Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Complete Genome Sequence of *Salmonella*Phage vB_SenS_Horemheb

Bishoy Maher Zaki
MARC for Medical Services and Scientific Research
Hadeer Ghonim
MARC for Medical Services and Scientific Research
Mark G. Emil
MARC for Medical Services and Scientific Research
Sohaila A. Al-Abraq
MARC for Medical Services and Scientific Research
Suha A. Farraj
MARC for Medical Services and Scientific Research
Ramy Karam Aziz
ramy.aziz@pharma.cu.edu.eg

MARC for Medical Services and Scientific Research

Data Note

Keywords:

Posted Date: July 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10182402/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Foodborne salmonellosis remains a public health concern, especially in children and immunocompromised patients. Here, we report the isolation and genome analysis of *Salmonella* phage vB_SenS_Horemheb, isolated from wastewater collected in Cairo, Egypt. Illumina platform was used to sequence the phage genome, and the generated reads were assembled into a single 43,278 bp contig with 283x coverage and a GC content of 49.89%. Genome annotation predicted 62 coding sequences, with no tRNA genes or genes associated with lysogeny, toxin production, or antimicrobial resistance. Manual curation improved the automated annotation by identifying two missed coding sequences and adjusting the start codon of a gene encoding an Rz-like spanin. The genome encodes capsid- and tail-associated proteins related to siphovirus-like phages and a predicted lysis system with separated spanin– and holin/endolysin–encoding regions. The classification program, TaxMyPhage, assigned vB_SenS_Horemheb as a member of a new species within the genus *Jerseyvirus*, family *Sarkviridae*.

Full Text

Genome Report:

Diseases caused by foodborne pathogens are a growing public health threat, affecting nearly one tenth of the global population each year. *Salmonella* speciesare major causative agents of foodborne diseases and are also among the top four causes of diarrheal diseases [1]. *Salmonella* is a genus of Gram-negative, rod-shaped, non-lactose fermenting bacteria in the family Enterobacteriaceae. *S. enterica* and *S. bongori* are the two main species of *Salmonella*. These species include thousands of serovars, and within *S. enterica,* several serovars can infect both animals and humans [2].

Salmonella species commonly colonize domestic and wild animals, and they are widespread in food animals including cattle, poultry, and pigs. *Salmonella* species can pass through the entire food chain and are transmitted to humans through contaminated food. Children and immunocompromised patients are among the vulnerable groups who may be severely affected by these infections, with higher morbidity and mortality rates reported in developing countries [3]. Control measures are therefore needed to reduce diarrheal diseases caused by *Salmonella*. In this effort, we isolated a new phage against *Salmonella enterica*.

Salmonella phage vB_SenS_Horemheb was isolated from a wastewater-treatment plant in Al-Sayeda Zeinab district in Cairo, Egypt. The wastewater was clarified by centrifugation (6,797 ×g/10 min) and filtration through a 0.45 µm polyethersulfone (PES) syringe filter. For phage enrichment, equal volumes of the clarified wastewater and exponentially growing *Salmonella enterica* Serovar Typhimurium(ATCC 14028) were incubated overnight. Next, the enriched phage lysate was centrifuged and filtered through a 0.22 µm PES syringe filter, then spotted onto a lawn of the exponentially growing bacterial host in 0.3% soft agar. After incubation, clear plaques were observed, indicating the presence of phages [4]. Agar plugs containing plaques were picked and incubated overnight in 1 mL of sterile broth at 4 °C to allow

phage elution. To confirm plaque formation, the elute was serially diluted 10-fold and tested by plaque assay . The phage plaques were purified by four rounds of single-plaque isolation [5].

The phage was amplified initially in a small-volume culture. Briefly, 100 µL of the phage lysate was incubated with 2 mL of exponentially growing bacterial host culture. Then, the amplified phage lysate was enumerated by spot titration. The undiluted lysate was estimated at 10^8 PFU/mL and six 20 µL aliquots were spotted onto a lawn of bacterial host to produce high-titer plaques. Agar plugs containing these plaques were excised, added to 45 mL of fresh host culture, and incubated overnight. Next, the phage lysate was filtered and serially diluted 10-fold for enumeration by spot titration, which quantified the final phage titer at 10^{10} PFU/mL. GeneJET™ Whole Blood Genomic DNA Purification Mini Kit was used to extract the phage DNA from 1 mL aliquot of the phage stock.

