UNIVERSITY OF CALIFORNIA SAN DIEGO

Nucleus-Forming Jumbo Phage:
Disrupting the Landscape of Phage Replication and Viral Speciation

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biology

by

Erica A. Birkholz

Committee in charge:

Professor Joseph Pogliano, Chair
Professor Kevin Corbett
Professor Arshad Desai
Professor Justin Meyer
Professor Elizabeth Villa

2022

The Dissertation of Erica A. Birkholz is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

ACKNOWLEDGEMENTS

First and foremost, I would like to express my unbounded gratitude for my thesis advisor, Joe Pogliano. You have known me for almost 9 years now, watching me grow into adulthood and greatly influencing my maturation into a bona fide biologist. I could not have imagined a more supportive and fruitful mentorship. You have never wavered in your sincerity, encouragement, and enthusiasm for cool science that exists outside of the mainstream. I could always count on you, as a mentor, a friend, and a goofy science-dad. I feel incredibly fortunate to be naturally drawn to cell biology, on which you happen to offer a unique perspective in the realm of my favorite model organism, bacteria. Many people write them off as simplistic unorganized organisms with no interesting cell biology. Boy, have we proved them wrong with the help of nucleus-forming jumbo phage! The phage nucleus system is really a dream project for me and with you as a dream mentor for me, I have had the most fantastic experience as a PhD student.

Next, I would like to thank all the members of the Pogliano Labs over the years who have supported me and enhanced the experience of grad school. Thank you Kit, Alan, Eammon, Roland, Kanika, Jelani, Gaya, and Joe Sugie. Thank you to the phage team; Amy for being a great rotation student turned colleague and officemate, and for all those long hours writing our first publication together with Joe; Emily for your shining smile, kind spirit, and the hard-working yet positive energy you bring to the lab; Eray for always being happy to help out and for your great scientific contributions; and Chase and Annie for joining in to carry the torch of the phage nucleus project. Thank you Avani for being such a cool and collected undergraduate researcher, and thank you Jina for your sweet notes and gifts, and for the unending smiles and love you put out into the world, besides your remarkable intellect and enthusiasm for science. Thank you Mike for your mentorship during my undergraduate work with BCP and thank you to Arthur for pioneering the

phage nucleus discovery and churning out data like a machine, also for your responsiveness and help as a collaborator. Thank you Katrina and Mac for initiating me into the phage nucleus team and showing me the ropes. Thank you Elizabeth M. for being a sincere friend and mentor through my time as an undergraduate and my return as a graduate. Thank you Krithika for your genuine connection and for being my fitness buddy. Thank you Hannah for everything you have done to keep the labs running smoothly, even after you graduated, and for your reliability as a beacon of happy energy, plus your dissertation materials which really helped this process to be more pleasant for me. And a big thank you to Ria for taking care of us and keeping us fed.

I would like to thank my cohort and my friends in the program who have helped to make this a special time in my life, particularly Maya, the founders of GradHAC and the Biotech Group, and my professional development partner, Brooke. I would also like to thank my Lindy Hop team, the Hang 10 Hoppers and our fearless leader Joel, my favorite dance partner and bestie Jeff, as well as everyone in the swing dancing, blues/fusion dancing, and beach volleyball communities who were a big part of feeding my soul throughout grad school.

I would like to express my great appreciation for the mentorship and support I received from my committee members, Justin Meyer, Elizabeth Villa, Kevin Corbett, and Arshad Desai. Thank you for your collaboration on great projects, your kindness when giving me feedback that helped me to improve as a scientist, and for continually making me feel like you were rooting for me. I am grateful that our committee meetings were always a positive and productive experience.

Finally, I would like to thank my family who has made this all possible. My grandma Adrienne and grandpa Terry, my grandma Joan and grandpa Jerry, thank you for loving me the way only grandparents can and being so good at it. My aunties Jenny, Julie, Ann, Amy, and Kim, my uncles Paul, Andy, Larry, Ken, John, and Rick, and my cousins Katie, Johnny, Connor,

Adrienne, Jake, Josh, Rick Jr., and Regan, thank you for being a fun, loving family with just the right amount of weirdness. Thank you to my boyfriend for brightening my world, I am grateful to share this milestone with you. Thank you Andria and Stephanie, my sisters from other mothers, you will always be deep in my heart. And last but most, my parents Lori and Matt. Your extraordinary support and encouragement of my education since before I can remember has made learning not only integral to my life but essential to my happiness. Your belief in me got me to grad school and your unconditional love got me through it. You inspire me to always do my best and never stop asking questions. Thank you. This PhD is as much for you as it is for me and I am proud of it and proud to have you as my parents and my foundation. I am eternally grateful for you.

Plus, a quick shoutout to my superstars ΦKZ, ΦPA3, and Goslar. Thanks for doing such mind-blowing things inside of bacteria that you will supply countless dissertations worth of awesome contributions to our understanding of biology. And thank you *P. aeruginosa* and *E. coli* for your sacrifices in this pursuit of epic bacterial cell biology.

Chapter 2, in entirety, is a reprint of the material in review as Birkholz, EA., Laughlin, TG., Suslov, S., Armbruster, E., Lee, J., Wittmann, J., Corbett, KD., Villa, E., & Pogliano, J., 2022, A cytoskeletal vortex drives phage nucleus rotation during jumbo phage replication in *E. coli* in review with Cell Reports. The dissertation author was the primary investigator and author of this material.

Chapter 3, in entirety, is a reprint of the material as it appears in Chaikeeratisak, V., Birkholz, EA., Prichard, A., Egan, ME., Mylvara, A., Nonejuie, N., Nguyen, KT., Sugie, J., Meyer, JR., & Pogliano, J., 2021, Viral speciation through subcellular genetic isolation and virogenesis incompatibility in Nature Communications. The dissertation author was a primary author and the secondary investigator of this material.

Chapter 4, in large part, is being formulated into a manuscript in preparation for the publication of the material. Erica A. Birkholz, Chase Morgan, Rebecca K. Lau, Amy Prichard, Sergey Suslov, Elizabeth Villa, Justin R. Meyer, Kevin D. Corbett, & Joe Pogliano, 2022. The dissertation author was the primary investigator and author of this material.

VITA

2016 Bachelor of Science, University of California San Diego

2022 Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Birkholz, E. A., Laughlin, T. G., Suslov, S., Armbruster, E., Lee, J., Wittmann, J., Corbett, K.D., Villa, E., & Pogliano, J. A cytoskeletal vortex drives phage nucleus rotation during jumbo phage replication in *E. coli*. *BioRxiv*. (2022) *In review with Cell Reports.*

Laughlin, T. G., Deep, A., Prichard, A. M., Seitz, C., Gu, Y., Enustun, E., Suslov, S., Khanna, K., **Birkholz, E. A.,** Armbruster, E., Amaro, R. E., Pogliano, J., Corbett, K. D., Villa, E. Architecture and self-assembly of the jumbo bacteriophage nuclear shell. *BioRxiv*. (2022). *In review with Nature.*

Chaikeeratisak, V., **Birkholz, E. A.**, & Pogliano, J. The Phage Nucleus and PhuZ Spindle: Defining Features of the Subcellular Organization and Speciation of Nucleus-Forming Jumbo Phages. *Frontiers in Microbiology*, *12*, 641317. (2021).

Nguyen, K. T., Sugie, J., Khanna, K., Egan, M. E., **Birkholz, E. A.**, Lee, J., Beierschmitt, C., Villa, E., & Pogliano, J. Selective transport of fluorescent proteins into the phage nucleus. *PLoS ONE*, *16*(6), e0251429. (2021).

Chaikeeratisak, V., **Birkholz, E. A.**, Prichard, A. M., Egan, M. E., Mylvara, A., Nonejuie, P., Nguyen, K. T., Sugie, J., Meyer, J. R., & Pogliano, J. Viral speciation through subcellular genetic isolation and virogenesis incompatibility. *Nature Communications*, *12*, 342. (2021).

Lau, R. K., Ye, Q., **Birkholz, E. A.**, Berg, K. R., Patel, L., Mathews, I. T., Watrous, J. D., Ego, K., Whiteley, A. T., Lowey, B., Mekalanos, J. J., Kranzusch, P. J., Jain, M., Pogliano, J., & Corbett, K. D. Structure and Mechanism of a Cyclic Trinucleotide-Activated Bacterial Endonuclease Mediating Bacteriophage Immunity. *Molecular Cell*, *77*(4), 723-733.e6. (2020).

Ye, Q., Lau, R. K., Mathews, I. T., **Birkholz, E. A.**, Watrous, J. D., Azimi, C. S., Pogliano, J., Jain, M., & Corbett, K. D. HORMA Domain Proteins and a Trip13-like ATPase Regulate Bacterial cGAS-like Enzymes to Mediate Bacteriophage Immunity. *Molecular Cell*, *77*(4), 709-722.e7. (2020).

ABSTRACT OF THE DISSERTATION

Nucleus-Forming Jumbo Phage:

Disrupting the Landscape of Phage Replication and Viral Speciation

by

Erica A. Birkholz

Doctor of Philosophy in Biology

University of California San Diego, 2022

Professor Joseph Pogliano, Chair

This dissertation investigates the diversity and competition of nucleus-forming jumbo phage. The first chapter provides a brief history of the discovery of bacteriophages and their early use to combat bacterial infections, the detection of the first jumbo phages, the first characterizations of nucleus-forming jumbo phage, current understanding of viral speciation, a background of phage-encoded mobile introns, the theory of Viral Eukaryogenesis, and the recent resurgence of phage therapy.

Chapter 2 describes the conservation and divergence of the nucleus-forming jumbo phage replication cycle in *Escherichia coli* phage Goslar. Fluorescent microscopy supported with cryo-electron tomography of focused ion beam-milled samples (cryo-FIB-ET) revealed the organization

and kinetics of Goslar infection. Genetics were used to deduce the function of the cytoskeleton found to assemble into a vortex-like array.

