eMicrobe
https://www.sciltp.com/journals/emicrobe

Communication

Genome Wide Analysis Reveals Diverse and Novel Prophages in *Tsukamurella*

Jordan Yik Hei Fong [1], Jie Liang [2,3], Lilong Jia [1,2,3], Patrick C.Y. Woo [4,5], Man Lung Yeung [2,3,6,7,*] and Jade Lee Lee Teng [1,*]

[1] Faculty of Dentistry, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[2] Department of Microbiology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[3] InnoHK (Centre for Virology, Vaccinology and Therapeutics), Hong Kong Science and Technology Park, Hong Kong SAR, China
[4] Doctoral Program in Translational Medicine and Department of Life Sciences, National Chung Hsing University, Taichung 402, Taiwan
[5] The iEGG and Animal Biotechnology Research Center, National Chung Hsing University, Taichung 402, Taiwan
[6] State Key Laboratory of Emerging Infectious Diseases, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
[7] Carol Yu Centre for Infection, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
* Correspondence: llteng@hku.hk (J.L.L.T.); pmlyeung@hku.hk (M.L.Y.)

How To Cite: Fong, J.Y.H.; Liang, J.; Jia, L.; et al. Genome Wide Analysis Reveals Diverse and Novel Prophages in *Tsukamurella*. *eMicrobe* **2026**, *2*(2), 10. https://doi.org/10.53941/emicrobe.2026.100010

Received: 31 December 2025
Revised: 22 May 2026
Accepted: 10 June 2026
Published: 22 June 2026

Abstract: Prophages play key roles in bacterial evolution, yet their diversity in the genus *Tsukamurella* remains poorly characterized. In this study, we analyzed 55 publicly available *Tsukamurella* genomes and identified eight high-quality prophages, including seven high-quality draft genomes and one complete circular genome, distributed among *T. asaccharolytica* (n = 1), *T. paurometabola* (n = 1), *T. tyrosinosolvens* (n = 2), *T. soli* (n = 1), and *T. ocularis* (n = 3). Comparative genomic analyses revealed three major clusters: the previously described *Tsukamurella* phages TIN2, TIN3, TIN4; the three *T. ocularis* prophages; and prophages from *T. paurometabola* and *T. tyrosinosolvens*. The remaining prophages showed limited similarities to known phages or prophages, highlighting their novelty. Despite substantial genomic divergence, key lysis-associated modules, particularly endolysin and holin families, were relatively conserved among phage and prophage genomes. These findings expand current understanding of *Tsukamurella* phage diversity and provide new insights into the evolutionary dynamics of actinobacteriophages.

Keywords: *Tsukamurella*; Actinomycetota, prophage diversity; phage evolution; genomic analysis

1. Introduction

The advent of high-throughput sequencing has dramatically accelerated bacterial genome analysis, enabling in-depth investigations into microbial diversity, evolution, and host-virus interactions [1]. Prophages, viral sequences integrated within bacterial chromosomes, are highly prevalent, with up to 75% of bacterial genomes containing prophage-derived sequences [2]. These prophages can influence bacterial physiology, virulence, antimicrobial resistance, and horizontal gene transfer, serving as indicators of past viral infections and contributors to genomic diversification, thereby driving growing interest in phage biology and evolution [3].

Despite their widespread biological significance, prophage diversity remains underexplored in many bacterial lineages. Systematic characterization of prophages across bacterial genomes is therefore essential for

understanding viral evolution and diversity [4]. Within Actinomycetota, an ecologically and clinically important bacterial phylum, investigations into phage research has primarily focused on *Mycobacterium,* driven by initiatives such as SEA-PHAGES that have generated over 3000 complete phage genomes, while recent large-scale meta-analyses of over 9000 mycobacterial genomes revealed 61 prophage clusters, including 18 singletons [5]. These studies have set a benchmark for understanding actinomycete viral diversity, yet they also highlight the vast "viral dark matter" residing in related, less-characterized genera.

Tsukamurella, a member of the Actinomycetota, with mycolic acid-rich cell walls similar to *Mycobacterium*, remains one such understudied genus. The genus occupies a unique position at the intersection of environmental and clinical microbiology. They are widely distributed in soils, activated sludge, arthropods, and are also increasingly recognized as opportunistic human pathogens causing bacteraemia, conjunctivitis, keratitis and catheter-related bloodstream infection [6–11]. Owing to their high prevalence in wastewater systems, to date phage research in this genus has been largely focusing on foaming control in wastewater treatment using a small handful of characterized isolates (TIN2, TIN3, TIN4, TPA2, and TPA4) [12–14]. However, the genomic diversity of prophages within clinically relevant *Tsukamurella* species remains almost entirely unknown.