A short-read sequencing library (paired-end, 150 bp) was prepared with the Illumina DNA Prep Kit and sequenced on an Illumina NextSeq 2000. Illumina DRAGEN was used for adapter trimming. On the BV-BRC server [6], the obtained reads passed the quality assessment by FastQC (v0.12.1) and quality-trimmed by Trim Galore (v0.6.5). The reads were *de novo* assembled by Unicycler (v0.4.8) into a single contig of size 43,278 bp with coverage of 283× and a GC content of 49.89%. Then, RASTtk was used to annotate the phage DNA contig, after the pipeline was customized to prioritize phage protein annotation [7, 8], followed by manual verification and curation on UGENE v52 [9]. Sixty coding sequences were predicted by Prodigal on RASTtk. After manual verification, two more CDSs (7 and 9) were identified, and a new start codon was assigned to CDS 8. All start codons were ATG, except for seven CDSs that used alternative codons (based on translation table 11): GTG, TTG, and ATT.

All predicted coding sequences (62 CDSs) were further analyzed to verify functional assignment by NCBI BLASTp [10], UniProt BLAST [11], InterProScan [12], and HHpred [13]. No tRNA genes or genes associated with lysogeny, toxigenicity, or antimicrobial resistance were identified. The phage genome was oriented to begin with the terminase-encoding genes. Proksee web-based tool was used to visualize the phage genome annotation (**Figure 1**).

The predicted virion (capsid and tail) proteins, particularly the major tail protein (CDS 21), tail assembly chaperones (CDSs 27 and 28), and tail length tape measure protein (CDS 29) had homologs in other siphoviruses, particularly phages infecting Gram-negative bacteria. Interestingly, the predicted host lysis-associated genes were not located in a single cassette. The spanin-encoding genes (CDSs 8 and 9, Rz- and Rz1-like, respectively) overlapped (**Figure 2A**) and were located at region of 6,860-7,373, while the two holin- encoding genes (CDSs 54 and 55) and endolysin- encoding gene (CDS 56) were located elsewhere (40,845-41,884, **Figure 3A**). The amino acid sequences of the two-component spanin system were analyzed by Phobius [14], DeepTMHMM [15], and SignalP6 [16] to identify potential transmembrane domains and signal peptides. These spanins had features of inner (i-) and outer (o-) spanins. The i-spanin contained an N-terminal transmembrane domain followed by a periplasmic coiled-coil region (**Figure 2B, C**), whereas the o-spanin had an N-terminal signal peptide and a lipobox cysteine motif

(**Figure 2D**) [17]. The holin sequences were also analyzed by DeepTMHMM, which identified three transmembrane domains in each protein, consistent with the topology of class I holins (**Figure 3B**) [18].

It is worth noting that manual verification enhanced the automated genome annotation, even at the gene-finding level. For example, manual verification allowed the adjustment of the start codon of CDS 8. Shifting the start of that open reading frame (ORF) by two codons upstream within the same reading frame resulted in a more confident transmembrane topology consistent with the Rz-like (i-) spanin structure (**Figure 2B-C**). UGENE ORF finder was also used to find missing CDS (i.e., gap filling), guided by the clinker genome alignment [19] and the comparison of phage vB_SenS_Horemheb gene clusters with closely related RefSeq phage genomes (**Figure 4**). Two other CDS were missed by automated gene finders: one encodes for a hypothetical protein (CDS 7) and the other encodes for an Rz1-like (o-) spanin (CDS 9). In addition, InterPro Scan analysis identified intein domains in two CDSs (5 and 38)-encoded proteins (**Figure 5**).

Finally, for phage taxonomy, TaxMyPhage [20] assigned vB_SenS_Horemheb as a new species within the genus *Jerseyvirus*, family *Sarkviridae*. With the potential discovery of more members of that genus, this taxonomic placement will be further confirmed.

Two different annotated versions of vB_SenS_Horemheb genome, RASTtk-generated and manually curated, are compared to one another and to the phage's closest RefSeq relatives (identified by BLASTn with a coverage threshold > 85% and identity threshold > 90%). Arrows indicate predicted CDSs and their direction of transcription. Homologous genes (defined with a clustering threshold of 0.3) have matching colors. Genes with higher sequence similarity (threshold = 0.95) are connected with shaded surfaces.