Chapter 3 analyzes the intracellular interactions between coinfecting phages ΦKZ and ΦPA3 in *Pseudomonas aeruginosa*. Two viral speciation factors were newly defined as Subcellular Genetic Isolation and Virogenesis Incompatibility which can be universally applied to viruses. Fluorescent microscopy illuminated the intracellular competition and long-term fitness against the host was assayed to demonstrate the effects of incompatibility.

Chapter 4 builds on Chapter 3 to investigate the mechanism of action of one of the virogenesis incompatibility factors that is produced by ΦPA3 and interferes with ΦKZ infection. Bioinformatic analyses combined with genetics and *in vitro* evidence uncovered the nature and mechanism of the ΦPA3 factor. The consequences for ΦKZ were clarified by cryo-FIB-ET.

Chapter 5 is a conclusion of the findings presented in this dissertation and a discussion of the way that nucleus-forming jumbo phage can address the antibiotic resistance crisis as well as the big question of how our earliest eukaryotic ancestor came into existence.

CHAPTER 1:

Bacteriophage discovery and application, speciation, mobile introns, and the recent characterization of nucleus-forming jumbo phage

1.1 Bacteriophage discovery and early use

The discovery of bacteriophages, or phages, occurred independently in 1915 by Frederick Twort and in 1917 by Felix d'Herelle. In a publication of The Lancet on December 4th, 1915, Twort recounts his discovery of a non-pathogenic "filter-passing virus" isolated from cultures of vaccinia-causing micrococcus infecting calves [1]. On September 3rd, 1917, at the French Academy of Sciences, Felix d'Herelle presented his investigations of "an invisible microbe antagonistic to dysenteric bacilli" isolated from the stool of patients recovering from dysentery [2]. In both discoveries, the microbiologists first observed the clearing of bacterial cells that were growing in culture. The cleared cultures could be diluted to very low concentrations and still clear a dense culture of bacteria. When bacteria were spread on a plate and the cleared culture was touched to one spot, a clearing in the plated bacteria would spread radially from that spot, forming a "plaque". Twort noted the preference of the virus for actively growing bacteria rather than old or dead bacteria. D'Herelle was even able to switch the host specificity of his virus through multiple passages, demonstrating that these microbes had the ability to adapt to new hosts, and fairly quickly. All of the observations and experiments by Twort and d'Herelle pointed to the discovery of self-replicating obligate parasites of bacteria, which d'Herelle termed bacteriophages.

Felix d'Herelle moved quickly to apply his bacteriophages to the treatment of deadly infections in humans by treating childhood dysentery at Hopital des Enfants-malades in 1919 [3]. At the Pasteur Institute in Paris 1918, d'Herelle met a Georgian microbiologist named George Eliava who had been tasked with establishing Georgia's first institute of microbiology to tackle their own battles with bacterial infections. The two scientists became close collaborators and in 1921, Eliava returned to Georgia with the knowledge of d'Herelle's treatment of dysentery in soldiers. In 1930, d'Herelle brought to Georgia two of his phage cocktails called *pyophage* and

intestiphage, originally marketed by the company that later became L'Oreal [4]. These phage therapeutics are still used today to treat purulent and enteric infections, respectively. Eliava and d'Herelle visited each other repeatedly and developed the idea of making the new Eliava Institute the world center of phage research and therapy, gaining the support of Joseph Stalin [3]. Due to the complicated political climate of the Soviet Union, in 1937, George Eliava was arrested and tragically executed without trial. D'Herelle never returned to Georgia but fortunately, he and Eliava laid a strong groundwork that allowed the Eliava Institute to survive the fall of the Soviet Union and persist to this day, remaining the focal point of experience in phage therapy, just as they had dreamed. Now the Institute relies on grants from international organizations including the US State Department which fund high quality clinical trials and basic research [3]. The newly structured Institute was able to spin-off the commercial venture Eliava BioPreparations which currently offers six phage cocktails that are included in the standard of care in Georgia.

The other hub of phage research and modern-day supplier of phage therapy is the Hirszfeld Institute of Immunology and Experimental Therapy located in Wroclaw, Poland [5]. This Institute has been publishing detailed scientific accounts since the early 1980s and in 2005 they established their own phage therapy clinic with rigorous studies being conducted in accordance with EU guidelines. Hirszfeld is a rich source of accessible literature addressing clinical phage application, phage purification, the immunomodulation and translocation of phage in the body, and even the economics of phage therapy [3, 5]. They have employed the use of a phage bank with a personalized medicine approach where each patient's bacterial isolates are tested against the phage bank and a cocktail is developed for that individual. According to Polish law, phage therapy is still considered experimental and therefore requires approval by an institutional review board. It remains restricted to life-threatening infections for which antibiotics have failed. Hirszfeld

physicians point out that the only infections they are allowed to treat with phage are the most advanced and difficult to penetrate so it is reasonable to think that they would have even greater success rates if phage therapy was the first line of defense or a prophylactic treatment [3]. These regulatory restrictions and the strategy of patient-specific phage bank querying have preserved phage therapy in Poland as a humanitarian effort rather than a commercial exploit.

1.2 Jumbo Phage

For decades after the discovery of bacteriophages, the only ones studied were those with a genome size similar to the best studied phage, λ, which infects *Escherichia coli* and has a genome of 48.5 kb. Those phages are small enough to pass through a 0.2 μm filter and to resist pelleting by centrifugation, the methods commonly used to remove bacteria from samples [6]. Larger phages were likely excluded from early studies due to their large virions being caught in the filter and the potential to pellet with bacteria. They are also elusive since they tend to make plaques that are barely visible in standard concentrations of agarose overlay, due to the limitation of diffusion for a large virion [7]. Nowadays, metagenome sequencing of crude samples prevents the exclusion of phages based on size or growth requirements, leading to a new understanding of the full spectrum of bacteriophage diversity. By 2007, only 6 phages with genomes over 200 kb had been sequenced [7], but in 2020, Al-Shayeb et. al reconstructed 348 phage genomes over 200 kb from a wide variety of human, animal, and environmental samples from around the world, and they were only looking for those encoding a CRISPR-Cas system [8]. They even found a phage genome predicted to be 735 kb, which would make it the largest on record. The term proposed for phage genomes larger than the somewhat arbitrary cutoff of 200 kb is ‘jumbo phages’ [9]. In 2019, Devoto et al. coined the term ‘mega phages’ for phages with a genome over 500 kb in size, upon their curation of 15 genomes over 540 kb, belonging to the ‘Lak phages’ infecting *Prevotella* in

human gut microbiomes [10]. While classifications based solely on genome size are not representative of phylogeny, they can be useful for discussion nonetheless. 518 phage genomes between 200 kb and 553 kb were pulled out of metagenomes from human gut samples in 2021 [11], so evidently, large phages are a common component of the human microbiome and understanding them is increasingly important.

Pseudomonas aeruginosa phage ΦKZ (280 kb) appears to be the first jumbo phage genome fully sequenced, published in 2002 [12]. The genus *Phikzvirus* [13] now contains 15 phages with genomes published to NCBI. ΦKZ was isolated in 1978 in Kazakhstan ("KZ"), from a patient with a chronic pulmonary infection [14] and has since been used in the commercial phage preparation PyoPhage by Eliava BioPreparations [15]. *Bacillus subtilis* phage PBS1 (252 kb) was isolated earlier, in 1961 [16], but not fully sequenced until 2017. ΦKZ is a major component of this dissertation so its background will now be summarized, with the earliest results accessible only by recent translation and summation from the original Russian studies. The genome of ΦKZ was found to have a low GC content which contrasts the high GC content of the host and this is unusual for phages [17]. It encodes a DNA polymerase that is highly divergent from any others known [18]. The plaques of ΦKZ were noted to display strong opalescence which appears as cloudiness with a blue hue [17]. This had been previously described for the T-even bacteriophages of *E. coli*; T2 (164 kb), T4 (169 kb), and T6 (169 kb) [19]. With the help of mutant strains of each T-even coliphage that lost the ability to create opalescent plaques and instead only made clear plaques, this property of opalescence was described as 'lysis inhibition', a type of quorum sensing where a large concentration of phage would inhibit continued lysis of the bacterial host. Further studies into the bacterial colonies that could evade lysis by ΦKZ and related phages led to the conclusion that 'pseudolysogeny' was achieved when multiplicity of infection was high [20].

Electron micrographs from 1983 of ΦKZ virions both intact and damaged revealed a cylindrical structure inside the phage head deemed the 'inner body' [21]. Further imaging showed a fibrous material 4 nm wide wound around the inner body, thought to be DNA with a supercoiled tertiary structure [22]. It was also observed that with disrupted phage particles, the inner body could slip outside the capsid but it was never found far away, insinuating a physical connection of some sort. Those escaped inner bodies were measured to be slightly longer than those still inside a capsid, suggesting that they were compressed when inside the virion. After the ejection of DNA, when the tail sheath was contracted to reveal the tail tube and there was no longer DNA in the capsid, the inner body also seemed to disappear. ΦKZ was reportedly the first specimen to be imaged with an inner body [22]. Since then, atomic force microscopy [23], cryo-EM [24, 25], over-exposed cryo-EM bubblegrams [26], and mass spectrometry with biochemistry on the cleavage of inner body proteins [27] have confirmed the presence of a proteinaceous inner body around which DNA is wound. It was also shown that the DNA spool is positioned at a 22° [26] or 32° [25] angle compared to the long axis of the capsid which was described as unexpected but in my opinion, it corroborates a spooling path with the least resistance.

