Article:
Alsharif, G., Pham, T.K., Connolly, A.N. et al. (5 more authors) (2025) Isolation of a E. coli flagellotrophic Jumbophage SHEFM2K that replicates in cytoplasmic putative assembly areas. Microbiological Research, 293. 128082. ISSN 0944-5013

https://doi.org/10.1016/j.micres.2025.128082

Isolation of an *E. coli* flagellotrophic Jumbophage SHEFM2K that replicates in cytoplasmic putative assembly areas

Ghadah Alsharif[3], Trong Khoa Pham[4], Alexandra N. Connolly[1], Karolina I Pyrszanowska[1], Elspeth M Smith[1], Alhassan Alrafaie[5], Carl Smythe[4] and Graham P. Stafford[1,2]

1 - School of Clinical Dentistry, University of Sheffield, Sheffield, S10 2TA
2 - Florey Institute for Host-pathogen Interactions, University of Sheffield, Sheffield, S10 2TN
3- Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, KSA
4 - School of Biosciences, Biological Mass Spectrometry Facility, The University of Sheffield, Sheffield, S10 2TN
5- Department of Medical Laboratory, College of Applied Medical Sciences in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia

*Corresponding author
email: g.stafford@sheffield.ac.uk

Keywords:
Jumbophage, Bacteriophage, flagella, *Escherichia coli*

Abstract
The bacteriophage SHEFM2K was isolated from unpasteurised dairy farm milk using a newly isolated *E. coli* ExPEC strain EcM2K (O23:H8, ST446). SHEFM2K is large contractile-tailed jumbophage with a genome of 348kb sharing homology with jumbophage from *E. coli* of the *Asteriusvirus* genus. SHEF-M2K host range testing indicated that it only makes clear plaques with EcM2K and a sepsis strain from our collection (G34590). Host-ranging assays indicated that it is able to suppress the growth of a range of *E. coli* strains in liquid culture assays: including EHEC O157:H7, K-12 (MC1000, MG1655) and *E.coli* B (BL21). TEM images of infection of EcM2K indicated association with flagella-like structures. An *E. coli* MC1000 mutant lacking the flagellin (*fliC*) gene was less sensitive to SHEFM2K infection, a phenotype restored by providing *fliC in trans*. These data illustrate M2K is a flagellotrophic phage that attaches to flagella as part of its infection cycle. We also present cross-sectional TEM images of the SHEFM2K infection cycle showing that it forms putative 'assembly areas' in the host cytoplasm cleared of ribosomes and other material with heads appearing within the periphery before tails appear and lysis occurs. We also present the proteome of mature SHEFM2K phage, highlighting proteins expressed and notable those no detected which might have a role in replication given their predicted function. Overall, we present a preliminary characterisation of a newly isolated jumbophage that interacts with the *E. coli* flagellum and uncover novel aspects of their biology by identifying an internal assembly area.

Introduction

Bacteriophages are arguably the most numerous biological entities on our planet and their diversity and variety are even broader than their bacterial hosts (Clokie *et al.*, 2011; Chanishvili, 2012). While phages were discovered over 100 years ago in the famous work of Twort, Hankin and d'Herelle (Chanishvili, 2012), it was not until 1978 that 'jumbo' or 'giant' phages were first identified(Krylov and Zhazykov, 1978). This term refers to phages with genomes larger than 200-250kb depending on opinion and often with capsid diameters of over 100 nm (Yuan and Gao, 2017; M Iyer *et al.*, 2021). Their genomes are both larger and more complex than other phages, with several tRNAs and various metabolic enzyme-encoding genes, e.g. ribonucleotide reductase, their own DNA polymerase and numerous DNA methyltransferases (Yuan and Gao, 2017).

The giant/ jumbo phage are relatively poorly understood biologically, but over 200 isolates exist that infect a range of marine bacteria as well as human pathogens, including *Synechococcus* (Hua *et al.*, 2017),*Pseudomonas* (Naknaen *et al.*, 2024), *Salmonella* (Xie *et al.*, 2021), *Klebsiella* (Hu *et al.*, 2023; Ranta, Skurnik and Kiljunen, 2024), *Bacillus (Yuan and Gao, 2016)*, *Erwinia* spp. (Prichard *et al.*, 2023) and other enteric spp. such as *Escherichia coli* (Yuan and Gao, 2017; Jo *et al.*, 2023). A general rule of thumb is that these are viruses whose genome is over 200-250kb in size and contain their own DNA polymerase as well as a range of characteristic replication and DNA replication enzymes such as RNA and DNA polymerases (Yuan and Gao, 2017; Prichard *et al.*, 2023).

The study of some jumbophage has revealed a novel mechanism of host-takeover and mechanism of replication inside self-assembling internal 'nuclei' (Chaikeeratisak *et al.*, 2017, 2017; Korf *et al.*, 2019; Guan and Bondy-Denomy, 2020; Birkholz *et al.*, 2022; Prichard and Pogliano, 2024). The best studied example are a group of jumbophage, best exemplified by *Goslar*-related phage (approx. 250kb) which produce a proteinaceous 'nucleus-like' compartment that is formed using a phage encoded protein known as Chimallin A (ChmA) (Laughlin *et al.*, 2022) or Phage Nuclear enclosure protein (PhuN)(Nieweglowska *et al.,* 2023). This is accompanied by the presence of a phage encoded tubulin type protein (PhuZ: Phage Tubulin/FtsZ) that produces an intra-'nuclear' treadmill involved in the production of phage particles that at this point is novel to biologyx(Kraemer *et al.*, 2012; Chaikeeratisak *et al.*, 2019). However, it is unclear how many other genera of giant/ jumbo phages employ similar mechanisms, partly due to a lack of studies. Phylogenetically the Goslar-like viruses form a jumbo phage clade known as *Goslarvirus,* although taxonomy is evolving with the name *Chimallivirus* now proposed for phage containing a homologue of the Chimallin protein (Prichard *et al.*, 2023).

Several other *E. coli* infecting jumbophages, such as pHAPEC6 (Wagemans *et al.*, 2020) are in a group known as the *Asteriusvirus* (Korf *et al.*, 2019). These have genomes in the region of 350kb, large heads (over 120nm) and the characteristic DNA replication genes mentioned above (Yuan and Gao, 2017). However, very little is known about *Asteriusvirus* biology in terms of their infective cycle or host cell receptor on *E,coli* strains that they infect.

Here, we report the isolation of a new member of the *Asteriusvirus* jumbophages with broad infectivity of a range of *E. coli* strains, SHEFM2K (348kb). We present data uncovering new aspects of its biology, including identifying, for the first time, a cellular receptor for an *Asteriusvirus*, gain insights into its replication cycle and set the foundation for further studies of these intriguing jumbophages.

Methods

Plasmids, bacterial strains, and growth

Plasmids used in this study are listed in **(Table S1)**. Bacterial strains are listed in **Table 1**. All strains were grown in Luria-Bertani (LB) or Brain heart infusion (BHI) broth and maintained on LB agar at 37°C. For phage infections, LB soft agar (0.35%) was employed.

Phage isolation protocol

Bacteriophage SHEFM2K was isolated by enrichment of 500µl of a raw milk sample from Cliffe House Farm, Dungworth in 4ml BHI (24h, 37°C), which was then centrifuged and filtered (0.45µm). This sample (100 µl) was then added to 400µl of EcM2K (OD600=1) and 4 ml 0.35% LB agar before overlaying on LB agar (Kropinski *et al.*, 2009). Clear plaques were picked into PBS and passaged 3 x from plate-lysates on 0.35% LB agar to purify. The collected lysate was then stored at 4°C. For liquid lysates and routine propagation, an exponential phage culture of host EcM2K was grown before infection with SHEFM2K in LB and diluted to A600=1 before being allowed to grow for 1-2 hours where lysis was observed before harvesting. Plate lysates were prepared as above using 100µl SHEFM2K (appropriate titre) and 400µl EcM2K in double layer agar assays.

Transmission electron microscopy of phage particles

Negative staining Transmission electron microscopy (TEM) of a PEG-concentrated phage sample of SHEFM2K was performed as previously (Yamamoto *et al.*, 1970; Al-Zubidi *et al.*, 2019). In summary, purified phage particles in SM buffer were placed onto carbon-coated copper grids and negatively stained with 2% (wt/vol) uranyl acetate for 1 min. Particles were visualized using a FEI Tecnai G2 Spirit transmission electron microscope at an accelerating voltage of 80 kV at the Electron Microscopy Unit in Sheffield. Electron micrographs were recorded using a Gatan Orius 1000 digital camera and Digital Micrograph software. To observe phages alongside bacteria, 1 ml of exponentially phase cells (OD_{600} = 1) were infected with phage at an MOI of 10 for 10 min, before centrifugation at 7,000 × *g* for 10 min. The pellet was resuspended in 1 ml of 3% glutaraldehyde for 1 h (room temperature) before examination by TEM as described above. PEG concentration was achieved by adding 15% PEG (w/v), 4°C, 1 h, before centrifugation (4°C 30 min 16,000 x g) and resuspension in PBS and PEG removal with chloroform ($CHCl_3$) and collection of the DNA containing upper layer.

Host range testing

Host range was determined using a modified plate assay, by diluting a phage stock (10^8 PFU/ml) and spotting 3 µl spots on double agar containing a host strain lawn and incubation for 24h at 37°C (Kutter and Sulakvelidze, 2005). Since SHEFM2K does not form clear plaques on most strains, the plate was visually examined for lysis and plaque formation and assigned a score (++ complete lysis, + indicates turbid lysis, - indicates no lysis). For liquid microtiter well assays, strains were grown to OD_{600} = 1 then diluted to OD_{600} = 0.05 in 96 well plate wells. SHEFM2K phage was added with varying MOI (1000, 100, and 10, calculated using EcM2K) and growth monitored overnight in a Tecan Sunrise plate reader (37°C, with shaking, measurements every 15 min). In all cases experiments were performed in technical triplicate for each condition and biological triplicate to confirm. Representative graphs from one biological replicate are shown and in all cases SEM shown.