Declarations

Acknowledgment:

We thank Prof. Andrew Kropinski for his support and guidance during phage genome annotation.

Competing interests: The authors declare no competing interests.

Funding: No funding was received for this work.

Ethics approval and consent to participate: Not applicable. This study did not involve human participants, human data, human samples, or animal experiments.

Use of generative AI: We used ChatGPT with OpenAI's GPT-5.5 Thinking model option for paragraph-by-paragraph spelling, grammar, and language-clarity checks, but not to generate text from scratch. We carefully reviewed and approved all edits suggested by the model.

Data availability:

The genomic sequence of *Salmonella* phage vB_SenS_Horemheb was deposited in the NCBI nucleotide database (GenBank) under accession number PZ583544 and the raw reads under Sequence Read Archive (SRA) accession number SRR38004690.

Author Contribution

B.M.Z. isolated the phage, performed all bioinformatics analyses, prepared the figures, and drafted the manuscript. H.G. extracted the phage DNA. S.A.A. and M.G.E. performed the sequencing workflow. S.A.F. supervised the sequencing process and reviewed the manuscript. R.K.A. supervised the entire work and critically reviewed the manuscript. All authors read and approved the final manuscript.

References

1. **Salmonella (non-typhoidal)** [https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal)]
2. Giannella RA: **Salmonella**. In: *Medical Microbiology.* Edited by S. B, 4th edn. Galveston (TX): University of Texas Medical Branch at Galveston; 1996.
3. Zizza A, Fallucca A, Guido M, Restivo V, Roveta M, Trucchi C: **Foodborne Infections and Salmonella: Current Primary Prevention Tools and Future Perspectives**. *Vaccines (Basel)* 2024, **13**(1).
4. Twest RV, Kropinski AM: **Bacteriophage enrichment from water and soil**. In: *Bacteriophages.* Springer; 2009: 15-21.
5. Carlson K: **Working with bacteriophages: common techniques and methodological approaches**, vol. 1: CRC press Boca Raton, FL; 2005.
6. Shukla M, Wattam AR, Aleman A, Bhattacharya R, Bowers N, Brettin T, Capria A, Chia N, Cucinell C, Davis JJ *et al*: **BV-BRC: a unified bacterial and viral bioinformatics resource with expanded functionality and AI integration**. *Nucleic Acids Res* 2026, **54**(D1):D715-d723.
7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al*: **The RAST Server: rapid annotations using subsystems technology**. *BMC Genomics* 2008, **9**:75.
8. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD *et al*: **RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes**. *Sci Rep* 2015, **5**:8365.
9. Okonechnikov K, Golosova O, Fursov M: **Unipro UGENE: a unified bioinformatics toolkit**. *Bioinformatics* 2012, **28**(8):1166-1167.
10. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**(3):403-410.
11. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I: **UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to**

use the entry view. In: *Plant bioinformatics.* Edited by Edwards D, vol. 1374: Humana Press, New York, NY; 2016: 23-54.

12. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L *et al*: **InterPro: the integrative protein signature database**. *Nucleic Acids Res* 2009, **37**(Database issue):D211-215.
13. Söding J, Biegert A, Lupas AN: **The HHpred interactive server for protein homology detection and structure prediction**. *Nucleic Acids Res* 2005, **33**(Web Server issue):W244-248.
14. Kall L, Krogh A, Sonnhammer EL: **A combined transmembrane topology and signal peptide prediction method**. *J Mol Biol* 2004, **338**(5):1027-1036.
15. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, Krogh A, Winther O: **DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks**. *bioRxiv* 2022.
16. Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H: **SignalP 6.0 predicts all five types of signal peptides using protein language models**. *Nat Biotechnol* 2022, **40**(7):1023-1025.
17. Kongari R, Rajaure M, Cahill J, Rasche E, Mijalis E, Berry J, Young R: **Phage spanins: diversity, topological dynamics and gene convergence**. *BMC Bioinformatics* 2018, **19**(1):326.
18. Reddy BL, Saier MH, Jr.: **Topological and phylogenetic analyses of bacterial holin families and superfamilies**. *Biochim Biophys Acta* 2013, **1828**(11):2654-2671.
19. Gilchrist CLM, Chooi YH: **clinker & clustermap.js: automatic generation of gene cluster comparison figures**. *Bioinformatics* 2021, **37**(16):2473-2475.
20. Millard A, Denise R, Lestido M, Nicholas MT, Webster D, Turner D, Sicheritz-Pontén T: **taxMyPhage: Automated taxonomy of dsDNA phage genomes at the genus and species level**. *Phage (New Rochelle)* 2025, **6**(1):5-11.