Micrographs of ultrathin sections of the intracellular infection of ΦKZ were also published in 1983 and 1984, revealing different phases of particle maturation where empty capsids containing only an inner body could be found next to capsids fully packed with DNA [17, 22]. These micrographs were interpreted to show a nucleoid of phage DNA which pushes the bacterial chromosome to the periphery [17], aligning with the hypothesis of pseudolysogeny achieved by sparing the host DNA from total degradation so that the cell may continue to divide while harboring the phage genome [20]. It was reported in 1982 that ΦKZ could transduce several bacterial markers [28]. This supports the idea that not all host DNA is degraded upon ΦKZ

infection. However, transduction was measured at a maximum average frequency of 7 x 10^{-6} [17]. In 2011, it was reported that *P. aeruginosa* phage ΦPA3, a *Phikzvirus,* is also a generalized transducing phage with a maximum frequency of 8 x 10^{-7} [29]. For comparison, *E. coli* phage λ and *Salmonella typhimurium* phage P22, which are capable of lytic and lysogenic cycles, were reported to achieve a maximum transduction efficiency of 2.2 x 10^{-2} [30] and 1.0 x 10^{-2} [31], respectively. This suggests that while host DNA may persist during a pseudolysogeny phase of *Phikzvirus* infection, these jumbo phages transduce with a frequency that is approximately 10,000-fold lower than the efficiently transducing phages. However, transduction and pseudolysogeny must be carefully considered when adapting jumbo phage for phage therapy.

1.3 Nucleus-Forming Jumbo Phage

Despite its ability to transfer pieces of bacterial plasmid DNA [28], ΦKZ protects its own genome by building a shell of protein around it, forming the phage nucleus [32]. This fascinating structure was first studied by our lab in a *Phikzvirus*, 201φ2-1 infecting *Pseudomonas chlororaphis* [33]. Leading up to that discovery, our lab was investigating the phage cytoskeleton formed by a tubulin-like protein PhuZ, encoded by 201φ2-1 with homologs found in many other phages with genomes of 186 – 316kb, though there were no prior reports of a cytoskeletal polymer [34]. When expressed *in vivo* at high levels in *P. chlororaphis*, filaments of GFP-PhuZ spontaneously polymerized in a majority of cells. *In vitro,* the observed lag phase in polymerization suggested a nucleation-extension mechanism with a critical concentration measured at 2.8 ± 0.1 μM [34]. Filaments did not form in the presence of GDP demonstrating a GTP hydrolysis-dependent mechanism of polymerization. Nonhydrolyzable GTP prevented filament disassembly so it allowed for imaging by negative stain EM, which showed twisted filaments [34], later revealed to be the first reported three-stranded tubulin filament [35]. The structure of the PhuZ monomer was

resolved to identify the nucleotide binding pocket and the C-terminal extension, reminiscent of other tubulin-like proteins [34]. Mutation of either of the conserved catalytic aspartic acid residues to alanine resulted in spontaneous polymerization at very low expression levels, supporting the hypothesis based on TubZ [36] and FtsZ [37] that GTP would be able to bind, allowing polymerization, but since it would not hydrolyze, the filament would not disassemble even below the critical concentration, as observed by Kraemer et al [34]. Further studies into PhuZ *in vitro* revealed that the filaments treadmill and undergo catastrophe that could depolymerize the entire filament, making PhuZ the first prokaryotic tubulin known to display dynamic instability [38]. This is thought to occur during stochastic loss of a stabilizing GTP cap when the PhuZ monomer at the (+) end hydrolyzes its nucleotide before another GTP-bound monomer joins the polymer.

Live cell time-lapse of *P. chlororaphis* expressing GFP-PhuZ and infected with 201φ2-1 revealed that the polymers organize into a bipolar spindle that appears anchored at each cell pole and dynamically extends toward the cell center without crossing into it [38]. Fluorescence recovery after photobleaching (FRAP) observed for wild-type PhuZ filaments showed steady movement of the bleached region towards the midcell while the mutant PhuZ filaments remained static, demonstrating the treadmilling function of PhuZ that is eliminated by the aspartic acid mutation [34]. 201φ2-1 infection caused the formation of a bulge at the center of the host cell, within which was located a dense ball of DAPI-stained phage DNA called the infection nucleoid [34, 38]. Fluorescent *in situ* hybridization (FISH) with probes for host DNA and phage DNA demonstrated that the host DNA is degraded by 40 minutes post infection (mpi) and the DNA in the infection nucleoid is exclusively phage DNA [38]. The phage nucleoid first appeared as a small punctum and usually near a cell pole, before growing in size while moving towards the midcell and oscillating there. GFP-PhuZ filaments were observed to form around 40 mpi, grow

and shrink, and remain assembled until cell lysis, when DAPI-stained capsids could be visualized exiting the cell [34]. DNase I treatment removed DAPI signal from the center of the nucleoid, leaving a halo of DAPI foci, most likely capsids packaged in an orderly fashion around the nucleoid. Mutation of PhuZ abolished the central positioning of the nucleoid and reduced burst size by 50% [34].

In order to find other proteins required for this novel phage replication pathway, mass spectrometry was used to look for early abundant proteins made during 201φ2-1 infection [33]. The most abundant early protein was tagged with GFP and observed to create a spherical structure around the phage DNA. This shell structure also separated DNA processing proteins (e.g. DNA helicase, DNA ligase, RecA) encoded by both the phage and the host, from the cytoplasmic host ribosomes and phage metabolic proteins like thymidylate kinase and synthase [33]. This protein shell therefore uncouples transcription from translation and is now called the phage nucleus. Simultaneous visualization of the tagged major phage nucleus protein and PhuZ confirmed that the bipolar spindle pushes the growing phage nucleus to the midcell. Amazingly, midway through infection, the phage nucleus rotates when the PhuZ filaments push against opposing sides [33]. This was the first observation of intracellular rotation of a large body inside of a prokaryote. Midcell positioning and rotation of the phage nucleus are abolished with expression of the catalytic mutant of PhuZ. Cryo-focused ion beam milling coupled with cryo-electron tomography (cryo-FIB-ET) confirmed the fluorescent observations including the presence of capsids containing varying amounts of DNA distributed around the periphery of the phage nucleus. Cryo-FIB-ET showed the phage nucleus to be an enclosed compartment surrounded by a regular chain of ~5 nm proteins, and revealed partially assembled capsids on the bacterial inner membrane as well as fully assembled virions in a group adjacent to the phage nucleus [33]. *P. aeruginosa* phages ΦKZ and

ΦPA3 were also found to replicate with a PhuZ spindle [39] and a phage nucleus composed of a homolog of the major phage nucleus protein, demonstrating the conservation of this pathway [40]. ΦKZ infection progression was also analyzed by a Russian team using TEM on fixed embedded cells [41]. They found that cells infected for only 5 minutes had round compartments, often multiple, containing nucleic acid and possible protein remnants. These compartments seemed to persist throughout the entire infection while the phage nucleus appeared after 15 mpi and grew in size, becoming more frequently found at the midcell. Two nuclei in a single cell were observed for 12% of the infections [41], consistent with prior observations [33]. Tomography revealed a filamentous network inside the phage nucleus at 30 minutes which appears to be higher order packing and organization of the genomes. Electron energy loss spectroscopy (EELS) for phosphorus signal at 15 mpi showed that the bacterial nucleoid vacated the pole of the cell where the round compartments were located, presumably the location of infection [41]. The bacterial nucleoid then receded to the cell periphery before becoming more diffuse but they did not see full degradation by 40 minutes, proposing that the remainder of the host DNA degrades very slowly as it is likely protected by stress-related DNA-binding proteins. The persistence of some of the bacterial genome throughout infection corroborates the findings on ΦKZ transduction [28] and pseudolysogeny [20].

The phage nucleus not only provides a DNA replication compartment free of ribosomes, but it also offers protection from host immune defenses [42-44]. ΦKZ and ΦPA3 could evade restriction modification (RM) enzymes (types I and II) as well as CRISPR-Cas systems including the types found in *Pseudomonas* (types I-C and I-F) and those not naturally occurring in the host (types II-A and V-A) [42]. *In vitro*, RM enzymes and Cas9 were able to cut ΦKZ DNA, but *in vivo* the components of those host immune systems were physically excluded from the phage

nucleus. Utilizing the naturally imported RecA homolog from the phage, attempts to localize the immunity enzymes to the inside of the phage nucleus resulted in successful reduction of phage titer by EcoRI but Cas9 accumulated at the periphery of the phage nucleus and did not produce a knockdown of phage replication [42]. While the phage nucleus provides physical protection from DNA-targeting host defenses, some immune systems are able to target RNA, which is not protected by the phage nucleus, such as the type VI-A CRISPR system. Cas13 targeting of ΦKZ reduced phage titer though it is a non-specific shredder which could limit phage infection by damaging the host. It was observed that the phage DNA can arrest as a punctum at the cell pole but this condition is yet to be investigated [42]. Jumbo phage PCH45 infecting *Serratia* was also found to evade DNA-targeting type I CRISPR systems while being susceptible to an RNA-targeting type III system [43]. This phage is highly divergent from the *Phikzvirus* genus but still contains homologs to the major phage nucleus protein and PhuZ. Tagging the putative phage nucleus protein confirmed the formation of a protein shell around the phage DNA, making up a phage nucleus that was centrally positioned 60% of the time. Bioinformatic analyses uncovered a significant enrichment of jumbo phage-targeting spacers in type III over type I systems. Type I spacers were even further depleted for jumbo phage encoding a homolog to the major phage nucleus protein, consistent with the physical protection of the genome conferred by the phage nucleus [43].

The newly assembled capsids of these nucleus-forming jumbo phage must reach the shell of the phage nucleus to be packaged with a genome, but they are assembled on the bacterial inner membrane [33]. Our lab investigated this capsid trafficking mechanism, first using two-second time-lapse fluorescence microscopy [45]. Fluorescently-labeled capsids appeared on the membrane for a few seconds before individually following a similar linear path to the phage

nucleus. Capsids traveled at a rate of ~50 nm/s, the same rate measured for PhuZ spindle treadmilling [45]. Co-expression of fluorescent capsid protein with fluorescent PhuZ protein displayed colocalization of the string of capsids with the PhuZ spindle, supporting the hypothesis that capsids are actively trafficked to the phage nucleus by the PhuZ spindle. The PhuZ catalytic mutant immobilized the capsids which still colocalized with the spindle. Cryo-FIB-ET of infected cells with mutant PhuZ showed a large number of capsids clustered around the PhuZ spindle distal to the phage nucleus. The majority were devoid of DNA at a time point at which they are filled under wild-type conditions [45].