Bacterial and phage genome sequencing and bioinformatics

Bacterial DNA extraction was performed using a Wizard Genomic DNA purification kit (Promega, UK). SHEFM2K bacteriophage DNA extraction was performed using a phenol-choloroform method (Al-Zubidi *et al.*, 2019). Genomes were sequenced either in-house using a MinION sequencer (v10, Flongle, Guppy 5.1.17, failed reads were filtered out automatically)

or using the Illumina NovaSeq 6000 platform and 250 bp paired-end technology (only reads passing QC were used in assemblies). For Q4724, G43590, U125544 and U126543, only illumina was performed, these were assembled using SPAdes version 3.7 and annotated using Prokka 1.13 (MicrobesNG (Birmingham UK)). For SHEFM2K and EcM2K, reads were hybrid assembled using Unicycler (Wick *et al.*, 2017) on Galaxy Europe (version 0.5.1 and default parameters). Annotation was via Pharokka (Bouras *et al.*, 2023) or Prokka (Seemann, 2014) on the Galaxy Europe server. PhageLeads (Yukgehnaish *et al.*, 2022) was used to check for lysogeny elements. PubMLST via Galaxy Europe was used for ST-typing alongside Plasmidfinder(25), PlasFlow (Krawczyk, Lipinski and Dziembowski, 2018) and Resfinder for AMR screening within StarAMR (Bharat *et al.*, 2022). Closest relative analysis was performed using PhageClouds(Rangel-Pineros *et al.*, 2021).

Assemblies are deposited in Genbank with the following accession numbers: SHEFM2K Phage: PQ390715; EcM2K: JBHIRZ000000000; Q47424: JBIEKA000000000, G43590: JBIEKA000000000, U125544: JBIEJZ000000000 and U125643: JBIEKA000000000. All other strain genome sequences are publicly available with accession numbers indicated in Figure 1.

Proteomic analysis

Sample Preparation: Bacteriophage were resuspended in Laemmli buffer (20% glycerol, 10% 2-mercaptoethanol, 0.125 M Tris-HCl pH 6.8, 4% SDS) before addition of Triethylammonium bicarbonate (TEAB 1M pH 8.0, Sigma-Aldrich) to obtain a final concentration of 50 mM. After quantification (BCA assay) 20 µg of proteins was used for proteomic analysis. Firstly, samples were reduced (20 mM Dithiothritol (DTT; Merck)) at 90°C (Thermomixer at 800 rpm, 10 mins). The sample was then cooled at room temperature (5 mins) before alkylation (50 mM of Iodoacetamide (IAA, Sigma-Aldrich), 25^{o}C, 5 mins, in the dark). The samples were then acidified using 12% phosphoric acid (1:10 v/v) followed by S-trap binding buffer (90% aqueous methanol, 0.1M TEAB, pH 7.1) at a 7:1 volume.

Samples were then transferred to an S-Trap column (Protifi, USA) and centrifuged for 60 seconds at 4000 x g. The S-Trap was then washed 3 times with 150 µL of binding buffer by centrifugation (60 seconds at 4000 x g), and was then transferred to a clean tube and proteins digested using 25 µL of MS grade Trypsin (Thermo Fisher) in 50 mM TEAB buffer (concentration of 0.1 µg/ µL at ratio of 1:10 (trypsin:protein)), and incubated at 47°C for 2 h. Peptides were then eluted by centrifugation (4000 x g, 60 s) using a series of solvents: 60µL of 50 mM TEAB, 60 µL of 0.2% aqueous formic acid (FA, Thermo Fisher), 60 µL of 50% acetonitrile (ACN) in 0.2% FA and 40µL of 80% ACN in 0.2%FA. Eluted peptides were collected and dried in a vacuum concentrator (Eppendorf, UK) before being reconstituted in 60 µL of 0.5% FA and loaded into a mass spec vial (Thermo Fisher).

MS Analysis: 4 µL of sample was injected on an Orbitrap Exploris E480 mass spectrometer (Thermo, UK) equipped with a nanospray source, coupled to a Vanquish LC System (Thermo, UK). Peptides were desalted online using a nano trap column(75 µm I.D.X 20 mm (Thermo, UK)) and then separated using an EASY-Spray column (50 cm × 50 µm ID, PepMap C18, 2 µm particles, 10 Å pore size (Thermo, UK)) using a gradient as follows: 3%- 20% buffer B (0.5% FA in 80% ACN) for 68 min; ramping up to 35% buffer B for 23 min; 99% buffer 10 min.

The Orbitrap Exploris was operated in positive mode with a cycle of 1 MS acquired at a resolution of 120,000 at m/z 400, with the top 20 most abundant multiply charged (2+ and higher) ions in a given chromatographic window subjected to MS/MS fragmentation in the linear ion trap with scan range (m/z) 375 – 1,200; normalised AGC target 300%; microscan 1. An FTMS target value of $1e^4$ and resolution of 15,000.

Data Analysis: Raw mass spectrometry data was analysed with MaxQuant version 1.6.10.43, and searched against the full predicted proteome of SHEFM2K and EcM2K (.fasta file format) using the following search parameters for standard protein identification: enzyme set to Trypsin/P (2 miss-cleavages), methionine oxidation and N-terminal protein acetylation as variable modifications, cysteine carbamidomethylation as a fixed modification. A protein FDR of 0.01 and a peptide FDR of 0.01 were used for identification level cut-offs based on a decoy database searching strategy. The coverage levels of 5% of the total protein length were used as cutoff values when identifying gene products as components of the viral particle (Casey *et al.,* 2014). The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol *et al.*, 2022) partner repository with the dataset identifier PXD056540.

Cross Sectional TEM

The EcM2K host strain was infected with SHEFM2K phages for 2.5 hrs and samples taken at the indicated time points and pelleted at 4500 x g. TEM embedding and preparation was performed at the electron microscope facility (University of Sheffield, School of Biosciences). Briefly, pellets were then fixed in 3% glutaraldehyde (o/n, 4°C) before washing 3 x PBS. Pellets were soaked in osmium tetroxide (OsO4, 1 hrs, RT) and washed (20 mins) twice in dH_2O before a wash in 75% ethanol, 95% ethanol and 100% ethanol, then dried (each wash 20 minutes). Samples were then soaked in propylene oxide for 20 minutes twice at room temperature. The pellets were then soaked in a 2 ml 50/50 reagent mixture of propylene oxide/ resin (5 ml sy212 resin, 5 ml DSA resin, 10 ml propylene oxide) in sealed containers overnight. Next, pellets were dried for 5 minutes before adding 100% resin (10 ml sy212+ 10 ml DSA + 10 drops R1062) for 4 hrs. Then pellets were then embedded in 100% resin containing 2 drops of R1062 to polymerize in a flat embedding silicon mould (60°C for 24-48 hrs). Embedded pellets were sectioned 75-80 nm and placed onto formvar-coated grids. These cuts were examined by TEM.

Results and Discussion

Phage SHEFM2K isolation, basic characteristics and genome sequence

The strain *E. coli* ExPEC EcM2K was isolated from raw bovine milk (Dairy farm, unpasteurised sample) and the genome sequence was hybrid assembled after short-read (Illumina) and long-read (nanopore) sequencing using Unicycler. This produced 32 assembled contigs with a total of 5.05 Mbp sequence and has been deposited under accession (JBHIRZ000000000). Multi-locus sequence Typing (MLST) and analysis using EcTyper (Bessonov *et al.,* 2021) determined that like many bovine isolates, *E. coli* M2K is an ST446 strain with an O23:H8 antigen profile (**Table 1**). It is predicted to contain a Tet resistance gene and contains two contigs that contain sequences corresponding to likely incF type plasmid replication origins (contigs 6(78.1kb):IncFII and 10(24.1kb): IncFIB(pB171)) while 10 other contigs were identified as plasmid DNA (see file **S1**). These data indicate that 4.93Mb is of chromosomal origin and there are potentially two plasmids in this strain totalling around 133kb. The plasmids do not contain potential AMR genes but do contain a range of virulence genes encoding colicin systems, and UV/Iron/ copper resistance. Whole genome phylogeny indicates that strain EcM2K clusters with other ExPEC strains typical of ST446 (**Fig. 1**). Using EcM2K we isolated a phage displaying clear plaques in double-layer agar from a raw milk BHI enrichment, which was then purified (**Fig. 2A**). This phage was imaged by TEM (**Fig. 2B**) revealing a large phage with a head diameter of 131 nm (measured from >20 images of non-contracted and contracted images). Image analysis revealed the presence of two forms of the phage, an extended form of 121.3 ± 1.8 nm (n = 44) (**Fig. 2C**) and a contracted form of 73.5 ± 5.5 nm (n = 16), with a 45 nm protruding section that is often attached to a piece of cellular material which in other jumbophage images has been suggested to be host membrane (Wagemans *et al.*, 2020). These TEM images also revealed the presence of thin baseplate and neck fibres that have been found in other *E. coli* jumbophage (Wagemans *et al.*, 2020).

Genome sequencing, assembly (hybrid Illumina/ Nanopore, Unicycler) and subsequent phylogenetic and bioinformatic analysis revealed SHEFM2K to be a new member of the *Asteriusvirus* genus with a genome size of 348 kb. SHEFM2K has 98.1 identity with the *E. coli* phage SP27 and is also closely related to PBECO-4(Kim *et al.*, 2013), PhAPEC6(Wagemans *et al.*, 2020), slurp01(Sazinas *et al.*, 2016) and others **(Fig. 3B, S1)**. The genome contains 663 predicted ORFs and contains clear head, tail and potential lysis modules **(Fig. 3A)**, 6 tRNAs and several predicted metabolic and DNA modification and RNA polymerases noted in other related phages (a full list of ORFs is included in Table S2). No integrase or other lysogeny genes were identified.

In terms of jumbo/ giant-phage the other genus that infects *Enterobacteria* is exemplified by phage Goslar and related Chimallinviridae phages(Prichard *et al.*, 2023), a jumbophage which have been shown to build an internal proteinaceous 'nuclear' compartment using the novel ChmA/ PhuN protein in concert with a network of accessory proteins (Chaikeeratisak *et al.*, 2017; Enustun *et al.*, 2023; Fossati *et al.*, 2023; Nieweglowska *et al.*, 2023). However, analysis of the SHEFM2K and other *Asteriusvirus* sequences does not reveal a homologue of Chimallin or the PhuZ-like tubulin and neither is there any knowledge of how these phages proceed through the infection cycle or if they share characteristics with *Goslarviruses*.

Phage SHEFM2K host range and infection

To begin to examine determinants of infection and broaden our knowledge of the *Asteriusviruses*, we examined the host range of SHEFM2K by infection of a range of *E. coli* clinical and laboratory strains by spot assay (**Table 1**). These data revealed that in addition

to the isolation of the host, with which SHEFM2K makes small clear plaques, SHEFM2K only produces clear (countable) plaques with a clinical sepsis strain from our collection (G43590, O75:H42(ST2223)). In the cases where clear plaquing was not observed, several strains displayed clearing in a spot assay with high titre preparations ($6.4X10^8$ PFU/ml)- including *E. coli* K-12 strain MC1000, *E. coli* BL21 (B-strain) and *E. coli* EHEC (O157:H7). In contrast, SHEFM2K did not infect EPEC E2348/69 or three clinical sepsis strains isolated from blood culture in Sheffield (Q4724 (ST127), U125544 (UPEC, ST73), and U125643 (UPEC, ST550) (**Table 1**). Pertinent details of ST-type and O- H- antigen typing are included in Table 1 and accession numbers in the methods section. Additionally, no infection was observed with K1 capsule strain EV36 (Møller-Olsen *et al.*, 2020).

