Title

Identifying the core genome of the nucleus-forming bacteriophage family and characterization of Erwinia phage RAY

Journal

Cell Reports, 42(5)

ISSN

2639-1856

Authors

Prichard, Amy

Lee, Jina

Laughlin, Thomas G

et al.

Publication Date

2023-05-01

DOI

10.1016/j.celrep.2023.112432

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2023 June 27.

Published in final edited form as:

Cell Rep. 2023 May 30; 42(5): 112432. doi:10.1016/j.celrep.2023.112432.

Identifying the core genome of the nucleus-forming bacteriophage family and characterization of *Erwinia* phage RAY

Amy Prichard[1], **Jina Lee**[1], **Thomas G. Laughlin**[1], **Amber Lee**[1], **Kyle P. Thomas**[1], **Annika E. Sy**[1], **Tara Spencer**[1], **Aileen Asavavimol**[1], **Allison Cafferata**[1], **Mia Cameron**[1], **Nicholas Chiu**[1], **Demyan Davydov**[1], **Isha Desai**[1], **Gabriel Diaz**[1], **Melissa Guereca**[1], **Kiley Hearst**[1], **Leyi Huang**[1], **Emily Jacobs**[1], **Annika Johnson**[1], **Samuel Kahn**[1], **Ryan Koch**[1], **Adamari Martinez**[1], **Meliné Norquist**[1], **Tyler Pau**[1], **Gino Prasad**[1], **Katrina Saam**[1], **Milan Sandhu**[1], **Angel Jose Sarabia**[1], **Siena Schumaker**[1], **Aaron Sonin**[1], **Ariya Uyeno**[1], **Alison Zhao**[1], **Kevin D. Corbett**[2], **Kit Pogliano**[1], **Justin Meyer**[1], **Julianne H. Grose**[3], **Elizabeth Villa**[1,4], **Rachel Dutton**[1], **Joe Pogliano**[1,5,*]

[1]School of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA

[2]Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA 92093, USA

[3]Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602, USA

[4]Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA 92093, USA

[5]Lead contact

SUMMARY

We recently discovered that some bacteriophages establish a nucleus-like replication compartment (phage nucleus), but the core genes that define nucleus-based phage replication and their phylogenetic distribution were still to be determined. Here, we show that phages encoding the major phage nucleus protein chimallin share 72 conserved genes encoded within seven gene blocks. Of these, 21 core genes are unique to nucleus-forming phage, and all but one of these genes encode proteins of unknown function. We propose that these phages comprise a novel viral

*Correspondence: jpogliano@ucsd.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, A.P., T.S., J.P., and R.D.; data curation, A.P., J.L., and T.G.L.; formal analysis, A.P., J.L., T.G.L., and A.L.; funding acquisition, J.P. and E.V.; investigation, A.P., J.L., T.G.L., A.L., K.P.T., and A.E.S.; methodology, A.P., J.L., and R.D.; project administration, J.P. and R.D.; resources, J.H.G., E.V., J.P., and R.D.; software, J.L.; supervision, J.P. and R.D.; validation, A.P. and J.L.; visualization, A.P., J.L., T.G.L., A.L., and K.P.T.; writing – original draft, A.P., J.L., T.G.L., A.L., K.P.T., A.E.S., T.S., A.A., A.C., M.C., N.C., D.D., I.D., G.D., M.G., K.H., L.H., E.J., A.J., S.K., R.K., A.M., M.N., T.P., G.P., K.S., M.S., A.J.S., S.S., A.S., A.U., A.Z., R.D., and J.P.; writing – review & editing, A.P., J.L., T.G.L., K.D.C., K.P., J.M., J.H.G., E.V., R.D., and J.P.

DECLARATION OF INTERESTS

K.P. and J.P. have an equity interest in Linnaeus Bioscience Incorporated and receive income. The terms of this arrangement have been reviewed and approved by the University of California, San Diego, in accordance with its conflict-of-interest policies.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

SUPPLEMENTAL INFORMATION

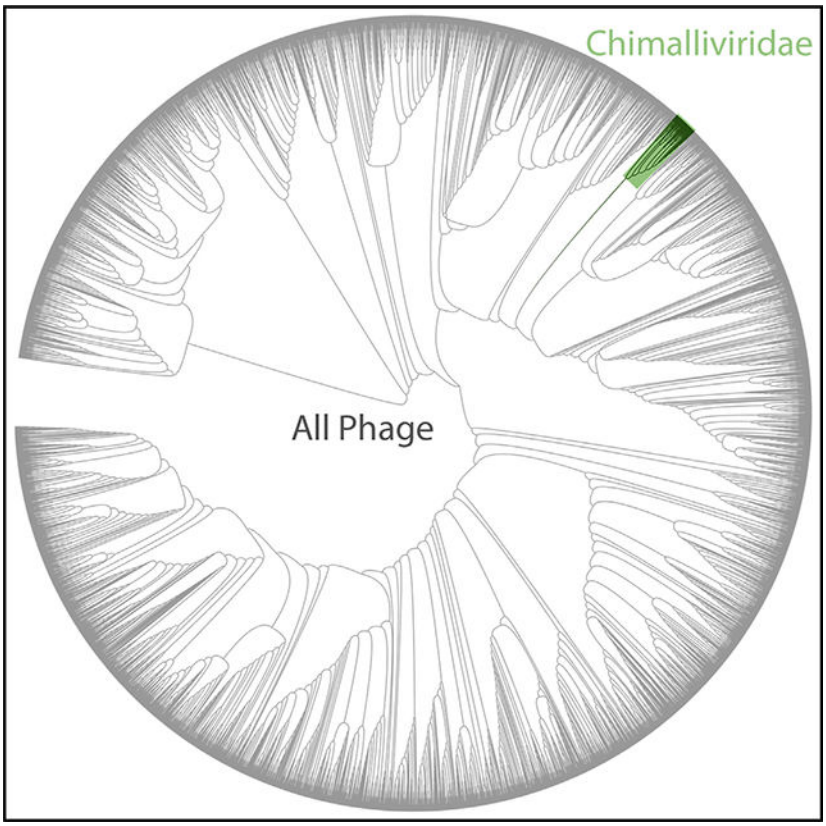
Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112432.

family we term Chimalliviridae. Fluorescence microscopy and cryoelectron tomography studies of *Erwinia* phage vB_EamM_RAY confirm that many of the key steps of nucleus-based replication are conserved among diverse chimalliviruses and reveal variations on this replication mechanism. This work expands our understanding of phage nucleus and PhuZ spindle diversity and function, providing a roadmap for identifying key mechanisms underlying nucleus-based phage replication.

In brief

Some bacteriophages have been shown to make a nucleus-like structure (the phage nucleus) during their infection cycle. These phages are widespread and infect numerous hosts. Prichard et al. find that these phages cluster together in phylogenetic trees, suggesting a common ancestry of all known nucleus-forming phages.

Graphical abstract



INTRODUCTION

The ability to establish and maintain subcellular organization is fundamental to cellular function. Even many viruses remodel their host cells, setting up their own complex compartments to suit their unique needs for viral replication.[1–6] We recently discovered that some bacteriophages replicate by creating a nucleus-like proteinaceous structure (the phage nucleus) that compartmentalizes the bacterial host cell during phage infection much in the same way a membranous nucleus compartmentalizes a eukaryotic cell.[6–10] Although

the phage nucleus is structurally different from the eukaryotic nucleus, it performs many similar functions. The phage nucleus, which is made of a protein called chimallin (ChmA), separates transcription from translation, exports mRNA, selectively imports proteins, and shields the phage DNA from cytoplasmic nucleases.[7–13]

Nucleus-forming phages belong to a larger group of phages that encode rifampicin-resistant multi-subunit RNA polymerases (msRNAP).[14–22] Nucleus-forming phages, typified by *Pseudomonas aeruginosa* phage ΦKZ, encode two msRNAPs composed of 4–5 subunits. One of these msRNAPs is the virion RNAP (vRNAP), which is packaged within the capsid and is likely injected into the cell along with the DNA upon infection.[15–17] The non-vRNAP (nvRNAP) is expressed by the vRNAP during infection.[17,19,20] While all nucleus-forming phages studied to date encode these unique msRNAPs, this feature is also observed in several jumbo phages that do not form phage nuclei.[14]

Another feature shared by currently characterized nucleus-forming phages is a phage-encoded tubulin homolog called PhuZ. Members of the PhuZ family studied thus far share a common filament structure and assembly mechanism, producing a dynamic three-stranded filament.[23,24] In *Pseudomonas* phages 201φ2–1, ΦKZ, and ΦPA3, the phage nucleus is centered at the host midcell and rotated by a bipolar PhuZ spindle,[7,8,25–27] and in *Escherichia coli* phage Goslar, the PhuZ filaments form a vortex that rotates the nucleus without positioning it at mid-cell.[10] Capsids assemble on the membrane and traffic along PhuZ filaments to reach the nucleus in the *Pseudomonas* phages.[7,8,27] Capsids then dock to initiate DNA packaging at the surface of the phage nucleus.[7,8,10,27] Filled capsids assemble with tails, forming cytoplasmic bouquet structures at rates varying from phage to phage prior to cell lysis.[10,28]

These recent discoveries prompt intriguing questions: how widespread is nucleus formation among phages that infect different hosts? Do all nucleus-forming phages share a common set of core genes? If so, which viral genes are part of the core genome versus the accessory genome? To answer these questions, we began by studying *Erwinia* phage vB_EamM_RAY (hereafter called RAY), which encodes a chimallin homolog. RAY infects *Erwinia amylovora*, an important agricultural pathogen and the causative agent of fire blight.[29,30] Many chimallin-encoding *Erwinia* phages are distantly related to each other and to known nucleus-forming phages, making *Erwinia amylovora* an enticing host for studying possible nucleus formation in diverse phages.[29,31,32]

Core genomes of phage and bacteria define their conserved components, while accessory genomes provide genes that are specific to the individual species, strains, or variants.[33–37] While genes required for making viral particles such as virion structural components, polymerases, and chimallin can be identified bioinformatically by comparison to sequence and structural homologs, the genes that comprise the nucleus-forming phage core genome have not been previously identified. Studying the core genome will allow us to identify and understand the genes required for the nucleus-based replication mechanism, including nuclear shell assembly, PhuZ spindle formation, mRNA export, selective protein import, capsid docking, and bouquet formation. While previous work identified some core genome components common of chimallin-encoding *Pseudomonas* phages 201φ2–1, ΦKZ, ΦPA3,

EL, and OBP and *Vibrio* phage JM-2012, this work predates the discovery of the phage nucleus replication mechanism, and only 6 phage genomes were included at the time.[38] Additionally, core genomes of large groups of jumbo phages have been analyzed but did not focus on viral replication mechanisms to identify the core genomes of phage that replicate by forming a nucleus.[39]

Here, we investigated the conservation of nucleus-based phage replication by focusing on *Erwinia* phage RAY and comparing it with previously studied nucleus-forming phages. We analyzed its phylogenetic relationships with other phages, identified the core and accessory genes, and used fluorescence microscopy and bioinformatics to characterize the putative functions of 15 core and 9 accessory RAY proteins. We show that RAY is a nucleus-forming phage by fluorescence microscopy and cryoelectron tomography and describe how its replication mechanism varies compared with previously characterized nucleus-forming phages.

RESULTS

Defining the core genome of chimallin-encoding phages

To identify the key, conserved genes required to replicate via the phage nucleus pathway, we first identified all phages in the NCBI database that encode a homolog of the major nuclear shell protein chimallin (ChmA). We found 66 unique phages encoding chimallin homologs and made whole genome trees to compare them (Figures 1A and 1B), showing that all 66 phages form a monophyletic group (Figure 1A) when compared with phages that encode msRNAPs and other related phages (see STAR Methods). Members of this clade infect a wide range of Gram-negative bacteria and one Gram-positive *Bacillus* species, and they vary greatly in genome size, from 167 to 322 kb. This suggests that the chimallin-encoding group of phages arose only once from a common ancestor, has a wide host range, and is not restricted only to “jumbo” phages larger than 200 kb. We created phylogenetic trees based on 6 conserved proteins. The whole genome tree (Figure 1B) is generally congruent with all of the protein-based trees, including chimallin (Figure 1C), major capsid protein (Figure 1D), terminase large subunit (Figure S1A), DNA polymerase (Figure S1B), replicative helicase (Figure S1C), and an RNAP subunit (Figure S1D). Each of these trees contained 16 groups of related phage species composed of the same individuals (color coded in Figures 1, S1, and S2). The phylogeny of the groups closely matched the whole genome tree (Figure S2). The similarity across different phylogenetic trees suggests that these proteins have been co-evolving, with little evidence of horizontal gene transfer between divergent phages for these proteins. This supports previous findings that the phage nucleus may help to limit recombination between nucleus-forming phages.[40] We then determined the set of core genes that are conserved among chimallin-encoding phages and whether this core genome is conserved in other msRNAP-encoding phages.

We classified a gene as core if it is present in more than 90% of the chimallin-encoding phages to accommodate potential sequencing errors and variability due to sampling[42] and found 72 conserved genes (Figures 2A–2C) (see STAR Methods for details). The 72 core genes included the major capsid protein, terminase large subunit, and msRNAP subunits, but surprisingly, the majority of the highly conserved core genes (53, 73.6%) had no

predicted function. A PhuZ homolog was present in only 66.2% of chimallin-encoding phages, which is notable given its well-characterized role in nucleus-based phage replication yet consistent with PhuZ function being dispensable for replication.[10,25,27,40,43] The core genes identified here likely contain many of the key proteins required for nucleus-forming phages to replicate.

The core genes occur in seven distinct blocks (of three or more genes) whose general order is conserved across *Pseudomonas aeruginosa* phage ΦKZ, *Escherichia coli* phage vB_EcoM_Goslar (Goslar), and *Erwinia amylovora* phage RAY (Figure 2C). These three phages were chosen for our analysis because they represent a diverse group of chimallin-encoding phages and are well studied (ΦKZ), infect highly tractable hosts (Goslar), or formed the basis of our bioinformatic studies (RAY). The core genome blocks are often rich with certain types of genes, such as block 7, which contains the terminase and a handful of virion structural genes, and block 5, which contains nineteen genes, including seven structural genes, the major capsid protein, one helicase, and an RNAP β subunit (Table S3). Gene order within the blocks is also conserved across chimallin-encoding phages. Upon mapping the level of conservation across the RAY, ΦKZ, and Goslar genomes, we found that there is a conserved region that is dense with core genome blocks and a variable region with putative accessory genes specific to each phage (Figures 2A and 2B). The majority of RAY's 317 genes are either part of this core conserved genome (23%, 72 genes) or are accessory genes only detectable in 10% or fewer of the phages we used in our analysis (178 genes, 56%) (Figure 2C). This is a similar organization to other previously reported phage core genomes.[35,36]

We next identified which of the core genes are specific to chimallin-encoding phages and likely represent the key genes required to replicate via the phage nucleus pathway versus core genes that are broadly conserved among phage in general. We found that 51 out of the 72 genes are also present in other phages lacking chimallin homologs (Table S1). These broadly conserved proteins include DNA polymerases, RNAP subunits, the major capsid protein, and other virion structural proteins. The remaining 21 genes include chimallin and 20 hypothetical proteins with no predicted structure or function. This unique set of 21 genes identified by our bioinformatic screen likely encodes the proteins specifically required to form and maintain a phage nucleus, but further study on each proposed core gene will be required to determine if it has a conserved role across different nucleus-forming phages. We propose that all phages encoding chimallin and the associated core genome belong to one viral family, Chimalliviridae, named after the shared protein that makes up the phage nucleus lattice.

