Title: Genome sequence and characterization of coliphage môr ffagbaw

Running title: phage môr ffagbaw

Authors: Nerissa E. Thomas1#, Nzubechukwu Innocent Ugokwe2#, Andrew Kinsella2, Edouard Galyov2, Willem van Schaik3, Naomi Joyce1, Franklin L. Nobrega4, Andrew Millard2 & Rachael C. Wilkinson1\*

1 Swansea University Medical School, Swansea University, Swansea, SA2 8PP

2 Centre for Phage Research, University of Leicester, University Road, Leicester, LE1 7RH, UK

3 Institute of Microbiology and Infection, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

4 University of Southampton, East Highfield Campus, University Road, SO17 1BJ

#These authors contributed equally.

\*corresponding author

Email: rachael.wilkinson@path.ox.ac.uk

Keywords: coliphage, phages, genomes

**Authorship statement:**

RW and AM conceived the idea. RW and AM secured resources to support the research. NT, NIU, AK were involved in the practical aspects of sample collection, phage isolation, genome sequencing, phenotypic characterisation, and analysis of raw data. RW, NT, NIU, AK and AM wrote the manuscript. All authors have reviewed and approved the manuscript prior to submission.

**Author Disclosure statement:**

No competing financial interests exist for any authors.

**Abstract**

**Background:** The quality of coastal waters around the United Kingdom is an area of increasing concern following sewer overflow, where wastewater is discharged into the environment. Coliphages, viruses which infect coliform bacteria, are associated with water quality in aquatic systems yet remain largely uncharacterized at the genomic level.

**Materials and Methods:** Phage môr ffagbaw was isolated from seawater against *E. coli* by enrichment and plaque assays. Whole genome sequencing, transmission electron microscopy and host range analysis against the ECOR collection were used to characterise the phage.

**Results:** The virion had a siphovirus morphology and genomic analysis placed it within the family *Drexlerviridae*, subfamily *Tempevirinae* and forms a new species within the genus *Hanrivervirus*. Spot assays revealed that phage môr ffagbaw could form plaques on 6 out of 72 ECOR strains (8%).

**Conclusions:** Môr ffagbaw represents a new species of phage within the genus *Hanrivervirus*, with a narrow host range.

**Introduction**

Bacteriophages have received significant recent attention as potential therapeutics due to their specificity and sensitivity to specific bacteria, which have also made them useful tools for the monitoring of water quality. Bacteriophages that target and infect coliforms, coliphages, are of particular interest for studying water quality, along with *Enterococcus* phages [(1)](https://paperpile.com/c/h34ufm/oB9d) and crAssphage [(2)](https://paperpile.com/c/h34ufm/rD9j). Coliphages are used as a proxy for the presence and abundance of their bacterial hosts, which serve as indicators of faecal contamination. Much of this work has focused on the abundance of coliphages that are found in seawater, rather than the genetic diversity of phages that are found within seawater. Despite over 600 coliphage genomes being publicly available [(3)](https://paperpile.com/c/h34ufm/mF21a) very few of these have been isolated from seawater samples, which largely remains an under-sampled environment for coliphages.

Swansea Bay is classed as a bathing water area, where water quality is monitored by Natural Resources Wales (NRW) by determining colony forming units of *Enterococcus spp.* and *E. coli* within a 100 ml water sample. In 2022 this area was reported as “good” quality by NRW [(4)](https://paperpile.com/c/h34ufm/ccOe). Following a period of high rainfall within September 2022 there was an increase in reported coliforms isolated from seawater samples (10-fold increase in *Enterococcus* isolated and 8-fold increase in *E. coli* isolated) [(4)](https://paperpile.com/c/h34ufm/ccOe). It was during this time we collected water for the isolation of coliphages.

Here we describe the isolation and characterization of a new species of virus within the genus *Hanrivervirus*, phage môr ffagbaw, from a seawater sample collected from Swansea Bay following the period of high rainfall in September 2022.

