1 **Genome sequences of bacteriophages that infect *Salmonella* Typhi from**

2 **Bangladesh**

3 Shuborno Islam[1], Rathindranath Kabiraj[1], Himadree Sarkar[1], Preonath Chondrow Dev[1],

4 Afroza Akter Tanni[1], Deb Purna Keya[1], Apurba Rajib Malakar[1], Arif Mohammad Tanmoy[1],

5 Samir K Saha[1,2], Yogesh Hooda[1], Senjuti Saha[1]

6 1. Child Health Research Foundation, Bangladesh

7 2. Department of Microbiology, Bangladesh Shishu (Children) Hospital and Institute,

8 Bangladesh

9

10 **Abstract**

11 This report presents near-complete genome sequences of 14 bacteriophages that infect

12 *Salmonella* Typhi, identified through environmental surveillance in Bangladesh between

13 August 2021 and June 2022. The bacteriophages, belonging to the genera *Kayfunavirus*,

14 *Macdonaldcampvirus*, and *Teseptimavirus*, exhibit high degrees of sequence similarity and

15 conserved genetic features with previously reported Typhi bacteriophages.

16

17 **Announcement**

18 *Salmonella enterica* serovar Typhi causes typhoid fever and remains a major public health

19 concern in low- and middle-income countries (1). Historically, *Salmonella* Typhi-specific

20 bacteriophages were used to characterize strains of *Salmonella* Typhi (2–4). Despite a decline

21 in their use post-1980s, recent findings from Nepal (5) and Bangladesh (6) confirmed the

22 persistence and genomic diversity of Typhi phages in endemic settings. This study reports

23 sequences of 14 Typhi phages isolated in Bangladesh, expanding the global database.

24

25    Here, bacteriophages were isolated from surface water samples (Table 1) using *Salmonella*

26    Typhi BRD948 strain grown at 37°C as the host, as described previously (6). Briefly, phages

27    were amplified from single plaques and pure phage lysates were treated with DNase, RNase

28    and Proteinase K, followed by genomic DNA extraction using the QIAamp DNA Mini Kit

29    (Qiagen, Germany). Sequencing libraries were prepared using the NEBNext® Ultra™ II FS

30    DNA Library Prep Kit (Massachusetts, USA), and sequenced on Illumina iSeq100 or

31    NextSeq2000 platforms with 150-bp paired-end reads (7) (Table 1).

32

33    **Table 1.** Details of the 14 bacteriophages sequenced in this study. CHRF_PH_0001 to

34    CHRF_PH_0007 were sequenced in an Illumina iSeq100 and the CHRF_PH_0009, 0011-

35    0016 were sequenced in a NextSeq2000 platform.

| Sample ID | Accession no./ (SRA accession no.) | Data 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) | 1,046x (51%) | *Kayfunavirus* |
| CHRF_PH_0002 | PQ336902 (SRR28112639) | 21/08/2021 | Dhaka ( 23.75 N 90.40 E ) | Lake | 45,880 (87) | 2,245x (46%) | *Macdonalcampvirus* |
| CHRF_PH_0003 | PQ336903 (SRR28112634) | 21/08/2021 | Dhaka ( 23.75 N 90.40 E ) | Lake | 37,939 (58) | 273x (51%) | *Kayfunavirus* |
| CHRF_PH_0004 | PQ336904 (SRR28112633) | 22/08/2021 | Dhaka ( 23.82 N 90.34 E) | Lake | 38,164 (53) | 1,013x (49%) | *Teseptimavirus* |
| CHRF_PH_0005 | PQ336905 (SRR28112632) | 22/08/2021 | Dhaka ( 23.86 N 90.36 E ) | River | 38,546 (56) | 881x (49%) | *Teseptimavirus* |
| CHRF_PH_0006 | PQ336906 (SRR28112631) | 23/08/2021 | Dhaka ( 23.77 N 90.36 E) | Sewage | 44,975 (84) | 1,174 x (46%) | *Macdonalcampvirus* |
| CHRF_PH_0007 | PQ336907 (SRR28112630) | 23/08/2021 | Dhaka ( 23.76 N 90.36 E ) | Sewage | 39,928 (63) | 884x (51%) | *Kayfunavirus* |
| CHRF_PH_0009 | PQ336908 (SRR28112629) | 03/11/2021 | Mirzapur ( 24.11 N 90.10 E ) | Stagnant Water | 39,181 (61) | 18,097x (51%) | *Kayfunavirus* |