Additionally, we searched metagenomic databases, and we were able to identify many more phages that encode chimallin and cluster within the Chimalliviridae. These phages were found in numerous environments from multiple locations across North America (Data S1). We also looked for phage metagenomes on NCBI, and when analyzing their whole-genome similarity, compared with known phages using VIPTree, we found that they also cluster within the Chimalliviridae clade (Figure S3). Taken together with the whole-phage genome sequence alignments, our results suggest that chimalliviruses are likely common and found throughout the world.

Phage nucleus formation by *Erwinia* phage RAY

Since RAY is a member of the Chimalliviridae and shares the core genome, we predicted that it replicates by forming a phage nucleus. We tested this prediction by imaging infected *Erwinia* cells and observed a bright zone of DAPI fluorescence that was often positioned at midcell, similar to *Pseudomonas* nucleus-forming phages (Figure 3A). We created an N-terminal fusion of GFPmut1 to gp222 (RAY chimallin homolog [$ChmA_{RAY}$]) (Figure S4A) and examined its localization in the absence and presence of phage infection. Fluorescence from GFP-$ChmA_{RAY}$ was uniformly distributed throughout uninfected cells but formed a ring in the center of the cell enclosing DNA during RAY infections (Figure 3B). Zones of DAPI fluorescence consistent with host chromosomal DNA were also present outside of the nucleus in every RAY infection, unlike previously characterized nucleus-forming phages, which degrade the host DNA,[7,8,10] but similar to *Serratia* phage PCH45.[11] To determine whether the extranuclear DAPI staining was due to host DNA, we tagged the *Erwinia amylovora* H-NS protein with GFP, which coats DNA, and used the H-NS-GFP fusion as a marker to visualize host DNA during phage infection, as H-NS is not imported into the phage nucleus. When we expressed H-NS-GFP in *Erwinia* cells and infected them with RAY, the H-NS-GFP fluorescence co-localized with the DAPI outside of the phage nucleus (Figure 3C). In time-lapse microscopy, the bacterial DNA was coated with H-NS before infection and was then pushed to the poles and compacted upon infection as the phage nucleus formed (Figure 3D; Video S1). These results suggest that RAY forms a nucleus-like structure without detectably degrading the host DNA.

Identification of RAY nuclear and cytoplasmic proteins

To gain further insight into the subcellular organization of RAY's lytic cycle, we used fluorescence microscopy to observe the localization of 24 RAY core and accessory proteins fused to GFP. To gain insights into both the conserved Chimalliviridae replication mechanism as well as RAY-specific variations, we tagged 15 proteins that were part of the Chimalliviridae core genome and 9 that were accessory components. We were able to tentatively assign potential protein families, to which most of these proteins belong based on *in silico* predictions; however, outside of the RNAP subunits, few of these phage proteins have been studied in detail, and their functions in the phage life cycle are unknown.

We identified 10 proteins that co-localize with the phage DNA, likely inside the phage nucleus (Figures 4A and 4B). Eight of these nucleus-associated RAY proteins (gp002, gp150, gp220, gp223, gp248, gp249, gp250, gp315) are part of the core genome, and two are not (gp116, gp153). Most of these proteins are likely involved in the core functions of DNA replication (gp220, gp315) (Figures 4B and S4C), transcription (gp002, gp248, gp223, gp249) (Figure S4D), or recombination (gp150, gp153) (Figures S4E and S4F) based on sequence and structural homology to previously characterized or annotated phage proteins. Two other proteins could be assigned to a protein family, but their functions are less clear (core gene gp250, SWI/SNF helicase family [Figure S4G] and accessory gene gp116, HslUV protease family [Figure S4H]).

We identified 4 proteins (gp039, gp048, gp064, and gp311) among our set of 24 that display diffuse localization in the cytoplasm outside of the phage nucleus and one protein

(gp094) that localized on the host chromosome (Figure 4C). Of these, only RAY gp311, a thymidylate kinase homolog (Figure S4I), is part of the core genome. Homologs of gp311 have also been shown to localize in the cytoplasm during replication of other phages,[7,10] where they likely participate in nucleotide metabolism. The remaining 4 proteins included a putative SspB-like ClpXP adaptor protein (gp039; Figure S4J), a putative tRNA ligase (gp048; Figure S4K), a putative exonuclease (gp064; Figure S4L), and a putative XRE superfamily transcriptional regulator (gp094; Figure S4M) whose localization mimics that of bacterial H-NS by binding to host DNA (Figures 3C and 4C). Taken together, these studies suggest that RAY forms a proteinaceous shell that encloses phage DNA and compartmentalizes proteins according to their functions.

Virion-associated RAY proteins

To follow the assembly process of RAY virions, we created GFP fusions to RAY's tail sheath (gp179; Figure S4N) and major capsid (gp317; Figure S4O) proteins. Fluorescence microscopy showed that the major capsid protein localized around the nucleus (Figure 4D), suggesting that capsids dock on the phage nucleus to initiate DNA packaging as occurs in ΦKZ-like phages and Goslar. The tail protein localizes in the cytoplasm, often forming foci near the phage nucleus (Figure 4D); however, we did not observe bouquet-like clusters of virions. This suggests that capsids dock on the phage nucleus to be packaged with DNA, but after assembling virions, bouquet-like structures are not formed like those observed in other nucleus-forming phages.[10,28]

We identified 5 proteins in our set of 24 that localized at the periphery of the phage nucleus during part of the lytic cycle, similar to the capsids (Figures 4D and 4E). Three of these (gp154, gp163, and gp270; Figure S4P) are vRNAP subunits. The other two are an SWI/SNF family helicase (gp131; Figure S4Q) and a putative head protein with unknown function (gp299; Figure S4R). Our localization of vRNAP and these other proteins during RAY infection is consistent with previous mass spectrometry results, which showed that RNAP subunits are packaged within the viral capsid.[16,29] These results also suggest that GFP fusions and fluorescence microscopy can be used as an additional approach to identify proteins that accumulate within capsids.

Phage tubulin PhuZ

RAY encodes a PhuZ homolog ($PhuZ_{RAY}$; gp210) (Figure S5A) that contains conserved domains of the tubulin superfamily, including the tubulin T4 signature motif (Figure 5A). In ΦKZ-like phages, PhuZ forms filaments that use dynamic instability and treadmilling to center and rotate the phage nucleus,[7,8,25–27] and in Goslar, PhuZ forms a vortex that rotates the phage nucleus without centering it.[10] To determine whether $PhuZ_{RAY}$ also forms filaments, we created a GFP-$PhuZ_{RAY}$ fusion and assessed its ability to form polymers at different expression levels in both infected and uninfected cells. In uninfected cells, at low levels of arabinose (≤0.01%), GFP-$PhuZ_{RAY}$ is uniformly distributed throughout the cell, but it spontaneously forms filaments at ≥0.05% arabinose (Figures 5B and 5C). During infections, we studied GFP-$PhuZ_{RAY}$ during infections under conditions where it did not spontaneously assemble filaments (0% arabinose) and found visible filaments in all infected

cells (Figures 5D and 5E), suggesting that the preexpressed GFP fusion in the host cell is able to co-assemble with phage-expressed native, untagged PhuZ.

To determine whether PhuZ centers the nucleus in RAY infections, we expressed a GTP-hydrolysis-deficient mutant, $PhuZ_{RAY}$D198A, as a dominant negative to inhibit the dynamic properties of the endogenous filaments and test whether the lack of filament dynamics interferes with nucleus positioning.[8,25,26] In uninfected and infected cells, GFP-$PhuZ_{RAY}$ D198A formed filaments at all arabinose concentrations, likely due to its inability to depolymerize (Figures 5B, 5C, and 5E). We measured the distance from the phage nucleus to the cell pole when GFP-$PhuZ_{RAY}$ and GFP-$PhuZ_{RAY}$D198A were expressed, taking into account whether the presence and location of bacterial nucleoids contributed to phage nucleus positioning. When a single bacterial nucleoid was present, the positioning of the RAY nucleus had a broad distribution with a slight bias toward midcell (Figure 5F, blue bars). The single nucleoid in these cells frequently occluded the center of the cell, thereby preventing the nucleus from being centered (Figure 5D, cell iii, white arrow). When two bacterial nucleoids were present, the RAY nucleus was usually (97%, n = 103) positioned between them near the cell midpoint (Figures 3A, 6D, cells i, ii, and iv, and 6G). When $PhuZ_{RAY}$D198A was expressed, the positioning of the RAY nucleus was less biased toward midcell when two nucleoids were present and had no specific localization when there was one nucleoid (Figures 5F and 5G). This is the first time that phage nucleus positioning was measured in the presence of the bacterial genome, and these results suggest that while the PhuZ filaments likely play a role in positioning the nucleus in phages that do not degrade host DNA, the phage nucleus must also compete for space with the bacterial DNA when it is not degraded.

In addition to centering the phage nucleus within the host cell, PhuZ also rotates the phage nucleus in both the ΦKZ-like phages and Goslar. We therefore performed time-lapse microscopy and measured nucleus rotation in RAY-infected cells expressing GFP-$ChmA_{RAY}$ (Figure 3E; Video S2). We observed very few of the phage nuclei actively rotating (3.7%, n = 858) compared with Goslar (97%)[10] and 201φ2–1 (46%).[27] Those that did rotated for 27.7 ± 5.8 s with an average linear rotation speed of 54.8 ± 15.6 nm/s and an angular velocity of 10.1° ± 3.2°/s (n = 25) (Figure 3F). We also measured nucleus rotations using GFP-tagged major capsid protein gp317 (MCP_{RAY}-GFP) docked on the surface of the nucleus as a fiduciary mark. The nuclei with docked capsids displayed a similar linear rotation speed of 57.5 ± 11.9 nm/s (n = 17). This rotation speed is only marginally faster when compared with Goslar (49.7 ± 12.5 nm/s) and 201φ2–1 (43.6 ± 7.6 nm/s) nucleus rotation speeds,[10,27] suggesting that this rotation rate is conserved, possibly to allow the even spacing of capsids docking on the phage nucleus surface for efficient DNA encapsidation.

In situ structural analysis of RAY phage replication in *Erwinia amylovora*

In order to investigate the structural conservation of chimallivirus replication components, we performed cryo-focused ion beam milling and electron tomography (cryo-FIB-ET) of *Erwinia amylovora* cells infected with RAY (Figure 6). The tomograms revealed the presence of the RAY phage nuclear shell as an irregularly shaped, protein-based compartment enclosing nucleic acid and excluding host ribosomes (Figures 6A and 6B).

Viral capsids were observed at different stages of maturation, docked on the phage nucleus where they package DNA, or in the cytoplasm with or without tails attached. We did not observe virion particles assembled into structured bouquets, in agreement with fluorescence microscopy (Figures 6A and 6B).

To further assess the components of RAY replication, we performed subtomogram analysis of virion components, the phage nucleus shell, and PhuZ filaments. Through averaging of assembled and isolated virion components, we obtained reconstructions for the RAY capsid (~19 Å), collar (~38 Å), tail sheath (~10 Å), and baseplate (~36 Å). Dimensions of assembled virions measured from the tomograms and overlapping segments of the component maps allowed reconstruction of a composite RAY virion (Figure 6C). The virion is largely similar to the composite structure of the isolated ΦKZ virion.[44] The RAY capsid exhibits a triangulation number of 27 (T = 27) and is approximately 142.5 nm in diameter, which is similar to the ΦKZ capsid. However, the RAY baseplate reconstruction appears in a retracted state, as opposed to the wider expanded state observed for the ΦKZ virion.[44]

Subtomogram analysis of the phage nuclear shell resulted in a ~20 Å reconstruction which revealed that the RAY nuclear shell is principally composed of a square chimallin lattice with ~11.5 nm repeat distance, similar to that assembled by the 201φ2–1 and Goslar chimallin (Figure 6D).[9] The conserved higher-order structure of the RAY chimallin lattice indicates a conserved underlying protomer structure. Indeed, the AlphaFold 2.0-predicted structure of RAY chimallin is largely consistent with the experimental chimallin structures determined for 201φ2–1 (root-mean-square deviation [RMSD]: 2.2 Å) and Goslar (RMSD: 2.2 Å).[9] The $ChmA_{RAY}$ protomer prediction predominantly differs from experimental structures in the low-confidence placement of the extended N- and C-terminal segments, which mediate interactions across ChmA protomers in higher-order assemblies (Figure S6).

Finally, we performed subtomogram analysis (StA) of long, hollow, filamentous structures present in RAY-infected *Erwinia* cells assumed to be the phage-encoded tubulin PhuZ (Figures 6A, 6B, and 6E). The resulting StA map achieved an estimated resolution of ~25 Å and revealed a hollow, five-stranded filament of approximately 13 nm in diameter (Figure 6F), which is in stark contrast to the three-stranded filaments made by $PhuZ_{201φ2–1}$ obtained *in vitro*.[23,25] We did not observe other filamentous structures in our tomogram set, thus the five-stranded structures are tentatively assigned as RAY PhuZ. StA of PhuZ filaments observed in our previously published 201φ2–1 and ΦKZ cryo-FIB-ET datasets revealed that these PhuZ variants form three-stranded filaments *in situ* (Figures S5B–S5D). The structure of the $PhuZ_{RAY}$ protomer as predicted by AlphaFold is highly similar (RMSD = 1.069 Å) to the experimentally determined crystal structure of $PhuZ_{201φ2–1}$ (Figure 5H),[23,25] thus indicating that subtle differences are likely responsible for differences in higher-order assembly among PhuZ variants. Of note, residues R217:D305 and E225:R290 form essential salt bridges for self-assembly of $PhuZ_{201φ2–1}$ and are conserved in $PhuZ_{ΦKZ}$ (R230:D316 and D238:R301) but not in $PhuZ_{RAY}$ (D223:D315 and L235:D299). Further work will be necessary to unambiguously determine whether the five-stranded filaments are indeed $PhuZ_{RAY}$ and the molecular bases of their higher-order assembly.