The late-stage organization of maturing virions into structures adjacent to the phage nucleus was recently characterized by our lab and termed 'phage bouquets' due to the near-spherical arrangement of DNA-packed capsids with tails pointing inward like the stems of a bouquet [46]. ΦPA3 was the focus of this study since ΦKZ and 201φ2-1 form more disordered bouquets and at a significantly lower frequency, suggesting this mechanism is conserved but not essential. Tagged capsid protein colocalized with DAPI staining in the bouquets which appears as a ring in 2D. As the bouquets gained DAPI signal, the phage nucleus lost DAPI signal, showing the transfer of DNA. Tagged tail proteins localized inside the ring of DAPI demonstrating the orientation of virions in the bouquet. The general exclusion of cytoplasmic proteins from the bouquet suggested that the structure is very densely packed with phage structural proteins. Using cryo-FIB-ET, a large bouquet with tails attached to capsids was visualized as well as a small bouquet that did not yet contain tails but partially assembled tails could be found near the bouquet [46]. Infection in the presence of the PhuZ catalytic mutant showed that bouquets retain their position in relation to the phage nucleus so they do not depend on the PhuZ spindle for positioning. However, the PhuZ mutation causes the filament to become static so capsids are not efficiently

trafficked to the phage nucleus for packaging, reducing the size and number of bouquets observed [46].

A discovery was stumbled upon in our lab that is very useful as a tool for studying these phages. It was found that one type of GFP in particular, GFPmut1, is naturally imported into the phage nucleus of ΦKZ, but not ΦPA3 or 201φ2-1 [44]. This was intriguing as there was no explanation for why such closely related phages would import proteins differently. However, this observation had the potential to shed light on the mechanism of selective import of DNA processing proteins into the phage nucleus, since it did not seem to involve a signal sequence. To investigate this, 3 amino acids that differed between GFPmut1 and two closely related GFPs that are excluded from the ΦKZ nucleus, were mutated. V163, tucked inside GFPmut1, was mutated to alanine to match the cytoplasmic GFPs, and this had no effect on localization. M153, also on the surface of the protein, was mutated to threonine and resulted in a mixed phenotype with some ΦKZ nuclei excluding it and others importing it. F99, on the surface of GFPmut1, was mutated to serine accordingly and this mutation was able to switch GFPmut1 localization from 100% in the phage nucleus to 100% excluded from it [44]. A deeper structural analysis is required to detail the mechanisms of selective import. Cryo-FIB-ET showed no physical differences in the shell structures composing the phage nuclei of ΦKZ, ΦPA3, and 201φ2-1, therefore offering no further structural explanation for the phenomenon of GFPmut1 import by ΦKZ [44]. To demonstrate whether GFPmut1 could be used as a tool to synthetically import proteins into the ΦKZ nucleus, GFPmut1 was fused to mCherry and resulted in the import of both active fluorophores. GFPmut1 also facilitated the import of a DNA nuclease, which resulted in a 20% reduction of DAPI intensity, demonstrating that the nuclease was not cleaved from the GFPmut1 or inactivated as it was imported [44].

The characterization of the nucleus-forming jumbo phage replication cycle has demonstrated the strict uncoupling of transcription from translation as well as the genome organization by a tubulin-based spindle that are reminiscent of the eukaryotic nucleus and spindle [32-34, 38, 40]. These investigations also provided the first observations of the rotation of a large structure inside a bacterium, and virion maturation in phage bouquets [32, 45, 46]. The prevalence of homologs to the major phage nucleus protein and PhuZ in phage infecting diverse hosts including *Pseudomonas, E. coli*, *Serratia, Salmonella, Ralstonia, Cronobacter, Erwinia,* and *Vibrio* suggests the relevance of nucleus-forming jumbo phage studies to the broader understanding of the phage that inhabit this Earth.

1.4 Viral Speciation Factors

The discovery of the intricate organization of nucleus-forming jumbo phages raised the question of how they would interact during a coinfection of the same host cell. Coinfection happens commonly in nature, including in the human microbiome [47-49], and often allows for genetic exchange between viruses, which was thought to be unrestricted [50-52]. If that were the case, we would expect viral genomes to cluster into homogenous gene pools separated strictly by host tropism, and while that is a significant determinant, it is more complex [53, 54]. Virus phylogeny is officially classified largely by comparative sequence analyses [55-59]. The idea of biologically defining viral species has only recently been tackled using The Biological Species Concept [60, 61]. This concept says that organisms belong to different species when they belong to separate gene pools. Their mating, or recombination of genetic traits, is no longer possible or productive. A limitation to genetic recombination caused by a reproductive isolating mechanism can lead to the accumulation of genetic variations that are incompatible with each other, called

Dobzhansky-Muller genetic incompatibilities, which further isolate the gene pools by preventing certain recombinants from being viable [60, 62-64].

Since viral genomes are only accessible to recombination when replicating inside a cell, viruses that do not infect the same cell are strongly isolated from recombining their genes. With respect to the Biological Species Concept, viral speciation has been addressed for viral populations that are physically separated by either geography or by host tropism [65-67]. The ability of viruses to coinfect a cell relies more on host ecology than on host phylogeny [47, 68, 69], meaning that viruses tend to figure out ways to infect diverse host cells occupying the same microbiome, pointing to the importance of geographic separation in the speciation of viruses. Though even when two viruses share the same environment and are able to infect the same host, they may not achieve coinfection due to superinfection exclusion, which prevents infection after an initial infection has already occurred [70, 71], demonstrating another strong mechanism of reproductive isolation.

Since coinfection is the major opportunity for viral "mating" [72, 73] and experimental and ecological observations point to the abundance of coinfections in nature, including in the human microbiome [49, 68, 74-76], investigating the intracellular interactions of viruses is vital to a more thorough understanding of viral speciation and evolution. Once a coinfection is successfully established, the viruses may recombine to produce hybrids with either increased or decreased virulence as demonstrated for influenza [77-81]. Beneficial recombination provides a selection for continued coinfection and genetic mixing while unproductive recombination reinforces a separation of those two gene pools. The intracellular separation of viral factories has been implicated in reproductive isolation between vaccinia viruses [82-85] and between HSV-1 viruses [86] but no universally applicable speciation factor was proposed.

The intracellular complexity of nucleus-forming jumbo phage presents an opportunity to observe new types of reproductive isolating mechanisms between phages that establish an infection within the same host cell. Generalized definitions of the reproductive isolating mechanisms discovered in these nucleus-forming phage will allow them to be applied to any virus or other intracellular parasite.

1.5 Mobile introns

Coinfections between divergent phages offers the potential for the spread of mobile genetic elements between speciating phage that maintain an overlapping host range. Phage commonly contain self-splicing introns [87, 88] encoding homing endonucleases that mobilize the intron to invade the unoccupied locus in a related genome, the intron(-) allele. Homing endonucleases target the DNA surrounding their open reading frame (ORF) or intron, causing a DNA break in a related intron(-) gene, triggering recombination which results in unilateral gene conversion with the loss of the homing site [88-90]. While these homing endonucleases can be freestanding, they are also often found within group I or group II introns as well as inteins [88, 91-93]. Those self-splicing elements allow the homing endonuclease to invade highly conserved coding sequences by ensuring the essential genes of the host remain functional [88, 89]. The introns invaded essential genes at conserved residues such as enzyme active sites, so that any imperfect attempt to delete the intron would lead to a nonfunctional essential protein [90, 91, 94]. It has been proposed that next the homing endonucleases, which were also targeting highly conserved sequences to optimize spread to related populations, invaded the introns that protected the host from the disruption of essential genes [90, 91, 95-99]. This created a composite mobile genetic element consisting of an intron and a protein that promotes the intron's mobility.

In phage, the most common introns are group I with a homing endonuclease and they often interrupt highly conserved regions of genes involved in DNA metabolism, likely because the minimal genomes of phage are most conserved in those essential genes [90]. In every domain of life, group I introns are found interrupting mRNA, rRNA, and tRNA genes [89, 94, 100]. A stop codon often occupies the intron near the 5' splice site, to reduce RNA folding interference by ribosomes, which can also serve as RNA folding chaperones [89, 90]. The ribozyme formed by group I intron RNA is highly conserved in structure while the nucleotide sequence can be highly divergent [94]. The ribozyme core is 250-500 nucleotides long and the architecture can tolerate large insertions in the terminal loops, ripe for invasion by homing endonuclease ORFs [89, 94, 99]. Later mutations in stop and start codons has resulted in the 'creep' of nuclease ORFs into the RNA core structure [89]. A homing endonuclease is sufficient to mobilize the intron for efficient unilateral gene conversion so it is curious that certain members of a family of coinfecting viruses have avoided intron acquisition [90, 99].

Homing endonucleases in phage instigate intron mobility by employing the recombination machinery of phage replication [88-90, 101]. They can do so by creating a double strand break [101-104] or by nicking a single strand of the target DNA [99, 105-108], with nicking producing a less toxic effect [109], suggesting an advantage in nicking to protect the host. When these introns invade a related genome, the dependence on homologous recombination leads to less efficient repair between more divergent genomes [101, 110]. As a reference, Redβ_{λ} loses 100-fold efficiency when the recombining sequences are 22% diverged and recombination does not occur at 52% divergence [111].

These nucleases housed in the introns of phage are often elongated proteins that enable recognition of a DNA target region over 30 bp long, despite the small number of residues making

up the protein [88, 90, 106, 112, 113]. Their loosely connected DNA-binding (NUMOD [114]) and catalytic domains allow non-uniform contacts in the major and minor groove of the target region, resulting in mismatch tolerant endonucleases that retain high specificity at only the most conserved residues [88-90, 112]. Homing endonucleases with an H-N-H/N domain utilize a leading histidine to activate a water molecule as a nucleophile, and an asparagine or histidine to coordinate the metal (magnesium or zinc), stabilizing the transition state [88, 106, 115]. The phage-encoded nicking homing endonucleases referenced above all belong to the HNH family. The HNH motif is found in the homing endonucleases of group I and II introns, transposases, restriction endonucleases, DNA packaging enzymes [116], and maturases [106]. Maturases are often degenerate homing endonucleases that bind RNA or protein to assist with the splicing of an intron that either encodes the maturase or is elsewhere in the genome [117].

