Complete Genome Sequence of *Salmonella enterica* Bacteriophage PRF-SP1

Prasanna Mutusamy,ᵃ Sasireigga Jaya Jothi,ᵃ Su Yin Lee,ᵃ,ᵇ Bent Petersen,ᶜ,ᵃ Thomas Sicheritz-Ponten,ᶜ,ᵃ Martha R. J. Clokie,ᵈ Stella Loke,ᵉ Andrew Millard,ᵈ [iD]Sivachandran Parimannan,ᵃ,ᵉ [iD]Heera Rajandasᵃ,ᵉ

ᵃCentre of Excellence for Omics-Driven Computational Biodiscovery (COMBio), AIMST University, Kedah, Malaysia
ᵇDepartment of Biotechnology, Faculty of Applied Sciences, AIMST University, Semeling, Kedah, Malaysia
ᶜSection for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
ᵈDepartment of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom
ᵉDeakin Genomics Centre, School of Life and Environmental Sciences, Faculty of Science, Engineering and Built Environment, Deakin University, Victoria, Australia

ABSTRACT We characterized the complete genome sequence of the lytic *Salmonella enterica* bacteriophage PRF-SP1, isolated from Penang National Park, a conserved rainforest in northern Malaysia. The novel phage species from the *Autographiviridae* family has a 39,966-bp double-stranded DNA (dsDNA) genome containing 49 protein-encoding genes and shares 90.96% similarity with *Escherichia* phage DY1.

Members of the genus *Salmonella* are common foodborne bacteria worldwide that cause infections requiring antibiotic treatments (1). Lately, however, the emergence of multidrug-resistant (MDR) *Salmonella* strains has triggered interest in alternative treatments like phage therapy (2). To our knowledge, no prior reports on phage isolation from tropical rainforests are available. Considering the vast diversity of microenvironments they harbor, we anticipated the presence of diverse phages against bacterial pathogens. In line with this, we successfully isolated and sequenced a novel species of *Salmonella* phage, PRF-SP1, from a dry sandy soil sample (GPS coordinate, 5.4620°N, 100.1900°E) obtained 1 to 3 cm deep in a rainforest in Penang. The complete genome sequence of the phage is reported here.

Bacteriophage PRF-SP1 was isolated using the enrichment method (3) with *Salmonella enterica* Paratyphi A as its host, and it formed clear, circular plaques. The phage was then propagated to a high titer using the double overlay agar technique (4). The phage morphology was visualized using 1% (wt/vol) uranyl acetate and observed under a transmission electron microscope (TEM) at 40 kV (Fig. 1a). It possesses an isometric head (diameter, 58 ± 2.5 nm) and a cone-shaped tail stub (length, 23 ± 2.2 nm).

Phage DNA was extracted using phenol-chloroform (5) and quantified using a Qubit fluorometer. The DNA was subjected to a Nextera DNA Flex library preparation kit and sequenced using the Illumina MiSeq platform, yielding 83,303 reads with 300-bp paired-end sequences. The raw reads were assessed using FastQC v0.11.9 (6), before trimming using Trimmomatic v0.39 (SLIDINGWINDOW:4:28 HEADCROP:10 CROP:200 MINLEN:200) (7). Trimming resulted in 30,791 reads with a mean length of 200 bp. The genome was assembled using SPAdes v3.15.3 with default settings, before assessment using QUAST v5.0.2 (8). The final length of the assembled genome was 39,966 bp, with a GC content of 50.26%. Output from Bowtie2 v2.4.4 (9) revealed that a total of 97.36% of reads mapped back to the genome, with an average coverage of 308×. A PHACTS (10) analysis showed that PRF-SP1 is a lytic phage, while ResFinder v4.1 and VirulenceFinder v2.0 confirmed that no lysogenic factors or antibiotic resistance genes were found in the genome (11).