Taken together, the structures seen in the cryo-FIB-ET corroborate the fluorescence microscopy results and confirm that RAY is a nucleus-forming phage that has unique and intriguing differences when compared with previously published Chimalliviridae Goslar, PCH45, and ΦKZ-like phages.

DISCUSSION

The phage nucleus is a remarkable and provocative cell biological structure, but the viral genes required for phage to replicate via this pathway are unknown. Here, we bioinformatically identified the core genome that defines the nucleus-forming Chimalliviridae family. The Chimalliviridae core genome consists of a set of 72 genes encoded within seven distinct blocks. The order of the core blocks and genes within the blocks was generally conserved among divergent members of the family, as was their phylogeny. This suggests that these phages evolved from a common ancestor and that their genes have been co-evolving without significant horizontal gene transfer. We propose that the core genome identified here includes many of the key genes required for nucleus-based phage replication and that the less conserved accessory genes may be important for different subfamilies of phage or allow specialization for infecting different hosts. The core genes encode proteins that participate in many key processes such as genome replication, gene expression, and formation of virion particles. However, 74% of the core genes (53/72) are hypothetical proteins of unknown function and origin; moreover, 29% (21/72) make up the core genes that are unique to the Chimalliviridae, and of these, only chimallin has a known function.[7,9] This analysis agrees well with previous studies that identified a core gene set shared by different groups of jumbo phages.[38,39] This core gene set likely encodes numerous unknown components required for phage nucleus formation, standing in stark contrast to the core genome of the phage T4 family (Tequatroviridae), where the function of the majority of the core genes are known and genes of unknown function are primarily accessory.[35] Thus, the Chimalliviridae core genome is rich with potential for discovery of novel biological functions due to its numerous uncharacterized genes.

Phage can be classified into families based on genetic data.[45] Our phylogenetic analysis suggests that all phages encoding this core genome form a single family, the Chimalliviridae, that only arose once and developed a unique nucleus-based replication mechanism. Within this family are clades of phages with specific accessory genomes, suggesting that different Chimalliviridae clades have adapted the phage nucleus replication mechanism in a variety of diverse ways. The core genome of nucleus-forming phages described here can be used to study the basic requirements for nucleus-based replication, discover new members of the Chimalliviridae, and guide future studies in synthetic biology aimed at building a phage encoding only the minimal components required for phage nucleus-based replication.

Chimalliviruses infect a broad range of bacteria and have been found in many locations throughout the world. By searching metagenomic databases, we identified many more Chimalliviridae members from a variety of sources (rhizosphere, brine, fracking water, peat, uranium contaminated flood plain, bog, agave, subway wood, and deep shale) and many distant locations (Michigan, Oregon, Alaska, Oklahoma, Iowa, New York, Ohio, Alaska, Nebraska, and Wyoming, USA, and Guanajuato, Mexico) (Data S1). Indeed, phages closely

related to ΦKZ have been isolated from South Korea, Germany, Poland, Denmark, the USA, Australia, Japan, Thailand, Iran, India, and Kazakhstan.[46–52] We also found that phage metagenomes on NCBI cluster within the clade of chimalliviruses as part of this same family (Figure S3). We conclude that Chimalliviridae form a single clade of phages that are widespread and found in habitats throughout the world.

We performed a detailed investigation into RAY to determine how conserved the phage nucleus replication mechanism is in a divergent Chimalliviridae family member. By examining the intracellular localization of 15 core and 9 non-core RAY proteins via GFP tagging and fluorescence microscopy, we discovered that these proteins have conserved subcellular localizations across Chimalliviridae, with DNA processing proteins being found inside the phage nucleus, virion proteins being found around the periphery of the phage nucleus, and other proteins being found in the cytoplasm (Figures 3, 4, 5, and 6).[7,8,10,11,13,27] We show that chimallin surrounds the phage DNA, and cryo-FIB-ET revealed it assembles an enclosed square lattice, forming a 6-nm-thick shell that separates phage DNA from ribosomes and virion structural proteins in the cytoplasm, similar to 201φ2–1, ΦKZ, ΦPA3, and Goslar.[7,8,10,13,27] The putative functions of these proteins have been updated in the NCBI annotation (NC_041973.1).

We also found several additional surprises (unconserved processes) that add to our knowledge of the diversity of chimallivirus replication. First, the host chromosome is not degraded and is excluded from the phage nucleus. This has also previously been observed with phage PCH45[11] and suggests that the host chromosome is not a barrier to building a phage nucleus. Second, the PhuZ spindle and host chromosomes both likely impact phage nucleus position. Third, unlike previously studied PhuZ proteins from 201φ2–1, ΦKZ, and ΦPA3, which form three-stranded filaments,[23] $PhuZ_{RAY}$ appears to form five-stranded filaments *in vivo*, which make $PhuZ_{RAY}$ the first phage-encoded tubulin-based structure that possesses a lumen (Figures 6F and S5). This is not only distinct from other PhuZ homologs but also from most bacterial tubulins, which assemble either single filaments (FtsZ)[53–55] or two- or four-stranded filaments (TubZ).[56–58] However, it is similar to *Prosthecobacter* tubulins BtubA/BtubB, which assemble five-stranded tubules *in vivo*, and a variety of single, double, and polymeric structures *in vitro* and are more similar to eukaryotic tubulins than bacterial tubulins.[59–61] Future studies will be needed to confirm that PhuZ has these assembly properties *in vitro*. Fourth, unlike *Pseudomonas* and *Escherichia coli* nucleus-forming phages, whose nuclei are more consistently rotated by the PhuZ spindle, we rarely observed nucleus rotation with RAY. Finally, unlike *Pseudomonas* and *Escherichia coli* phages,[10,28] RAY viral particles do not form bouquets.

We described the characterization of ten proteins that had not previously been studied in any chimallivirus. For example, we discovered one core and one accessory protein that both contain a domain belonging to the SWI/SNF helicase family, but their functions are unknown. Surprisingly, we found one of these proteins (gp250) localized inside the nucleus, while the other one (gp131) localized inside the capsids. Another example is gp094, an accessory protein predicted to be an XRE family transcriptional repressor. gp094 binds to bacterial DNA, prompting us to speculate that it might regulate host gene expression. Other proteins described here include a HslUV-like protease (gp116), a ClpXP protease adapter

protein (gp039), and a tRNA ligase (gp048), which could potentially be involved in phage replication or overcoming host defense systems.

Our results suggest a model for RAY replication that is distinct from other previously characterized nucleus-forming phages (Figure 7). Upon infection, RAY injects its DNA and a suite of proteins for viral replication, including four vRNAP proteins, as well as proteins of unknown function (such as putative SF2 helicase gp131 and internal head protein gp299). The injected RNAP expresses chimallin,[17] which forms a proteinaceous shell surrounding the phage DNA (Figure 3B). Unlike with 201φ2–1, ΦKZ, ΦPA3, and Goslar, the host chromosome is not degraded, similar to PCH45 infection,[11] and the cell does not undergo substantial swelling. The growing phage nucleus therefore must compete for limited space with the chromosomal DNA. If the cell contains two well-separated nucleoids, the nucleus usually localizes between them (Figures 3B, 6D, and 6E; Video S1). Capsids dock on the phage nucleus (Figures 4D, 4E, 7A, and 7B), package DNA, and assemble mature viral particles (Figure 7). $PhuZ_{RAY}$ forms filaments that appear to contribute to RAY nucleus positioning (Figures 5F and 5G), but whether they contribute to capsid trafficking as speculated in Figure 7 is unknown. It is also unclear if they contribute to nuclear rotation: since the RAY nucleus was only observed to rotate in ~3.7% of infected cells, it was not possible to measure a significant difference between expressing live versus catalytically dead PhuZ. After assembly of mature phage particles, RAY does not form bouquets (Figures 6A and 6B). Given RAY's inability to degrade the chromosome or cause the cell to bulge, there may be no room for such structures to assemble.

The phage-encoded tubulin homolog PhuZ is absent from one-third of known Chimalliviridae yet displays conserved function in phage nucleus rotation in 201φ2–1, ΦKZ, ΦPA3, and Goslar[10,27] and in nuclear positioning in 201φ2–1, ΦKZ, ΦPA3, and RAY[7,8,25,26] (Figures 5F and 5G). PhuZ is an example of an accessory protein that has likely adapted to make different types of cytoskeletal structures for different phages. Among the subset of Chimalliviridae we have examined thus far, we have observed PhuZ to form a bipolar spindle or a vortex and either three- or five-stranded tubes reminiscent of Verrucomicrobial BtubA/BtubB.[59–61] Additionally, prior studies have reported only a 50% decrease in phage titers when expressing dominant negative PhuZ mutants despite severe phage nucleus positioning defects.[8,25,26] In agreement with these studies, the PhuZ gene can be completely deleted from ΦKZ, and while phage nucleus position is again strongly affected, the phage is still viable.[43] Taken together, these studies suggest that PhuZ is an accessory protein that may confer improved infection efficiency in some nucleus-forming phages, but it is not essential for this viral replication pathway.

Many phages package proteins within their capsids that they are thought to inject into the cell along with their DNA.[14,16,62–64] These proteins have previously been discovered by performing mass spectrometry on purified viral particles.[16] Using GFP fusions, we have shown localizations consistent with virion packaging of these proteins, corroborating previous mass spectroscopy data and showing that visualization of internal capsid proteins is possible using fluorescence microscopy, as found in another a recent study.[43] For example, the vRNAP subunits localized around the phage nucleus where the capsids accumulated, in agreement with the previous mass spectrometry data that these vRNAPs are encapsidated.[16]

This work highlights how, despite the unique features of each nucleus-forming phage studied to date, many of the basic components are conserved across the entire diverse family of Chimilliviridae. Here, we document differences in phage tubulin assembly, nucleus positioning, nucleus rotation, chromosome degradation, and cell bulging that likely represent just a few of the many phenotypic consequences that result from continual phage-host evolutionary conflicts. Many of these differences are likely due to the continually evolving accessory genome. Now that the core and accessory genomes of this unique phage family have been identified, future studies can be better directed toward understanding their functions in order to elucidate the intricacies of the nucleus-based phage replication mechanism.

Limitations of the study

The core genome determination is based on genes identified using BLAST. Proteins with similar structures and common ancestry can be so highly divergent as to appear unrelated solely based on sequence identity. To counter this, we performed four iterations of PSI-BLAST. Another limitation is the arbitrary 90% conservation cutoff we used to define the core genome. We used this number to avoid excluding some genes from the core genome due to sequencing errors. Thus, the list of core proteins identified here is based on currently available phage sequence space but does not represent all genes important for nucleus-based replication. Finally, GFP tagging can alter protein localization. To address this possibility, we tagged proteins at both the N and C termini.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Joe Pogliano (jpogliano@ucsd.edu).

Materials availability—This study did not generate new unique reagents.

Data and code availability

- All raw fluorescence microscopy images used in this paper are deposited in a Mendeley dataset (Mendeley Data: https://doi.org/10.17632/jd4yj84463.1). Tilt-series frames and alignment metadata for the RAY-infected *E. amylovora* cells are deposited with the Electron Microscopy Public Image Archive with accession number EMPIAR-11198. All subtomogram averaging maps from this study are deposited with the Electron Microscopy Data Bank with the following accession numbers: *E. amylovora* 70S (EMD-27973), *E. amylovora* 50S (EMD-27993), RAY capsid vertex (EMD-28003), RAY collar (EMD-28004), RAY tail sheath (EMD-28005), RAY baseplate (EMD-28006), RAY chimallin (EMD-28007), putative RAY PhuZ (EMD-28008), 201φ2–1 PhuZ (28009), and ΦKZ PhuZ (28010). The composite RAY virion map is deposited as an additional map with the RAY capsid vertex (EMD-28003).

- All original code used for core genome determination PSI-BLAST has been deposited on GitHub (GitHub: https://github.com/jina-leemon/core-genome-proj). Raw data of the PSI-BLAST results used for core genome determination can be found in Data S2.
- Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial growth and phage infection conditions—*Erwinia amylovora* ATCC 29780 was grown on LB plates at room temperature or in LB liquid cultures at 30°C. To collect RAY lysates, 0.5 mL of dense *E. amylovora* culture grown in liquid LB media at 30°C overnight was incubated with 10 μL of RAY serial dilutions, incubated for 15 minutes at room temperature, and mixed with 4.5 mL molten (approximately 55°C) LB 0.35% top agar. This mixture was quickly poured over standard LB plates and incubated overnight at room temperature. The following day, 5 mL phage buffer was poured over plates showing web lysis and incubated at room temperature for 5 hours. The phage lysate was collected by aspiration and centrifuged for 10 min at 3220 rcf to pellet cell debris. The resulting clarified phage lysate was filtered through a 0.45 μm filter to remove any residual bacterial contamination and stored at 4°C.

Expression of phage proteins—To visualize phage proteins in *Erwinia amylovora*, the proteins were fused to GFPmut1 and expressed from the pHERD-30T plasmid under the inducible control of the AraBAD promoter. GFP fusions were designed using Benchling with the GFP tag attached at either the N- or C-terminus depending on the protein. The plasmids were then synthesized by GenScript. Plasmids were transformed into electrocompetent *Erwinia amylovora* cells via electroporation. The cells were plated on LB plates with 15 μg/mL gentamicin sulfate as a selectable marker for transformants.

METHOD DETAILS

Genome alignments and phylogenetic analysis—To find close and distant relatives of RAY, we used ViPTree[65] to find phage genomes related to RAY (Figure S8) and Position-Specific Interative Basic Local Alignment Search Tool (PSI-BLAST) to find phage encoding chimallin homologs (Table S1) and msRNAP homologs since all chimallin-encoding phages also encode msRNAPs[14,38,39] PSI-BLAST iterations were performed until no new sequences were found above the 0.005 E-value threshold (for chimallin, the highest E-value in the results was 1e-08). These results were checked for validity by multiple sequence alignments with previously identified homologs to ensure they were not false positives. *Bacillus* phage PBS1, *Staphylococcus* phage PALS_2, and *Yersinia* phage ΦR1–37 were used as the representative phages that encode msRNAPs but not chimallin homologs, and well-studied *Escherichia* phage T4, giant *Bacillus* phage G, and *Erwinia* phage vB_EamM_Alexandra were also included in our analysis for comparison. VICTOR[66] predicted the species assignments, which were used to determine which phage could be discluded to keep the trees readable; genus assignments, which were used to color-code the trees; and subfamily and family assignments, which were used along with core genome

analysis to define the family Chimalliviridae (Table S2). Phylogenetic trees were made using VICTOR (Figure 1A) and colored using Photoshop (Adobe) or Clustal Omega (Figures 1B–1D, S1, S2, and S5) and colored using iTOL.[41] Data on metagenomes was found by BLASTing $ChmA_{RAY}$ in IMG/VR and collecting the high confidence results.