Given these issues, we used liquid growth infection assays (**Fig. 4**) to examine infection dynamics in more detail. These data confirmed infection of the EHEC strain O157:H7 (NCTC12900) as well as G43590 (Fig. S3), while U125643 (UPEC, ST550) displayed no growth suppression.

Examination of the growth characteristics of EcM2K in liquid killing assays with SHEFM2K also revealed suppression of growth at an MOI of 10 and 100 at 4 hrs. A similar OD to untreated EcM2K is reached after 13 hrs, with sustained growth suppression only occurring with an MOI of 1000 (**Fig. 4A**). A similar situation occurs with the NCTC12900 (EHEC) and BL21 strain (**Fig. 4B and D**). However, despite not plaquing, the K12 strain MC1000 was suppressed for a longer period than EcM2K at MOI=10 and completely at MOI=100 (**Fig. 4**). N.B. we obtained similar data for K12 strain MG1655 (data not shown).

Overall, these growth curves may indicate the possibility of early onset of resistance or induction of phage defence systems, although no phage defence systems for *E. coli* jumbophage have yet been identified. We are currently investigating these scenarios. Of note, SHEFM2K does share some host range similarities with its close relatives pHAPEC6 and 121Q- which were isolated from an APEC (exPEC) strain (Buttimer *et al.*, 2017; Wagemans *et al.*, 2020) and Slurp01 was isolated using MG1655 (Sazinas *et al.*, 2016), while PBECO-4 (Kim *et al.*, 2013) (Kim *et al.,* 2013) with EHEC (O157:H7), although a broader host range was not tested in those studies. Finally, we cannot find any obvious commonalities between phylogeny, ST-type or O/H antigens derived from the genome sequences of these strains and infectivity by M2K, it is of note that in general *E. coli* jumbophages do not infect all *E. coli* and that their recognition determinant remains unknown.

Is Phage SHEFM2K flagellotrophic?

During our characterisation of purified SHEFM2K phage, we frequently observed phage associated with fibres that we surmised could be flagella (**Fig. 2B, Fig. 5A**). To probe this further and examine the receptor for SHEFM2K, we visualised its interactions with EcM2K cells and imaged them directly on TEM grids with negative staining during the early stages of infection (0-15 mins). These images are striking in that the phage seems to be lined up in contact with flagella filaments that are attached to the bacterial cell (**Fig. 5A, Fig. S2**). In some cases, it appears that the interaction occurs via loose association with the head, and in others, it appears that small 'legs' associated with the baseplate are interacting with the flagella fibres or the shaft fibres (**Fig. 5A**). Of note, we also saw this interaction in cross-sectional TEM experiments (**Fig. 7**).

To further test how the potential interaction with flagellin influenced infection of SHEFM2K, we tested infection of an MC1000 (K12) strain lacking *fliC* (flagellin) (Green *et al.*, 2019). Unlike the wild-type MC1000, growth of the Δ*fliC* strain was not suppressed by SHEFM2K at MOI of 10 or 100 **(Fig. 5B)**. Furthermore, deletion of the flagella hook (Δ*CKLDE*) was also insensitive to infection at the same MOI **(Fig. S3)**. To further probe the determinants of

interaction with flagella, we also tested infection of a strain expressing the entire FliC protein and a version of FliC lacking the central antigenic domain (D2-D3, ∆237-331) (previously characterised as being fully motile) under the control of an IPTG inducible promoter(Green *et al.*, 2019). Importantly, sensitivity to infection was restored by complementation of the *fliC* gene *in trans* **(Fig. 5B, S3)**. Surprisingly, infection of cells expressing the ∆137-331 FliC were even more sensitive to infection than cells expressing wild-type FliC **(Fig. 5B, Fig. S3)**, indicating that binding may involve the core D0-D1 domains with the central (D2-3) domain not being required. These data collectively suggest that the flagella binding of SHEFM2K is involved in SHEFM2K adsorption and infection, but is not the only determinant of binding for productive infection. While this was unexpected, it is known that other flagellotrophic phages use fibres attached to the head to access the flagellum, e.g. phage CbK from *Caulobacter crescentus* (Guerrero-Ferreira *et al.*, 2011). In terms of jumbophage, a recent study highlighted that resistant strains of *Klebsiella aerogenes* to the jumbophage fENko-Kae01 (370kb genome) were less motile and contained mutations in flagella synthesis genes (Ranta, Skurnik and Kiljunen, 2024). However, in all known cases, it is thought that flagellotrophic phage have a secondary receptor on the host cell surface, for example Salmonella phage χ binds flagella before outer membrane protein TolC (Esteves *et al.*, 2021; Esteves and Scharf, 2022). Our data would suggest the SHEFM2K also has a secondary receptor given some infection still occurs in a *ΔfliC* mutant. Of note, we often observed uncontracted phage bases seemingly interacting with flagella fibres or tangled up in flagella bundles (see **Fig. S2**), while contracted particles were often attached via the head. All of these interaction mechanisms are displayed by different phages, but at present, we have no data confirming this with SHEFM2K.

Examination of the life-cycle of SHEF2K jumbophage

As outlined above, one of the features of the lifecycle of jumbophage of the *Goslarvirus* group is the formation of a distinct intracellular compartment bounded by a protein shell composed of the protein chimallin. This forms at the centre of the bacterial cell and is accompanied by a tubulin-like treadmill structure that is the location of phage assembly (Guan and Bondy-Denomy, 2020; Prichard and Pogliano, 2024).

To investigate the life cycle of the SHEFM2K *Asteriusvirus* jumbophage, we performed infections of the EcM2K host and stopped the infection by fixing in glutaraldehyde between 15 and 150 minutes before embedding in resin, thin-sectioning and staining and visualisation by TEM. As already intimated, SHEFM2K is a flagellotrophic phage, an observation reinforced by our discovery of several phage attached with flagella fibres to the outside of EcM2K (**Fig 5A, Fig. S2**). After initial attachment we observed several phage more intimately attached, which we assume were in the process of injecting their DNA (**Fig. 6-2**). From 30 minutes onwards we see phage heads forming in cleared areas of the cytoplasm which we surmise to be a putative phage assembly area (**Fig. 6-3**), indeed the images suggest heads at different levels of maturity (darker or lighter). These areas are depleted of ribosomes and seem to have 'clear' areas around the phage heads, but at the resolution of our images we cannot tell if any firm boundary is present. There are also thin dark 'fibres' of web-like material in a number of images in these areas (**Fig. 6-4, Fig. S4**). In **Fig. 6-4** we also see clear head-and-tail structures that look like maturing full virions. Finally, from 30-60 minutes we observe the appearance of lysed cells (**Fig. 6-5**).

The data gained here clearly show that SHEFM2K follows a typical pattern of infection through lysis, accompanied by formation of a cleared putative 'assembly area' in the cytoplasm. In contrast to Goslar-type jumbophage we have no evidence so far of cell bulging at mid-cell or of formation of a defined 'phage-nucleus'; however the resolution of our sectioned-resin embedded TEM images are not as high nor result in the same preservation levels as the cryo-EM employed in studies of goslar-type viruses (Birkholz *et al.*, 2022; Prichard *et al.*, 2023).

When comparing the SHEFM2K images to those of other phage reproducing inside cells (e.g. Shigella -phage sfk20 (Mallick, Mondal and Dutta, 2021), T3(Serwer, Hunter and Wright, 2020), phiKz(Krylov *et al.*, 2021), Bp7(Zhang *et al.*, 2013)) several features are notable: a) the number of phage heads assembling seems to be lower, at 6-10 vs 20-50 for smaller phage (although we acknowledge we have not established burst size) and b) SHEFM2K causes significant cytosolic clearing around the phage head. Of note, a recent resin-embedded image of the nucleus forming phiKZ (Danilova *et al.*, 2020; Naknaen *et al.*, 2024) is reminiscent of SHEFM2K images shown here. However, it remains to be seen whether Cryo-EM would reveal more about the internal architecture of this replication area. Furthermore, it is not known which SHEFM2K proteins are expressed during infection or clearing of the cytoplasm to direct this reorganisation.

The proteome of SHEFM2K phage particle

To investigate SHEFM2K biology further and establish the composition of the mature SHEFM2K phage particle we performed an infection (2h) of the EcM2K host strain before removing whole bacteria by centrifugation (4000 x g), filtration (0.45µm filter) and concentration of whole phage particles by ultracentrifugation at (25,000 rpm) (1h). Proteomic analysis identified a total of 1,728 proteins (≥ 2 peptides) with 277 able to be mapped to Phage proteins (42% of putative proteome, suppl file 2) and 1,451 EcM2K host proteins with their abundance determined by iBAQ value (Tyanova, Temu and Cox, 2016) (**Fig. 7A**). Of the phage proteins we were able to estimate that 57% of the protein was made of predicted head proteins with a further 20% attributable to Tail or Tail/lysin functions (**Fig. 7B**). The most abundant proteins were M2K_0492 (16%), M2K_0507(T4-capsid like,10%) and M2K_0493 (8%), which likely make up the capsid/ portal. Of note, M2K0445, a predicted tail lysin made up 8% of counts, suggesting it is a high abundance protein, possibly forming part of the base/ tail structures (**Fig. 7C**). Also of note, a putative ribosomal methylase (M2L_ 0014, 1.5%) was the only non-structural protein detected above 1% as well as the putative tail/spike colanidase protein (M2K_0543) identified in phage PBECO4 (Kim *et al.*, 2013) (0.4%), suggesting it is part of the final virion and may be involved in interactions with colanic acid capsules- common in *E. coli*. Our data add to those from the related phAPEC6 phage in which a chloroform extracted phage proteomics experiment detected 62 proteins but without relative quantification (Wagemans *et al.*, 2020).