The assembled genome was annotated using Prokka v1.12 (12), which predicted the presence of 49 protein-coding genes; 23 had putative functions, and 26 were hypothetical proteins, with no tRNAs. Further analysis using BLASTn indicated that the phage had <95%
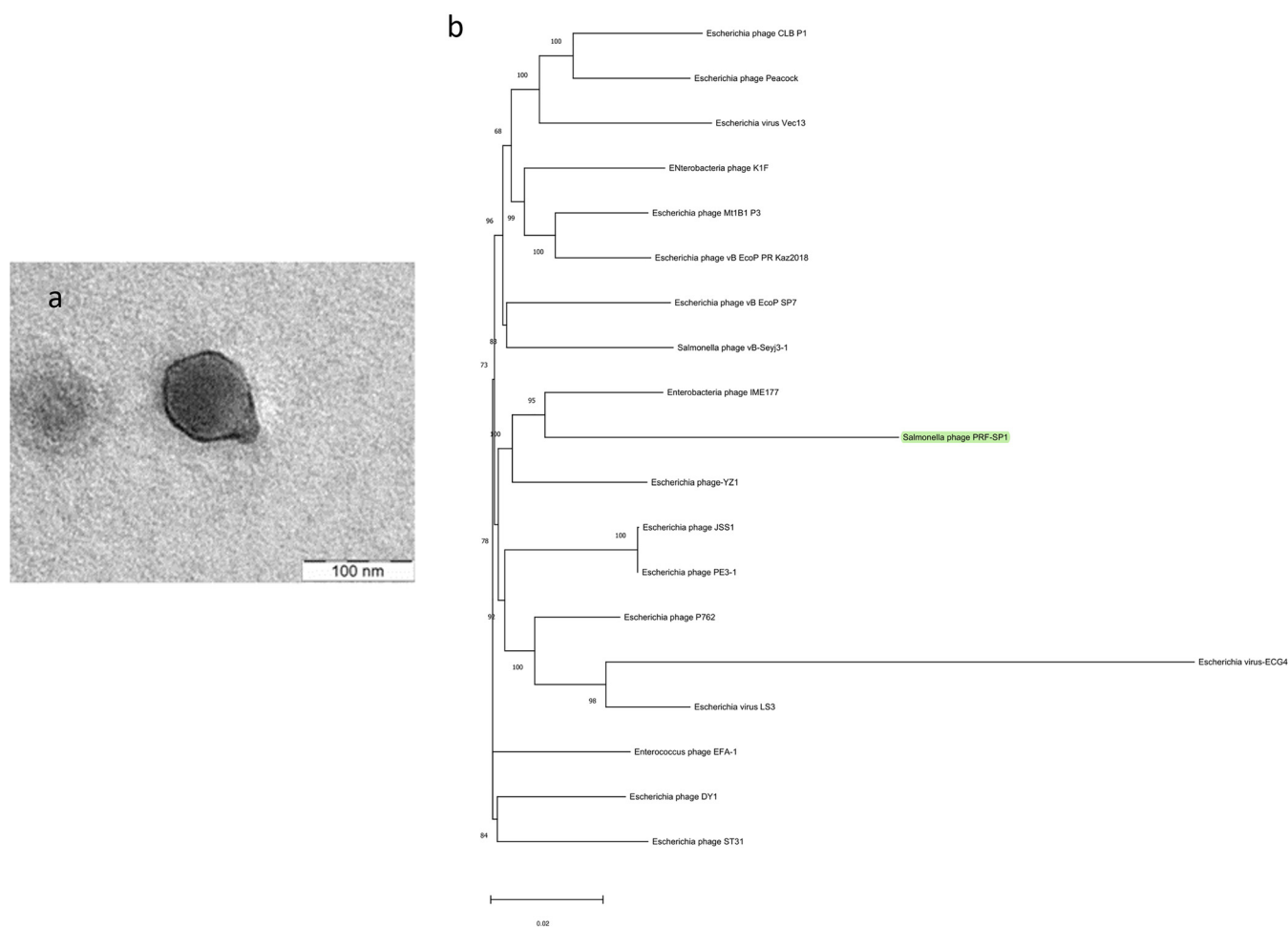
FIG 1 (a) Transmission electron microscopy image of phage PRF-SP1. The scale bar represents 100 nm. The size estimate of the virion was based on an average of 10 particles. (b) Phylogenetic analysis of *Salmonella* phage PRF-SP1. The top hits from BLASTn were used to identify related phages. The phylogeny was based on aligned core genes identified using Roary, with trees built using the neighbor-joining method with FastTree v2.1 (15). The bootstrap value was replicated 1,000 times (parameters, -boot 1000 -noml -nt), and the FastTree output was then visualized using MEGA v11.0.8. All phages identified were classified into the family *Autographiviridae*, subfamily *Studiervirinae*, and genus *Kayfunavirus*. *Salmonella* phage PRF-SP1 is highlighted in green.