Systematically characterizing the prophage landscape of *Tsukamurella* is critical for understanding the genus's pathobiology. In clinical isolates, prophages may harbor genes that enhance host fitness during infection, while in environmental strains, they likely mediate the exchange of metabolic genes within the wastewater microbiome. As our research group has recently identified and characterized multiple novel *Tsukamurella* species associated with human infections [7–10,15], a comprehensive analysis of the "prophage-ome" is warranted to determine how these viral elements contribute to the genomic diversification of this genus.

We hypothesized that the *Tsukamurella* pangenome harbors a vast, largely uncharacterized reservoir of prophages that are evolutionarily distinct from currently sequenced lytic phages. To test this, we performed a systematic comparative genomic analysis of prophages across all publicly available *Tsukamurella* genomes to define the taxonomic diversity and distribution of prophages across the genus and to quantify the genomic divergence between these integrated prophages and previously characterized *Tsukamurella* phage isolates.

2. Methods

2.1. Genome Data, Prophage Identification and Annotation

Publicly available *Tsukamurella* genomes (n = 55), representing 12 validly published species and two additional genomes from strains without validated species names, were retrieved from the NCBI GenBank database. Genome of previously characterized bacteriophages, TIN2 (NC_028865.1), TIN3 (NC_028966.1), TIN4 (NC_041962.1), TPA2 (NC_015210.1), and TPA4 (NC_030916.1), were downloaded for comparative analysis.

Prophage sequences were identified using VIBRANT (v1.2.0) with default parameters [16]. Candidate viral sequences were classified by VIBRANT as low-, medium-, or high-quality draft, or as complete circular genomes, based on minimum viral criteria, including scaffold length >1 kb and at least four predicted open reading frames (ORFs). Draft quality categories were assigned according to Virus Orthologous Groups (VOG)-based annotations. High-quality draft genomes were defined by VIBRANT as sequences likely to contain most of the complete viral genome and to include annotations informative for downstream analyses, such as phylogenetic inference and true-positive verification [16]. For subsequent analyses, only viral elements classified as high-quality draft or complete circular genomes were selected, as these were considered to represent near-complete or complete viral genomes. These sequences were subsequently annotated using Pharokka v1.8.2 [17].

2.2. Virus Clustering and Comparative Genomics Analysis

To assess the genomic relatedness of identified *Tsukamurella* prophages and previously described *Tsukamurella* phages (TIN2, TIN3, TIN4, TPA2, and TPA4), we applied both nucleotide- and protein-level clustering approaches. Average nucleotide identity (ANI) was calculated using skani (v0.3.1) [18]. For protein-level analyses, GenBank files generated through Pharokka (default parameters) were used as input for PhaMMseqs v1.0.4 to assign proteins into phamilies (phams). Resulting phams were subsequently clustered with PhamClust v1.3.3 [19,20], with default parameters, including a minimum cluster size of 6 for sub-clustering, a similarity threshold of 0.6 for sub-clustering using single linkage, and a clustering similarity threshold of 0.25.

Representative phamilies were examined through synteny analysis with visualization of clinker v0.031 to assess structural organization [21]. Sequence divergence within phamilies was evaluated in MEGA v12 using the overall mean distance method with the p-distance substitution model and 1000 bootstrap replicates [22].

Comparative visualization of prophage relatedness and gene content was performed in RStudio [23]. Heatmaps and bar plots were manually curated and generated using ANI values, PhamClust clustering results, and protein family

counts obtained from the annotated prophage genomes. These data were integrated to visualize genomic similarity, clustering patterns, and variation in protein family composition among the identified prophages.

3. Results and Discussion

3.1. Prevalence of Prophage Distribution

The 55 publicly available *Tsukamurella* genomes ranged from 3.1 to 5.7 Mb in size, with an average of 83 contigs and an N50 of 1,703,034 bp. Their GC content varied between 68.4% and 71.6% (Supplementary Table S1).