| CHRF_PH_0011 | PQ336909 (SRR28112628) | 09/11/2021 | Mirzapur ( 24.12 N 90.15 E ) | Pond | 38,099 (61) | 11,083 (51%) | *Kayfunavirus* |
|---|---|---|---|---|---|---|---|
| CHRF_PH_0012 | PQ336910 (SRR28112627) | 08/01/2022 | Chattogram ( 22.38 N 91.81 E ) | Sewage | 40,368 (65) | 9,607 (51%) | *Kayfunavirus* |
| CHRF_PH_0013 | PQ336911 (SRR28112638) | 17/01/2022 | Chattogram ( 22.35 N 91.85 E ) | Sewage | 39,889 (60) | 6,989 (51%) | *Kayfunavirus* |
| CHRF_PH_0014 | PQ336912 (SRR28112637) | 06/03/2022 | Chattogram ( 22.37 N 91.85 E ) | Sewage | 40,094 (61) | 6,906 (51%) | *Kayfunavirus* |
| CHRF_PH_0015 | PQ336913 (SRR28112636) | 14/03/2022 | Chattogram ( 22.37 N 91.80 E ) | Sewage | 39,834 (62) | 8,555 (51%) | *Kayfunavirus* |
| CHRF_PH_0016 | PQ336914 (SRR28112635) | 14/03/2022 | Chattogram ( 22.38 N 91.80 E ) | Sewage | 39,834 (62) | 10,259 (51%) | *Kayfunavirus* |

36

37

38  The quality of the sequencing reads was assessed using FastQC (V.0.11.5). Adapter

39  sequences were trimmed using Trimmomatic (V.0.39) before assembled into single contig

40  using Unicycler (V.0.4.9)  (8). Assembled genomes were used for genus identification with

41  Kraken2 (V 2.1.2) (9) and for protein function annotation with Pharokka (V 1.2.1) (10). The

42  tail fiber and terminase genes were individually extracted for comparison of all Typhi phage

43  genomes sequenced globally to date using a Biopython script (https://github.com/CHRF-

44  Genomics/Phage_Analysis). These sequences were then aligned using Clustal Omega (11),

45  and two separate phylogenetic trees were generated for each gene. All tools were run with

46  default parameters.

47

48  All assembled phages resulted in near-complete genomes with an average coverage depth of

49  over 1000x, assessed using QUAST (V.5.2) (12). Ten of the 14 phages belong to the

50  *Kayfunavirus* genus with 60-65 open reading frames (ORFs), two were categorized under

51  *Macdonaldcampvirus* with 84-87 ORFs, and the remaining two under *Teseptimavirus*,

52  containing 53-56 ORFs (Table 1).

53

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). 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 (Figure

1), indicating that either gene can be used for comparative genomic analysis.

60

**Figure 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 separately phylogenetic trees were generated using Clustal
Omega (online) (11). The resulting trees were visualized with iTOL (V 6.0) (15).

68

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.

72

**Data Availability**

The raw data of all 14 Typhi phages in this study are available on NCBI under the BioProject

PRJNA1081195. The phage genome sequences have been deposited in GenBank (accession

numbers: PQ336901-14). Sequences of the 26 phages from Nepal are available under the

BioProject PRJNA933946 (accession numbers: OQ860962-87). The sequences of the 7

phages from USA, Canada and Germany are available through NCBI (accession:

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.