**Materials and Methods**

Bacteriophage, môr ffagbaw (pronunciation: more faj-b-ow) was isolated from seawater within Swansea Bay, South West Wales in September 2022 following a period of flooding. Water samples (50 ml) were collected and cleared of debris via centrifugation (5000 rpm, 10 min) and then filter-sterilised using a 0.2 µm pore size filter. A phage enrichment step was achieved by the addition of 5 ml of the processed water sample with 5 mL culture of *E. coli* K-12 MG1655 in Luria Bertani (LB) Broth at mid-log growth, followed by incubation at 37 °C overnight. The double-agar overlay method was used for phage isolation and the subsequent three rounds of purification, as has been described previously [(5)](https://paperpile.com/c/h34ufm/BCNgV). High titre lysate was produced by infection of ~50 ml of exponentially growing *E. coli* MG1655 and incubation at 37 °C with shaking at 300 rpm, until lysis had occurred. Host range was determined by screening against the *Escherichia coli* reference (ECOR) collection [(6)](https://paperpile.com/c/h34ufm/s26x4) via a spot test assay. Briefly, phage môr ffagbaw was diluted to a concentration of 106 PFU/ml, with 10 µl of serially diluted phage added to the top of a lawn of *E. coli.* The 106 PFU/ml stock was diluted six times with 10-fold dilutions. The observed titre on the ECOR strains were compared to the initial host of *E. coli* MG1655, in triplicate. The virulence index was calculated using a method described previously [(by 7)](https://paperpile.com/c/h34ufm/MAYKN/?prefix=by%20) using a SPECTROstar Omega (BMG) plate reader. For genome sequencing, DNA was extracted from 1 ml of bacteriophage lysate as previously described [(8)](https://paperpile.com/c/h34ufm/2vJJm). Sequencing was carried out by the SeqCentre (Pittsburgh, USA) with Nextera XT library preparation, using 2 × 150 on a NextSeq 200. For assembly reads were first trimmed using trim\_galore with the `paired` option [(9)](https://paperpile.com/c/h34ufm/aYNBV), followed by sub sampling of reads using bbnorm.sh with the settings of `target=150`. Assembly and annotation was carried out based on previously described methods [(10)](https://paperpile.com/c/h34ufm/p49JW). Briefly, SPAdes v3.15.5 [(11)](https://paperpile.com/c/h34ufm/C2jBj) using the options of `assembler-only` was used for assembly. The resultant phage contig was identified, and the closest relative identified with MASH using database of known phage genomes [(3)](https://paperpile.com/c/h34ufm/mF21a), the genome was reordered against phage JakobBernoulli (Accession number: MZ501079) and assembly errors detected and corrected with pilon using default settings [(12)](https://paperpile.com/c/h34ufm/5PMfL). Annotation was carried out with Prokka using the PHROGS database [(13,14)](https://paperpile.com/c/h34ufm/YJazc+Kwj1C). The top 50 genomes with highest nucleotide identity were identified using the INPHARED database [(3)](https://paperpile.com/c/h34ufm/mF21a). Virus Intergenomic Distance Calculator (VIRIDIC) default settings [(15)](https://paperpile.com/c/h34ufm/HGV4J) was used to calculate distance similarity. Core-gene analysis was carried out with ROARY using settings --e --mafft -p 32 –i 90. Phylogenetic trees were constructed using IQTree2 [(16)](https://paperpile.com/c/h34ufm/6sZuq) using the *terL* with a GTR+F model, using --alrt 1000 -B 1000.

Transmission Electron Microscopy

TEM was carried out by the University of Leicester Core Biotechnology Services Electron Microscopy Facility. Briefly, 5 µl of sample was applied to a freshly glow-discharged carbon film copper grid for two minutes, washed with two drops of pure water and stained with two 5 µl drops of 1% uranyl acetate (w/v). Excess stain was blotted with filter paper and grids allowed to dry before viewing on TEM. Samples were viewed on a JEOL JEM-1400 TEM with an accelerating voltage of 120 kV. Digital images were collected with a EMSIS Xarosa digital camera with Radius software. Virion images were processed in ImageJ [(17)](https://paperpile.com/c/h34ufm/FyGQd) using the measure tool, with the scale bar present used as a calibration to measure phage particle size. The data presented is the mean of 10 virions.