In addition to phage proteins, we also detected 1451 EcM2K host proteins, which the iBAQ data indicates are predominantly outer membrane proteins such as OmpA, OmpC and OmpF (**Supplementary file 2A-E**). These proteins are also present in Outer membrane derived vesicles(Berlanda Scorza *et al.*, 2008; Kim *et al.*, 2018) which we would also expect to precipitate and co-purify with the phage using the centrifugation based protocol as employed here. However, the recent finding that Goslar-like jumbophages encapsulate their injected genomes in bacterial membrane derived lipid vesicles could mean this finding is significant(Antonova *et al.*, 2024; Mozumdar *et al.*, 2024). Hence, we do not suggest these are receptors for SHEFM2K, rather are co-incidentally enriched with SHEFM2K particles.

The proteomics data also reveal over 380 proteins that are not detected and a further 185 detectable at less than 0.1% of total phage protein detected (see **Supplementary file 2E**). We postulate that these may be involved at other stages of the phage life-cycle, although at this stage we do not know when or for what purpose. Of note, 54 of these proteins have a predicted function in nucleotide metabolism or replication and regulation (S**upplementary** file 2E) and another 19 predicted to be involved in phage assembly, suggesting a possible role in SHEFM2K replication and internal assembly. Another 495 of the undetected proteins have no known function- highlighting the potential for the discovery of new biology involved in the replication of these unique group of jumbophages. Of note, the core genome of goslar-type (chimallivirus) has 7 proposed core segments (Prichard *et al.*, 2023), with the nuclear shell

proteins contained in gene loci with RNA-polymerase and other replicative genes. However, examination of the genome of SHEFM2K reveals no putative chimallin proteins, and furthermore alignment of genomes with phage-goslar reveals no homology across the whole genome, indicating the potential for new biology in the *Asteriusvirus* genus.

Conclusions:
In this paper we report the isolation of a new member of the *Asteriusvirus* genus, SHEFM2K. Our data further emphasize the widespread nature of jumbophages in the environment and establish their ability to infect a range of environmental, hospital and well-characterised lab strains across a range of flagellin (H) and LPS (O) serotypes, including EHEC, ExPEC and K12 lab strains, but not K1 capsule types or several clinical strains tested. We uncovered the intriguing possibility that Asteriusvirus defence systems may also be present in some strains given our data on growth suppression being overturned during culture for some strains. We also establish, for the first time, that an Asteriusvirus phage utilises flagella fibres during its attachment to the host cell surface, but that this is not the sole determinant of infection. Finally, host infection by SHEFM2K locally clears the cytoplasm of the host, forming putative assembly areas in the cytoplasm before lysis. Overall, our study accelerates research on this important group of jumbophage and sets a platform for future work on their biology.

Acknowledgements:
GA was supported by a postdoctoral scholarship from Prince Sattam Bin Abdulaziz University ES is funded by an NIHR-BRC PhD and KP by a BBSRC studenship (BB/T007222/1). AA, was supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1446). Mass spectrometry analysis was performed by TKP at the Biological Mass Spectrometry Facility at the University of Sheffield. We also acknowledge the team from 'Our Cow Molly' at Cliffe House farm for providing samples and guidance in herding cattle (https://www.ourcowmolly.co.uk/) as well as the team from Sheffield TEM facilities, esp. Mr Chris Hill.

Author contributions: CRediT
GAS: Investigation, Data curation, formal analysis, methodology, Writing- original draft, Visualization,Writing- review and editing.
TKP: Investigation, Data curation, formal analysis, methodology, Writing- review and editing
AC: Investigation, Data curation, Writing- review and editing
KP: Investigation, Data curation, Writing- review and editing
ES: Investigation, Data curation, Writing- review and editing
AA: Investigation, Data curation, Writing- review and editing
CS: Data curation, Writing- review and editing
GS: Conceptualisation, Data curation, formal analysis, methodology, Writing- original draft, Visualization,Writing- review and editing, Project administration.

References:

Figure legends:

Figure 1: Phylogenetic tree of *E.coli* EcM2K with indicated related strains, including clinical sequenced as part of this study. Tree was generated using TYGS (https://tygs.dsmz.de/) and was inferred with FastME 2.1.6.1 [7] from GBDP distances calculated from whole

genome sequences and drawn using FigTree. The branch lengths are scaled in terms of GBDP distance formula d_5. The tree is rooted at the midpoint.

Figure 2: A. Plaques identified in lawns of E. coli Strain M2K in double layer agar. B. TEM of Phage SHEFM2K negatively stained using 2% uranyl acetate and imaged using an FEI Tecnai G2 Spirit transmission electron microscope, revealing a contractile tail morphology. C. Measurement and distribution of native (121.3±1.8nm, n=44) and contracted (73.5 ± 5.5 nm, n=16) forms of Phage M2K and displayed as a box plot with t-test results
.

Figure 3: A) Circular representation of SHEFM2K phage genome. Colour key indicated. Genbank accession: PQ390715
B) Proteome-based VICTOR tree of SHEFM2K compared with 8 related *E.coli* jumbophage from Asteriusvirus/ Goslarvirus viral groups. Scale bar indicates interproteomic distance inferred using the distance formula d4.

Figure 4: Infection curves for indicated *E.coli* strains with SHEFM2K phage at an MOI as indicated in brackets. All assays used triplicate cultures at A600 of 0.05, grown in LB with SHEFM2K added at t=0. One biological replicate from three is shown. SEM of technical triplicates is shown.

Figure 5: A) (Top left panel) Images of Phage SHEFM2K with co-purified flagella filaments after PEG precipitation of liquid lysates and (lower) Phage M2K incubated for 30 minutes with *E.coli* M2K showing association of PhageM2K with flagella fibres.
B) Infection curves for indicated *E.coli* strains with SHEFM2K phage at an MOI as indicated in brackets. All assays used cultures at A600 of 0.05, grown in LB with SHEFM2K added at t=0, and IPTG as indicated. SEM is shown.

Figure 6: Ultra-thin section TEM (80nm) capturing SHEFM2K lifecycle of infection of EcM2K. Samples were taken over 2.5 hours and fixed in glutaraldehyde before staining (see methods). Images 1-7 illustrate the stages from Attachment (1., 15mins) and Infection (2. 15 mins) of DNA through to Assembly (3-4, 30 mins) and lysis (5, 60 mins).

Figure 7: A) Schematic summary of Phage proteome data by number of Phage and Bacterial host proteins to which ≥2 peptides were mapped, and intensity Based Absolute Quantification (iBAQ) %. B) Mapping of iBAQ counts for phage proteins according to prediction outlined in C) Table summarizing all phage proteins detected at >0.4% of total iBAQ for Phage proteins.

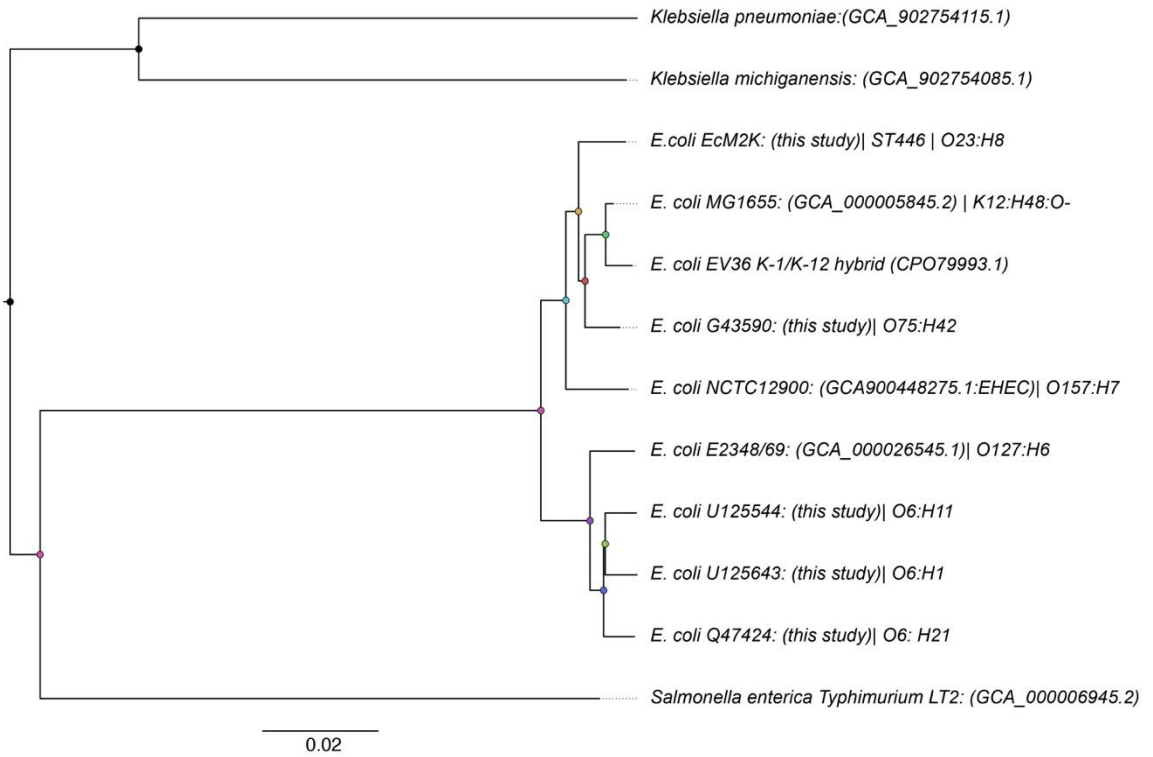


Figure 1. Phylogenetic tree of E.coli EcM2K with indicated related strains, including clinical sequenced as part of this study. Tree was generated using TYGS (https://tygs.dsmz.de/) and was inferred with FastME 2.1.6.1 [7] from GBDP distances calculated from whole genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_5. The tree is rooted at the midpoint.