Figures

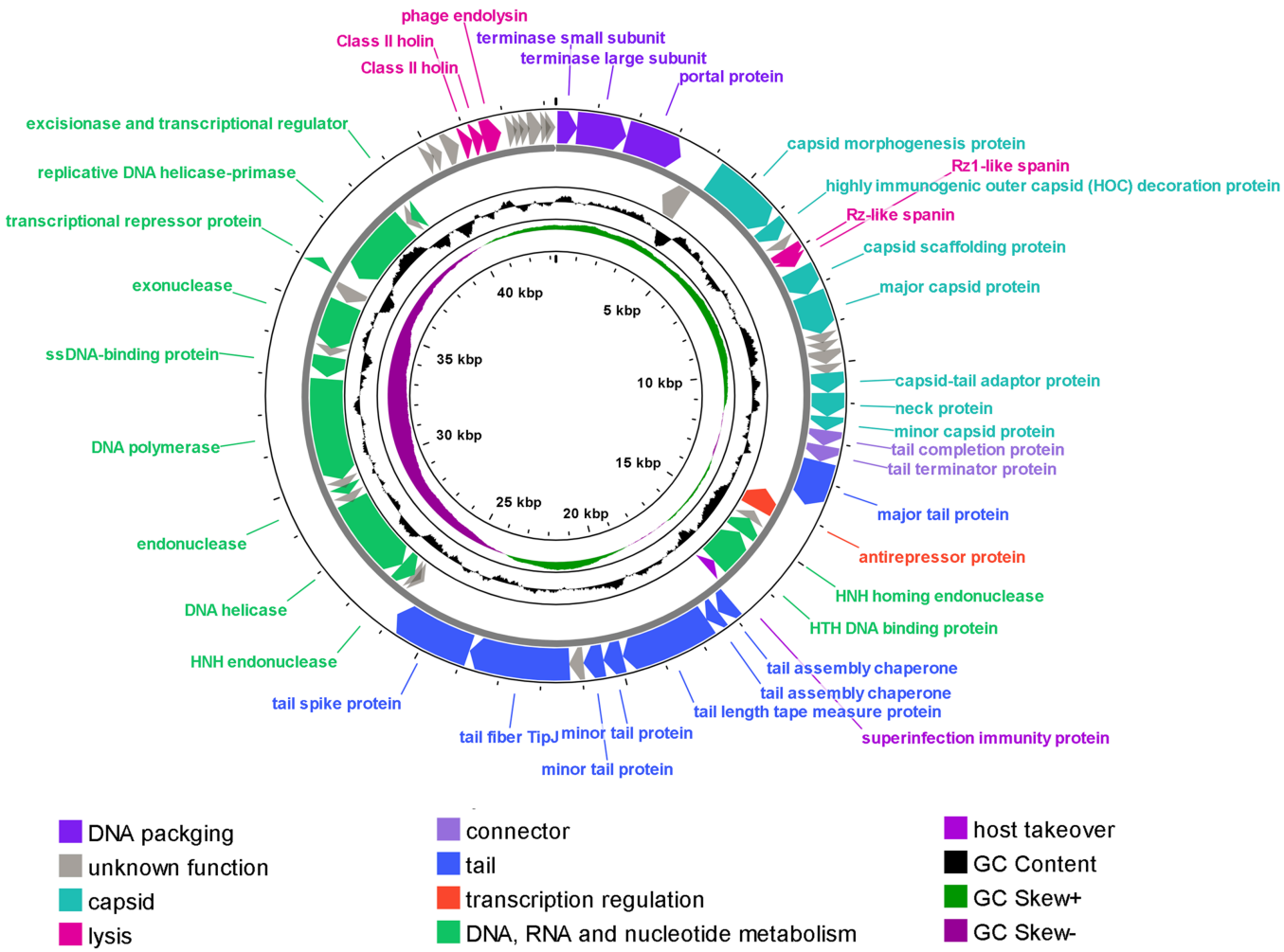


Figure 1

Genome map of *Salmonella* phage vB_SenS_Horemheb.