**Results**

Bacteriophage môr ffagbaw was isolated from seawater within Swansea Bay, South West Wales using *E. coli* MG1655 as a host and could only be isolated after enrichment. It was thus not possible to determine the titre of coliphages in the water sample. Imaging of phage môr ffagbaw using transmission electron microscopy revealed that the virion had a siphovirus morphology (Figure 1A) with a mean head length of 60 +/- 3 nm and width of 59 +/- 3 nm. The tail was estimated to be 113 +/- 16 nm in length, with a width of 12 +/-1 nm (Figure 1). Genomic analysis of phage môr ffagbaw revealed a genome of 52463 bp in length with a G+C content of 44.2% with 83 predicted genes and no tRNAs. The use of NexteraXT prevented the identification of the termini, thus the genome was ordered against phage JakobBernoulli (accession number MZ501079). Analysis with CheckV v1.0.1 [(18)](https://paperpile.com/c/h34ufm/gYY1I) suggested with high confidence the genome was complete based on AAI metric and identification of direct terminal repeats. Only 47% of genes could be ascribed a predicted function, the majority of which were linked to phage morphogenesis proteins. There were 16 genes for which no PHROG number could be ascribed, suggesting they are not common in other bacteriophage genomes.

Comparing the genome sequence of phage môr ffagbaw against known phage genomes with MASH demonstrated that it had the greatest similarity to the phage grams (MN850567, MASH distance 0.0519203) and was similar several other phages within the genus *Hanrivervirus.* The relationship of phage môr ffagbaw to other species within the genus *Hanrivervirus* and other closely related phages was determined by phylogenetic analysis of the *terL* gene, given the conserved function of this gene. The resultant phylogeny confirms phage môr ffagbaw is a member of the *Hanrivervirus* genus and is most closely related to *Shigella* phage ESh4. All phages that are classified as *Hanrivervirus* by the ICTV form a monophyletic clade (Figure 2), sister to this clade is a group of phages that were isolated on *Campylobacter* and are described as *Hanrivervirus* in GenBank but have yet to be formally classified. Intriguingly, the phage A16 which was isolated on *Campylobacter,* does not fit with other phages isolated on *Campylobacter* and instead sits within the clade that contains phages infecting *Escherichia*, *Salmonella* and *Shigella* (Figure 2). To further classify phage môr ffagbaw, we utilised VIRIDIC, confirming phage môr ffagbaw as a member of the genus *Hanrivervirus*, and a representative of a new species, having < 95% similarity with any other phage [(19,20)](https://paperpile.com/c/h34ufm/Nlltk+yXWVQ) (Figure S1).

Furthermore, it confirmed the phages infecting *Campylobacter* are also members of the genus *Hanrivervirus.* However, whether the genomes of all phages infecting *Campylobacter* are entirely complete, as described in their Genbank submissions, is currently unclear. Based on genome length compared to other members of *Hanrivervirus*, it appears two are incomplete or have undergone substantial genome reduction (Figure 2). Further, analysis of the synteny of the genomes between selected members of the *Hanrivervirus* also suggest some of the *Campylobacter* phages are incomplete. There is clear synteny in the genomes of phages within the *Hanrivervirus* (Figure 3), yet some of the *Campylobacter* phage genomes lack genes that would otherwise be considered core.

To determine the host range of phage môr ffagbaw, we tested it against the ECOR collection that has been widely used as a representative collection of diverse *E. coli* strains [(6)](https://paperpile.com/c/h34ufm/s26x4). Phage môr ffagbaw was only capable of infecting six of the 72 ECOR strains. Of these strains it had a similar efficiency of plating (EOP) on ECOR 13, ECOR 21 and 35, and reduced EOP on three strains (ECOR 16, 26, and 57). Of the ECOR strains infected five of six are thought to be commensals (ECoR clusters A, B1 and B2) with only one pathogenic strain, ECOR35 (Table 1.). All the strains have different O antigen serotypes, how this links to infection directly will only become clear when the receptor of the phage is elucidated. Phage môr ffagbaw was further characterised by calculation of the virulence index (Vp) [(7)](https://paperpile.com/c/h34ufm/MAYKN) which was found to be 0.75 against *E. coli* K12 MG1655 when grown at 37 °C in LB medium (Figure 1B), with phage môr ffagbaw effective at killing across a range of MOIs (Figure 1B)

