Arif Mohammad Tanmoy,<sup>1</sup> Samir K. Saha,<sup>1,2</sup> Yogesh Hooda,<sup>1</sup> Senjuti Saha<sup>1</sup>

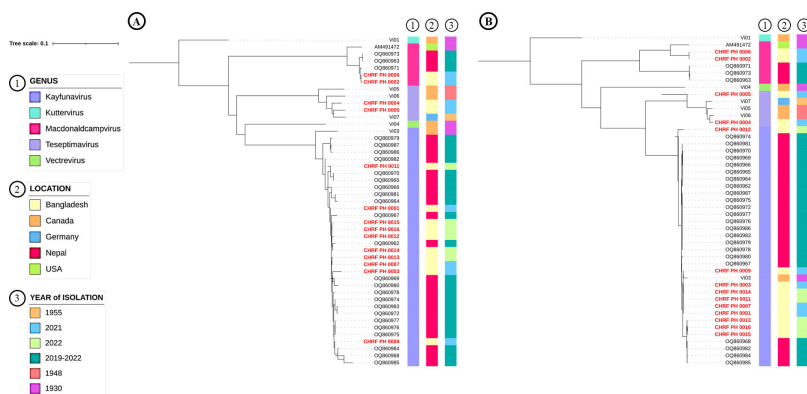
**AUTHOR AFFILIATIONS** See affiliation list on p. 3.

**ABSTRACT** This report presents near-complete genome sequences of 14 bacteriophages that infect *Salmonella* Typhi, identified through environmental surveillance in Bangladesh between August 2021 and June 2022. The bacteriophages, belonging to the genera *Kayfunavirus*, *Macdonaldcampvirus*, and *Teseptimavirus*, exhibit high degrees of sequence similarity and conserved genetic features with previously reported Typhi bacteriophages.

**KEYWORDS** *Salmonella*, bacteriophages, typhoid, genomics

*Salmonella enterica* serovar Typhi causes typhoid fever and remains a major public health concern in low- and middle-income countries (1). Historically, *Salmonella* Typhi-specific bacteriophages were used to characterize strains of *Salmonella* Typhi (2–4). Despite a decline in their use post-1980s, recent findings from Nepal (5) and Bangladesh (6) confirmed the persistence and genomic diversity of Typhi phages in endemic settings. This study reports sequences of 14 Typhi phages isolated in Bangladesh, expanding the global database.

Here, bacteriophages were isolated from surface water samples (Table 1) using *Salmonella* Typhi BRD948 strain grown at 37°C as the host, as described previously (6). Briefly, phages were amplified from single plaques, and pure phage lysates were treated with DNase, RNase, and Proteinase K, followed by genomic DNA extraction using the



**FIG 1** Phylogenetic tree of *Salmonella* Typhi bacteriophages. The 14 phages (highlighted in red) sequenced in this study were contextualized against 7 phages from the USA (13), Canada, and Germany (14) and 26 phages from Nepal (5). The genus, location, and year of isolation for the phage sequences are also shown. (A) Tail-fiber nucleotide sequences and (B) terminase nucleotide sequences of respective phages were extracted separately. These sequences were aligned, and separate phylogenetic trees were generated using Clustal Omega (online) (11). The resulting trees were visualized with iTOL (version 6.0) (15).

**Editor** Simon Roux, DOE Joint Genome Institute, Berkeley, California, USA

Address correspondence to Senjuti Saha, senjutisaha@gmail.com.

The authors declare no conflict of interest.

See the funding table on p. 3.

