
Supplementary Information

Prophages in Aquatic Environments: From Silent Passengers to Keystone Regulators of Biogeochemical Cycles, Biofilm Dynamics, and Antibiotic Resistance Dissemination

(Running title: Aquatic Prophages: Drivers of Biogeochemistry, Biofilms, Antibiotic Resistance)

Ling Li ^a, Siyu Huang ^a, Litao Wei ^a, Xinyi Li ^a, Junya Zhang ^b, Xiang Li ^c, Haining Huang ^{a, *}, Yinguang Chen ^{a, *}

^a State Key Laboratory of Water Pollution Control and Green Resource Recycling, School of Environmental Science and Engineering, Tongji University, 1239 Siping Road, Shanghai 200092, China

^b State Key Joint Laboratory of Environmental Simulation and Pollution Control, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

^c College of Environmental Science and Engineering, Donghua University, Songjiang, Shanghai 201620, China

* Corresponding author

Tel.: +86 21 65981263; fax: +86 21 65986313

E-mail address: Haining Huang: haininghuang@tongji.edu.cn

Yinguang Chen: yinchentj@163.com; yinguangchen@tongji.edu.cn

Journal: *Frontiers of Environmental Science & Engineering*

Number of pages: 3

Number of tables: 1

Table S1. Summary of Bioinformatics Identification Methods for Prophages

Method name	Characteristic	Detect the target substance	Function	Limit	References
CheckV	Tools for evaluating the quality of viral genomes	Bacterial or viral gene sequences	Integrity threshold of high-quality prophages $\geq 90\%$	It is impossible to directly detect prophages.	(Nayfach et al., 2021)
PhageScope	Phage similarity alignment and annotation	Bacterial genomes or metagenomic contigs	Only provides basic homology-matching analysis	Unable to customize the reference sequence.	(Wang et al., 2024)
MARVEL	Predict prophage regions in metagenomic data	Contigs for metagenomic assembly of bacterial or prophage genes	Prophage region prediction tool specialized for metagenomic data	The prediction effect of a single bacterial genome is poor.	(Amgarten et al., 2018)
ProphageHunter	Predicting the insertion sites and evolution of prophages helps to understand the relationship between prophages and their hosts.	Bacterial genome sequence	Unique prophage activity assessment; supports insertion site prediction, one-stop annotation & novel phage mining	The speed is slow and it is not suitable for large-scale genomic analysis.	(Song et al., 2019)
PHASTER	Predict prophages based on BLAST and HMMER alignment with phage databases, and provide automated annotation and visualization.	Bacterial genome sequence	PPV = 91% (default parameters); provides automatic annotation & visualization	Only the binary results of "yes/no" for prophages.	(Arndt et al., 2016)
PhiSpy	Use machine learning combined with sequence features to identify prophage regions, especially good at detecting highly repetitive or partially degenerate prophages.	Bacterial genome sequence.	Strong detection for highly repetitive/partially degraded prophages Identifies prophage regions via machine learning + sequence features	Unable to distinguish between lysogenic phages and horizontally transferred viral elements	(Akhter et al., 2012)
Phage Finder	A Perl-based tool, which uses BLAST to compare with the phage protein database and combines sequence characteristics to predict prophage regions. .	Complete bacterial genome sequence	Predicts prophage regions via BLAST (phage protein database) + sequence features	Insensitive to the characteristics of novel phages; fails to detect prophages without integrase..	(Fouts, 2006)
VIBRANT	Identify viral sequences from metagenomes and distinguish between lysogenic/lytic phases based on protein features and neural networks	Assembled contigs	Distinguishes lysogenic/lytic phages; identifies viral sequences in metagenomes	Detection of prophages without typical phage genes failed.	(Kieft et al., 2020)

RaFAH	Infer the potential host range by comparing the phage genome with the host CRISPR array, protospacer sequence, and homologous regions.	Prophage region sequence	Function: Infers prophage host range	Only applicable to annotated genomes	(Coutinho et al., 2021)
DEPhT	Tools based on deep learning to predict the boundaries of prophages.	Short-read metagenomic data of bacteria or prophages	Specialized deep learning tool; only predicts prophage boundaries accurately	A complete host genome sequence is required.	(Gauthier et al., 2022)
ProPhinder	Adopt machine learning to integrate sequence composition, structural features (gene density), and homology alignment, which is applicable to complex microbial community data.	Bacterial or prophage gene sequences	ML tool (integrates sequence composition, structural features [gene density] & homology alignment); adapted for complex microbial community data	The algorithm has a high complexity; the prediction error is large.	(Lima-Mendez et al., 2008)

Reference:

- Akhter S, Aziz R K, Edwards R A (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Research*, 40(16): e126
- Amgarten D, Braga L P P, Da Silva A M, Setubal J C (2018). MARVEL, a tool for prediction of bacteriophage sequences in metagenomic bins. *Frontiers in Genetics*, 9: 304–322
- Arndt D, Grant J R, Marcu A, Sajed T, Pon A, Liang Y, Wishart D S (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, 44(W1): W16–W21
- Coutinho F H, Zaragoza-Solas A, López-Pérez M, Barylski J, Zieleszinski A, Dutilh B E, Edwards R, Rodriguez-Valera F (2021). RaFAH: host prediction for viruses of bacteria and archaea based on protein content. *Patterns*, 2(7): 100274
- Fouts D E (2006). Phage Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. *Nucleic Acids Research*, 34(20): 5839–5851
- Gauthier C H, Abad L, Venbakkam A K, Malnak J, Russell D A, Hatfull G F (2022). DEPhT: a novel approach for efficient prophage discovery and precise extraction. *Nucleic Acids Research*, 50(13): e75
- Kieft K, Zhou Z, Anantharaman K (2020). VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*, 8(1): 90
- Lima-Mendez G, Van Helden J, Toussaint A, Leplae R (2008). Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics*, 24(6): 863–865
- Nayfach S, Camargo A P, Schulz F, Eloie-Fadrosch E, Roux S, Kyrpides N C (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, 39(5): 578–585
- Song W, Sun H-X, Zhang C, Cheng L, Peng Y, Deng Z, Wang D, Wang Y, Hu M, Liu W, Yang H, Shen Y, Li J, You L, Xiao M (2019). Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Research*, 47(W1): W74–W80
- Wang R H, Yang S, Liu Z, Zhang Y, Wang X, Xu Z, Wang J, Li S C (2024). PhageScope: a well-annotated bacteriophage database with automatic analyses and visualizations. *Nucleic Acids Research*, 52(D1): D756–D761