Proksee-generated circular genome map of *Salmonella* phage vB_SenS_Horemheb. Predicted coding sequences are represented on the outer ring and categorized according to PHROGs functional categories. The middle and inner rings represent the genome's GC content (baseline, 0.5) and GC skew (baseline, 0), respectively.

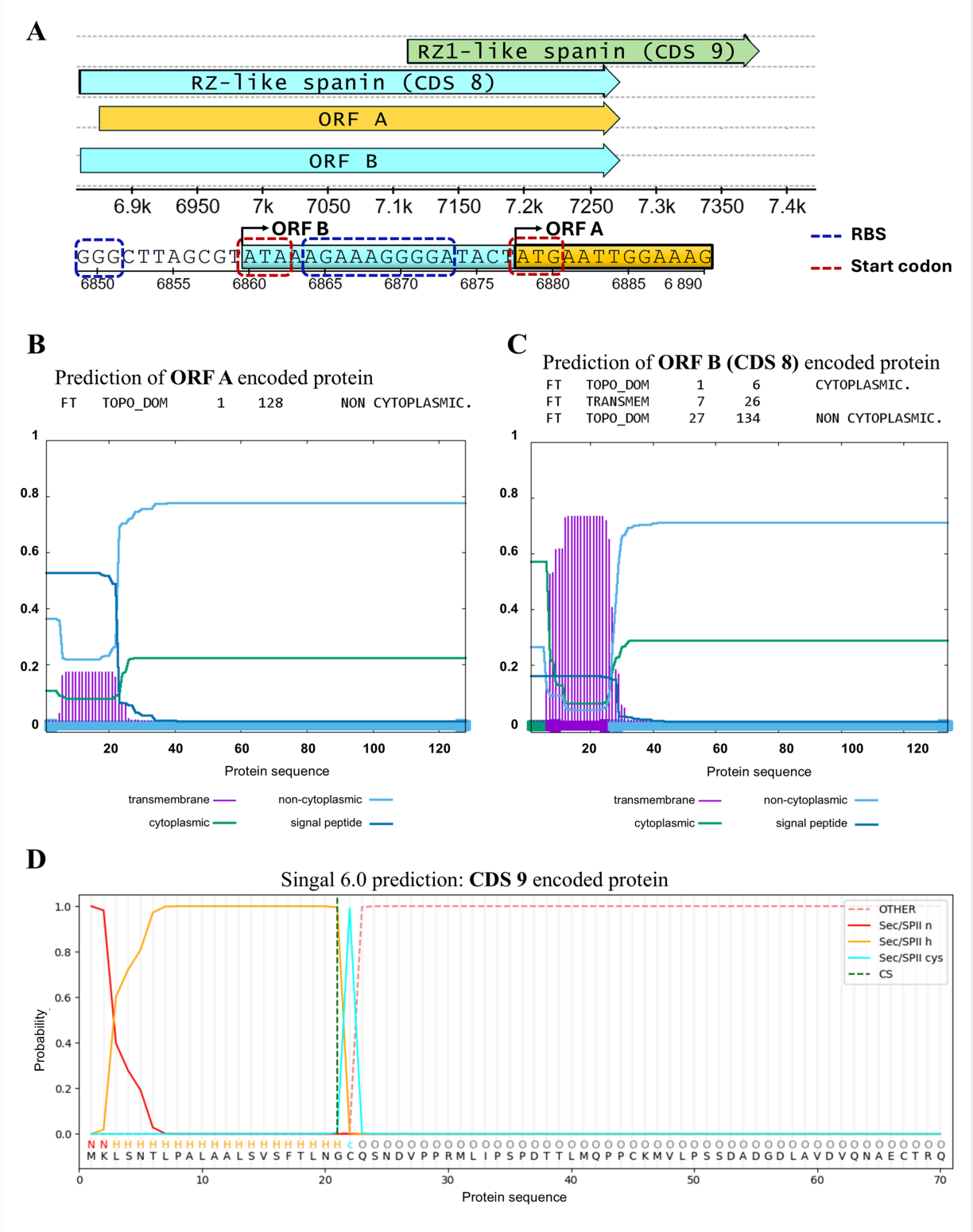


Figure 2

Genomic organization and predicted topology of the spanin-encoding region of *Salmonella*phage vB_SenS_Horemheb.