**Received** 2 May 2024

**Accepted** 20 November 2024

**Published** 17 December 2024

Copyright © 2024 Islam et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

TABLE 1 Details of the 14 bacteriophages sequenced in this study<sup>a,b</sup>

Sample ID	Accession no./ (SRA accession no.)	Date of collection (DD/MM/YYYY)	Sample location in Bangladesh (Latitude longitude)	Sample source	Size in bp (ORFs)	Genome coverage (G + C%)	Genus
CHRF_PH_0001	PQ336901 (SRR28112640)	21/08/2021	Dhaka (23.76 N 90.39 E)	Sewage	39,761 (65)	1,046× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0002	PQ336902 (SRR28112639)	21/08/2021	Dhaka (23.75 N 90.40 E)	Lake	45,880 (87)	2,245× (46%)	<i>Macdonaldcampvi- rus</i>
CHRF_PH_0003	PQ336903 (SRR28112634)	21/08/2021	Dhaka (23.75 N 90.40 E)	Lake	37,939 (58)	273× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0004	PQ336904 (SRR28112633)	22/08/2021	Dhaka (23.82 N 90.34 E)	Lake	38,164 (53)	1,013× (49%)	<i>Teseptimavirus</i>
CHRF_PH_0005	PQ336905 (SRR28112632)	22/08/2021	Dhaka (23.86 N 90.36 E)	River	38,546 (56)	881× (49%)	<i>Teseptimavirus</i>
CHRF_PH_0006	PQ336906 (SRR28112631)	23/08/2021	Dhaka (23.77 N 90.36 E)	Sewage	44,975 (84)	1,174× (46%)	<i>Macdonaldcampvi- rus</i>
CHRF_PH_0007	PQ336907 (SRR28112630)	23/08/2021	Dhaka (23.76 N 90.36 E)	Sewage	39,928 (63)	884× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0009	PQ336908 (SRR28112629)	03/11/2021	Mirzapur (24.11 N 90.10 E)	Stagnant water	39,181 (61)	18,097× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0011	PQ336909 (SRR28112628)	09/11/2021	Mirzapur (24.12 N 90.15 E)	Pond	38,099 (61)	11,083× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0012	PQ336910 (SRR28112627)	08/01/2022	Chattogram (22.38 N 91.81 E)	Sewage	40,368 (65)	9,607× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0013	PQ336911 (SRR28112638)	17/01/2022	Chattogram (22.35 N 91.85 E)	Sewage	39,889 (60)	6,989× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0014	PQ336912 (SRR28112637)	06/03/2022	Chattogram (22.37 N 91.85 E)	Sewage	40,094 (61)	6,906× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0015	PQ336913 (SRR28112636)	14/03/2022	Chattogram (22.37 N 91.80 E)	Sewage	39,834 (62)	8,555× (51%)	<i>Kayfunavirus</i>
CHRF_PH_0016	PQ336914 (SRR28112635)	14/03/2022	Chattogram (22.38 N 91.80 E)	Sewage	39,834 (62)	10,259× (51%)	<i>Kayfunavirus</i>

<sup>a</sup>CHRF\_PH\_0001–CHRF\_PH\_0007 were sequenced in an Illumina iSeq100 and the CHRF\_PH\_0009, 0011–0016 were sequenced in a NextSeq2000 platform.

<sup>b</sup>ORFs, open reading frames.

QIAamp DNA Mini Kit (Qiagen, Germany). Sequencing libraries were prepared using the NEBNext Ultra II FS DNA Library Prep Kit (MA, USA) and sequenced on Illumina iSeq100 or NextSeq2000 platforms with 150-bp paired-end reads (7) (Table 1).

The quality of the sequencing reads was assessed using FastQC (version 0.11.5). Adapter sequences were trimmed using Trimmomatic (version 0.39) before assembled into single contig using Unicycler (version 0.4.9) (8). Assembled genomes were used for genus identification with Kraken2 (version 2.1.2) (9) and for protein function annotation with Pharokka (version 1.2.1) (10). The tail fiber and terminase genes were individually extracted for comparison of all Typhi phage genomes sequenced globally to date using a Biopython script ([https://github.com/CHRF-Genomics/Phage\\_Analysis](https://github.com/CHRF-Genomics/Phage_Analysis)). These sequences were then aligned using Clustal Omega (11), and two separate phylogenetic trees were generated for each gene. All tools were run with default parameters.

All assembled phages resulted in near-complete genomes with an average coverage depth of over 1,000×, assessed using QUAST (version 5.2) (12). Of the 14 phages, 10 phages belong to the *Kayfunavirus* genus with 60–65 open reading frames (ORFs), 2 were categorized under *Macdonaldcampvirus* with 84–87 ORFs, and the remaining 2 under *Teseptimavirus*, containing 53–56 ORFs (Table 1).

The phages sequenced in this study from Bangladesh were compared with 26 phages from Nepal (5) and 7 phages from the USA (13), Canada, and Germany (14). The majority of the phages from both Nepal and Bangladesh were identified as *Kayfunavirus*, covering

three of the five genera of Typhi phages reported to date (5, 13, 14). Phylogenetic analyses based on the terminase and tail-fiber nucleotide sequences exhibited similar clustering patterns (Fig. 1), indicating that either gene can be used for comparative genomic analysis.

This study presents the near-complete genome sequences of *Salmonella* Typhi bacteriophages from Bangladesh, shedding light on the phage landscape in a region critically affected by typhoid fever.

# ACKNOWLEDGMENTS

We thank Prof. Jason R. Andrews of Stanford University for his guidance and collaboration during the project.

This study was funded by the Bill and Melinda Gates Foundation (INV-003717).

# AUTHOR AFFILIATIONS

<sup>1</sup>Child Health Research Foundation, Dhaka, Bangladesh

<sup>2</sup>Department of Microbiology, Bangladesh Shishu (Children) Hospital and Institute, Dhaka, Bangladesh

# AUTHOR ORCIDs

Shuborno Islam  <http://orcid.org/0000-0001-8421-3115>

Yogesh Hooda  <http://orcid.org/0000-0002-7181-9254>

Senjuti Saha  <http://orcid.org/0000-0001-6087-6766>

# FUNDING

Funder	Grant(s)	Author(s)
Bill and Melinda Gates Foundation (GF)	INV-003717	Senjuti Saha