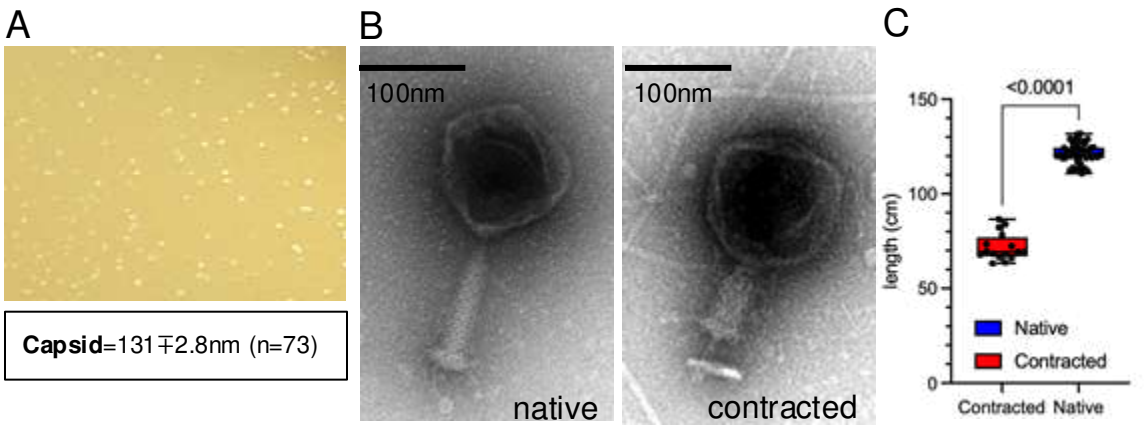


Figure 2: A. Plaques identified in lawns of E. coli Strain M2K in double layer agar. B. TEM of Phage SHEFM2K negatively stained using 2% uranyl acetate and imaged using an FEI Tecnai G2 Spirit transmission electron microscope, revealing a contractile tail morphology. C. Measurement and distribution of native (121.3±1.8nm, n=44) and contracted (73.5 ± 5.5 nm, n=16) forms of Phage M2K and displayed as a box plot with t-test results.

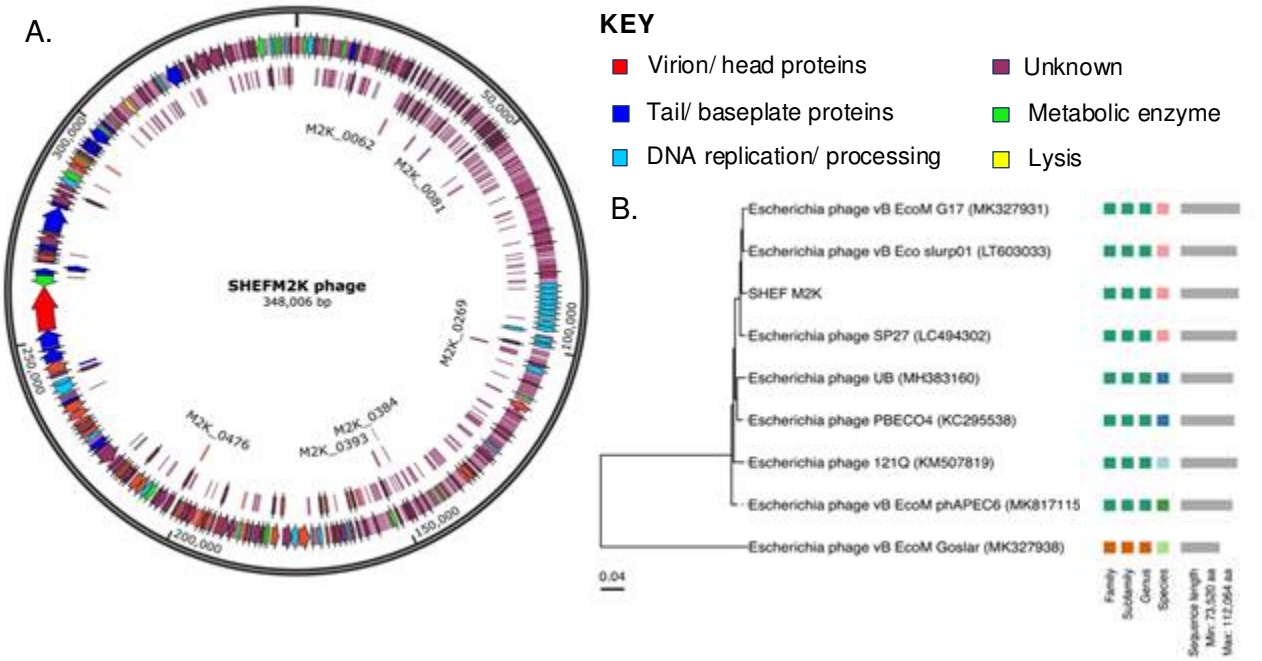


Figure 3: A) Circular representation of SHEFM2K phage genome. Colour key indicated. Genbank accession: PQ390715

B) Proteome-based VICTOR tree of SHEFM2K compared with 8 related *E.coli* jumbophage from Asteriusvirus/ Goslarvirus viral groups. Scale bar indicates interproteomic distance inferred using the distance formula d4.

Fig. 4

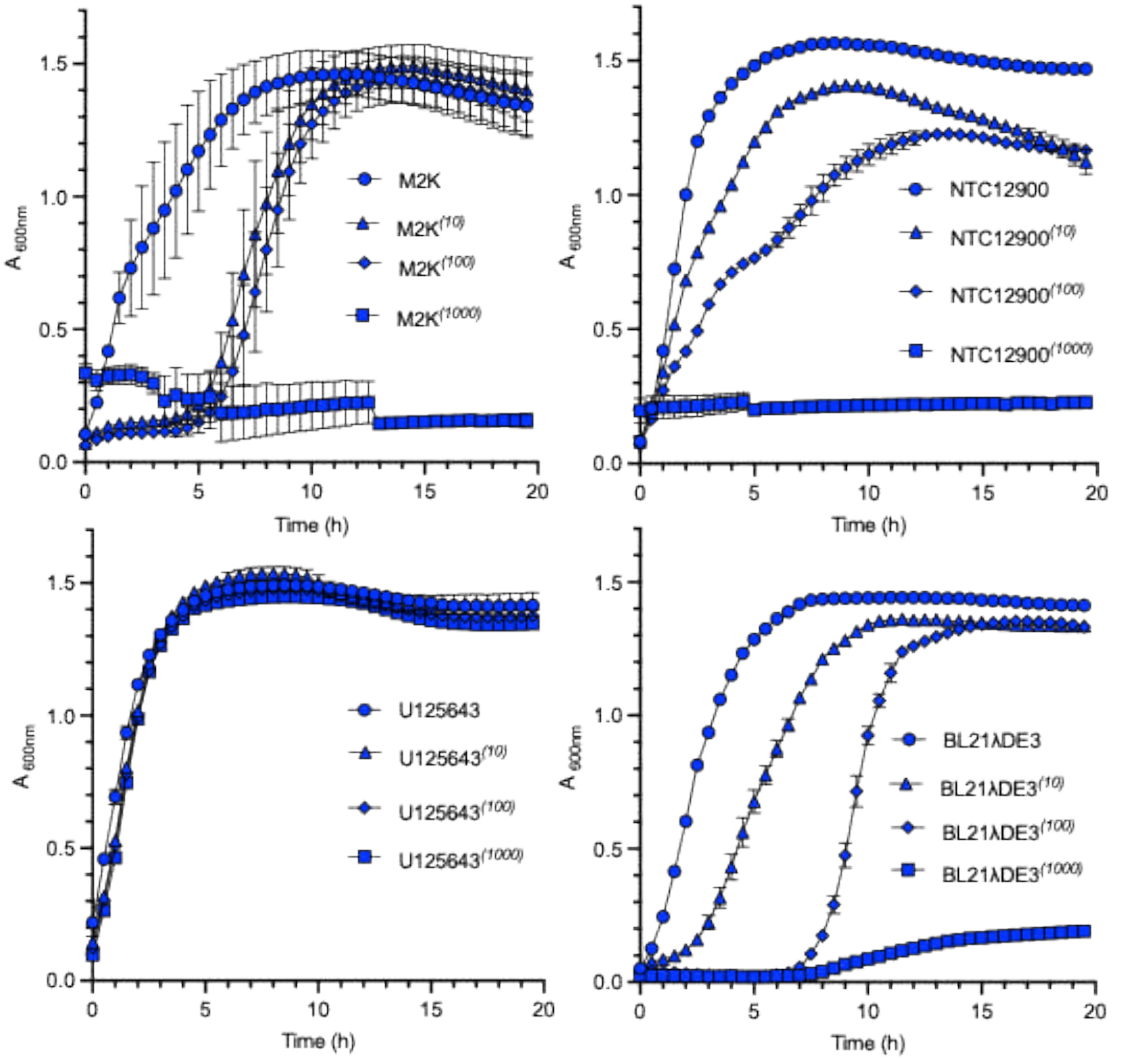


Figure 4. Infection curves for indicated *E.coli* strains with SHEFM2K phage at an MOI as indicated in brackets. All assays used cultures at A600 of 0.05, grown in LB with SHEFM2K added at t=0. SEM is shown.