Group I introns containing a homing endonuclease have been studied in the context of coinfecting phage. Nicking HNH endonuclease I-PfoP3I from the group I intron in the DNA polymerase (DNAP) of cyanophage Pf-WMP3 is an example of a typical homing endonuclease. It prefers the intron(-) allele of the homologous phage over related phage Pf-WMP4 and cuts both genomes 4 nucleotides upstream of the intron insertion site (IIS) with the recognition region spanning only 14 bp [105]. *Bacillus subtilis* phages SPO1 and SP82 also contain DNAPs interrupted by group I introns with nicking HNH homing endonucleases, I-HmuI and I-HmuII respectively [99, 106, 118]. However, unlike many homing endonucleases, these two are able to cut both the intron(+) and intron(-) alleles of the DNAP and they each prefer the allele of the heterologous phage [99]. Interestingly, I-HmuII cuts SPO1 DNA and excludes the I-HmuI intron while causing the co-conversion of the flanking genes. I-HmuII cleaves DNA 52 nucleotides from the IIS while I-HmuI cleaves 4 nucleotides away, more like a typical homing endonuclease, so this

feature of I-HmuII seems to provide a selective advantage to the SP82 alleles of the flanking genes which replace the SPO1 alleles [99] though there was no evidence that this advantage is bestowed upon the entire host phage. Bastille is another phage infecting *B. subtilis* and it too contains an HNH homing endonuclease, I-BasI, in a group I intron interrupting its DNAP [107, 112]. I-BasI is able to cut the intron(-) alleles of both Bastille and SPO1 3 nucleotides upstream of the IIS, but I-HmuI cannot cut Bastille so its ability to cut the intron(+) DNA of SPO1 and SP82 may have come at the cost of being able to cut other divergent intron(-) alleles [112].

These mobile intron systems are 'selfish' mobile elements but sometimes their nuclease confers a benefit to the host genome, such as intron splicing and transcriptional repression of an intron, as well as stress-induced ectopic insertions with a potential for productive genetic diversity [89, 119, 120]. As far as any selective advantage for an intron(+) virus over an intron(-) virus, the only benefit reported is for the individual genes surrounding the I-HmuII intron that facilitates co-conversion of alleles [99]. The influence of eukaryotic [121] and archaeal [122, 123] selfish genetic elements on speciation has been described. However, a homing endonuclease-mediated competitive advantage for an entire virus in a single coinfection has yet to be reported [89, 124]. Usually, the cleavage of related phage DNA is repaired by homologous recombination, eventually resulting in intron integration since other repairs reconstitute the nuclease recognition site. But the efficiency of recombination decreases as the sequences diverge, even for mismatch tolerant phage recombinases [111, 125] which may lead to the destruction of phage genomes that cannot be repaired to evade the nuclease of the related phage, but this has yet to be demonstrated experimentally. The cross-activity of homing endonucleases inhabiting diverging viral genomes presents potential consequences for the speciation of coinfecting viruses.

1.6 Viral Eukaryogenesis

Introns must be spliced from transcripts before they are translated, which is difficult when transcription and translation are coupled as in prokaryotes which lack a nucleus to exclude ribosomes from the DNA. It has therefore been theorized that intron invasion could have been the driving factor for the emergence of the eukaryotic nucleus [126, 127]. This idea converges the major evolutionary events of endosymbiosis of the ancestral mitochondrion and the advent of the nucleus and subsequent intracellular organization during the formation of the first eukaryotic cell, or eukaryogenesis [126-129].

All eukaryotic cells have a nucleus with a double membrane scaffolded by a protein lamina and the chromosomes are organized by a tubulin-based spindle, suggesting that the acquisition of those structures was critical to the early evolution of the first proto-eukaryotic cells. Since the eukaryotic spindle and nucleus are very complex structures, even in the simplest eukaryotic organisms, it has always been a mystery how they might have evolved. The Viral Eukaryogenesis theory [130, 131] proposes that the first eukaryotic cell came into existence when a viral compartment took on the role of the nucleus 2 billion years ago. However, no intermediates between eukaryotes with a nucleus and prokaryotes without one were known, prior to our discovery of nucleus-forming jumbo phage [32, 33, 40]. These newly characterized phages provide concrete evidence that these two structures, which are common to all eukaryotes, also evolved in viruses. Therefore, while the phage nucleus and spindle likely evolved due to advantages for viral replication, it is possible that other evolutionary pressures subsequently selected for their acquisition by ancient archaeal cells, giving rise to the first proto-eukaryotic cells containing a nucleus-like structure and a tubulin-based spindle.

That evolutionary pressure may have been the invasion of introns into the genome of the evolving proto-eukaryote. Mounting evidence builds an evolutionary pathway whereby the ancestral archaea, related to the extant Asgard archaea [132-135], became dependent on the metabolism of engulfed α-proteobacteria that evolved into mitochondria [136-151]. This symbiotic α-proteobacterium could have contained the ancestors of bacterial group II mobile introns [132, 144, 147, 152-158] which would have been released into the cellular environment when the symbiont lysed [127]. This could expose the host archaeal genome to the introns which would become the eukaryotic introns we see today [132, 159-171]. Intron invasion resulting in non-functional genes being translated [132], would provide evolutionary pressure to uncouple transcription and translation, by the compartmentalization of the genome into a nucleus [127]. This feat was already accomplished by viruses [131, 172] that could have been infecting the archaea that were becoming dependent on α-proteobacteria for metabolism [130, 133]. With the archaeal genome and an encapsidated virus genome competing for relevancy, the barrage of introns from the bacterial symbiont could have provided the selective advantage to the physically protected genetic system with uncoupled transcription and translation [127, 173], leading to the viral genome enslaving the archaeon and its symbiotic bacteria. The separation of transcription from translation allows time for the splicing of transcripts prior to translation but it also requires the export of mRNA to the cytosol for translation. These eukaryotic hallmarks of a DNA compartment that uncouples transcription from translation and caps mRNA for export are shared by eukaryotes and viruses to the exclusion of bacteria and archaea [130, 131]. Further support for the viral origin of the eukaryotic nucleus [174, 175] includes the analysis of DNA polymerases [176] and the characterization of complex nucleocytoplasmic large DNA viruses (NCLDVs) and giant bacterial viruses [177-181]. The separation of transcription and translation likely opened the

door for evolution through exon shuffling and alternative splicing [152, 182], leading to an explosion of cellular complexity [183-185] that eventually brought humans into existence.

1.7 Resurgence of Phage Therapy

As we explore the intricate replication cycle, diversity, and evolution of these recently discovered nucleus-forming phage, our broader understanding of the phage world will contribute to the global adoption of phage therapy in the battle against antibiotic resistance and microbiome dysbiosis. Since the early 1900s, antibiotics have been saving lives [186-188] and overshadowing phage therapy. Aside from the lack of standardized, well-controlled phage therapy clinical trials, and the patentability and ease of antibiotic use and production, there was an intense decade-long controversy over the nature of phage between Felix d'Herelle and Nobelist Jules Bordet [189]. Bordet won his Nobel Prize for the discovery of antibody-mediated bacterial lysis and he argued that the lysis observed by Frederick Twort [1] and d'Herelle [2] was not a bacteria-targeting virus, but instead a self-perpetuating lytic enzyme. Bordet's theory was accepted by the scientific community and their textbooks, but in 1939, the first electron micrograph of a phage was obtained in Germany, though the 1940 publication of this finding was engulfed by the second World War. Although we now know that Bordet was wrong, his influence overshadowed d'Herelle and fueled the early stagnation of phage therapy [189]. The politics surrounding phage therapy, particularly its use by the enemies of the United States during the World Wars, also did no favors for the progress of phage therapy. This is exemplified by a passage published by Gunther Stent in 1963, which discredits the history and validity of phage therapy [189, 190].

Our love of antibiotics is believed to have improved quality of life in the modernizing world, but unfortunately, the bacteria are catching up quickly in this arms race. Since many antibiotics are natural compounds produced by bacteria that are in competition with one another,

it is no surprise that genes for resistance are quickly discovered. Through horizontal gene transfer (HGT), these natural resistance genes can quickly dominate a population under the pressure of antibiotics in the human body or in the ecosystem [191]. The first resistance enzyme, penicillinase, was discovered even before penicillin was used widely as a therapeutic [192, 193]. Resistance of *M. tuberculosis* to streptomycin was seen during the course of treatment in an individual patient [193]. The low costs and wide usage of antibiotics quickly distributed them into the biosphere where they have helped to drive the evolution of better resistance genes and the transfer of those genes into many species of pathogens [193]. The CDC's 2019 Antibiotic Resistance Threats Report lists 4 bacteria and 1 fungus as the most urgent threats; Carbapenem-resistant *Acinetobacter* and *Enterobacterales*, Drug-resistant *Neisseria gonorrhoeae*, *Clostridioides difficile*, and *Candida auris*. That report now estimates that more than 2.8 million antibiotic-resistant infections occur in the U.S. each year and over 35,000 people die every year from those resistant infections, not including the 12,800 that died from standard *C. difficile* in 2017 [194].

This epidemic of antibiotic resistance is a call to action for phage therapy. While there are a few downfalls of phage therapeutics including the requirement for fully lytic phage that can withstand the immune system long enough to penetrate infections, and the issue of narrow host range, which can be surmounted by cocktails of multiple phages, there are a good number of advantages predicted for mainstream implementation of phage therapy [195]. Lytic phages are destructive to the bacterial host and do not allow for recovery, unlike bacteriostatics such as tetracycline. Phages also self-replicate where there are higher concentrations of host and naturally clear from the body when the host is eliminated [196]. Since phages are the natural predator to bacteria, they are able to adapt to bacterial resistance and play an active role in this arms race. A group at the Eliava Institute recently published in 2018 that they could increase the therapeutic

potential of a traditional phage cocktail by first passaging the phage with the UTI-causing pathogen, allowing the phage a head start in that race [197]. In terms of toxicity, phages are a natural component of our microbiome and are not inherently toxic, as reflected by the FDA's GRAS (generally recognized as safe) classification of 10 phage cocktails as of 2018 [198] and their current use in food products [199], though high-quality lysates are necessary to avoid contamination by bacterial endotoxins [195]. While the mounting antibiotic resistance crisis requires continual examination of novel antibiotic compounds, bringing phage therapy to the West is becoming increasingly vital to the successful treatment of deadly resistant infections [200-204].