VIBRANT identified 210 viral fragments across the 55 *Tsuakmurella* genomes. Of these, 73.1% of scaffolds met minimum viral criteria (>1 kbp and ≥4 ORFs), producing a total of 162 putative phages. The number of putative phages per genome varied substantially, ranging from 0 to 16. Among the 162 putative phages, seven were classified as high-quality draft prophages and one as a complete circular prophage (Table 1). These high-quality elements were found in *T. asaccharolytica* (n = 1), *T. paurometabola* (n = 1), *T. tyrosinosolvens* (n = 2), *T. soli* (n = 1), and *T. ocularis* (n = 3) (Table 1). The prophage from *T. paurometabola* strain 3OW was classified as lytic, whereas the others were lysogenic. Genome sizes of high-quality prophages ranged from 47,404 to 71,033 bp, encoding 72 to 150 predicted ORFs, with GC contents from 59.5 to 70.2%. In addition, 56.9 to 83.9% of ORFs encoded proteins with no known functional annotations, underscoring the novelty of *Tsukamurella*-associated viral elements (Table 1).

3.2. Virus Clustering and Taxonomic Classification of Tsukamurella Phages

ANI analysis revealed strong clustering among the known lytic *Tsukamurella* phages TIN2, TIN3, and TIN4, which shared 88 to 100% ANI, consistent with previous observations of their recent common ancestry (Figure 1A). However, these phages showed no detectable nucleotide similarity to VIBRANT-identified prophages or to other known phages, suggesting distinct evolutionary trajectories. Similarly, phages TPA2 and TPA4 showed no ANI-based relationships to other phages, reflecting their unique genomic backgrounds.

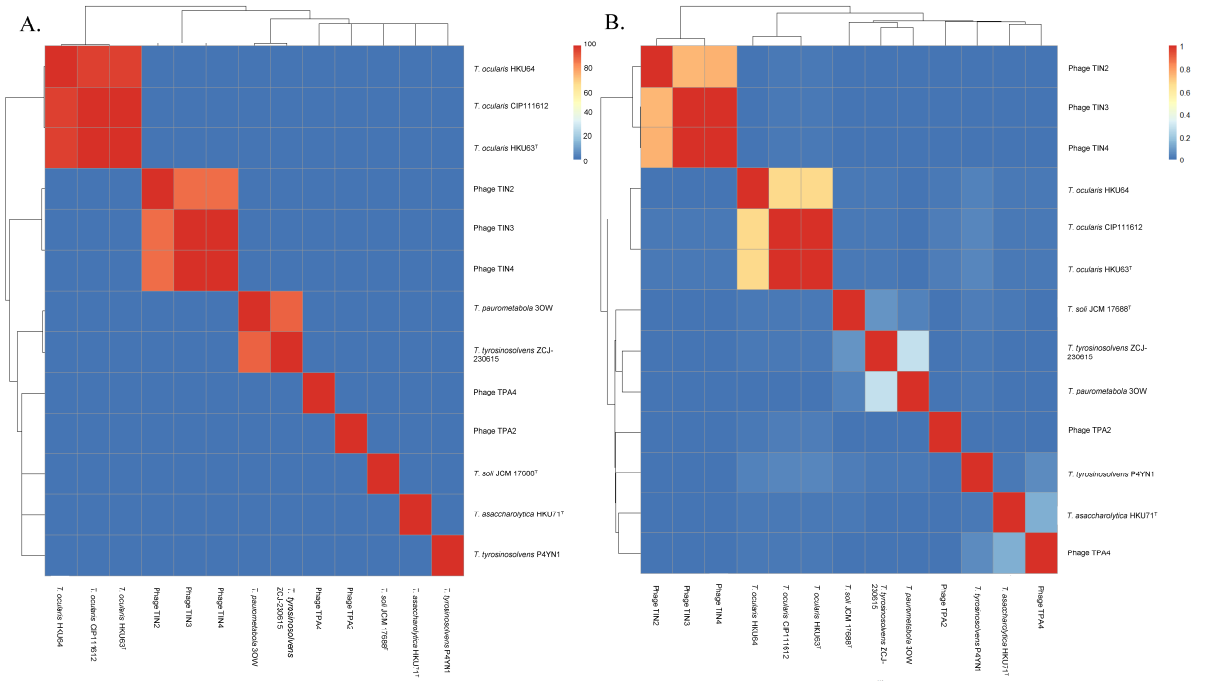

Figure 1. Comparative genomic analyses of *Tsukamurella* phages and prophages. (**A**) Heatmap of ANI values among known *Tsukamurella* phages (phages TIN2, TIN3, TIN4, TPA2, TPA4). (**B**) Proteomic clustering of phages and prophages using PhamClust. Three major phage/prophage clusters are observed.