Figure 5

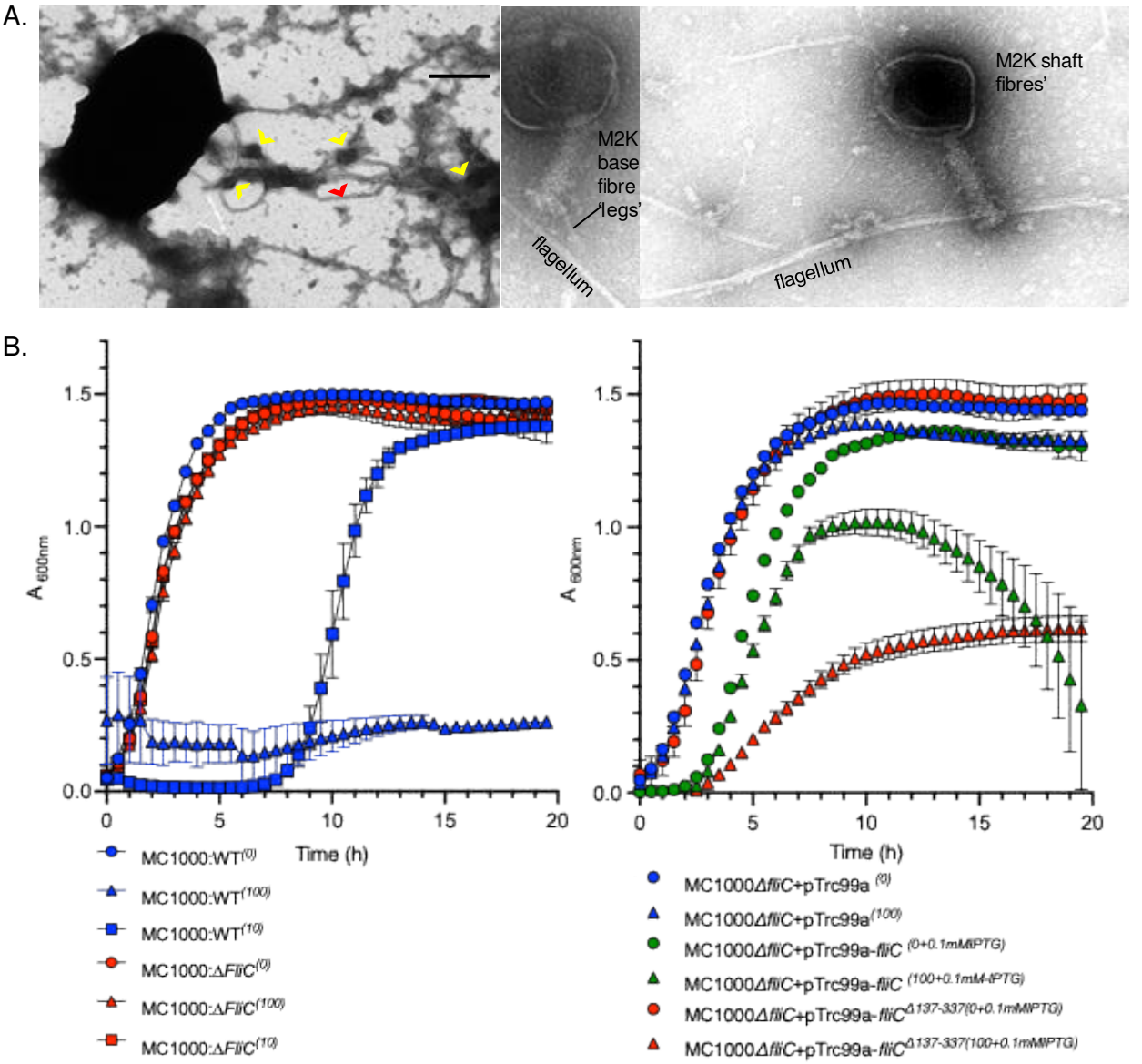


Figure 5 A) (Top lef panel) Images of Phage SHEFM2K with co-purified flagella filaments after PEG precipitation of liquid lysates and (lower) Phage SHEFM2K incubated for 30 minutes with E.coli EcM2K showing association of PhageM2K with flagella fibres, highlighted with yellow arrowheads. B) Infection curves for indicated E.coli strains with SHEFM2K phage at an MOI as indicated in brackets. All assays used cultures at A600 of 0.05, grown in LB with SHEFM2K added at t=0, and IPTG as indicated. SEM is shown.

Fig. 6

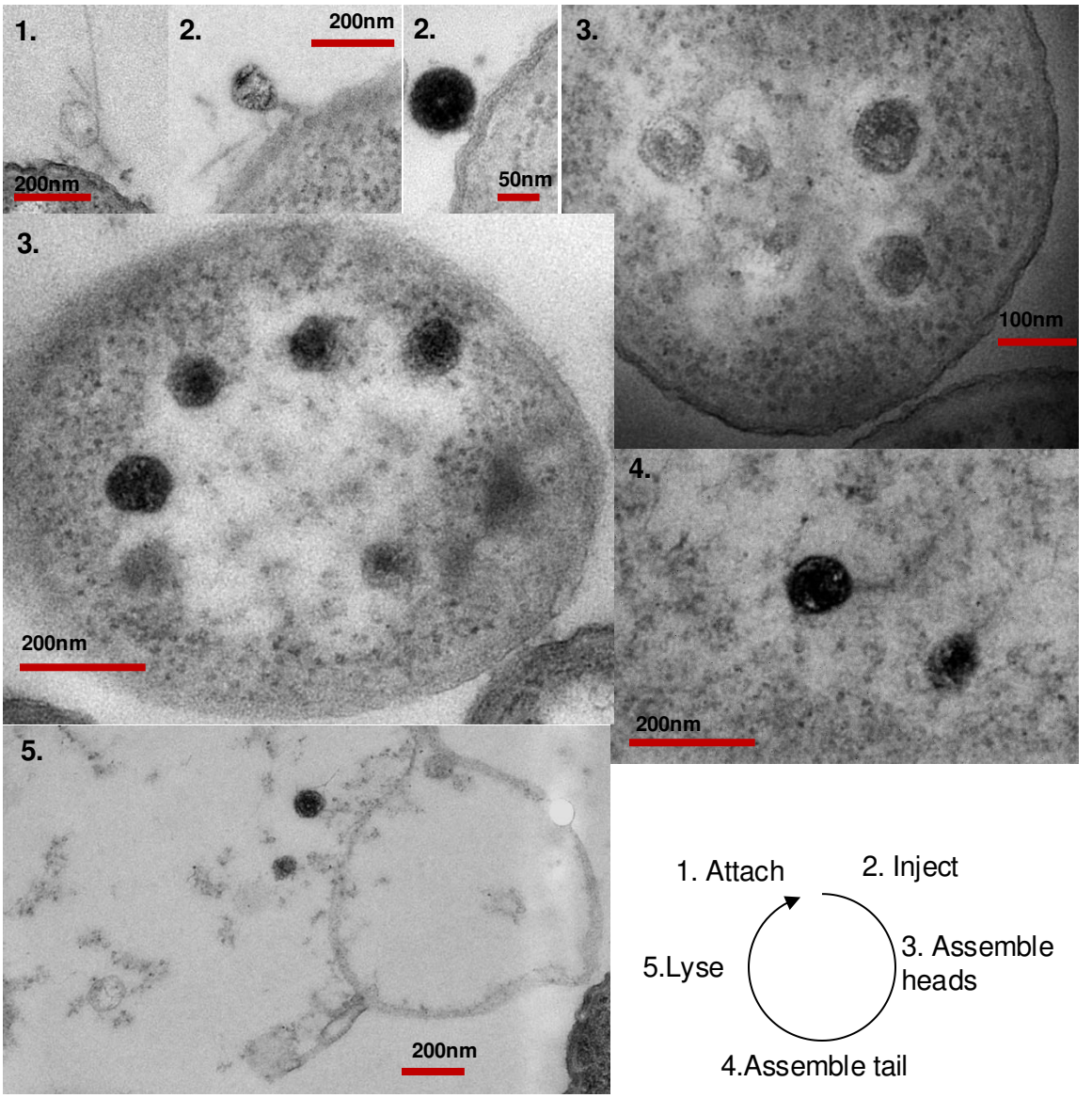
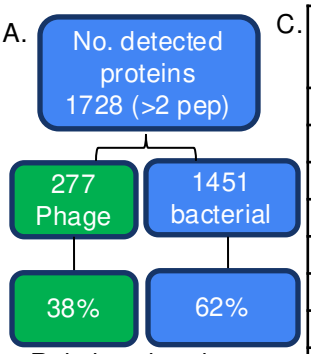


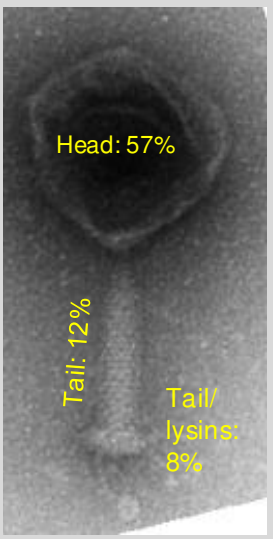
Figure 6. Ultra-thin section TEM (80nm) capturing SHEFM2K lifecycle of infection of EcM2K. Samples were taken over 2.5 hours and fixed in glutaraldehyde before staining (see methods). Images 1-7 illustrate the stages from Attachment (1., 15mins) and Infection (2. 15 mins) of DNA through to Assembly (3-4, 30 mins) and lysis (5, 60 mins).

A.



B.

IBAQ counts (% phage peptides) by Phannotate prediction/ location



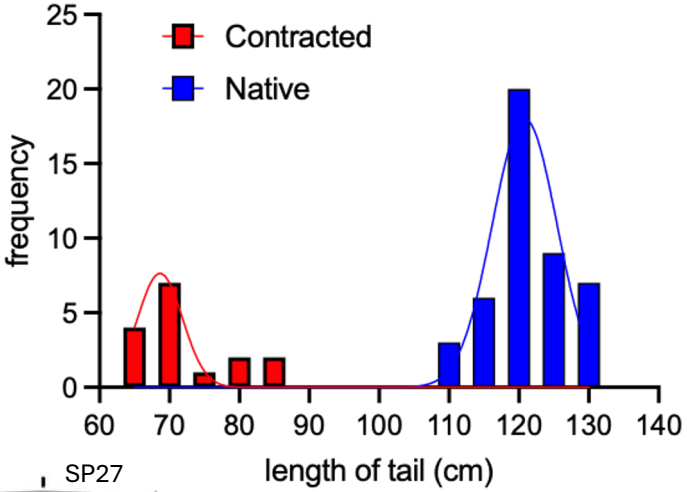
Replication: 1.45%
Assembly: 1.83%
Metabolic: 1.45%
Unknown: 2.50%

C.

ID	Predicted function	IBAQ % Phage peptides	Phannotate prediction: location	Mw [kDa]
M2K_0492	Virion structural protein	15.98	Structure:Head	20.9
M2K_0507	Major head protein, T4 capsid homology	10.48	Structure:Head	42.2
M2K_0493	Virion structural protein	8.20	Structure:Head	24.3
M2K_0445	Putative amidase	7.06	Structure:Tail lysin	20.6
M2K_0656	Virion structural protein	5.73	Structure:Head	14.1
M2K_0510	Virion structural protein	3.83	Structure:Head	31.2
M2K_0545	Tail fiber protein	3.75	Structure:Tail	180.4
M2K_0298	Virion structural protein	3.36	Structure:Head	14.8
M2K_0565	Tail tube protein	3.02	Structure:Tail	25.1
M2K_0567	Tail sheath	2.29	Structure:Tail	97.1
M2K_0559	Hsp70 heat shock protein	1.83	Assembly	20.0
M2K_0544	Virion structural protein	1.59	Structure:Tail	31.4
M2K_0014	prmA, 50S ribosomal methyltransferase	1.45	Replication	22.6
M2K_0560	Virion structural protein	1.44	Structure:Head	20.6
M2K_0542	Tail protein	1.10	Structure:Tail	41.0
M2K_0322	Putative DNA-binding	1.01	Unknown	8.8
M2K_0296	Virion structural protein	0.94	Structure:Head	56.6
M2K_0428	Virion structural protein	0.92	Structure:Head	21.4
M2K_0346	Virion structural protein	0.87	Structure:Head	42.9
M2K_0529	Baseplate hub subunit and tail lysozyme	0.86	Structure:Tail lysin	18.7
M2K_0460	Virion structural protein	0.83	Structure:Head	21.5
M2K_0464	Virion structural protein	0.80	Structure:Head	23.4
M2K_0420	Virion structural protein	0.80	Structure:Head	21.0
M2K_0646	PhoH-like phosphate starvation-inducible	0.78	Metabolic	53.2
M2K_0539	Hypothetical protein- (head module)	0.76	Unknown:(head?)	22.0
M2K_0538	Virion structural protein	0.74	Structure:Head	40.3
M2K_0627	Anaerobic ribonucleoside reductase	0.68	Metabolic	67.8
M2K_0547	Hypothetical protein	0.68	Unknown	70.2
M2K_0509	Head maturation T4 protease	0.60	Structure:Head	23.0
M2K_0512	Portal protein	0.57	Structure:Head	64.6
M2K_0557	Virion structural protein	0.56	Structure:Head	35.8
M2K_0543	Colanic acid degradation spike lysin	0.42	Structure:Tail /spike	77.7
M2K_0503	Tail fiber protein	0.42	Structure:Tail	56.7
M2K_0558	Virion structural protein	0.40	Structure:Head	56.8

Figure 7 A) Schematic summary of Phage proteome data by number of Phage and Bacterial host proteins to which >2 peptides were mapped, and Isometric Absolute Quantification counts (IBAQ) %. B) Mapping of IBAQ counts for phage proteins according to prediction outlined in C) Table summarizing all phage proteins detected at >0.4% of total IBAQ for Phage proteins.