average nucleotide identity with all other phages reported in the NCBI nucleotide (nt) database (top hit of ∼91% identity and 80% coverage with *Escherichia* phage DY1) and thus is representative of a new species based on the current standards (13). Further phylogenetic analysis was carried out with related phages which were identified by BLASTn through a fast core-gene alignment with MAFFT (used with -e and -n) using Roary v3.13.0 (14). As shown in Fig. 1b, the closest relatives of PRF-SP1 are *Enterobacteria* phage IME177 and *Escherichia* phage YZ1. All the identified related phages are part of the genus *Kayfunavirus*. The combination of genomic analysis, the presence of a gene encoding RNA polymerase, and phylogenetic analysis places PRF-SP1 as a new species in the family *Autographiviridae*, subfamily *Studiervirinae*, and genus *Kayfunavirus*. Based on this, we propose the new species "Kayfunavirus combio."

Data availability. The complete genome sequence of phage PRF-SP1 has been deposited in the GenBank database under the accession number MZ923531. The associated BioProject, SRA, and BioSample accession numbers are PRJNA760259, SRR15809652, and SAMN21357400, respectively.

REFERENCES

1. Cianflone NFC. 2008. Salmonellosis and the GI tract: more than just peanut butter. Curr Gastroenterol Rep 10:424–431. https://doi.org/10.1007/s11894-008-0079-7.
2. Sulakvelidze A, Alavidze Z, Morris JG. 2001. Bacteriophage therapy. Antimicrob Agents Chemother 45:649–659. https://doi.org/10.1128/AAC.45.3.649-659.2001.
3. Twest R, Kropinski AM. 2009. Bacteriophage enrichment from water and soil, p 15–21. In Clokie MRJ, Kropinski AM (ed), Bacteriophages. Methods in molecular biology, vol 501. Humana Press, Totowa, NJ. http://link.springer.com/10.1007/978-1-60327-164-6_2.
4. Jakočiūnė D, Moodley A. 2018. A rapid bacteriophage DNA extraction method. Methods Protoc 1:27. https://doi.org/10.3390/mps1030027.
5. Hyman P, Abedon ST. 2009. Practical methods for determining phage growth parameters. Methods Mol Biol 501:175–202. https://doi.org/10.1007/978-1-60327-164-6_18.
6. Trivedi UH, Cézard T, Bridgett S, Montazam A, Nichols J, Blaxter M, Gharbi K. 2014. Quality control of next-generation sequencing data without a reference. Front Genet 5:111. https://doi.org/10.3389/fgene.2014.00111.
7. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
8. 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.
9. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
10. McNair K, Bailey BA, Edwards RA. 2012. PHACTS, a computational approach to classifying the lifestyle of phages. Bioinformatics 28:614–618. https://doi.org/10.1093/bioinformatics/bts014.
11. Kleinheinz KA, Joensen KG, Larsen MV. 2014. Applying the ResFinder and VirulenceFinder Web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4:e27943. https://doi.org/10.4161/bact.27943.
12. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
13. Turner D, Kropinski AM, Adriaenssens EM. 2021. A roadmap for genome-based phage taxonomy. Viruses 13:506. https://doi.org/10.3390/v13030506.
14. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421.
15. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.