Core genome determination—Bulk PSI-BLAST in the standalone BLAST+ Suite[67] was performed on the tailed phages taxon (taxid: 28883) with RAY proteins as the query, a max iteration of 4, and an E-value cutoff of 0.05. Then two scripts were run on the result to search for proteins from the phages of interest as well as output the ones present in more than 95% of the phages through a text mining approach. Scripts are available on GitHub (https://github.com/jina-leemon/core-genome-proj). Blocks were defined as 3 or more conserved genes, tolerating a maximum of 3 non-conserved genes between. The core genome numbers (Table S3) were assigned depending on the homologs that showed up through PSI-BLAST in ΦKZ, Goslar, and RAY.

Phage protein characterization—PSI-BLAST was used in order to identify potential homologs of each unknown RAY protein.[67] The PSI-BLASTs were performed with the non-redundant protein database, limited to tailed phages (taxid: 28883), excluding uncultured and environmental sample sequences, and with a maximum of 5000 sequences. Iterations were run excluding results with E-values lower than the default PSI-BLAST threshold, stopping when results converged. For each protein studied, Phyre^2 was used in order to predict secondary structure and to identify known proteins with similar structures.[68] The amino acid sequences of each protein were uploaded to Phyre^2 with the normal modeling mode. Phyre^2 results with confidence higher than 70% were compared to PSI-BLAST results as an independent method for predicting the putative functions of unannotated RAY proteins. Using the potential homologs of each unknown RAY protein identified by PSI-BLAST, multiple sequence alignments were created using Clustal Omega[69] to align homologs from previously studied nucleus-forming phages 201φ2–1, ΦKZ, ΦPA3, PCH45, and Goslar. If one or more of these phages were missing a homolog of the query protein (as was the case with some of the non-core genome proteins), a close relative of RAY (*Erwinia* phage AH06) and/or a distant relative of RAY (*Klebsiella* phage Miami) were supplemented to fill out the alignment when possible. The resulting multiple sequence alignments were uploaded to ESPript for visualization using the default settings and downloaded as PDF files.[70]

Fluorescence microscopy—*E. amylovora* cells were inoculated onto imaging pads in welled microscopy slides. Pads were made up of 1% agarose, 25% LB, 2 μg/mL FM4–64, and 0.1 μg/mL DAPI. Between 0–1% arabinose was used to induce expression from the pHERD-30T plasmid, depending on the construct. The slides were incubated at 30°C for 3 hours, then moved to room temperature for infection with 10 μL undiluted RAY lysate. Slides were imaged using the DeltaVision Elite deconvolution microscope (Applied Precision) and deconvolved using the aggressive algorithm in the DeltaVision softWoRx program (Applied Precision). Image analysis was performed on images prior to deconvolution.

Cryo-electron microscopy of RAY-infected cells—*E. amylovora* cells were infected on agarose pads as previously described[7] with a few changes. Briefly, cells were grown on agarose pads as described above for fluorescence microscopy and infected at room temperature with 10 μL undiluted RAY lysate. 60 minutes post infections, 25 μL 25% LB was added to each pad and cells were gently scraped off with the bottom of a 1.7 mL eppendorf tube. A droplet containing cells was collected from each pad and an aliquot was saved as an unconcentrated sample. The remainder was centrifuged at 8000 rcf for 30 sec, and resuspended in a portion of the supernatant to concentrate the cells.

Infected cell suspension was mixed 9:1 with 50% (w/v) trehalose solution to mitigate crystalline ice formation and 4 μL of this mixture immediately applied to a R2/1 Cu 200 grid (Quantifoil), which had been glow-discharged for 1 min at 0.19 mbar and 20 mA in a PELCO easiGlow device shortly before use. The grid was then mounted in a custom-built manual plugning device (MPI-Biochemistry, Martinsreid, Germany) and excess media wicked away with filter paper (Whatman #1) from the backside of the grid for 4–7 seconds prior to plunging the grid into a 50:50 ethane:propane mixture (Airgas) cooled by liquid nitrogen. The data presented in this manuscript were collected from two independent preparations of RAY-infected *E. amylovora* at approximately 100–110 mpi.

Frozen grids of infected cells were mounted into notched Autogrids (TFS) compatible with cryo-focused ion beam milling. Samples were loaded into an Aquilos 2 cryo-focused ion beam/scanning electron microscope (TFS) with a Gallium ion source and milled to generate lamellae using progressively lower milling currents from 0.5 nA to 10 pA as previously described.[71]

For tilt-series collection, samples were transferred to a Titan Krios G3 transmission electron microscope (TFS) operated at 300 kV, configured for fringe-free illumination, and equipped with a K2 direct electron detector (Gatan) mounted post a Quantum 968 LS imaging filter (Gatan). The microscope was operated in EFTEM mode with a slit-width of 20 eV and using a 70 μm objective aperture. Automated data acquisition was performed using SerialEM-v3.8.7[72] and all images were acquired using the K2 in counting mode. Tilt-series were acquired at 4.27 Å/pixel following a dose-symmetric scheme to span a nominal range of either +/− 57° (dataset-1) or +/− 64° (dataset-2) in 2° steps. The exposure was uniformly distributed across tilt-images, achieving totals of approximately 150 e/Å^{-2} (dataset-1) or 170 e/Å^{-2} (dataset-2). Target defoci for tilt-series were between 4 and 6 μm. A total of 32 tilt-series were acquired, of which 26 (13 from each session) were deemed suitable for subsequent processing.

Tomogram segmentation—To aid in segmentation and for display purposes, tomograms were missing-wedge corrected using IsoNet-v0.1[73] on tomograms reconstructed at 20 Å/px and devolved with Warp-v1.09. Membranes were initially segmented using TomoSegMemTV[74] and patched manually using Amira3D-v2021.2 (TFS). Subtomogram averages were placed at refined particle positions using Dynamo-v1.1514. The segmentation was rendered using ChimeraX-v1.3.[75]

Structural predictions—AlphaFold v2.1.0[76] was used in order to predict the tertiary structure of unsolved RAY proteins. We ran $ChmA_{RAY}$ (gp222) and $PhuZ_{RAY}$ (gp210) through AlphaFold v2.1.0 to obtain the predicted structure of the proteins. These predictions were then imported into PyMOL Version 2.5.2[77] in order to visualize the structures. Structural alignments between previously published protein structures and RAY protein structural predictions were made using PyMOL's align function, and the RMSD values of these alignments were found using the same method.

QUANTIFICATION AND STATISTICAL ANALYSIS

Nucleus rotation analysis—Time lapses (2 minutes long with 4-second intervals) were obtained, generally between 60–90 mpi. The time lapses that included actively rotating nuclei were analyzed using FIJI. The segmented line tool was used to measure all distances. To determine the nucleus rotation speed, a distinct point on the surface of the nucleus was tracked frame by frame to measure the total distance traveled. The number of frames used and the time lapse interval size were used to determine the rotation time. The rotation speed was then calculated by dividing the total distance traveled by the rotation time. The total number of cells with nuclei and the number of nuclei that were actively rotating were manually quantified to determine the percentage of rotating nuclei. The angular velocity was determined by dividing the total distance traveled by the length of the nucleus' radius (half of the nucleus' diameter).

PhuZ filamentation and nucleus positioning analysis—For the filamentation analysis, the number of cells with filaments versus without filaments was manually quantified from microscopy images, and the percentage of cells with filaments at different arabinose concentrations was recorded. A filament was defined as a visible line that was at least double the level of background GFP fluorescence. The data were graphed using Prism (GraphPad), and a nonlinear [Agonist] vs. response curve was fit to represent the [Arabinose] vs. filamentation data.

For the nucleus positioning analysis, the length of the cell and the distance between the center of the phage nucleus and the cell pole was measured using FIJI. The nucleus position was calculated as the ratio of the distance from the phage nucleus to the pole over the entire cell length. The data were graphed as a relative frequency histogram using Prism (GraphPad).

Subtomogram analysis—All pre-processing steps were performed using Warp-v1.09[78] unless otherwise specified. Tilt-movies were corrected for whole-frame motion, their defocus values estimated, and stacked into tilt-series. Tilt-series were aligned via patch-tracking with Etomo (IMOD-v4.10.28).[79] Tomograms were reconstructed with the default deconvolution filter settings for visualization and without for template-matching and subsequent processing, when necessary.

First, an *ab initio* host 70S ribosome reference was generated from the data by manually picking particles and aligning them in RELION-v3.1.3.[80,81] A low-pass filtered (50 Å) reference was used for template-matching across the entire dataset and the hits curated by Figure-Of-Merit score to remove false-positives. The 51,512 template-matching hits were

classified in RELION-v3.1.3 without a starting reference to further remove false positives. Alignment of the 28,032 selected particles resulted in a 18 Å map of the 70S ribosome. Further refinement of particle poses and tilt-series parameters in M-v1.09[82] improved the map to 10.3 Å. This map was low-pass filtered and used for a second round of template-matching against the tomograms, which resulted in 74,713 hits after curation. The new particle set was coarsely aligned with RELION-v3.1.3 and followed by M-v1.0.9, which resulted in a 9.5 Å map. Further alignment with RELION-v3.1.3 improved the map to 9.2 Å. Reference-free classification without alignment in identified 45,708 70S particles which refined to 8.9 Å and 2,551 50S particles which refined to 13.4 Å.

For virion capsids, 156 particle positions were manually picked and aligned while enforcing icosahedral symmetry in RELION-v3.1.3. The vertices were subsequently extracted to create 1,872 sub-particles, which aligned with C5 symmetry and then classified in C1 without a starting reference nor additional alignment. The 1,056 selected particles were further classified without a starting reference using an ellipsoid mask, C6 symmetry with relaxed C5 symmetry, and local-searches to separate pentameric vertices from the portal vertices. The 1,023 pentameric vertices were further aligned with C5 symmetry to yield a 19 Å map. The 22 portal vertices were aligned with C6 symmetry to yield a 38 Å map.

For virion tails, the start and endpoints of nine tails exhibiting clear polarity in the tomograms were picked and filament cropping models generated using Dynamo-v1.1514.[83] Subtomograms were extracted every 2 nm to yield 900 segments and the azimuth angles randomized. An initial reference was generated by reconstruction of the segments and smoothing of the resulting map using a Gaussian filter of 3 pixel width. The smoothed map was used to align all segments without enforced symmetry and restricted angular search range. The resulting map exhibited clear polar and C6 symmetry. Next, the start and endpoints of all tails were picked, filament models generated, and subtomograms extracted every 2 nm with randomized azimuth angles. A copy of the metadata table was generated in which the polarity of each filament was flipped. Both tables were aligned for a single iteration against the previously generated reference with C6 symmetry enforced while limiting the alignment to 40. Duplicate particles which had converged on the same position due to the initial oversampling and subsequent alignment step were removed. In order to determine the polarity of segments, the cross-correlation (CC) values for analogous tails between the two tables were compared. For each tail, the orientation leading to the higher median CC across all its segments was chosen for subsequent steps. For tails exhibiting similar, low median CC values in both orientations were discarded. A CC threshold was selected for the entire dataset to remove these ambiguous tails and other low CC segments. This was typical of tails only partially contained within the lamellae, for which a majority was either FIB-ablated or extended outside of the field-of-view. The resulting 3,291 segments were converted for processing in RELION-v3.1.3 using dynamo2m-v0.2.2[84] and halfsets were assigned on a per-tail basis. Refinement in RELION-v3.1.3 while enforcing C6 symmetry resulted in a 10.4 Å map. Reference-free classification with C6 symmetry and without alignment separated tail segments from baseplates. Alignment of the selected 3,257 tails segments with C6 symmetry resulted in a 10 Å map which exhibits a rise of 37.4 Å and twist of 22.3°. The 34 baseplates were aligned with C6 symmetry to yield 36 Å map.

For ChmA, the surfaces of the twenty phage nuclei were coarsely contoured to generate surface cropping models with Dynamo-v1.1514. Subtomograms were extracted every 4 nm and oriented normal to the surface models. An initial reference was generated by reconstruction of the segments and smoothing of the resulting map using a Gaussian filter of 3 pixel width. A subset of 3,910 segments from two tomograms were aligned against the smoothed map with a restricted angular search range to prevent flipping of sidedness and without enforced symmetry. The resulting map exhibited apparent p442 lattice symmetry with an approximate 11.5 nm spacing, which was corroborated by inspection of 'neighbor plots'.[85,86] This new reference was C4 symmeterized and used to align the entire dataset of 123,416 oversampled segments for a single iteration while limiting alignment to 40 Å. Duplicate particles which had converged on the same position due to the initial oversampling and subsequent alignment step were removed to leave 52,646 segments. Misaligned and edge segments were then removed by selecting only segments with at least three neighboring segments with 10–13 nm, which retained 23,509 segments. The metadata was converted for processing in RELION-v3.1.3 using dynamo2m-v0.2.2. Further alignment in RELION-v3.1.3 enforcing C4 symmetry yielded a 20 Å map.

For PhuZ, the start and endpoints of 14 filaments were manually picked from the tomograms and cropping models generated with Dynamo-v1.1514. The polarity of the filaments was uniformly assigned from the cell pole to the mid-cell. Segments were extracted every 1 nm along the filaments and the azimuth angles were randomized, which resulted in 5,000 subtomograms. Sub-tomograms were aligned to the unaligned average reconstructed from the initial cropping points with Dynamo-v1.1514 while limiting the alignment resolution to 40 Å and preventing subtomograms from flipping polarity. Duplicate particles which had converged on the same position due to the initial oversampling and subsequent alignment step were removed, which resulted in 3,080 remaining subtomograms. The metadata was converted for processing in RELION-v3.1.3 using dynamo2m-v0.2.2 and halfsets were assigned on a perfilament basis. Alignment in RELION-v3.1.3 resulted in a 25 Å map. Individual protomers are not resolved in this map and attempts at estimating reliable helical parameters were unsuccessful.