Possibly one of the most exciting benefits of phage therapy is the sparing of commensal bacteria, which can play a critical role in eliminating and resisting a recurrent infection [205, 206]. This final point is a strong motivation for the pursuit of phage therapeutics for standard treatments of minor bacterial infections, in light of the body of work revealing the harmful effects of antibiotics on overall human health and the consequences of indiscriminately killing large populations of bacteria in our guts [207-209]. Antibiotics clearly alter the composition of the infant microbiome [210-214] and the adult microbiome [207, 215] with long-term effects regularly observed [216-218]. Antibiotic exposure has also been shown to affect the gut microbiome and the central nervous system in mice [217, 219-221]. These discoveries combined with the observations of an altered microbiome in humans presenting with obesity and type 2 diabetes, cancer, immune dysregulation, and altered cognitive function to name a few [222-231], imply the importance of an unperturbed microbiome.

These problems of antibiotic resistance and antibiotic-mediated microbiome dysbiosis [232, 233] may be improved by the mitigation of antibiotic use in agriculture [198, 234], and the increased implementation of phage therapeutics in food as well as in humans. Phage therapy is

gaining traction in the American academic system [203] as well as the commercial biotech sphere so the study of these bacteriophage contributes to the deeper understanding of the phages that are, and will increasingly be, saving and improving human lives.

1.8 References

1. Twort, F.W., *An Investigation on the Nature of Ultra-Microscopic Viruses.* The Lancet, 1915. **186**(4814): p. 1241-1243.

2. D'Herelle, F., *On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917.* Res Microbiol, 2007. **158**(7): p. 553-4.

3. Kutter, E., De Vos, D., Gvasalia, G., Alavidze, Z., Gogokhia, L., Kuhl, S., and Abedon, S.T., *Phage therapy in clinical practice: treatment of human infections.* Curr Pharm Biotechnol, 2010. **11**(1): p. 69-86.

4. Kutateladze, M. and Adamia, R., *Phage therapy experience at the Eliava Institute.* Med Mal Infect, 2008. **38**(8): p. 426-30.

5. Zaczek, M., Weber-Dabrowska, B., Miedzybrodzki, R., Lusiak-Szelachowska, M., and Gorski, A., *Phage Therapy in Poland - a Centennial Journey to the First Ethically Approved Treatment Facility in Europe.* Front Microbiol, 2020. **11**: p. 1056.

6. Yuan, Y. and Gao, M., *Jumbo Bacteriophages: An Overview.* Front Microbiol, 2017. **8**: p. 403.

7. Serwer, P., Hayes, S.J., Thomas, J.A., and Hardies, S.C., *Propagating the missing bacteriophages: a large bacteriophage in a new class.* Virol J, 2007. **4**: p. 21.

8. Al-Shayeb, B., Sachdeva, R., Chen, L.X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y., He, C., Meheust, R., Brooks, B., Thomas, A., Lavy, A., Matheus-Carnevali, P., Sun, C., Goltsman, D.S.A., Borton, M.A., Sharrar, A., Jaffe, A.L., Nelson, T.C., Kantor, R., Keren, R., Lane, K.R., Farag, I.F., Lei, S., Finstad, K., Amundson, R., Anantharaman, K., Zhou, J., Probst, A.J., Power, M.E., Tringe, S.G., Li, W.J., Wrighton, K., Harrison, S., Morowitz, M., Relman, D.A., Doudna, J.A., Lehours, A.C., Warren, L., Cate, J.H.D., Santini, J.M., and Banfield, J.F., *Clades of huge phages from across Earth's ecosystems.* Nature, 2020. **578**(7795): p. 425-431.

9. Hendrix, R.W., *Jumbo bacteriophages.* Curr Top Microbiol Immunol, 2009. **328**: p. 229-40.

10. Devoto, A.E., Santini, J.M., Olm, M.R., Anantharaman, K., Munk, P., Tung, J., Archie, E.A., Turnbaugh, P.J., Seed, K.D., Blekhman, R., Aarestrup, F.M., Thomas, B.C., and Banfield, J.F., *Megaphages infect Prevotella and variants are widespread in gut microbiomes.* Nat Microbiol, 2019. **4**(4): p. 693-700.

11. Nayfach, S., Paez-Espino, D., Call, L., Low, S.J., Sberro, H., Ivanova, N.N., Proal, A.D., Fischbach, M.A., Bhatt, A.S., Hugenholtz, P., and Kyrpides, N.C., *Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome.* Nat Microbiol, 2021. **6**(7): p. 960-970.

12. Mesyanzhinov, V.V., Robben, J., Grymonprez, B., Kostyuchenko, V.A., Bourkaltseva, M.V., Sykilinda, N.N., Krylov, V.N., and Volckaert, G., *The genome of bacteriophage phiKZ of Pseudomonas aeruginosa.* J Mol Biol, 2002. **317**(1): p. 1-19.

13. Krylov, V.N., Dela Cruz, D.M., Hertveldt, K., and Ackermann, H.W., *"phiKZ-like viruses", a proposed new genus of myovirus bacteriophages.* Arch Virol, 2007. **152**(10): p. 1955-9.

14. Krylov, V.N. and Zhazykov, I., *[Pseudomonas bacteriophage phiKZ--possible model for studying the genetic control of morphogenesis].* Genetika, 1978. **14**(4): p. 678-85.

15. Villarroel, J., Larsen, M., Kilstrup, M., and Nielsen, M., *Metagenomic Analysis of Therapeutic PYO Phage Cocktails from 1997 to 2014.* Viruses, 2017. **9**(11): p. 328.

16. Takahashi, I., *Genetic transduction in Bacillus subtilis.* Biochem Biophys Res Commun, 1961. **5**: p. 171-5.

17. Krylov, V., Bourkaltseva, M., Pleteneva, E., Shaburova, O., Krylov, S., Karaulov, A., Zhavoronok, S., Svitich, O., and Zverev, V., *Phage phiKZ-The First of Giants.* Viruses, 2021. **13**(2).

18. Kazlauskas, D. and Venclovas, C., *Computational analysis of DNA replicases in double-stranded DNA viruses: relationship with the genome size.* Nucleic Acids Res, 2011. **39**(19): p. 8291-305.

19. Doermann, A.H., *Lysis and Lysis Inhibition with Escherichia coli Bacteriophage.* J Bacteriol, 1948. **55**(2): p. 257-76.

20. Pletnev, E.A., Krylov, S.V., Shaburova, O.V., Burkal'tseva, M.V., Miroshnikov, K.A., and Krylov, V.N., *[Pseudolysogeny of Pseudomonas aeruginosa bacteria infected with phiKZ-like bacteriophages].* Genetika, 2010. **46**(1): p. 26-32.

21. Smirnova, T.A., Minenkova, I.B., Khrenova, E.A., Plotnikova, T.G., and Krylov, V.N., *[Electron microscope study of the intracellular development of the Pseudomonas aeruginosa bacteriophage phi KZ].* Zh Mikrobiol Epidemiol Immunobiol, 1983(5): p. 25-8.

22. Krylov, V.N., Smirnova, T.A., Minenkova, I.B., Plotnikova, T.G., Zhazikov, I.Z., and Khrenova, E.A., *Pseudomonas bacteriophage phi KZ contains an inner body in its capsid.* Can J Microbiol, 1984. **30**(6): p. 758-62.

23. Matsko, N., Klinov, D., Manykin, A., Demin, V., and Klimenko, S., *Atomic force microscopy analysis of bacteriophages phiKZ and T4.* J Electron Microsc (Tokyo), 2001. **50**(5): p. 417-22.

24. Fokine, A., Kostyuchenko, V.A., Efimov, A.V., Kurochkina, L.P., Sykilinda, N.N., Robben, J., Volckaert, G., Hoenger, A., Chipman, P.R., Battisti, A.J., Rossmann, M.G.,

and Mesyanzhinov, V.V., *A three-dimensional cryo-electron microscopy structure of the bacteriophage phiKZ head.* J Mol Biol, 2005. **352**(1): p. 117-24.

25. Fokine, A., Battisti, A.J., Bowman, V.D., Efimov, A.V., Kurochkina, L.P., Chipman, P.R., Mesyanzhinov, V.V., and Rossmann, M.G., *Cryo-EM study of the Pseudomonas bacteriophage phiKZ.* Structure, 2007. **15**(9): p. 1099-104.

26. Wu, W., Thomas, J.A., Cheng, N., Black, L.W., and Steven, A.C., *Bubblegrams reveal the inner body of bacteriophage phiKZ.* Science, 2012. **335**(6065): p. 182.

27. Thomas, J.A., Weintraub, S.T., Wu, W., Winkler, D.C., Cheng, N., Steven, A.C., and Black, L.W., *Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant Pseudomonas aeruginosa phage phiKZ.* Mol Microbiol, 2012. **84**(2): p. 324-39.

28. Dzhusupova, A.B., Plotnikova, T.G., and Krylov, V.N., *[Detection of transduction by virulent bacteriophage phi KZ of Pseudomonas aeruginosa chromosomal markers in the presence of plasmid RMS148].* Genetika, 1982. **18**(11): p. 1799-802.

29. Monson, R., Foulds, I., Foweraker, J., Welch, M., and Salmond, G.P.C., *The Pseudomonas aeruginosa generalized transducing phage phiPA3 is a new member of the phiKZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients.* Microbiology (Reading), 2011. **157**(Pt 3): p. 859-867.

30. Sternberg, N., *The production of generalized transducing phage by bacteriophage lambda.* Gene, 1986. **50**(1-3): p. 69-85.