(A) Genomic organization of the predicted spanin region visualized on UGENE. The nucleotide sequence below the map indicates the ribosome-binding sites (RBS) and start codons of ORF A, which was

predicted by automated annotation, and ORF B, which was manually identified as an alternative candidate ORF for CDS 8. (B, C)Phobius-based topology predictions of the encoded proteins from the candidate ORFs (A and B), respectively. The probability of membrane localization is plotted on the y-axis, while the amino acid positions are plotted on the x-axis=. (D) SignalP 6 prediction of the CDS 9-encoded protein, in which each type of signal peptides is plotted with a different color. Abbreviations: Sec/SPII, Sec-dependent lipoprotein signal peptide cleaved by signal peptidase II; n, N-terminal region; h, hydrophobic region; cys, conserved cysteine/lipobox region; CS, cleavage site; OTHER, region not assigned to a signal peptide.

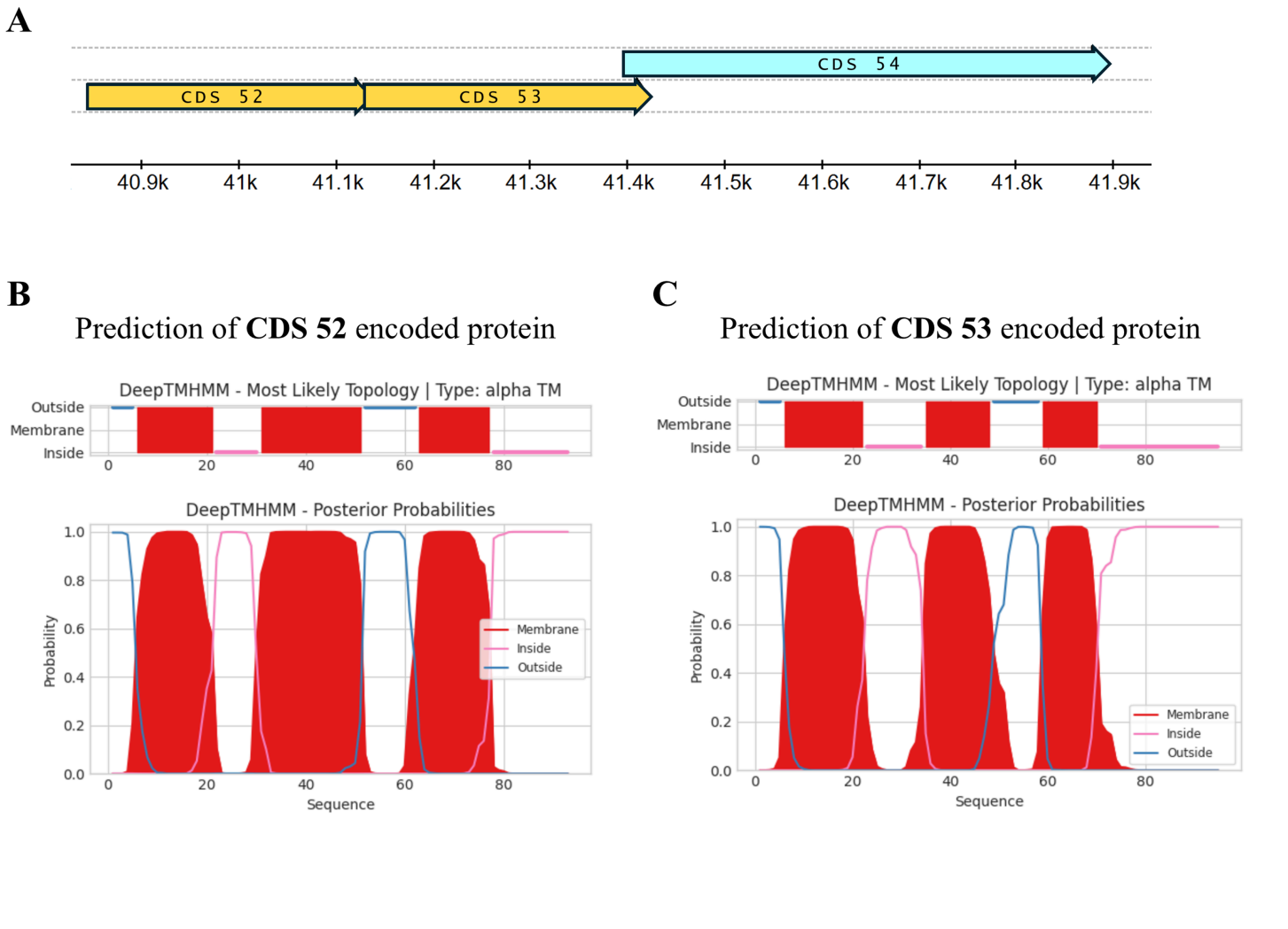


Figure 3

Genomic organization of holin–endolysin encodingcassette and the predicted transmembrane topology of the identified holins.