All resolution estimates are based on the 0.143-threshold of the Fourier shell correlation between masked half-maps using high-resolution noise-substitution to mitigate masking artifacts.[87] Local-resolution estimates were commuted using RELION. The virion 'frakenmap' was generated by the *vop maximum* command in UCSF-Chimera-v1.15[88] to combine the various components for display purposes. Maps were rendered using ChimeraX-v1.3 or v1.4.[75]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank B. Dennis and K. Smith of UCSD Physics Computing for computational support. Electron microscopy data were collected at the UCSD Cryo-Electron Microscopy Facility, which was built and equipped with funds from the UCSD and an initial gift from the Agouron Institute. The authors acknowledge funding from the National

Institutes of Health grant R01 GM129245 (to J.P. and E.V.), the National Science Foundation grant DBI 1920374 (to E.V.), and the Howard Hughes Medical Institute Emerging Pathogens Initiative grant (to E.V., J.P., K.P., J.M., and K.D.C.). T.G.L. is a Simons Foundation Awardee of the Life Sciences Research Foundation. E.V. is a Howard Hughes Medical Institute Investigator.

REFERENCES

1. Paszkowski P, Noyce RS, and Evans DH (2016). Live-cell imaging of vaccinia virus recombination. PLoS Pathog. 12, e1005824. [PubMed: 27525721]
2. Tomer E, Cohen EM, Drayman N, Afriat A, Weitzman MD, Zaritsky A, and Kobiler O (2019). Coalescing replication compartments provide the opportunity for recombination between coinfecting herpesviruses. FASEB J 33, 9388–9403. [PubMed: 31107607]
3. Kieser Q, Noyce RS, Shenouda M, Lin Y-CJ, and Evans DH (2020). Cytoplasmic factories, virus assembly, and DNA replication kinetics collectively constrain the formation of poxvirus recombinants. PLoS One 15, e0228028. [PubMed: 31945138]
4. Trinh JT, Shao Q, Guan J, and Zeng L (2020). Emerging heterogeneous compartments by viruses in single bacterial cells. Nat. Commun. 11, 3813. [PubMed: 32732913]
5. Labarde A, Jakutyte L, Billaudeau C, Fauler B, López-Sanz M, Ponien P, Jacquet E, Mielke T, Ayora S, Carballido-López R, and Tavares P (2021). Temporal compartmentalization of viral infection in bacterial cells. Proc. Natl. Acad. Sci. USA 118, e2018297118. 10.1073/pnas.2018297118. [PubMed: 34244425]
6. Knipe DM, Prichard A, Sharma S, and Pogliano J (2022). Replication compartments of eukaryotic and bacterial DNA viruses: common themes between different domains of host cells. Annu. Rev. Virol. 9, 307–327. [PubMed: 36173697]
7. Chaikeeratisak V, Nguyen K, Khanna K, Brilot AF, Erb ML, Coker JKC, Vavilina A, Newton GL, Buschauer R, Pogliano K, et al. (2017). Assembly of a nucleus-like structure during viral replication in bacteria. Science 355, 194–197. [PubMed: 28082593]
8. Chaikeeratisak V, Nguyen K, Egan ME, Erb ML, Vavilina A, and Pogliano J (2017). The phage nucleus and tubulin spindle are conserved among large Pseudomonas phages. Cell Rep. 20, 1563–1571. [PubMed: 28813669]
9. Laughlin TG, Deep A, Prichard AM, Seitz C, Gu Y, Enustun E, Suslov S, Khanna K, Birkholz EA, Armbruster E, et al. (2022). Architecture and self-assembly of the jumbo bacteriophage nuclear shell. Nature 608, 429–435. [PubMed: 35922510]
10. Birkholz EA, Laughlin TG, Armbruster E, Suslov S, Lee J, Wittmann J, Corbett KD, Villa E, and Pogliano J (2022). A cytoskeletal vortex drives phage nucleus rotation during jumbo phage replication in E. coli. Cell Rep. 40, 111179. [PubMed: 35977483]
11. Malone LM, Warring SL, Jackson SA, Warnecke C, Gardner PP, Gumy LF, and Fineran PC (2020). A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. Nat. Microbiol. 5, 48–55. [PubMed: 31819217]
12. Mendoza SD, Nieweglowska ES, Govindarajan S, Leon LM, Berry JD, Tiwari A, Chaikeeratisak V, Pogliano J, Agard DA, and Bondy-Denomy J (2020). A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. Nature 577, 244–248. [PubMed: 31819262]
13. Nguyen KT, Sugie J, Khanna K, Egan ME, Birkholz EA, Lee J, Beierschmitt C, Villa E, and Pogliano J (2021). Selective transport of fluorescent proteins into the phage nucleus. PLoS One 16, e0251429. [PubMed: 34111132]
14. Sokolova ML, Misovetc I, and Severinov K,V. (2020). Multisubunit RNA polymerases of jumbo bacteriophages. Viruses 12, 1064. 10.3390/v12101064. [PubMed: 32977622]
15. Thomas JA, Rolando MR, Carroll CA, Shen PS, Belnap DM, Weintraub ST, Serwer P, and Hardies SC (2008). Characterization of Pseudomonas chlororaphis myovirus 201varphi2–1 via genomic sequencing, mass spectrometry, and electron microscopy. Virology 376, 330–338. [PubMed: 18474389]
16. Thomas JA, Weintraub ST, Wu W, Winkler DC, Cheng N, Steven AC, and Black LW (2012). Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant Pseudomonas aeruginosa phage φKZ. Mol. Microbiol. 84, 324–339. [PubMed: 22429790]

17. Ceyssens P-J, Minakhin L, Van den Bossche A, Yakunina M, Klimuk E, Blasdel B, De Smet J, Noben J-P, Bläsi U, Severinov K, and Lavigne R (2014). Development of giant bacteriophage φKZ is independent of the host transcription apparatus. J. Virol. 88, 10501–10510. [PubMed: 24965474]

18. Matsui T, Yoshikawa G, Mihara T, Chatchawankanphanich O, Kawasaki T, Nakano M, Fujie M, Ogata H, and Yamada T (2017). Replications of two closely related groups of jumbo phages show different level of dependence on host-encoded RNA polymerase. Front. Microbiol. 8, 1010. [PubMed: 28659872]

19. Yakunina M, Artamonova T, Borukhov S, Makarova KS, Severinov K, and Minakhin L (2015). A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage. Nucleic Acids Res. 43, 10411–10420. [PubMed: 26490960]

20. Orekhova M, Koreshova A, Artamonova T, Khodorkovskii M, and Yakunina M (2019). The study of the phiKZ phage non-canonical non-virion RNA polymerase. Biochem. Biophys. Res. Commun. 511, 759–764. [PubMed: 30833081]

21. Krylov V, Bourkaltseva M, Pleteneva E, Shaburova O, Krylov S, Karaulov A, Zhavoronok S, Svitich O, and Zverev V (2021). Phage phiKZ-the first of giants. Viruses 13, 149. 10.3390/v13020149. [PubMed: 33498475]

22. de Maríın Garrido N, Orekhova M, Lai Wan Loong YTE, Litvinova A, Ramlaul K, Artamonova T, Melnikov AS, Serdobintsev P, Aylett CHS, and Yakunina M (2021). Structure of the bacteriophage PhiKZ non-virion RNA polymerase. Nucleic Acids Res. 49, 7732–7739. [PubMed: 34181731]

23. Zehr EA, Kraemer JA, Erb ML, Coker JKC, Montabana EA, Pogliano J, and Agard DA (2014). The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA. Structure 22, 539–548. [PubMed: 24631461]

24. Aylett CHS, Izoré T, Amos LA, and Löwe J (2013). Structure of the tubulin/FtsZ-like protein TubZ from Pseudomonas bacteriophage ΦKZ. J. Mol. Biol. 425, 2164–2173. [PubMed: 23528827]

25. Kraemer JA, Erb ML, Waddling CA, Montabana EA, Zehr EA, Wang H, Nguyen K, Pham DSL, Agard DA, and Pogliano J (2012). A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell. Cell 149, 1488–1499. [PubMed: 22726436]

26. Erb ML, Kraemer JA, Coker JKC, Chaikeeratisak V, Nonejuie P, Agard DA, and Pogliano J (2014). A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells. Elife 3, e03197. 10.7554/eLife.03197. [PubMed: 25429514]

27. Chaikeeratisak V, Khanna K, Nguyen KT, Sugie J, Egan ME, Erb ML, Vavilina A, Nonejuie P, Nieweglowska E, Pogliano K, et al. (2019). Viral capsid trafficking along treadmilling tubulin filaments in bacteria. Cell 177, 1771–1780.e12. [PubMed: 31199917]

28. Chaikeeratisak V, Khanna K, Nguyen KT, Egan ME, Enustun E, Armbruster E, Lee J, Pogliano K, Villa E, and Pogliano J (2022). Subcellular organization of viral particles during maturation of nucleus-forming jumbo phage. Sci. Adv. 8, eabj9670. [PubMed: 35507660]

29. Sharma R, Pielstick BA, Bell KA, Nieman TB, Stubbs OA, Yeates EL, Baltrus DA, and Grose JH (2019). A novel, highly related jumbo family of bacteriophages that were isolated against Erwinia. Front. Microbiol. 10, 1533. [PubMed: 31428059]

30. Vanneste JL (2000). Fire Blight: The Disease and its Causative Agent, Erwinia Amylovora (CABI).

31. Arens DK, Brady TS, Carter JL, Pape JA, Robinson DM, Russell KA, Staley LA, Stettler JM, Tateoka OB, Townsend MH, et al. (2018). Characterization of two related Erwinia myoviruses that are distant relatives of the PhiKZ-like Jumbo phages. PLoS One 13, e0200202. [PubMed: 29979759]

32. Esplin IND, Berg JA, Sharma R, Allen RC, Arens DK, Ashcroft CR, Bairett SR, Beatty NJ, Bickmore M, Bloomfield TJ, et al. (2017). Genome sequences of 19 novel Erwinia amylovora bacteriophages. Genome Announc. 5, e00931–17. 10.1128/genomeA.00931-17. [PubMed: 29146842]

33. Konstantinidis KT, and Tiedje JM (2005). Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. USA 102, 2567–2572. [PubMed: 15701695]

34. Park S-C, Lee K, Kim YO, Won S, and Chun J (2019). Large-scale genomics reveals the genetic characteristics of seven species and importance of phylogenetic distance for estimating Pan-genome size. Front. Microbiol. 10, 834. [PubMed: 31068915]

35. Comeau AM, Bertrand C, Letarov A, Tétart F, and Krisch HM (2007). Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. Virology 362, 384–396. [PubMed: 17289101]
36. Cazares A, Mendoza-Hernández G, and Guarneros G (2014). Core and accessory genome architecture in a group of Pseudomonas aeruginosa Mu-like phages. BMC Genom. 15, 1146.
37. Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, Koehrsen M, Rokas A, Yandava CN, Engels R, Zeng E, et al. (2008). Dynamics of Pseudomonas aeruginosa genome evolution. Proc. Natl. Acad. Sci. USA 105, 3100–3105. [PubMed: 18287045]
38. Jang HB, Fagutao FF, Nho SW, Park SB, Cha IS, Yu JE, Lee JS, Im SP, Aoki T, and Jung TS (2013). Phylogenomic network and comparative genomics reveal a diverged member of the ΦKZ-related group, marine vibrio phage ΦJM-2012. J. Virol. 87, 12866–12878. [PubMed: 24067958]
39. Iyer LM, Anantharaman V, Krishnan A, Burroughs AM, and Aravind L (2021). Jumbo phages: a comparative genomic overview of core functions and adaptions for biological conflicts. Viruses 13, 63. 10.3390/v13010063. [PubMed: 33466489]
40. Chaikeeratisak V, Birkholz EA, Prichard AM, Egan ME, Mylvara A, Nonejuie P, Nguyen KT, Sugie J, Meyer JR, and Pogliano J (2021). Viral speciation through subcellular genetic isolation and virogenesis incompatibility. Nat. Commun. 12, 342. [PubMed: 33436625]
41. Letunic I, and Bork P (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49, W293–W296. [PubMed: 33885785]
42. van Tonder AJ, Mistry S, Bray JE, Hill DMC, Cody AJ, Farmer CL, Klugman KP, von Gottberg A, Bentley SD, Parkhill J, et al. (2014). Defining the estimated core genome of bacterial populations using a Bayesian decision model. PLoS Comput. Biol. 10, e1003788. [PubMed: 25144616]
43. Guan J, Oromí-Bosch A, Mendoza SD, Karambelkar S, Berry J, and Bondy-Denomy J (2022). RNA targeting with CRISPR-Cas13a facilitates bacteriophage genome engineering. Preprint at bioRxiv. 10.1101/2022.02.14.480438.
44. Fokine A, Battisti AJ, Bowman VD, Efimov AV, Kurochkina LP, Chipman PR, Mesyanzhinov VV, and Rossmann MG (2007). Cryo-EM study of the Pseudomonas bacteriophage phiKZ. Structure 15, 1099–1104. [PubMed: 17850749]
45. Barylski J, Enault F, Dutilh BE, Schuller MB, Edwards RA, Gillis A, Klumpp J, Knezevic P, Krupovic M, Kuhn JH, et al. (2020). Analysis of spounaviruses as a case study for the overdue reclassification of tailed phages. Syst. Biol. 69, 110–123. [PubMed: 31127947]
46. Krylov VN, and Zhazykov IZ (1978). Pseudomonas bacteriophage phiKZ–possible model for studying the genetic control of morphogenesis. Genetika 14, 678–685. [PubMed: 95987]
47. Blazanin M, Lam WT, Vasen E, Chan BK, and Turner PE (2022). Decay and damage of therapeutic phage OMKO1 by environmental stressors. PLoS One 17, e0263887. [PubMed: 35196336]
48. Latz S, Krüttgen A, Häfner H, Buhl EM, Ritter K, and Horz H-P (2017). Differential effect of newly isolated phages belonging to PB1-like, phiKZ-like and LUZ24-like viruses against multi-drug resistant Pseudomonas aeruginosa under varying growth conditions. Viruses 9, 315. 10.3390/v9110315. [PubMed: 29077053]
49. Danis-Wlodarczyk K, Vandenheuvel D, Jang HB, Briers Y, Olszak T, Arabski M, Wasik S, Drabik M, Higgins G, Tyrrell J, et al. (2016). A proposed integrated approach for the preclinical evaluation of phage therapy in Pseudomonas infections. Sci. Rep. 6, 28115. [PubMed: 27301427]
50. Olsen NS, Hendriksen NB, Hansen LH, and Kot W (2020). A new high-throughput screening method for phages: enabling crude isolation and fast identification of diverse phages with therapeutic potential. Phage (New Rochelle) 1, 137–148. [PubMed: 36147828]
51. Vaitekenas A, Tai AS, Ramsay JP, Stick SM, Agudelo-Romero P, and Kicic AWAERP; AREST CF (2022). Complete genome sequences of four Pseudomonas aeruginosa bacteriophages: Kara-mokiny 8, Kara-mokiny 13, Kara-mokiny 16, and Boorn-mokiny 1. Microbiol. Resour. Announc. 11, e0096022. [PubMed: 36374083]
52. Aghaee BL, Khan Mirzaei M, Alikhani MY, Mojtahedi A, and Maurice CF (2021). Improving the inhibitory effect of phages against Pseudomonas aeruginosa isolated from a burn patient using a combination of phages and antibiotics. Viruses 13, 334. 10.3390/v13020334. [PubMed: 33670028]