**References**

1. Garrett DO, Longley AT, Aiemjoy K, Yousafzai MT, Hemlock C, Yu AT, Vaidya K,

Tamrakar D, Saha S, Bogoch II, Date K, Saha S, Islam MS, Sayeed KMI, Bern C,

Shakoor S, Dehraj IF, Mehmood J, Sajib MSI, Islam M, Thobani RS, Hotwani A,

Rahman N, Irfan S, Naga SR, Memon AM, Pradhan S, Iqbal K, Shrestha R, Rahman H,

Hasan MM, Qazi SH, Kazi AM, Saddal NS, Jamal R, Hunzai MJ, Hossain T, Marks F,

Carter AS, Seidman JC, Qamar FN, Saha SK, Andrews JR, Luby SP. 2022. Incidence of

typhoid and paratyphoid fever in Bangladesh, Nepal, and Pakistan: results of the

Surveillance for Enteric Fever in Asia Project. The Lancet Global Health 10:e978–e988.

2. Craigie J, Felix A. 1947. Typing of Typhoid Bacilli with Vi Bacteriophage. Suggestions

for its Standardisation. Lancet 823–7.

3. Felix A. 1955. World survey of typhoid and paratyphoid-B phage types. Bull World

Health Organ 13:109–170.

102     4.     Freytag B, Plochmann E. 1955. Phage Typing of Typhoid Strains. Zeitschrift fur

103         Hygiene und Infektionskrankheiten 142:188–96.

104     5.     Shrestha S, Silva KED, Shakya J, Yu AT, Katuwal N, Shrestha R, Shakya M, Shahi SB,

105         Naga SR, LeBoa C, Aiemjoy K, Bogoch II, Saha S, Tamrakar D, Andrews JR. 2024.

106         Detection of Salmonella Typhi bacteriophages in surface waters as a scalable approach

107         to environmental surveillance. PLOS Neglected Tropical Diseases 18:e0011912.

108     6.     Hooda Y, Islam S, Kabiraj R, Rahman H, Sarkar H, Silva KE da, Raju RS, Luby SP,

109         Andrews JR, Saha SK, Saha S. 2024. Old tools, new applications: Use of environmental

110         bacteriophages for typhoid surveillance and evaluating vaccine impact. PLOS Neglected

111         Tropical Diseases 18:e0011822.

112     7.     Islam S, Kabiraj R, Amin A, Pranto SH, Dey S, Malaker AR, Keya DP, Tanmoy AM,

113         Saha SK, Hooda Y, Saha S. 2024. Protocol for Genomic DNA Extraction and

114         Sequencing Library Preparation from Phage Stock.

115     8.     Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome

116         assemblies from short and long sequencing reads. PLOS Computational Biology

117         13:e1005595.

118     9.     Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2.

119         Genome Biology 20:257.

120     10.     Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald P-J, Vreugde S. 2022. Pharokka: a

121         fast scalable bacteriophage annotation tool. Bioinformatics 39:btac776.

122     11.     Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many

123         protein sequences. Protein Sci 27:135–145.

124    12.  Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for

125          genome assemblies. Bioinformatics 29:1072–1075.

126    13.  Pickard D, Thomson NR, Baker S, Wain J, Pardo M, Goulding D, Hamlin N, Choudhary

127          J, Threfall J, Dougan G. 2008. Molecular Characterization of the Salmonella enterica

128          Serovar Typhi Vi-Typing Bacteriophage E1. Journal of Bacteriology 190:2580–2587.

129    14.  Pickard D, Toribio AL, Petty NK, van Tonder A, Yu L, Goulding D, Barrell B, Rance R,

130          Harris D, Wetter M, Wain J, Choudhary J, Thomson N, Dougan G. 2010. A Conserved

131          Acetyl Esterase Domain Targets Diverse Bacteriophages to the Vi Capsular Receptor of

132          Salmonella enterica Serovar Typhi. Journal of Bacteriology 192:5746–5754.

133    15.  Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for

134          phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.

135