31. Kwoh, D.Y. and Kemper, J., *Bacteriophage P22-mediated specialized transduction in Salmonella typhimurium: high frequency of aberrant prophage excision.* J Virol, 1978. **27**(3): p. 519-34.

32. Chaikeeratisak, V., Birkholz, E.A., and Pogliano, J., *The Phage Nucleus and PhuZ Spindle: Defining Features of the Subcellular Organization and Speciation of Nucleus-Forming Jumbo Phages.* Front Microbiol, 2021. **12**: p. 641317.

33. Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., Villa, E., Agard, D.A., and Pogliano, J., *Assembly of a nucleus-like structure during viral replication in bacteria.* Science, 2017. **355**(6321): p. 194-197.

34. Kraemer, J.A., Erb, M.L., Waddling, C.A., Montabana, E.A., Zehr, E.A., Wang, H., Nguyen, K., Pham, D.S., Agard, D.A., and Pogliano, J., *A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell.* Cell, 2012. **149**(7): p. 1488-99.

35. Zehr, E.A., Kraemer, J.A., Erb, M.L., Coker, J.K., Montabana, E.A., Pogliano, J., and Agard, D.A., *The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA.* Structure, 2014. **22**(4): p. 539-48.

36. Larsen, R.A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P., and Pogliano, J., *Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis.* Genes Dev, 2007. **21**(11): p. 1340-52.

37. Mukherjee, A. and Lutkenhaus, J., *Dynamic assembly of FtsZ regulated by GTP hydrolysis.* EMBO J, 1998. **17**(2): p. 462-9.

38. Erb, M.L., Kraemer, J.A., Coker, J.K., Chaikeeratisak, V., Nonejuie, P., Agard, D.A., and Pogliano, J., *A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells.* Elife, 2014. **3**.

39. Aylett, C.H., Izore, T., Amos, L.A., and Lowe, J., *Structure of the tubulin/FtsZ-like protein TubZ from Pseudomonas bacteriophage PhiKZ.* J Mol Biol, 2013. **425**(12): p. 2164-73.

40. Chaikeeratisak, V., Nguyen, K., Egan, M.E., Erb, M.L., Vavilina, A., and Pogliano, J., *The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages.* Cell Reports, 2017. **20**(7): p. 1563-1571.

41. Danilova, Y.A., Belousova, V.V., Moiseenko, A.V., Vishnyakov, I.E., Yakunina, M.V., and Sokolova, O.S., *Maturation of Pseudo-Nucleus Compartment in P. aeruginosa, Infected with Giant phiKZ Phage.* Viruses, 2020. **12**(10).

42. Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J., *A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases.* Nature, 2020. **577**(7789): p. 244-248.

43. Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gumy, L.F., and Fineran, P.C., *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, 2020. **5**(1): p. 48-55.

44. Nguyen, K.T., Sugie, J., Khanna, K., Egan, M.E., Birkholz, E.A., Lee, J., Beierschmitt, C., Villa, E., and Pogliano, J., *Selective transport of fluorescent proteins into the phage nucleus.* PLoS One, 2021. **16**(6): p. e0251429.

45. Chaikeeratisak, V., Khanna, K., Nguyen, K.T., Sugie, J., Egan, M.E., Erb, M.L., Vavilina, A., Nonejuie, P., Nieweglowska, E., Pogliano, K., Agard, D.A., Villa, E., and Pogliano, J., *Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria.* Cell, 2019. **177**(7): p. 1771-1780.e12.

46. Chaikeeratisak, V., Khanna, K., Nguyen, K. T., Egan, M. E., Enustun, E., Armbruster, E., Pogliano, K., Villa, E., & Pogliano, J. , *Subcellular Organization of Viral Particles During Maturation of Nucleus-Forming Jumbo Phage*. 2021: BioRxiv.

47. Diaz-Munoz, S.L., *Viral coinfection is shaped by host ecology and virus-virus interactions across diverse microbial taxa and environments.* Virus Evol, 2017. **3**(1): p. vex011.

48. Domingo-Calap, P., Mora-Quilis, L., and Sanjuan, R., *Social Bacteriophages.* Microorganisms, 2020. **8**(4).

49. Luque, A. and Silveira, C.B., *Quantification of Lysogeny Caused by Phage Coinfections in Microbial Communities from Biophysical Principles.* mSystems, 2020. **5**(5).

50. Mavrich, T.N. and Hatfull, G.F., *Bacteriophage evolution differs by host, lifestyle and genome.* Nat Microbiol, 2017. **2**: p. 17112.

51. Casjens, S.R., *Diversity among the tailed-bacteriophages that infect the Enterobacteriaceae.* Res Microbiol, 2008. **159**(5): p. 340-8.

52. Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., and Hendrix, R.W., *Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages.* J Mol Biol, 2000. **299**(1): p. 27-51.

53. Bobay, L.M. and Ochman, H., *Biological species in the viral world.* Proc Natl Acad Sci U S A, 2018. **115**(23): p. 6040-6045.

54. Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B., *Viral tagging reveals discrete populations in Synechococcus viral genome sequence space.* Nature, 2014. **513**(7517): p. 242-5.

55. International Committee on Taxonomy of Viruses Executive, C., *The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks.* Nat Microbiol, 2020. **5**(5): p. 668-674.

56. Calisher, C.H., Briese, T., Brister, J.R., Charrel, R.N., Durrwald, R., Ebihara, H., Fulhorst, C.F., Gao, G.F., Groschup, M.H., Haddow, A.D., Hyndman, T.H., Junglen, S., Klempa, B., Klingstrom, J., Kropinski, A.M., Krupovic, M., LaBeaud, A.D., Maes, P., Nowotny, N., Nunes, M.R.T., Payne, S.L., Radoshitzky, S.R., Rubbenstroth, D., Sabanadzovic, S., Sasaya, T., Stenglein, M.D., Varsani, A., Wahl, V., Weaver, S.C., Zerbini, F.M., Vasilakis, N., and Kuhn, J.H., *Strengthening the Interaction of the Virology Community with the International Committee on Taxonomy of Viruses (ICTV) by Linking Virus Names and Their Abbreviations to Virus Species.* Syst Biol, 2019. **68**(5): p. 828-839.

57. Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., and Smith, D.B., *Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV).* Nucleic Acids Res, 2018. **46**(D1): p. D708-D717.

58. Adriaenssens, E. and Brister, J.R., *How to Name and Classify Your Phage: An Informal Guide.* Viruses, 2017. **9**(4).

59. Krupovič, M. and Bamford, D.H., *Order to the Viral Universe.* Journal of Virology, 2010. **84**(24): p. 12476-12479.

60. Dobzhansky, T., *Genetics and the Origin of Species*. 1937: Columbia University Press.

61. Mayr, E., *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. 1942: New York: Columbia Univ. Press.

62. Muller, H.J., *Reversibility in evolution considered from the standpoint of genetics.* Biological Reviews of the Cambridge Philosophical Society, 1939. **14**: p. 261-280.

63. Muller, H.J. and Pontecorvo, G., *Recombinants between Drosophila Species the F1 Hybrids of which are Sterile.* Nature, 1940. **146**(3693): p. 199-200.

64. Muller, H.J., *Isolating Mechanisms, Evolution, and Temperature.* Biology Symposium, 1942. **6**: p. 71–125.

65. Duffy, S., Burch, C.L., and Turner, P.E., *Evolution of host specificity drives reproductive isolation among RNA viruses.* Evolution, 2007. **61**(11): p. 2614-22.

66. Meyer, J.R., Dobias, D.T., Medina, S.J., Servilio, L., Gupta, A., and Lenski, R.E., *Ecological speciation of bacteriophage lambda in allopatry and sympatry.* Science, 2016. **354**(6317): p. 1301-1304.

67. Saxenhofer, M., Schmidt, S., Ulrich, R.G., and Heckel, G., *Secondary contact between diverged host lineages entails ecological speciation in a European hantavirus.* PLoS Biol, 2019. **17**(2): p. e3000142.

68. Diaz-Munoz, S.L., Tenaillon, O., Goldhill, D., Brao, K., Turner, P.E., and Chao, L., *Electrophoretic mobility confirms reassortment bias among geographic isolates of segmented RNA phages.* BMC Evol Biol, 2013. **13**: p. 206.

69. O'Keefe, K.J., Silander, O.K., McCreery, H., Weinreich, D.M., Wright, K.M., Chao, L., Edwards, S.V., Remold, S.K., and Turner, P.E., *Geographic differences in sexual reassortment in RNA phage.* Evolution, 2010. **64**(10): p. 3010-23.

70. Dulbecco, R., *Mutual exclusion between related phages.* J Bacteriol, 1952. **63**(2): p. 209-17.

71. Abedon, S.T., *Bacteriophage secondary infection.* Virol Sin, 2015. **30**(1): p. 3-10.

72. Refardt, D., *Within-host competition determines reproductive success of temperate bacteriophages.* ISME J, 2011. **5**(9): p. 1451-60.

73. Worobey, M. and Holmes, E.C., *Evolutionary aspects of recombination in RNA viruses.* J Gen Virol, 1999. **80 (Pt 10)**: p. 2535-2543.

74. Diaz-Munoz, S.L. and Koskella, B., *Bacteria-phage interactions in natural environments.* Adv Appl Microbiol, 2014. **89**: p. 135-83.

75. Silander, O.K., Weinreich, D.M., Wright, K.M., O'Keefe, K.J., Rang, C.U., Turner, P.E., and Chao, L., *Widespread genetic exchange among terrestrial bacteriophages.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 19009-14.

76. DaPalma, T., Doonan, B.P., Trager, N.M., and Kasman, L.M., *A systematic approach to virus-virus interactions.* Virus Res, 2010. **149**(1): p. 1-9.

77. Turner, P.E., Burch, C.L., Hanley, K.A., and Chao, L., *Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage phi6.* J Virol, 1999. **73**(3): p. 2420-4.