**Discussion**

The study of bacteriophages infecting *Escherichia coli (E. coli)* in seawater has largely focused on their role as a proxy for faecal coliforms and as a marker as drinking water quality [(21,22)](https://paperpile.com/c/h34ufm/X3tX+5oCV). Previous work has shown a wide diversity of phage morphotypes can be found in seawater, including siphovirus, myoviruses and podoviruses [(23)](https://paperpile.com/c/h34ufm/lcvYu). Additionally, genomic analysis has revealed phages of the families; *Demerecviridae, Straboviridae* and *Drexlerviridae* [*(5)*](https://paperpile.com/c/h34ufm/BCNgV)*.* Here we extend this to include phages of a further subfamily (*Tempevirinae*) in the family *Drexlerviridae*. Through comparative genomic analysis we identified phage môr ffagbaw is representative of a new viral species, and as the first representative of this species, we propose the species name *“Hanrivervirus bawmor”*.

Although coliphages are some of the most abundant phages isolated to date, several high throughput studies have identified phages that represent new diversity at the species, genus and family level, combined with theoretical analysis suggesting there is still much coliphage diversity to be discovered [(23,24)](https://paperpile.com/c/h34ufm/lcvYu+Mupu). Sewage overflow into the sea, following periods of heavy rainfall, is predicted to result in higher levels of coliforms within this environment. Thus, it was unsurprising to isolate a new species of coliphage from seawater, an environment that is relatively under-sampled for the isolation of coliphages. As larger numbers of phages are isolated from seawater, following periods of high and low rainfall, and characterised it will be possible to further evaluate how valid coliphages are as a marker of coliform contamination. Previous studies have identified a variety of morphotypes, with some enrichment of particular morphotypes [(23)](https://paperpile.com/c/h34ufm/lcvYu). Currently, it is not known how long different species of coliphages are stable in seawater at the genomic level. The isolation of new model phage-host systems like this one, will allow such questions to be answered in the future. Moving towards a more systematic study of the different types of phages found in seawater and if they simply mirror the diversity of phages found in wastewater, that is the likely source, or if particular types of coliphages are able to reside for longer periods than others.

Previous reports of phages isolated from seawater have suggested they have a broad host range [(5)](https://paperpile.com/c/h34ufm/BCNgV), yet phage môr ffagbaw was only infective against six of 72 ECOR strains tested. Comparison of host ranges between studies is always problematic as it is often not clear whether the tested strains are representative of the species. The ECOR collection for *E. coli* goes some way to alleviate this problem and provides a like-for-like comparison, with previous studies that have also utilised this strain collection. In comparison to these previously described coliphages, phage môr ffagbaw has a narrow host range- only being able to infect six ECOR isolates, in comparison to phages JK1 and JK08 which were capable of infecting 16 strains and 41 ECOR strains respectively [(25)](https://paperpile.com/c/h34ufm/u9Sag), and phage AC3 which can absorb to 43 and replicated on 23 ECOR strains [(26)](https://paperpile.com/c/h34ufm/6sPUO).

With increased use of standardised collections of strains such as the ECOR collection, it will be possible to begin to determine the underlying genetic differences between phages that control this host range. Currently the small number of phages that have been tested on the ECOR prevent this. It is worth noting that AC3 and JK08 are phages of the family *Straboviridae* (formerly “T4-like”). If having narrow host ranges is characteristic of phages in the genus *Hanrivervirus*, remains to be determined empirically. However, the isolation of closely related phages on *Salmonella*, *Escherichia*, *Shigella* and *Camplyobacter*, suggests they may not. In a similar way using a standard set of strains allows easier comparison between studies, we utilised the virulence index to compare the effectiveness of phage môr ffagbaw at killing its host. With a virulence index of 0.75, phage môr ffagbaw is highly virulent compared to other coliphages that have been assessed using this quantitative method. This includes T4, T5, T7, SLUR29 and vB\_EcoM-4HA13, with only phage T7 having a higher virulence index, compared to môr ffagbaw, of 0.84. The increased use of standardised host sets, such as the ECOR collection or a recently curated set for *Klebsiella* [(27)](https://paperpile.com/c/h34ufm/DaGG) alongside the use of quantitative methods for assessing infectivity such as the virulence index across studies, in conjunction with genome sequencing, will allow further comparison between studies to decipher the link between genotype and phenotype.