In contrast, high ANI values (97 to 100%) were detected among the three *T. ocularis* prophages from strains HKU63^T, HKU64, and CIP111612, with the HKU63^T and CIP111612's prophage genomes exhibiting complete identity, suggesting recent prophage integration events. A moderately distant relationship (90.2% ANI) was identified between the prophages of *T. paurometabola* 3OW and *T. tyrosinosolvens* ZCJ-230615. In contrast, prophages from *T. asaccharolytica* HKU71^T, *T. tyrosinosolvens* P4YN1, and *T. soli* JCM 17688^T showed no close ANI-based similarity to any known phages or prophages.

Table 1. Genomic features and prophage characteristics of *Tsukamurella* strains containing high-quality prophages identified by VIBRANT.

Bacterial Genome						Prophage Genome					
Accession Number	**Species/Strain**	**Number of Contigs**	**Genome Sizes (bp)**	**N50 (bp)**	**GC (%)**	**Prophage Completeness**	**Predicted Lifestyle**	**Prophage Size (bp)**	**GC (%)**	**Total Number of ORFs**	**Hypothetic Proteins (Number, %)**
GCF_018335505	*T. paurometabola* 3OW	906	5,203,618	51,091	70.4	High-quality draft	Lytic	47,404	59.5	137	115, 83.9
GCF_035871755	*T. tyrosinosolvens* P4YN1	57	5,021,494	201,463	71.0	Complete circular	Lysogenic	50,312	70.2	76	46, 60.5
GCF_040790855	*T. tyrosinosolvens* ZCJ-230615	1	5,100,639	5,100,639	70.9	High-quality draft	Lysogenic	71,033	60.7	150	106, 70.7
GCF_042666795	*T. soli* JCM 17688^{T}	2	5,699,607	5,648,486	70.1	High-quality draft	Lysogenic	70,017	64.4	120	70, 58.3
GCF_007858435	*T. asaccharolytica* HKU71^{T}	83	4,123,067	413,707	70.2	High-quality draft	Lysogenic	65,709	69.8	99	62, 62.6
GCF_965138335	*T. ocularis* CIP111612	72	4,927,858	131,669	71.3	High-quality draft	Lysogenic	49,153	69.6	72	41, 56.9
SAMN54270355	*T. ocularis* HKU63^{T}	58	4,964,078	1,018,061	71.3	High-quality draft	Lysogenic	49,153	69.6	72	41, 56.9
SAMN54270356	*T. ocularis* HKU64	99	5,069,553	1,102,516	71.3	High-quality draft	Lysogenic	48,652	69.2	77	47, 61.0

Because phage genomes are highly mosaic due to frequent gene rearrangements, ANI-based comparisons alone may underestimate their evolutionary relatedness. To complement this approach, we performed protein family-based comparative analyses. Using PhaMMseqs, all proteins encoded by the *Tsukamurella* prophages and phages were grouped into 899 phamilies (phams), with an average of 1.5 members per family, including 629 orphams (single member protein family lacking similar homolog). These results highlight the substantial protein-level novelty within this group.

Subsequent proteomic clustering using PhamClust revealed three major clusters (Figure 1B)—(1) the three *T. ocularis* prophages, (2) phages TIN2, TIN3, and TIN4, and (3) The prophages from *T. paurometabola* 3OW and *T. tyrosinosolvens* ZCJ-230615. All remaining viral elements formed singletons, further underscoring the high genomic diversity and limited relatedness among *Tsukamurella*-associated phages.

3.3. Conserved lysis Cassette in Tsukamurella Phages and Prophages

Manual annotation of the 10 most abundant protein families revealed that they predominantly encoded core viral functions (Figure 2A), including two endolysin families (Pham 1and Pham 10), holins (Pham 3), minor tail proteins (Pham 2), single-stranded DNA-binding proteins (Pham 4), membrane proteins (Pham 11), terminase large subunits (Pham 13), and integrases (Pham 6).