# AUTHOR CONTRIBUTIONS

Shuborno Islam, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Rathindranath Kabiraj, Formal analysis, Investigation, Methodology, Writing – review and editing | Himadree Sarkar, Visualization, Writing – review and editing | Preonath Chondrow Dev, Methodology, Visualization | Afroza Akter Tanni, Investigation, Methodology | Deb Purna Keya, Investigation, Methodology | Apurba Rajib Malakar, Investigation, Methodology | Arif Mohammad Tanmoy, Investigation, Methodology, Project administration, Visualization | Samir K. Saha, Funding acquisition, Supervision | Yogesh Hooda, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing | Senjuti Saha, Conceptualization, Funding acquisition, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing

# DATA AVAILABILITY

The raw data of all 14 Typhi phages in this study are available on NCBI under the BioProject [PRJNA1081195](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1081195). The phage genome sequences have been deposited in GenBank (accession numbers: [PQ336901-14](https://www.ncbi.nlm.nih.gov/nuccore/PQ336901-14)). Sequences of the 26 phages from Nepal are available under the BioProject [PRJNA933946](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA933946) (accession numbers: [OQ860962-87](https://www.ncbi.nlm.nih.gov/nuccore/OQ860962-87)). The sequences of the 7 phages from the USA, Canada, and Germany are available through NCBI (accession: [AM491472](https://www.ncbi.nlm.nih.gov/nuccore/AM491472)) and Sanger FTP (ID: Vi01, Vi03-Vi07, from <https://ftp.sanger.ac.uk/pub/project/pathogens/Phage/>). The script used for extracting phage-specific protein sequences has been uploaded to [https://github.com/CHRF-Genomics/Phage\\_Analysis](https://github.com/CHRF-Genomics/Phage_Analysis).

## REFERENCES

- Garrett DO, Longley AT, Aiemyjoy K, Yousafzai MT, Hemlock C, Yu AT, Vaidya K, Tamrakar D, Saha S, Bogoch II, et al. 2022. Incidence of typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the Surveillance for Enteric Fever in Asia Project. *Lancet Glob Health* 10:e978–e988. [https://doi.org/10.1016/S2214-109X\(22\)00119-X](https://doi.org/10.1016/S2214-109X(22)00119-X)
- Craigie J, Felix A. 1947. Typing of typhoid bacilli with Vi bacteriophage; suggestions for its standardisation. *Lancet* 1:823–827. [https://doi.org/10.1016/S0140-6736\(47\)91719-4](https://doi.org/10.1016/S0140-6736(47)91719-4)
- FELIX A. 1955. World survey of typhoid and paratyphoid-B phages types. *Bull World Health Organ* 13:109–170. <https://iris.who.int/handle/10665/265572>
- Freytag B, Plochmann E. 1955. Phage typing of typhoid strains. *Z Hyg Infekt Krankh* 142:188–196. <https://doi.org/10.1007/BF02150088>
- Shrestha S, Da Silva KE, Shakya J, Yu AT, Katuwal N, Shrestha R, Shakya M, Shahi SB, Naga SR, LeBoa C, Aiemyjoy K, Bogoch II, Saha S, Tamrakar D, Andrews JR. 2024. Detection of *Salmonella* Typhi bacteriophages in surface waters as a scalable approach to environmental surveillance. *PLoS Negl Trop Dis* 18:e0011912. <https://doi.org/10.1371/journal.pntd.0011912>
- Hooda Y, Islam S, Kabiraj R, Rahman H, Sarkar H, da Silva KE, Raju RS, Luby SP, Andrews JR, Saha SK, Saha S. 2024. Old tools, new applications: use of environmental bacteriophages for typhoid surveillance and evaluating vaccine impact. *PLoS Negl Trop Dis* 18:e0011822. <https://doi.org/10.1371/journal.pntd.0011822>
- Islam S, Kabiraj R, Amin A, Hussain Pranto S, Dey S, Rajib Malaker A, Purna Keya D, Mohammad Tanmoy A, K Saha S, Hooda Y, Saha S. 2024. Protocol for genomic DNA extraction and sequencing library preparation from phage stock v1. *Protocols.io*. <https://doi.org/10.17504/protocols.io.kqdg32477v25/v1>
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
- Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald P-J, Vreugde S. 2023. Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* 39:btac776. <https://doi.org/10.1093/bioinformatics/btac776>
- Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* 27:135–145. <https://doi.org/10.1002/pro.3290>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Pickard D, Thomson NR, Baker S, Wain J, Pardo M, Goulding D, Hamlin N, Choudhary J, Threfall J, Dougan G. 2008. Molecular characterization of the *Salmonella enterica* serovar Typhi Vi-typing bacteriophage E1. *J Bacteriol* 190:2580–2587. <https://doi.org/10.1128/JB.01654-07>
- Pickard Derek, Toribio AL, Petty NK, van Tonder A, Yu L, Goulding D, Barrell B, Rance R, Harris D, Wetter M, Wain J, Choudhary J, Thomson N, Dougan G. 2010. A conserved acetyl esterase domain targets diverse bacteriophages to the Vi capsular receptor of *Salmonella enterica* serovar Typhi. *J Bacteriol* 192:5746–5754. <https://doi.org/10.1128/JB.00659-10>
- Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 49:W293–W296. <https://doi.org/10.1093/nar/gkab301>