**Acknowledgements**

Bioinformatics analysis was carried out using MRC CLIMB Infrastructure MR/L015080/1. RW and NT were funded by the European Structural Fund Programme 2014-2020 via the European Regional Development Fund funded C81844/C81845 ACCELERATE (Welsh Health Innovation Technology Accelerator) project. AM was funded by MR/L015080/1. NIU and AK were in receipt of PhD studentships funded by the University of Leicester F100 and MRC-AIM DTP respectively.

**References**

1. Chyerochana N, Kongprajug A, Somnark P, et al. Distributions of enterococci and human-specific bacteriophages of enterococci in a tropical watershed. Int J Hyg Environ Health 2020; 226: 113482.

2. García-Aljaro C, Ballesté E, Muniesa M, et al. Determination of crAssphage in water samples and applicability for tracking human faecal pollution. Microb Biotechnol 2017; 10: 1775–1780.

3. Cook R, Brown N, Redgwell T, et al. INfrastructure for a PHAge REference Database: Identification of large-scale biases in the current collection of cultured phage genomes. PHAGE. Epub ahead of print 5 October 2021. DOI: 10.1089/phage.2021.0007.

4. Bathing water profile, https://environment.data.gov.uk/wales/bathing-waters/profiles/profile.html?site=ukl1800-36900 (accessed 17 August 2023).

5. Michniewski S, Redgwell T, Grigonyte A, et al. Riding the wave of genomics to investigate aquatic coliphage diversity and activity. Environ Microbiol 2019; 21: 2112–2128.

6. Patel IR, Gangiredla J, Mammel MK, et al. Draft Genome Sequences of the Escherichia coli Reference (ECOR) Collection. Microbiol Resour Announc; 7. Epub ahead of print October 2018. DOI: 10.1128/MRA.01133-18.

7. Storms ZJ, Teel MR, Mercurio K, et al. The Virulence Index: A Metric for Quantitative Analysis of Phage Virulence. PHAGE 2020; 1: 27–36.

8. Rihtman B, Meaden S, Clokie MRJ, et al. Assessing Illumina technology for the high-throughput sequencing of bacteriophage genomes. PeerJ 2016; 4: e2055.

9. Krueger. Trimgalore https://www. bioinformatics. babraham. ac. uk/projects/trim\_galore. Accessed October.

10. Shen A, Millard A. Phage genome annotation: Where to begin and end. PHAGE 2021; 2: 183–193.

11. Bankevich A, Nurk S, Antipov D, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 2012; 19: 455–477.

12. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 2014; 9: e112963.

13. Terzian P, Olo Ndela E, Galiez C, et al. PHROG: families of prokaryotic virus proteins clustered using remote homology. NAR Genom Bioinform 2021; 3: lqab067.

14. Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 2014; 30: 2068–2069.

15. Moraru C, Varsani A, Kropinski AM. VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. Viruses 2020; 12: 1268.

16. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 2020; 37: 1530–1534.

17. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012; 9: 671–675.

18. Nayfach S, Camargo AP, Schulz F, et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nat Biotechnol 2021; 39: 578–585.

19. Turner D, Shkoporov AN, Lood C, et al. Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. Arch Virol 2023; 168: 74.

20. Turner D, Kropinski AM, Adriaenssens EM. A roadmap for genome-based phage taxonomy. Viruses; 13. Epub ahead of print 18 March 2021. DOI: 10.3390/v13030506.