Figure S1



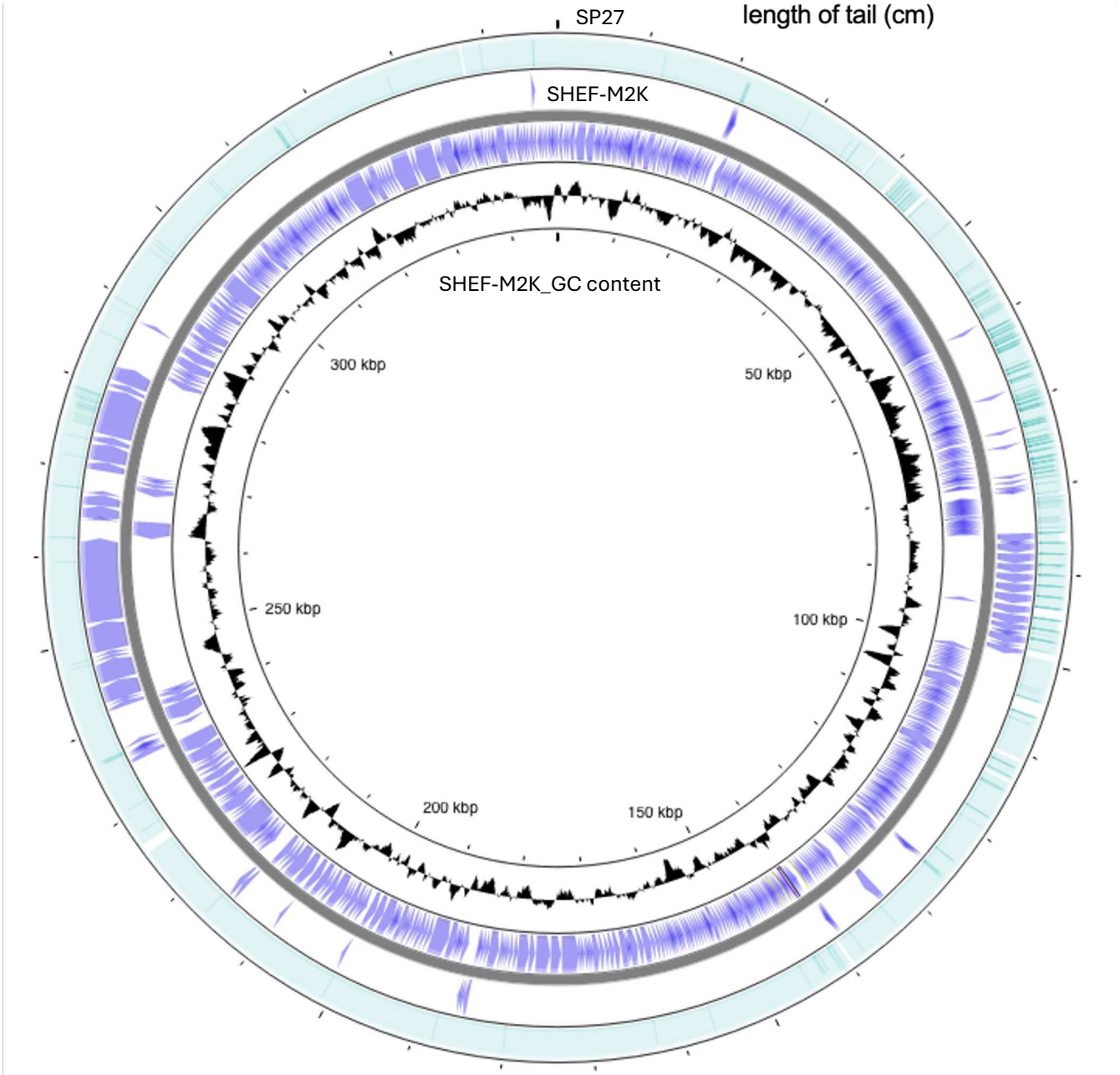


Figure S1: histogram showing distribution of tail length measured from TEM images (upper). Genome of SHEFM2K in alignment with it's closest relative, SP27, generated using Proksee using BLAST function.

Figure S2

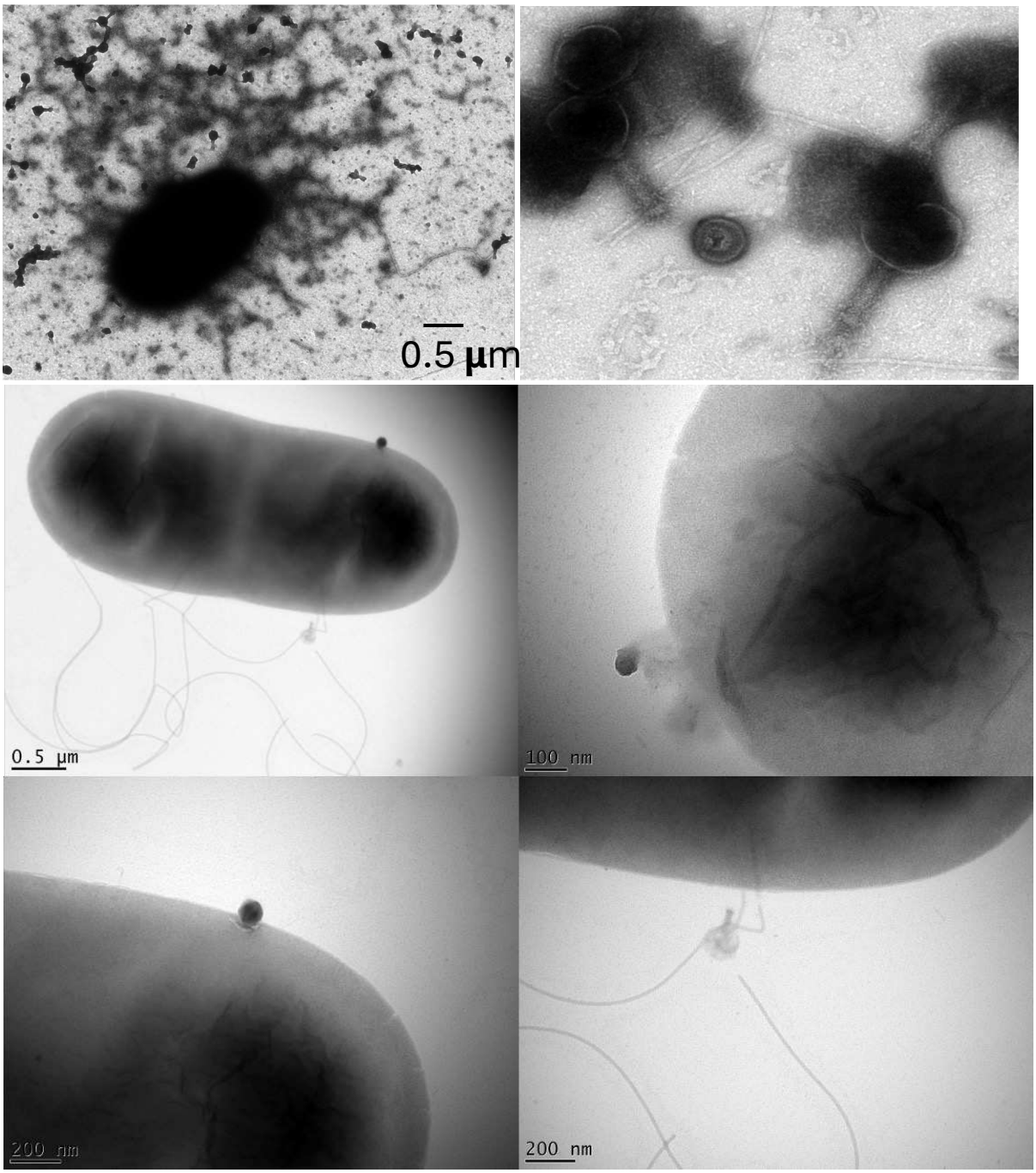
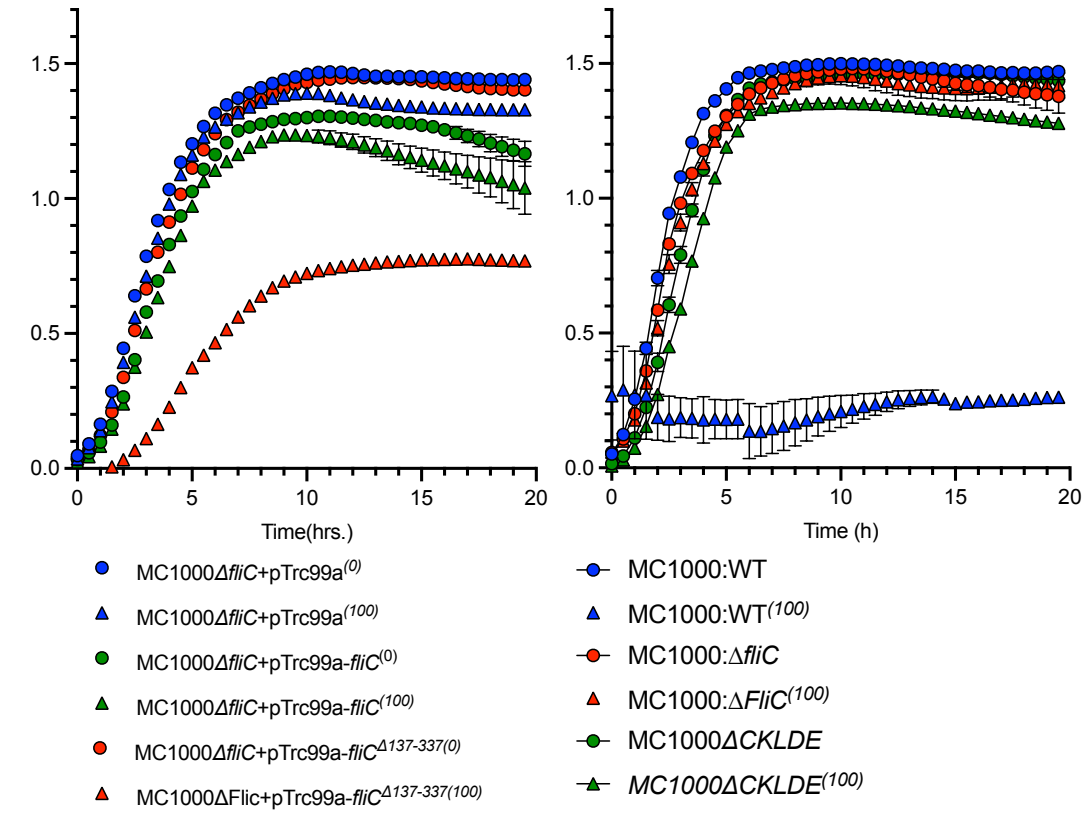