(A) Genomic organization of the predicted lysis cassette visualized was using UGENE, and it contained two holin-encoding genes, CDS 52 and CDS 53, followed by the endolysin-encoding gene, CDS 54. **(B, C)** DeepTMHMM topology plots for the proteins encoded by CDS 52 and CDS 53, respectively. The upper

plots show the most likely membrane topology, and the lower plots show posterior probabilities across the amino acid sequence. The y-axis indicates the prediction probability, and the x-axis indicates amino acid positions.

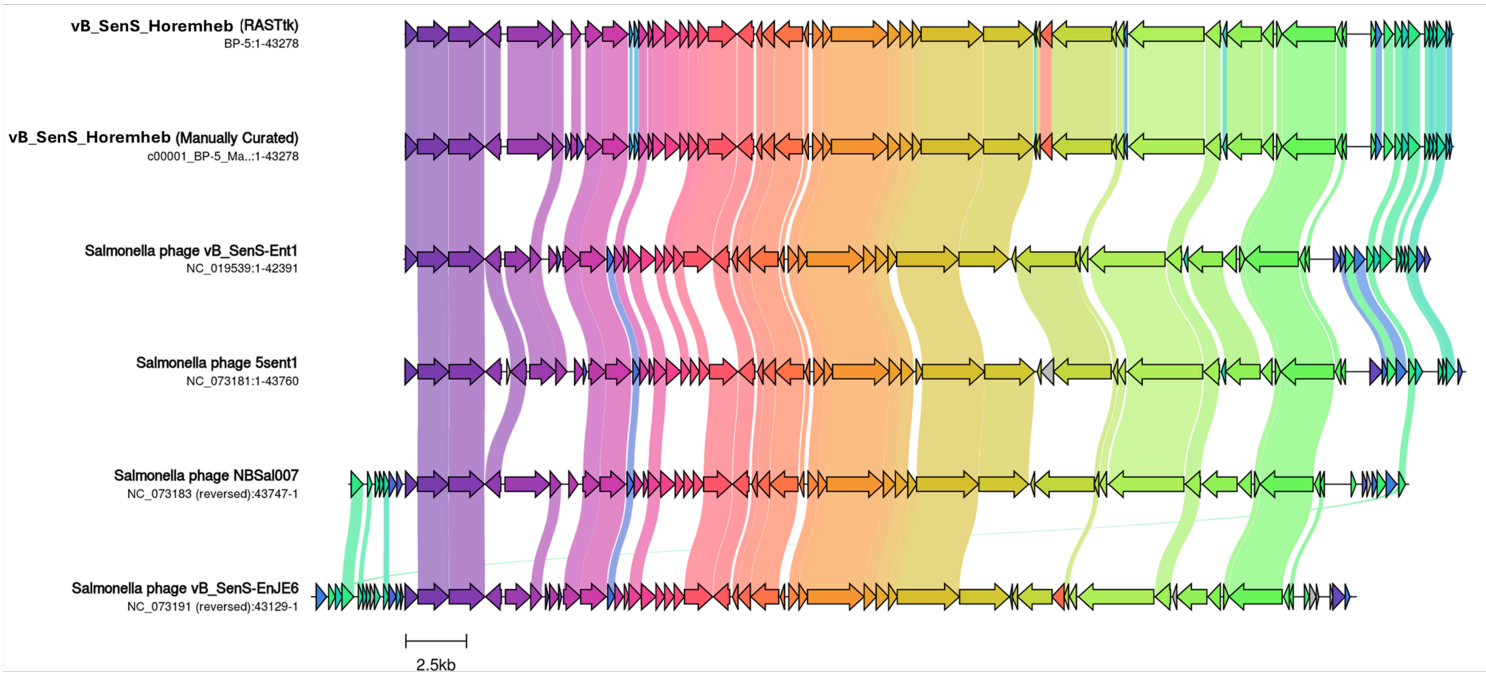


Figure 4

Clinker gene cluster comparison of *Salmonea* phage vB_SenS_Horemheb and closely related RefSeq phages.