53. Erickson HP, Anderson DE, and Osawa M (2010). FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. Microbiol. Mol. Biol. Rev. 74, 504–528. [PubMed: 21119015]
54. Du S, and Lutkenhaus J (2019). At the heart of bacterial cytokinesis: the Z ring. Trends Microbiol. 27, 781–791. [PubMed: 31171437]
55. Haeusser DP, and Margolin W (2016). Splitsville: structural and functional insights into the dynamic bacterial Z ring. Nat. Rev. Microbiol. 14, 305–319. [PubMed: 27040757]
56. Larsen RA, Cusumano C, Fujioka A, Lim-Fong G, Patterson P, and Pogliano J (2007). Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis. Genes Dev. 21, 1340–1352. [PubMed: 17510284]
57. Aylett CHS, Wang Q, Michie KA, Amos LA, and Löwe J (2010). Filament structure of bacterial tubulin homologue TubZ. Proc. Natl. Acad. Sci. USA 107, 19766–19771. [PubMed: 20974911]
58. Montabana EA, and Agard DA (2014). Bacterial tubulin TubZ-Bt transitions between a two-stranded intermediate and a four-stranded filament upon GTP hydrolysis. Proc. Natl. Acad. Sci. USA 111, 3407–3412. [PubMed: 24550513]
59. Schlieper D, Oliva MA, Andreu JM, and Löwe J (2005). Structure of bacterial tubulin BtubA/B: evidence for horizontal gene transfer. Proc. Natl. Acad. Sci. USA 102, 9170–9175. [PubMed: 15967998]
60. Yee B, Lafi FF, Oakley B, Staley JT, and Fuerst JA (2007). A canonical FtsZ protein in Verrucomicrobium spinosum, a member of the Bacterial phylum Verrucomicrobia that also includes tubulin-producing Prosthecobacter species. BMC Evol. Biol. 7, 37. [PubMed: 17349062]
61. Díaz-Celis C, Risca VI, Hurtado F, Polka JK, Hansen SD, Maturana D, Lagos R, Mullins RD, and Monasterio O (2017). Bacterial tubulins A and B exhibit polarized growth, mixed-polarity bundling, and destabilization by GTP hydrolysis. J. Bacteriol. 199, e00211–17. 10.1128/JB.00211-17. [PubMed: 28716960]
62. Falco SC, Zehring W, and Rothman-Denes LB (1980). DNA-dependent RNA polymerase from bacteriophage N4 virions. Purification and characterization. J. Biol. Chem. 255, 4339–4347. [PubMed: 6989837]
63. Molineux IJ, and Panja D (2013). Popping the cork: mechanisms of phage genome ejection. Nat. Rev. Microbiol. 11, 194–204. [PubMed: 23385786]
64. Jin Y, Sdao SM, Dover JA, Porcek NB, Knobler CM, Gelbart WM, and Parent KN (2015). Bacteriophage P22 ejects all of its internal proteins before its genome. Virology 485, 128–134. [PubMed: 26245366]
65. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, and Goto S (2017). ViPTree: the viral proteomic tree server. Bioinformatics 33, 2379–2380. [PubMed: 28379287]
66. Meier-Kolthoff JP, and Göker M (2017). VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33, 3396–3404. [PubMed: 29036289]
67. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, and Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. [PubMed: 9254694]
68. Kelley LA, Mezulis S, Yates CM, Wass MN, and Sternberg MJE (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858. [PubMed: 25950237]
69. Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, Madhusoodanan N, Kolesnikov A, and Lopez R (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 50, W276–W279. 10.1093/nar/gkac240. [PubMed: 35412617]
70. Robert X, and Gouet P (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42, W320–W324. [PubMed: 24753421]
71. Lam V, and Villa E (2021). Practical approaches for cryo-FIB milling and applications for cellular cryo-electron tomography. Methods Mol. Biol. 2215, 49–82.
72. Mastronarde DN (2005). Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol. 152, 36–51. [PubMed: 16182563]
73. Liu Y-T, Zhang H, Wang H, Tao C-L, Bi G-Q, and Zhou ZH (2022). Isotropic reconstruction for electron tomography with deep learning. Nat. Commun. 13, 6482. [PubMed: 36309499]

74. Martinez-Sanchez A, Garcia I, Asano S, Lucic V, and Fernandez J-J (2014). Robust membrane detection based on tensor voting for electron tomography. J. Struct. Biol. 186, 49–61. 10.1016/j.jsb.2014.02.015. [PubMed: 24625523]

75. Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, and Ferrin TE (2018). UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25. [PubMed: 28710774]

76. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, žídek A, Potapenko A, et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. [PubMed: 34265844]

77. Schrödinger LLC (2021). The PyMOL Molecular Graphics System, Version 2.5.2.

78. Tegunov D, and Cramer P (2019). Real-time cryo-electron microscopy data preprocessing with Warp. Nat. Methods 16, 1146–1152. [PubMed: 31591575]

79. Kremer JR, Mastronarde DN, and McIntosh JR (1996). Computer visualization of three-dimensional image data using IMOD. J. Struct. Biol. 116, 71–76. [PubMed: 8742726]

80. Scheres SHW (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530. [PubMed: 23000701]

81. Bharat TAM, Russo CJ, Löwe J, Passmore LA, and Scheres SHW (2015). Advances in single-particle electron cryomicroscopy structure determination applied to sub-tomogram averaging. Structure 23, 1743–1753. 10.1016/j.str.2015.06.026. [PubMed: 26256537]

82. Tegunov D, Xue L, Dienemann C, Cramer P, and Mahamid J (2021). Multi-particle cryo-EM refinement with M visualizes ribosome-antibiotic complex at 3.5 Å in cells. Nat. Methods 18, 186–193. [PubMed: 33542511]

83. Navarro P, Scaramuzza S, Stahlberg H, and Castaño-Díez D (2020). The dynamo software package for cryo-electron tomography and subtomogram averaging. Microsc. Microanal. 26, 3142–3145. 10.1017/s1431927620023958.

84. Burt A, Gaifas L, Dendooven T, and Gutsche I (2021). A flexible framework for multi-particle refinement in cryo-electron tomography. PLoS Biol. 19, e3001319. [PubMed: 34437530]

85. Pyle E, Hutchings J, and Zanetti G (2022). Strategies for picking membrane-associated particles within subtomogram averaging workflows. Faraday Discuss 240, 101–113. 10.1039/d2fd00022a. [PubMed: 35924570]

86. Kovtun O, Leneva N, Bykov YS, Ariotti N, Teasdale RD, Schaffer M, Engel BD, Owen DJ, Briggs JAG, and Collins BM (2018). Structure of the membrane-assembled retromer coat determined by cryo-electron tomography. Nature 561, 561–564. 10.1038/s41586-018-0526-z. [PubMed: 30224749]

87. Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SHW, and Henderson R (2013). High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. Ultramicroscopy 135, 24–35. 10.1016/j.ultramic.2013.06.004. [PubMed: 23872039]

88. Goddard TD, Huang CC, and Ferrin TE (2007). Visualizing density maps with UCSF Chimera. J. Struct. Biol. 157, 281–287. [PubMed: 16963278]

Highlights

- *Erwinia* phage RAY is a nucleus-forming phage
- Nucleus-forming phages cluster phylogenetically and share a core genome, including ChmA
- We propose that all phage encoding ChmA form a family of nucleus-forming phages

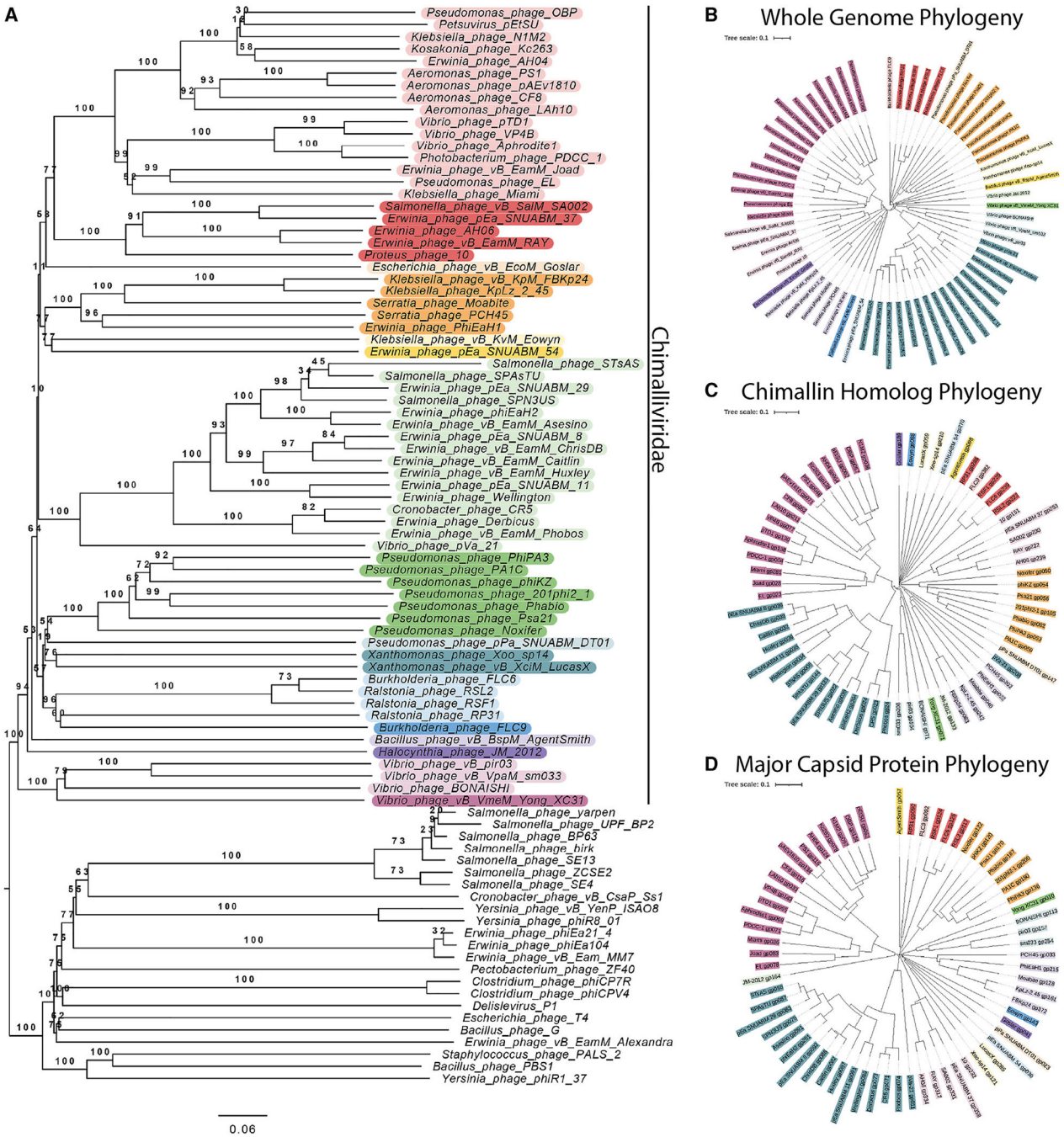


Figure 1. Phylogenetic comparison of Chimalliviridae

(A) Phylogenetic tree of the Chimalliviridae and related phages. The Chimalliviridae appear to form one clade, separate from msRNAP-encoding phages, ViPTree-predicted relatives, and other phages with large genomes color coded by predicted genus (Table S2) to improve readability.

(B) Phylogenetic tree based on whole-genome comparison of 66 representative chimallin-encoding phages (Table S1).

(C and D) Phylogenetic tree based on the protein sequences of the (C) chimallin homologs and (D) major capsid proteins from the 66 phages used in our analysis. Trees were colored in iTOL[41] by predicted genus (Table S2). Some of these predicted genera are not fully consistent with current ICTV classification, but for maximum readability and consistency of the colors, only the genera predicted by VICTOR are color coded.