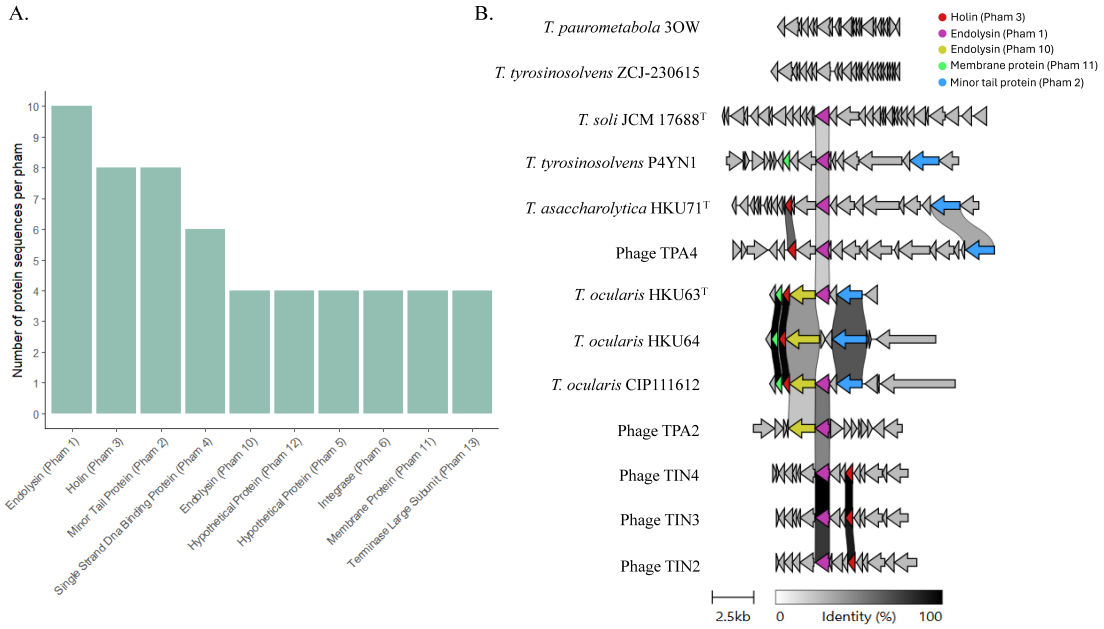

Figure 2. Conserved and divergent protein families in *Tsukamurella* phages and prophages. (**A**) Distribution of the top 10 most abundant protein families (phams) identified across all phages and prophages. Core viral functional categories dominate the protein landscape. (**B**) Synteny analysis of the lysis cassette region, revealing relatively conserved genomic organization across *Tsukamurella* phages and prophages. Endolysin (Pham 1 and Pham 10) and holin (Pham 3) are relatively conserved, whereas minor tail proteins (Pham 2) show higher variability, suggesting differential selective pressures between structural and lytic modules. Minor tail protein genes (Pham 2) are not shown for *T. tyrosinosolvens* ZCJ-230615 and *T. soli* JCM 17688^T because they are located outside the lysis cassette region.

Synteny analysis further showed that the lysis cassette was relatively conserved across both phages and prophages (Figure 2B). Endolysin (Pham 1) was detected in 10 of the 13 phage/prophage genomes analyzed (n = 10, Figure 2A), including phages TIN2, TIN3, TIN4, TPA2, and TPA4, and the prophages of *T. ocularis* HKU63^T and CIP111612, *T. tyrosinosolvens* P4YN1, *T. asaccharolytica* HKU71^T, and *T. soli* JCM 17688^T (Figure 2B). Notably, phages TPA2 and TPA4, despite lacking ANI similarity to other phages (Figure 1A), retained homologous endolysin phamilies (Pham1 and Pham 10 in TPA2; Pham1 in TPA4), mirroring the synteny observed in *T. ocularis* prophages HKU63^T and CIP111612's prophages (Figure 2B).

Holin (Pham 3), another key component of the lysis module, was detected in 8 of the 13 phage/prophage genomes, including phages TIN2, TIN3, TIN4, the three *T. ocularis* prophages, prophage of *T. asaccharolytica* HKU71^T, and phage TPA4. The relatively frequent detection of endolysin and holin phamilies suggests that shared lysis strategies may maintain among *Tsukamurella* phages/prophages despite substantial genomic divergence.