21. Ibarluzea JM, Moreno B, Serrano E, et al. Somatic coliphages and bacterial indicators of bathing water quality in the beaches of Gipuzkoa, Spain. J Water Health 2007; 5: 417–426.

22. Jofre J, Lucena F, Blanch AR, et al. Coliphages as model organisms in the characterization and management of water resources. Water 2016; 8: 1–21.

23. Burbano-Rosero EM, Ueda-Ito M, Kisielius JJ, et al. Diversity of somatic coliphages in coastal regions with different levels of anthropogenic activity in São Paulo State, Brazil. Appl Environ Microbiol 2011; 77: 4208–4216.

24. Olsen NS, Hendriksen NB, Hansen LH, et al. A New High-Throughput Screening Method for Phages: Enabling Crude Isolation and Fast Identification of Diverse Phages with Therapeutic Potential. Phage (New Rochelle) 2020; 1: 137–148.

25. Kaczorowska J, Casey E, Lugli GA, et al. In Vitro and In Vivo Assessment of the Potential of Escherichia coli Phages to Treat Infections and Survive Gastric Conditions. Microorganisms; 9. Epub ahead of print 3 September 2021. DOI: 10.3390/microorganisms9091869.

26. Farquharson EL, Nugen SR. Enterobacteria Phage Ac3’s Genome Annotation and Host Range Analysis Against the ECOR Reference Library. Phage (New Rochelle) 2022; 3: 165–170.

27. Martin MJ, Stribling W, Ong AC, et al. A panel of diverse Klebsiella pneumoniae clinical isolates for research and development. Microb Genom; 9. Epub ahead of print May 2023. DOI: 10.1099/mgen.0.000967.

28. Alikhan NF, Zhou Z, Sergeant MJ, et al. A genomic overview of the population structure of Salmonella. PLoS Genetics. Epub ahead of print 2018. DOI: 10.1371/journal.pgen.1007261.

**Figures**

**Figure 1.** Morphology of phage môr ffagbaw. Virions were imaged by TEM after negative staining with uranyl acetate, imaged at 60,000×. Scale bar is 100 nm. A siphovirus morphology of long non-contractile tail and icosahedral head was observed. B Virulence Index. The local virulence index (V*i*) was determined for MOIs of 1 to 1E-7, to calculate the Virulence index (Vp) of0.75 against *E. coli* K12 MG1655 when grown at 37 °C in LB medium.

**Figure 2:** Phylogenetic analysis of phage môr ffagbaw. A phylogenetic tree was constructed using TerL from complete phage genomes in the INPHARED databases. The amino acid sequences were aligned in MAFFT, and trees constructed in IQ-TREE2 with 1000 bootstraps. Bootstrap values above 75 are denoted as black circles, with the size proportional to the support value. The genome length is represented as a blue bar, with bars proportional to genome length. The host of each phage isolate is denoted by a coloured square. Orange: *Campylobacter*, Blue: *Escherichia*, Pink: *Shigella* and Green: *Salmonella*.

**Figure 3**: Comparative genome analysis of phage môr ffagbaw and representatives of the genus *Hanrivervirus.* Comparative genome analysis was carried out with clinker with default settings and the final figure edited with InkScape. Orthologues are coloured and joined with shaded lines.

**Table 1**: Efficiency of plating of phage môr ffagbaw on the ECOR collection. For brevity, only strains of ECOR that were infected are displayed. Serotype and ECoR cluster are derived from data within Enterobase [(28)](https://paperpile.com/c/h34ufm/z1j6). Efficiency of plating was calculated in respect to plating on the host of isolation, *E. coli* K12 MG1655.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ECOR Host | EOP | SD | Serotype | ECoR  Cluster |
| 13 | 1.233333 | 0.262467 | Nd:H25 | A |
| 16 | 0.043333 | 0.004714 | O9:H10 | A |
| 21 | 0.933333 | 0.309121 | O121:H11 | A |
| 26 | 0.046667 | 0.012472 | O104:H21 | B1 |
| 35 | 0.833333 | 0.286744 | O1:K1:NM | D |
| 57 | 0.4 | 0.08165 | Nd | B2 |