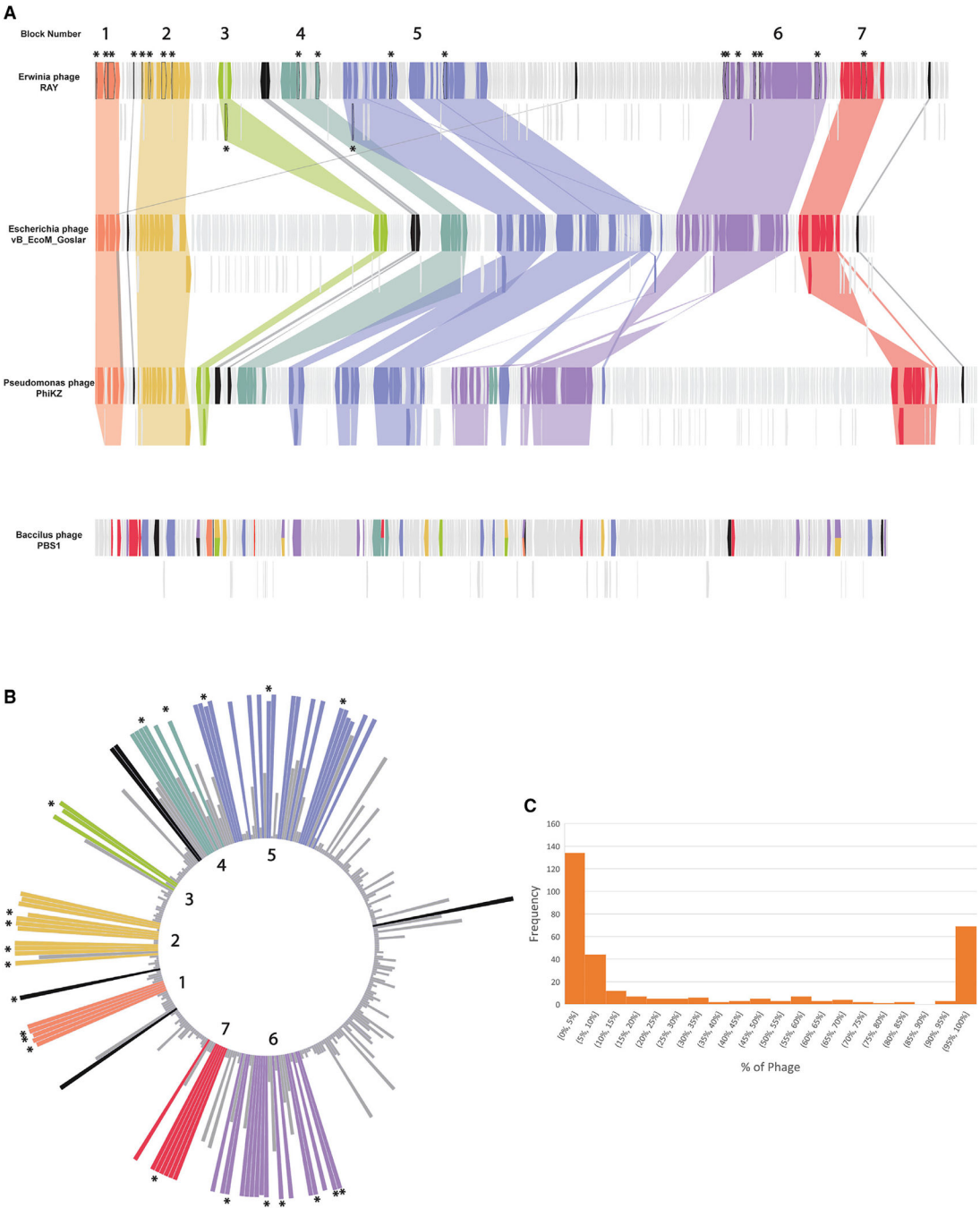


Figure 2. Core genome determination

(A) A genome alignment of *Erwinia* phage RAY, *Escherichia* phage Goslar, *Pseudomonas* phage ΦKZ, and *Bacillus* phage PBS1. PBS1 was included for comparison since it encodes the unique msRNAP but does not encode chimallin. The highly conserved regions with more than 3 genes are annotated in colors as different blocks. Core genes that are not part of blocks are colored black. While in chimalliviruses the core genes are conserved in seven blocks, in PBS1, some of these genes are present but are dispersed across its genome rather than being encoded in blocks. Goslar gp188 and ΦKZ gp055 are homologous to each other

and to two RAY genes, gp223 (orange) and gp070 (black). The full list of genes and blocks can be found in Table S3.
(B) A circular bar plot of RAY's genome with heights of bars corresponding to the percentage of phage that contain genes homologous to each RAY gene. Certain regions are more variable (right of circle), while other regions are highly conserved.
Colors denote blocks and asterisks denote unique genes in (A) and (B).
(C) Histogram showing the frequency of conservation of each RAY protein among chimalliviruses. The left peak indicates RAY-specific proteins (conserved in less than 10% of chimalliviruses), and the right peak indicates core proteins (conserved in 90% or more chimalliviruses).

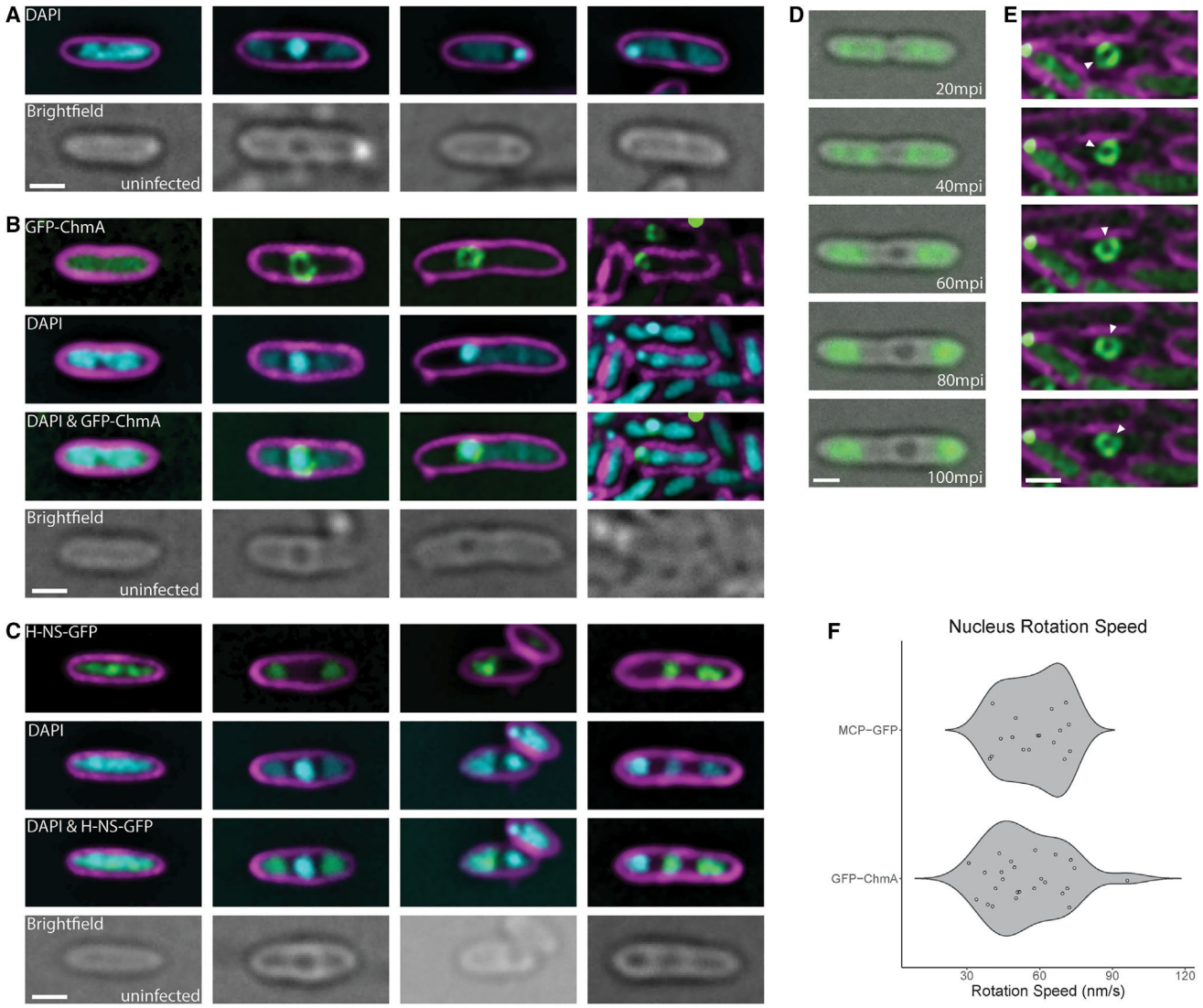


Figure 3. RAY phage nucleus formation

(A) DAPI localization showing the presence of a bright staining consistent with phage nucleus formation. Cells were imaged during midinfection between 60 and 75 minutes post infection (mpi).

(B) GFP-ChmA$_{RAY}$ surrounds the bright DAPI zones, consistent with the morphology of a phage nucleus.

(C) H-NS-GFP localizes to the DNA outside of the phage nucleus.

(D) H-NS-GFP moves away from midcell during the course of infection (Video S1).

(E) The phage nucleus rotates during infection (Video S2). Time-lapse images were taken 4 s apart. The white arrow shows a segment of the nuclear shell that was tracked to determine rotation speed.

(F) Violin plot showing distribution of rotation speeds of individual measured RAY nuclei using either a chimallin (ChmA) (n = 25) or major capsid protein (MCP) (n = 17) GFP tag to track rotation. Scale bar is 1 μm; magenta is membrane stain FM4–64, cyan is DNA stain

DAPI, green is GFP, and grayscale is brightfield. Cells were imaged between 60 and 75 mpi (midinfection) unless labeled otherwise.

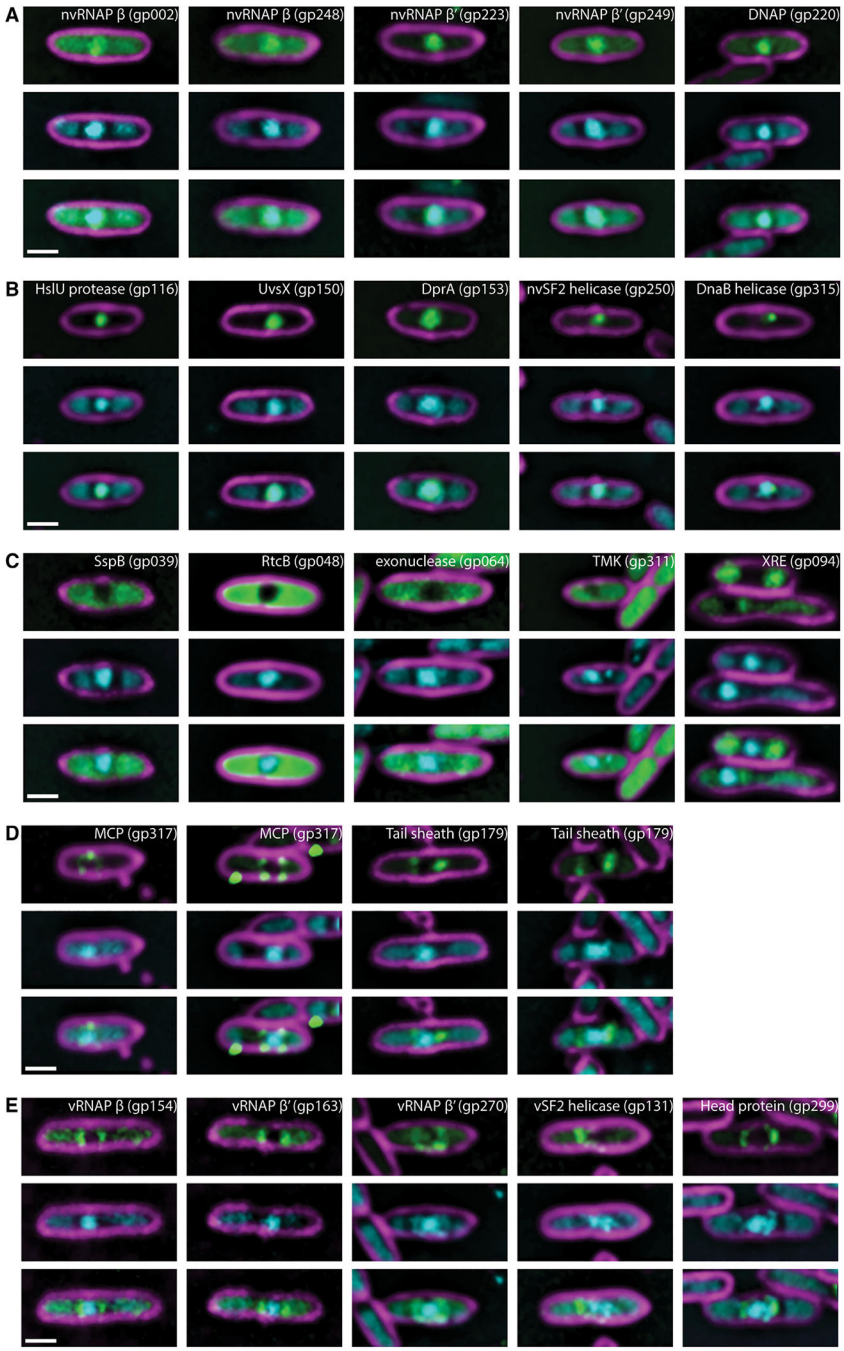


Figure 4. Localization of proteins encoded by RAY

(A) Putative RNA (nvRNAP β subunit 1 gp002, nvRNAP β subunit 2 gp248, nvRNAP β′ subunit 1 gp223, and nvRNAP β′ subunit 2 gp249) and DNA (DNA polymerase gp220) polymerases co-localize with phage DNA in the phage nucleus.

(B) RAY gp116 is a homolog of HslUV superfamily heat shock proteases, but other nuclear-localized phage proteins are homologs of known DNA-associated proteins such as proteins involved in recombination (UvsX gp150 and DprA gp153) and helicases (non-virion superfamily 2 helicase gp250 and DnaB-like replicative helicase gp315).

(C) Cytoplasmic proteins are not predicted to be involved with DNA replication, recombination, or transcription. These include an SspB homolog gp039, RtcB homolog gp048, putative exonuclease gp064, thymidylate kinase gp311, and putative XRE superfamily transcriptional regulator gp094.
(D) The MCP gp317 and tail sheath gp179 are structural components of the RAY virion that localize near the periphery of the phage nucleus. Puncta can be seen near the membrane in the MCP fusion, possibly because capsids are assembled there or, alternatively, as part of an aggregate formed due to overexpression of gp317-GFP.
(E) Virion proteins (vRNAP β subunit 2 gp154, vRNAP β′ subunit 1 gp163, vRNAP β′ subunit 2 gp270, virion-associated superfamily 2 helicase gp131, and head protein of unknown function gp299) can be seen with similar localizations as the MCP and sometimes (in the case of gp270 and gp299) have visible DAPI co-localized with them.
For all images, scale bar is 1 μm; magenta is membrane stain FM4–64, cyan is DNA stain DAPI, and green is GFP.