In contrast, the minor tail protein family (Pham 2) was detected in 8 of the 13 phage/prophage genomes analyzed (Figure 2A), including phage TPA4, and the prophages of *T. ocularis* HKU63^T, HKU64, and CIP111612, *T. tyrosinosolvens* P4YN1, ZCJ-230615, *T. asaccharolytica* HKU71^T, and *T. soli* JCM 17688^T. This protein family exhibited higher sequence diversity (average pairwise p-distance = 0.65, SE = 0.01) than the lysis-associated families (Pham 1, p-distance = 0.58 ± 0.02 SE; Pham 3, p-distance = 0.56 ± 0.02; Pham 10, p-distance = 0.40 ± 0.01). These results suggest stronger selective pressures may be associated with host interaction, recognition, and attachment.

4. Conclusions

This study provides a genome-based overview of prophage diversity in *Tsukamurella*. We identified eight high-quality prophages, including distinct clusters among *T. ocularis* isolates and several highly divergent elements with limited similarity to previously described *Tsukamurella* phages or prophages. Although genome-wide nucleotide similarity was generally low, lysis-associated protein families, particularly endolysins and holins, were relatively conserved across phages and prophages. Together, these findings expand the current catalogue of *Tsukamurella*-associated prophages and provide a foundation for future studies on their diversity and evolutionary relationships.

Supplementary Materials

The additional data and information can be downloaded at: https://media.sciltp.com/articles/others/260618 1630462627/eMicrobe-25120224-SM.zip.

Author Contributions

Conceptualization, J.L.L.T.; Methodology, J.Y.H.F., J.L. and L.J.; Formal Analysis, J.Y.H.F.; Writing—Original Draft, J.Y.H.F. and J.L.L.T.; Writing—Review & Editing, J.L., L.J., P.C.Y.W. and M.L.Y.; Funding Acquisition, J.L.L.T. and M.L.Y. All authors have read and agreed to the published version of the manuscript.

Funding

This study was partly supported by the start-up fund for new staff at the Faculty of Dentistry, The University of Hong Kong; the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government; the Feature Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE-113-S-0023-A), Taiwan, and a donation from the TE Health Consultant Company Limited. The sponsors have no role in the study design; collection, analysis, and interpretation of data; writing of the report; or decision to submit the article for publication.

Institutional Review Board Statement

The study was approved by the Institute Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster (Reference number: UW 16-365).

Informed Consent Statement

Not applicable.

Data Availability Statement

The seven high-quality draft prophage genomes and one complete prophage genome have been deposited in GenBank under accession numbers PZ428596-PZ428603.

Conflicts of Interest

The authors declare no conflict of interest. Given their editorial roles, Patrick C.Y. Woo (Editor-in-Chief) and Jade Lee Lee Teng (Editorial Board Member) had no involvement in the peer review of this paper and had no access to information regarding its peer-review process. Full responsibility for the editorial process of this paper was delegated to another editor of the journal.

Use of AI and AI-Assisted Technologies

No AI tools were utilized for this paper.