Figure S2. TEM images of negatively stained SHEF-M2K and host strain M2K. Phage SHEFM2K (MOI10) added to host strain M2K for ten minutes at room temperature TEM images show negatively stained using 2% uranyl acetate and imaged using an FEI Tecnai G2 Spirit transmission electron microscope. A. showing is an embedded phage head in the host membrane and another approaching phage attached to the flagella. B. Shown is a phage attaching to the membrane closer image in C. in D. shown is the another closer look of the phage attaching to the flagella.

Fig. S3



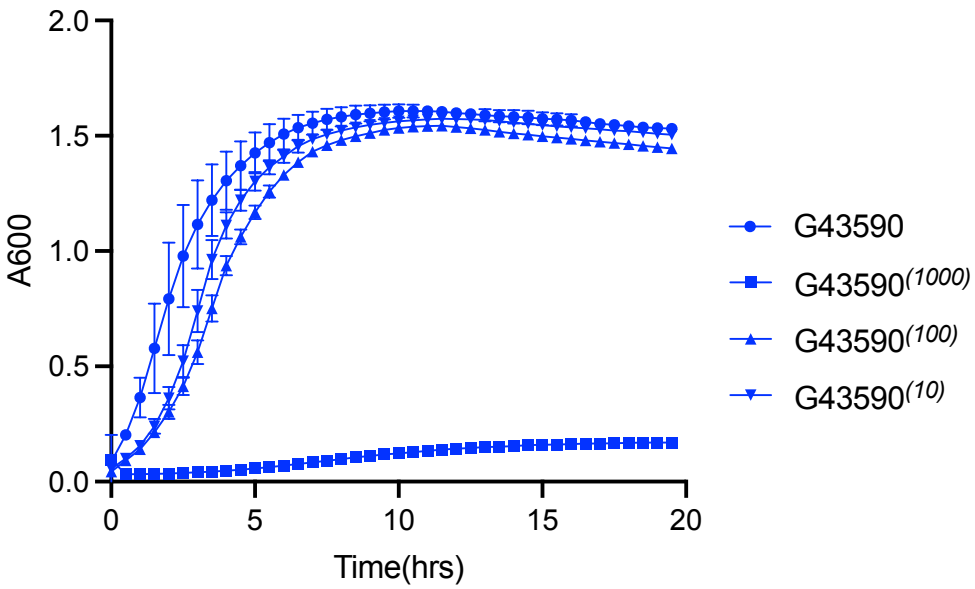
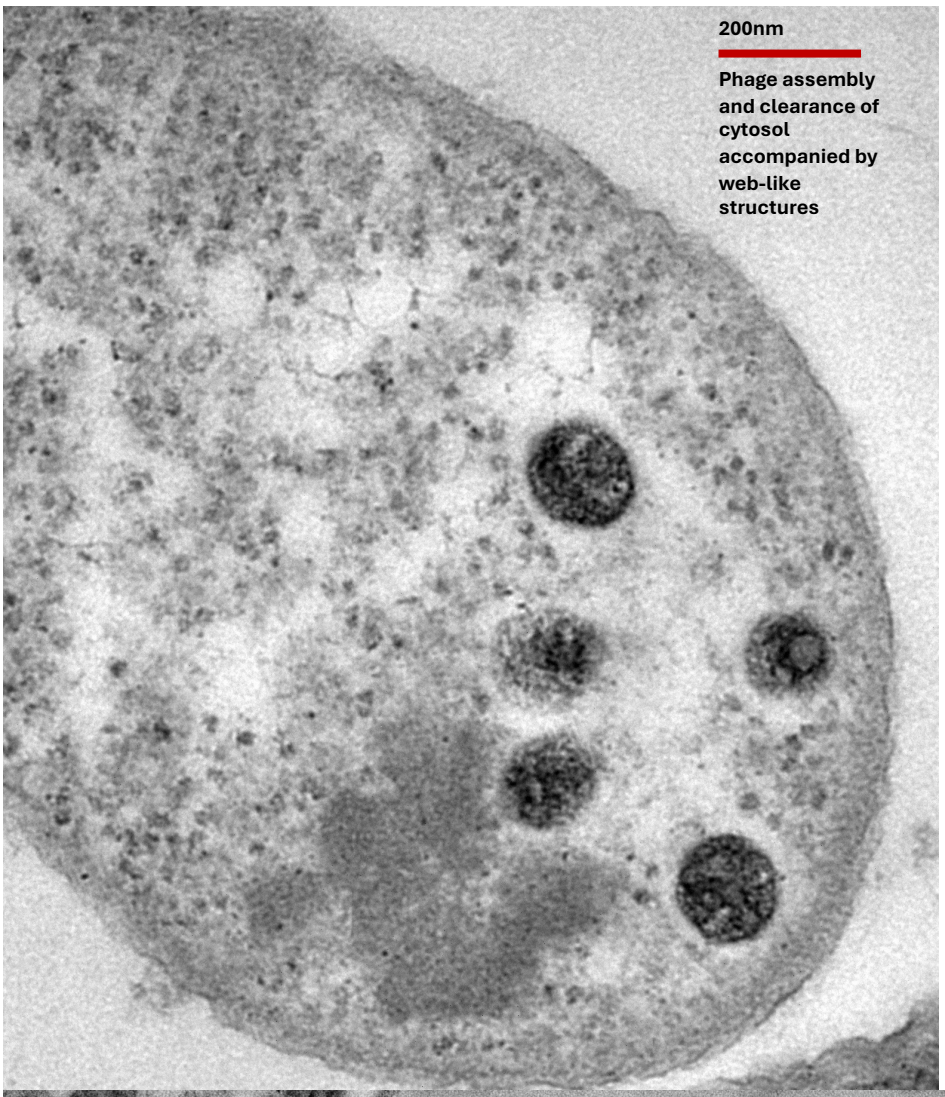


Figure S3 Infection curves for indicated *E.coli* strains with SHEFM2K phage at an MOI as indicated in brackets. All assays used cultures at A600 of 0.05, grown in LB with SHEFM2K added at t=0. SEM is shown.

Figure S4



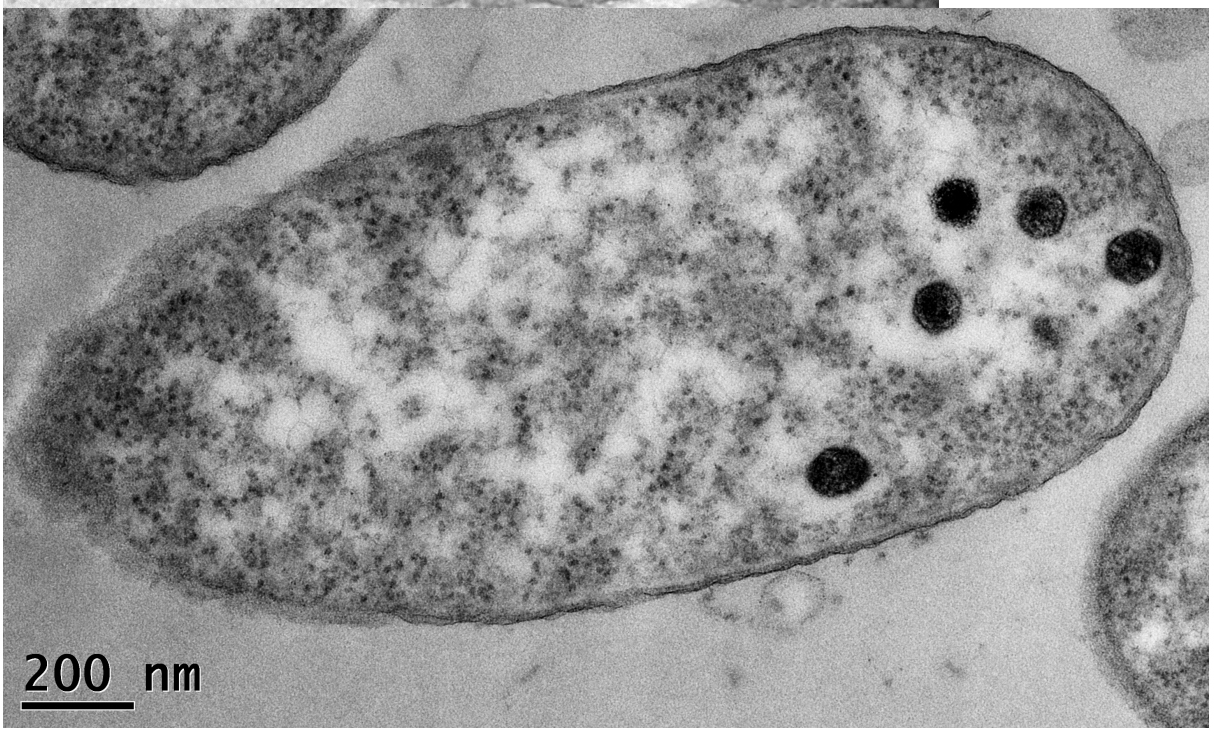


Figure 6. Ultra-thin section TEM (80nm) capturing SHEFM2K lifecycle of infection of EcM2K. Samples were taken over 2.5 hours and fixed in glutaraldehyde before staining (see methods). Images illustrate the intracellular compartments