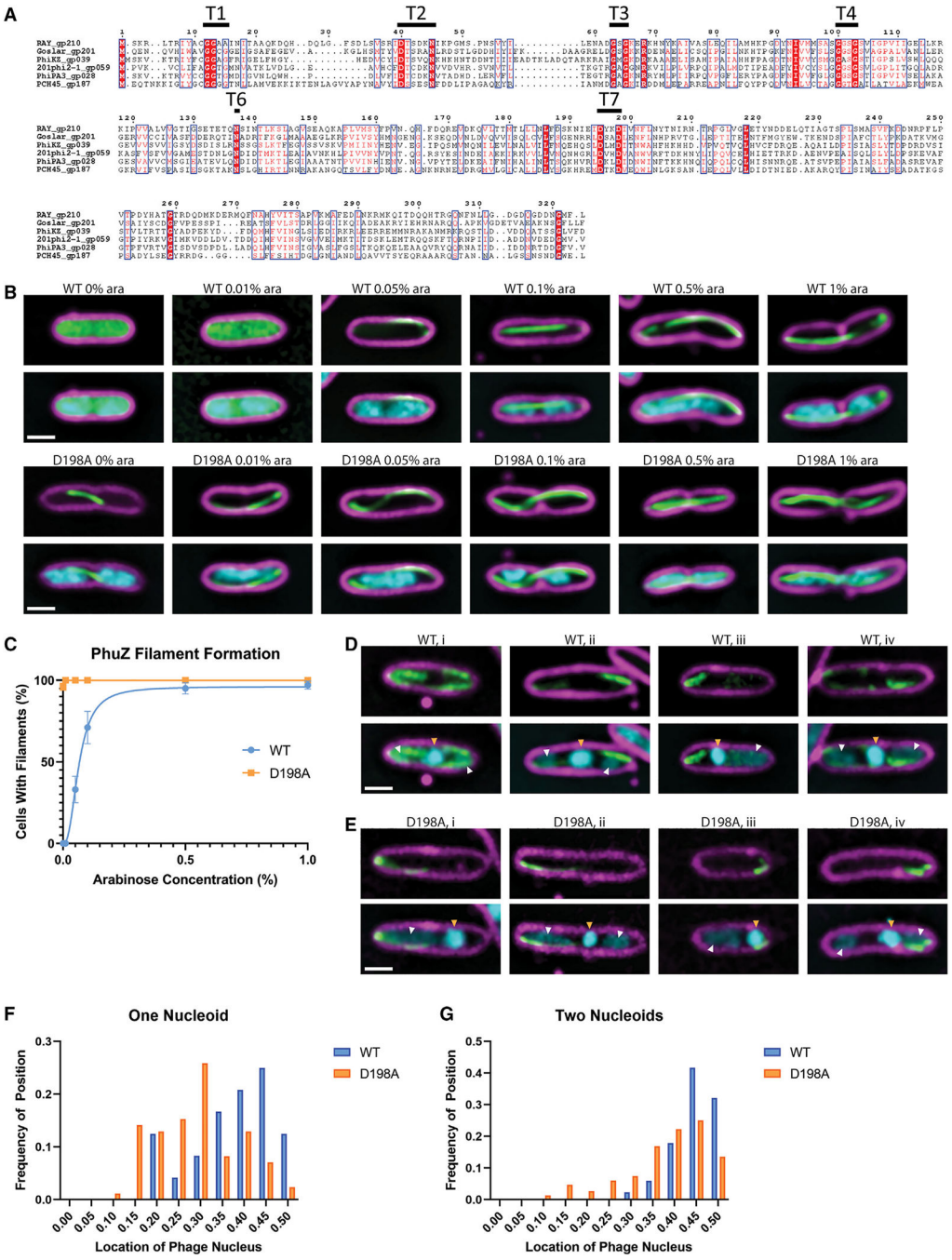


Figure 5. RAY PhuZ homolog

(A) $PhuZ_{RAY}$ has conserved tubulin motifs. T5 is a structural motif with a poorly conserved sequence and is not included.

(B) Wild-type $PhuZ_{RAY}$ does not polymerize spontaneously until it reaches a critical concentration, but the D198A mutant polymerizes at all levels of arabinose induction tested.

(C) The average percentage of uninfected cells with PhuZ filaments plotted versus arabinose concentration. Data are represented as mean ± SEM.

(D) Wild-type PhuZ filaments form with 0% arabinose during RAY infections. The nucleus is normally positioned near midcell (white arrows show bacterial DNA, and gold arrows show phage nuclear DNA).
(E) When the D198A mutant PhuZ is expressed, midcell localization of the phage nucleus during infection is not as common as with wild-type PhuZ (white arrows show bacterial DNA, and gold arrows show phage nuclear DNA).
(F) When only one bacterial nucleoid is present, the phage nucleus positioning histogram has a wide distribution for wild-type PhuZ (blue), and the D198A mutant PhuZ (orange) does not appear to strongly affect positioning.
(G) When two bacterial nucleoids are present, the wild-type PhuZ has a strong nucleus positioning bias toward midcell (blue), and the D198A mutant PhuZ has a weaker positioning bias toward midcell (orange).
For all microscopy images (B, D, and E), scale bar is 1 μm; magenta is membrane stain FM4–64, cyan is DNA stain DAPI, and green is GFP.

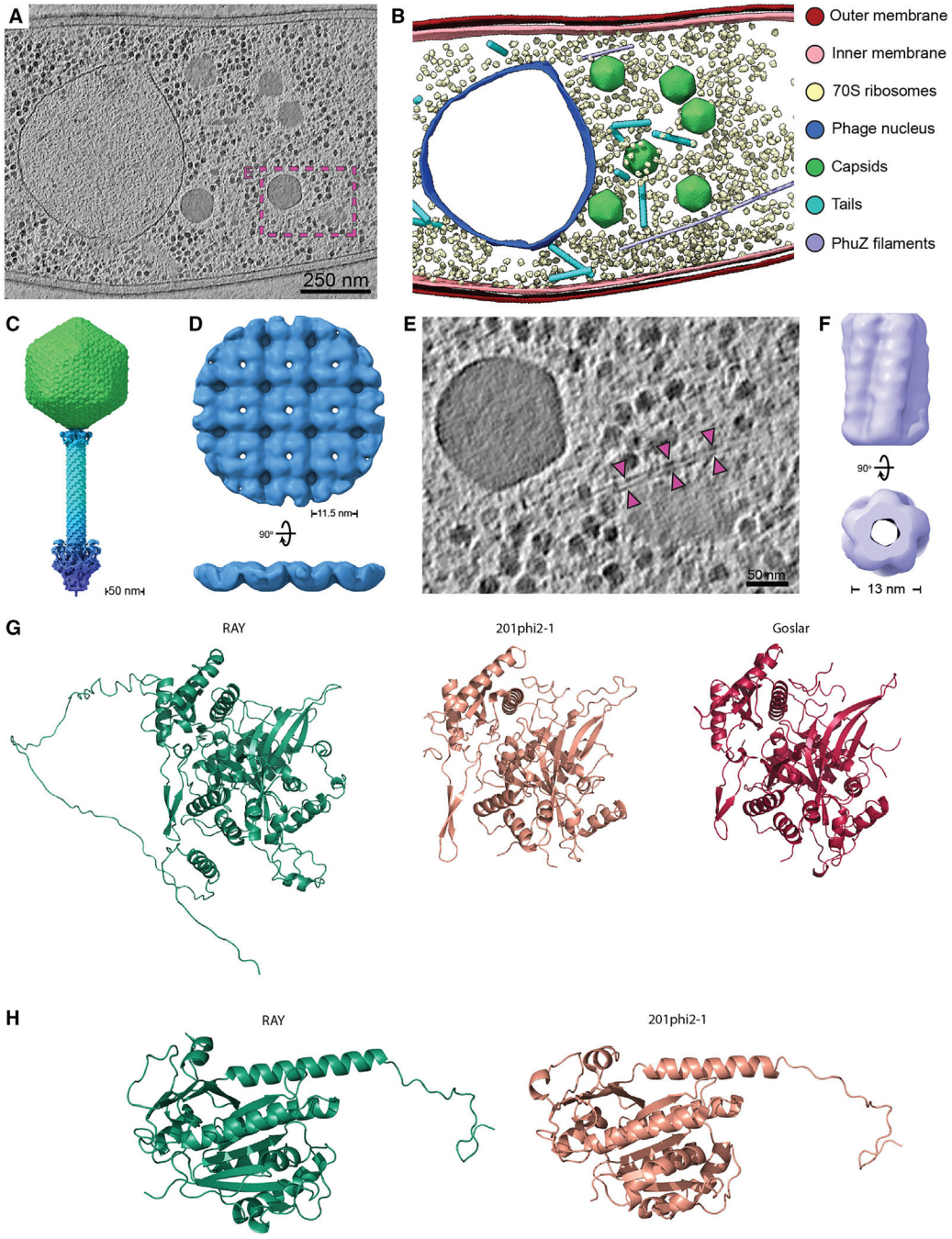


Figure 6. Cryoelectron tomography and structural analysis

(A) Slice through a cryoelectron tomogram of a RAY-infected *Erwinia amylovora* cell at approximately 105–110 mpi. Scale bar is 250 nm.

(B) Segmentation of the tomogram in (A).

(C) Composite RAY virion reconstruction.

(D) Orthogonal views of the RAY chimallin lattice reconstruction.

(E) Enlarged view of region boxed in (A) with magenta arrows pointing to a putative RAY PhuZ filament. Scale bar is 50 nm.

(F) Reconstruction of the putative RAY PhuZ filaments showing the five-stranded structure and hollow lumen.

(G) AlphaFold v.2.1.0 structure of RAY chimallin compared with experimentally determined 201φ2–1 and Goslar chimallin.[9]

(H) AlphaFold v.2.1.0 structure of RAY PhuZ compared with experimentally determined 201φ2–1 PhuZ.[23,25]

Accession numbers: (C) EMD-28003 (additional map), (D) EMD-28007, and (F) EMD-28008. See Figure S7 for subtomogram analysis workflows of RAY components.

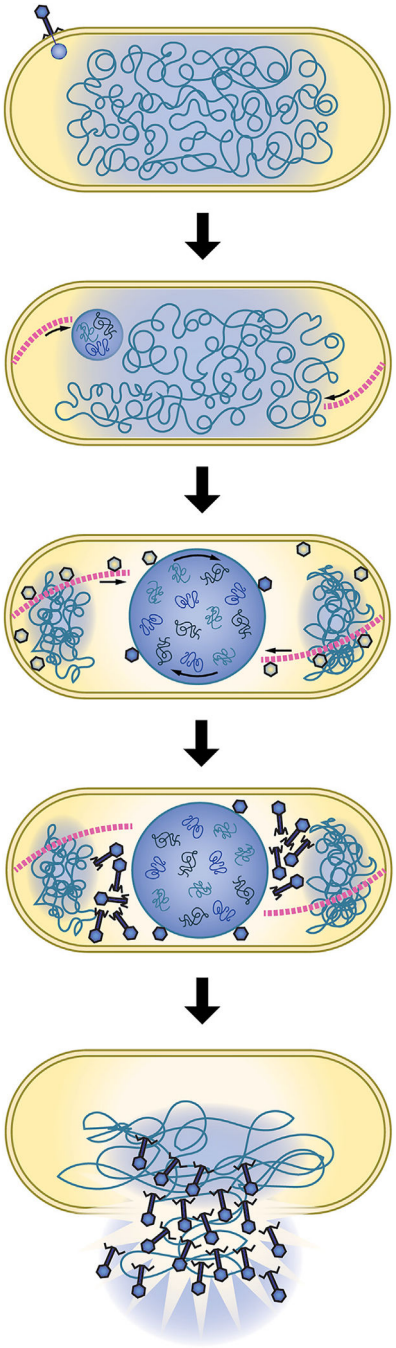


Figure 7. RAY infection model
See text for details.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Erwinia amylovora	American Type Culture Collection	ATCC 29780
Erwinia phage RAY	Julie Grose	vB_EamM_RAY
Chemicals, peptides, and recombinant proteins		
DAPI (4',6-diamidino-2-phenylindole dihydrochloride)	Life Technologies	Cat#D21490
FM4–64	Thermo Scientific	Cat#T13320
Gentamicin	MP Biomedicals, LLC	Cat#190057
Arabinose	Sigma Aldrich	Cat#A91906
Deposited data		
Fluorescence microscopy	This study	https://doi.org/10.17632/jd4yj84463.1
Cryo-ET tilt series	This study	EMPIAR-11198
Set of proteins encoded in the genomes of all tailed phages [taxid: 28883]	NCBI	https://www.ncbi.nlm.nih.gov/protein/?term=txid28883[organism:exp]
Recombinant DNA		
pHERD-30T plasmid	Qiu D, Damron FH, Mima T, Schweizer HP, Yu HD.	https://novoprolabs.com/vector/Vg4ydcni
Table S4	This study	N/A
Software and algorithms		
Core genome identification script	This study	https://doi.org/10.5281/zenodo.7787482
DeltaVision softWoRx 6.5.2	GE HealthCare	http://incelldownload.gehealthcare.com/bin/download_data/SoftWoRx/7.0.0/SoftWoRx.htm
FIJI (ImageJ 1.52p)	Wayne Rasband, National Institutes of Health, USA	http://imagej.nih.gov/ij
GraphPad Prism version 9.4.1	GraphPad Software, San Diego, California, USA	www.graphpad.com
Benchling	Benchling	benchling.com
Microsoft Excel	Microsoft Office	http://www.microsoft.com/en-us/microsoft-365/excel
Microsoft PowerPoint	Microsoft Office	http://www.microsoft.com/en-us/microsoft-365/powerpoint
Microsoft Word	Microsoft Office	http://www.microsoft.com/en-us/microsoft-365/word
Adobe Illustrator	Adobe	http://www.adobe.com/products/illustrator.html
Adobe Photoshop	Adobe	http://www.adobe.com/products/photoshop.html
R-V1.2.1335	CRAN	https://cran.r-project.org/bin/windows/base
Python-v3.9.11	Python	https://www.python.org
PSI-BLAST-v2.12.0	Bioconda	https://anaconda.org/bioconda/blast
IMG/VR	JGI	https://img.jgi.doe.gov/cgi-bin/vr/main.cgi
SerialEM-v3.8.7	PMID: 16182563	bio3d.colorado.edu/SerialEM/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IsoNet-v0.1	https://doi.org/10.1101/2021.07.17.452128	https://zenodo.org/record/7016051#.Y_fE8x_MKUk
Warp-v1.09	PMID: 31591575	www.warpem.com/warp/
TomoSegMemTV	https://doi.org/10.1016/jjsb.2014.02.015	http://sites.google.com/site/3demimageprocessing/tomosegmemtv
Amira3D-v2021.2	Thermo Fisher Scientific	http://www.thermofisher.com/us/en/home/electron-microscopy/products/software-em-3d-vis/amira-software.html
Dynamo-v1.1514	https://doi.org/10.1017/s1431927620023958	https://wiki.dynamo.biozentrum.unibas.ch/w/index.php/Main_Page
ChimeraX-v1.3 or v1.4	PMID: 28710774	http://www.rbvi.ucsf.edu/chimerax/
AlphaFold v2.1.0	DeepMind	https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb
PyMOL version 2.5.2	Schrodinger	https://pymol.org/2/
IMOD-v4.10.28	PMID: 8742726	http://bio3d.colorado.edu/imod/
RELION-v3.13	https://doi.org/10.1016/j.jsb.2012.09.006 and https://doi.org/10.1016/j.str.2015.06.026	http://relion.readthedocs.io/en/release-3.1/
M-v1.09	PMID: 33542511	https://warpem.com/warp/#
dynamo2m-v0.2.2	PMID: 34437530	http://github.com/alisterburt/dynamo2m
UCSF-Chimera-v1.15	PMID: 16963278	https://www.cgl.ucsf.edu/chimera/download.html