References

1. Rodino, K.G.; Simner, P.J. Status check: Next-generation sequencing for infectious-disease diagnostics. *J. Clin. Investig.* **2024**, *134*, e178003. https://doi.org/10.1172/JCI178003.
2. Lopez-Leal, G.; Camelo-Valera, L.C.; Hurtado-Ramirez, J.M.; et al. Mining of Thousands of Prokaryotic Genomes Reveals High Abundance of Prophages with a Strictly Narrow Host Range. *mSystems* **2022**, *7*, e0032622. https://doi.org/10.1128/msystems.00326-22.
3. Nanda, A.M.; Thormann, K.; Frunzke, J. Impact of spontaneous prophage induction on the fitness of bacterial populations and host-microbe interactions. *J. Bacteriol.* **2015**, *197*, 410–419. https://doi.org/10.1128/JB.02230-14.
4. Fortier, L.C.; Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **2013**, *4*, 354–365. https://doi.org/10.4161/viru.24498.
5. Abad, L.; Gauthier, C.H.; Florian, I.; et al. The heterogenous and diverse population of prophages in *Mycobacterium* genomes. *mSystems* **2023**, *8*, e0044623. https://doi.org/10.1128/msystems.00446-23.
6. Safaei, S.; Fatahi-Bafghi, M.; Pouresmaeil, O. Role of *Tsukamurella* species in human infections: First literature review. *New Microbes New Infect.* **2018**, *22*, 6–12. https://doi.org/10.1016/j.nmni.2017.10.002.
7. Teng, J.L.; Tang, Y.; Huang, Y.; et al. Phylogenomic Analyses and Reclassification of Species within the Genus *Tsukamurella*: Insights to Species Definition in the Post-genomic Era. *Front. Microbiol.* **2016**, *7*, 1137. https://doi.org/10.3389/fmicb.2016.01137.
8. Teng, J.L.L.; Tang, Y.; Wong, S.S.Y.; et al. *Tsukamurella ocularis* sp. nov. and *Tsukamurella hominis* sp. nov., isolated from patients with conjunctivitis in Hong Kong. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 810–818. https://doi.org/10.1099/ijsem.0.002589.
9. Teng, J.L.L.; Tang, Y.; Wong, S.S.Y.; et al. MALDI-TOF MS for identification of *Tsukamurella* species: *Tsukamurella tyrosinosolvens* as the predominant species associated with ocular infections. *Emerg. Microbes Infect.* **2018**, *7*, 80. https://doi.org/10.1038/s41426-018-0083-4.
10. Teng, J.L.L.; Fong, J.Y.H.; Fok, K.M.N.; et al. *Tsukamurella asaccharolytica* sp. nov., *Tsukamurella conjunctivitidis* sp. nov. and *Tsukamurella sputi* sp. nov., isolated from patients with bacteraemia, conjunctivitis and respiratory infection in Hong Kong. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 995–1006. https://doi.org/10.1099/ijsem.0.003861.
11. Teng, J.L.; Tang, Y.; Wong, S.S.; et al. Mycolyltransferase is important for biofilm formation and pathogenesis of *Tsukamurella keratitis*. *Emerg. Microbes Infect.* **2024**, *13*, 2373317. https://doi.org/10.1080/22221751.2024.2373317.
12. Petrovski, S.; Seviour, R.J.; Tillett, D. Genome sequence and characterization of the *Tsukamurella bacteriophage* TPA2. *Appl. Environ. Microbiol.* **2011**, *77*, 1389–1398. https://doi.org/10.1128/AEM.01938-10.
13. Dyson, Z.A.; Tucci, J.; Seviour, R.J.; et al. Three of a Kind: Genetically Similar Tsukamurella Phages TIN2, TIN3, and TIN4. *Appl. Environ. Microbiol.* **2015**, *81*, 6767–6772. https://doi.org/10.1128/AEM.01145-15.
14. Dyson, Z.A.; Brown, T.L.; Farrar, B.; et al. Locating and Activating Molecular 'Time Bombs': Induction of Mycolata Prophages. *PLoS ONE* **2016**, *11*, e0159957. https://doi.org/10.1371/journal.pone.0159957.
15. Teng, J.L.L.; Tang, Y.; Wong, S.S.Y.; et al. *Tsukamurella hongkongensis* sp. nov. and *Tsukamurella sinensis* sp. nov., isolated from patients with keratitis, catheter-related bacteraemia and conjunctivitis. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 391–397. https://doi.org/10.1099/ijsem.0.000733.
16. Kieft, K.; Zhou, Z.; Anantharaman, K. VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **2020**, *8*, 90. https://doi.org/10.1186/s40168-020-00867-0.
17. Bouras, G.; Nepal, R.; Houtak, G.; et al. Pharokka: A fast scalable bacteriophage annotation tool. *Bioinformatics* **2023**, *39*, btac776. https://doi.org/10.1093/bioinformatics/btac776.
18. Shaw, J.; Yu, Y.W. Fast and robust metagenomic sequence comparison through sparse chaining with skani. *Nat. Methods* **2023**, *20*, 1661–1665. https://doi.org/10.1038/s41592-023-02018-3.
19. Gauthier, C.H.; Cresawn, S.G.; Hatfull, G.F. PhaMMseqs: A new pipeline for constructing phage gene phamilies using MMseqs2. *G3* **2022**, *12*, jkac233. https://doi.org/10.1093/g3journal/jkac233.
20. Gauthier, C.H.; Hatfull, G.F. PhamClust: A phage genome clustering tool using proteomic equivalence. *mSystems* **2023**, *8*, e0044323. https://doi.org/10.1128/msystems.00443-23.
21. Gilchrist, C.L.M.; Chooi, Y.H. clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* **2021**, *37*, 2473–2475. https://doi.org/10.1093/bioinformatics/btab007.
22. Kumar, S.; Stecher, G.; Suleski, M.; et al. MEGA12: Molecular Evolutionary Genetic Analysis Version 12 for Adaptive and Green Computing. *Mol. Biol. Evol.* **2024**, *41*, msae263. https://doi.org/10.1093/molbev/msae263.
23. Team, R.C. *R: A language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.