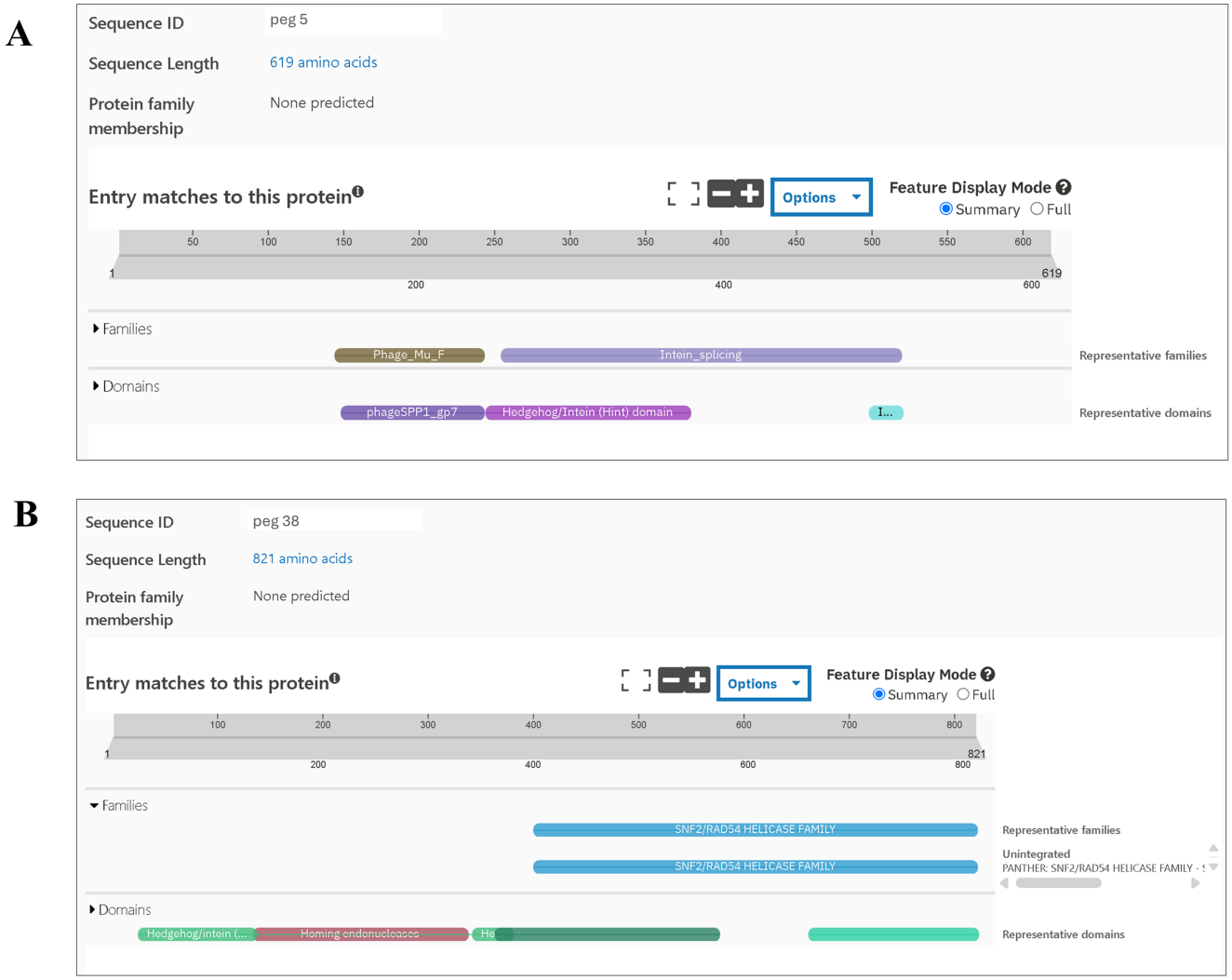


Figure 5

InterProScan domain predictions of proteins containing intein-related features in *Salmonella*phage vB_SenS_Horemheb.

(A, B) InterProScan graphical outputs for CDS 5 and CDS 38 encoded proteins, respectively.