78. Li, C., Hatta, M., Nidom, C.A., Muramoto, Y., Watanabe, S., Neumann, G., and Kawaoka, Y., *Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence.* Proc Natl Acad Sci U S A, 2010. **107**(10): p. 4687-92.

79. Li, C., Hatta, M., Watanabe, S., Neumann, G., and Kawaoka, Y., *Compatibility among polymerase subunit proteins is a restricting factor in reassortment between equine H7N7 and human H3N2 influenza viruses.* J Virol, 2008. **82**(23): p. 11880-8.

80. Song, M.S., Pascua, P.N., Lee, J.H., Baek, Y.H., Park, K.J., Kwon, H.I., Park, S.J., Kim, C.J., Kim, H., Webby, R.J., Webster, R.G., and Choi, Y.K., *Virulence and genetic compatibility of polymerase reassortant viruses derived from the pandemic (H1N1) 2009 influenza virus and circulating influenza A viruses.* J Virol, 2011. **85**(13): p. 6275-86.

81. Dudas, G., Bedford, T., Lycett, S., and Rambaut, A., *Reassortment between Influenza B Lineages and the Emergence of a Coadapted PB1–PB2–HA Gene Complex.* Molecular Biology and Evolution, 2015. **32**(1): p. 162-172.

82. Cairns, J., *The initiation of vaccinia infection.* Virology, 1960. **11**: p. 603-23.

83. Kieser, Q., Noyce, R.S., Shenouda, M., Lin, Y.J., and Evans, D.H., *Cytoplasmic factories, virus assembly, and DNA replication kinetics collectively constrain the formation of poxvirus recombinants.* PLoS One, 2020. **15**(1): p. e0228028.

84. Lin, Y.C. and Evans, D.H., *Vaccinia virus particles mix inefficiently, and in a way that would restrict viral recombination, in coinfected cells.* J Virol, 2010. **84**(5): p. 2432-43.

85. Paszkowski, P., Noyce, R.S., and Evans, D.H., *Live-Cell Imaging of Vaccinia Virus Recombination.* PLoS Pathog, 2016. **12**(8): p. e1005824.

86. Tomer, E., Cohen, E.M., Drayman, N., Afriat, A., Weitzman, M.D., Zaritsky, A., and Kobiler, O., *Coalescing replication compartments provide the opportunity for recombination between coinfecting herpesviruses.* FASEB J, 2019. **33**(8): p. 9388-9403.

87. Lambowitz, A.M. and Belfort, M., *INTRONS AS MOBILE GENETIC ELEMENTS.* Annual Review of Biochemistry, 1993. **62**(1): p. 587-622.

88. Stoddard, B.L., *Homing endonucleases from mobile group I introns: discovery to genome engineering.* Mob DNA, 2014. **5**(1): p. 7.

89. Edgell, D.R., Chalamcharla, V.R., and Belfort, M., *Learning to live together: mutualism between self-splicing introns and their hosts.* BMC Biol, 2011. **9**: p. 22.

90. Edgell, D.R., Belfort, M., and Shub, D.A., *Barriers to intron promiscuity in bacteria.* J Bacteriol, 2000. **182**(19): p. 5281-9.

91. Edgell, D.R., *Selfish DNA: homing endonucleases find a home.* Curr Biol, 2009. **19**(3): p. R115-7.

92. Belle, A., Landthaler, M., and Shub, D.A., *Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns.* Genes Dev, 2002. **16**(3): p. 351-62.

93. Belfort, M., *Bacteriophage introns: parasites within parasites?* Trends Genet, 1989. **5**(7): p. 209-13.

94. Nielsen, H., *Ribozymes: Methods and Protocols*. Group I Intron Ribozymes, ed. J.S. Hartig. 2012, Totowa, NJ: Humana Press. 73–89.

95. Zeng, Q., Bonocora, R.P., and Shub, D.A., *A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages.* Curr Biol, 2009. **19**(3): p. 218-22.

96. Bonocora, R.P. and Shub, D.A., *A likely pathway for formation of mobile group I introns.* Curr Biol, 2009. **19**(3): p. 223-8.

97. Loizos, N., Tillier, E.R., and Belfort, M., *Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease.* Proc Natl Acad Sci U S A, 1994. **91**(25): p. 11983-7.

98. Bechhofer, D.H., Hue, K.K., and Shub, D.A., *An intron in the thymidylate synthase gene of Bacillus bacteriophage beta 22: evidence for independent evolution of a gene, its group I intron, and the intron open reading frame.* Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11669-73.

99. Goodrich-Blair, H. and Shub, D.A., *Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers.* Cell, 1996. **84**(2): p. 211-21.

100. Katharios, P., Kalatzis, P.G., Kokkari, C., Sarropoulou, E., and Middelboe, M., *Isolation and characterization of a N4-like lytic bacteriophage infecting Vibrio splendidus, a pathogen of fish and bivalves.* PLoS One, 2017. **12**(12): p. e0190083.

101. Thaler, D.S., Stahl, M.M., and Stahl, F.W., *Tests of the double-strand-break repair model for red-mediated recombination of phage lambda and plasmid lambda dv.* Genetics, 1987. **116**(4): p. 501-11.

102. Robbins, J.B., Stapleton, M., Stanger, M.J., Smith, D., Dansereau, J.T., Derbyshire, V., and Belfort, M., *Homing endonuclease I-TevIII: dimerization as a means to a double-strand break.* Nucleic Acids Res, 2007. **35**(5): p. 1589-600.

103. Bell-Pedersen, D., Quirk, S., Clyman, J., and Belfort, M., *Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications.* Nucleic Acids Res, 1990. **18**(13): p. 3763-70.

104. Chu, F.K., Maley, G., Pedersen-Lane, J., Wang, A.M., and Maley, F., *Characterization of the restriction site of a prokaryotic intron-encoded endonuclease.* Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3574-8.

105. Kong, S., Liu, X., Fu, L., Yu, X., and An, C., *I-PfoP3I: a novel nicking HNH homing endonuclease encoded in the group I intron of the DNA polymerase gene in Phormidium foveolarum phage Pf-WMP3.* PLoS One, 2012. **7**(8): p. e43738.

106. Shen, B.W., Landthaler, M., Shub, D.A., and Stoddard, B.L., *DNA binding and cleavage by the HNH homing endonuclease I-HmuI.* J Mol Biol, 2004. **342**(1): p. 43-56.

107. Landthaler, M. and Shub, D.A., *The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the Bacillus thuringiensis phage Bastille.* Nucleic Acids Res, 2003. **31**(12): p. 3071-7.

108. Landthaler, M., Begley, U., Lau, N.C., and Shub, D.A., *Two self-splicing group I introns in the ribonucleotide reductase large subunit gene of Staphylococcus aureus phage Twort.* Nucleic Acids Res, 2002. **30**(9): p. 1935-43.

109. Metzger, M.J., McConnell-Smith, A., Stoddard, B.L., and Miller, A.D., *Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template.* Nucleic Acids Res, 2011. **39**(3): p. 926-35.

110. Quirk, S.M., Bell-Pedersen, D., and Belfort, M., *Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames.* Cell, 1989. **56**(3): p. 455-65.

111. Martinsohn, J.T., Radman, M., and Petit, M.A., *The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism.* PLoS Genet, 2008. **4**(5): p. e1000065.

112. Landthaler, M., Shen, B.W., Stoddard, B.L., and Shub, D.A., *I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities.* J Mol Biol, 2006. **358**(4): p. 1137-51.

113. Bell-Pedersen, D., Quirk, S.M., Bryk, M., and Belfort, M., *I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target.* Proc Natl Acad Sci U S A, 1991. **88**(17): p. 7719-23.

114. Sitbon, E. and Pietrokovski, S., *New types of conserved sequence domains in DNA-binding regions of homing endonucleases.* Trends Biochem Sci, 2003. **28**(9): p. 473-7.

115. Mate, M.J. and Kleanthous, C., *Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases.* J Biol Chem, 2004. **279**(33): p. 34763-9.

116. Kala, S., Cumby, N., Sadowski, P.D., Hyder, B.Z., Kanelis, V., Davidson, A.R., and Maxwell, K.L., *HNH proteins are a widespread component of phage DNA packaging machines.* Proc Natl Acad Sci U S A, 2014. **111**(16): p. 6022-7.

117. Gimble, F.S., *Degeneration of a homing endonuclease and its target sequence in a wild yeast strain.* Nucleic Acids Res, 2001. **29**(20): p. 4215-23.

118. Goodrich-Blair, H. and Shub, D.A., *The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames.* Nucleic Acids Res, 1994. **22**(18): p. 3715-21.

119. Robbins, J.B., Smith, D., and Belfort, M., *Redox-responsive zinc finger fidelity switch in homing endonuclease and intron promiscuity in oxidative stress.* Curr Biol, 2011. **21**(3): p. 243-8.

120. Coros, C.J., Piazza, C.L., Chalamcharla, V.R., Smith, D., and Belfort, M., *Global regulators orchestrate group II intron retromobility.* Mol Cell, 2009. **34**(2): p. 250-6.

121. Hurst, G.D. and Werren, J.H., *The role of selfish genetic elements in eukaryotic evolution.* Nat Rev Genet, 2001. **2**(8): p. 597-606.

122. Aagaard, C., Dalgaard, J.Z., and Garrett, R.A., *Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron- cells of Sulfolobus acidocaldarius.* Proc Natl Acad Sci U S A, 1995. **92**(26): p. 12285-9.

123. Naor, A., Altman-Price, N., Soucy, S.M., Green, A.G., Mitiagin, Y., Turgeman-Grott, I., Davidovich, N., Gogarten, J.P., and Gophna, U., *Impact of a homing intein on recombination frequency and organismal fitness.* Proc Natl Acad Sci U S A, 2016. **113**(32): p. E4654-61.

124. Edgell, D.R., Fast, N.M., and Doolittle, W.F., *Selfish DNA: the best defense is a good offense.* Curr Biol, 1996. **6**(4): p. 385-8.

125. De Paepe, M., Hutinet, G., Son, O., Amarir-Bouhram, J., Schbath, S., and Petit, M.A., *Temperate phages acquire DNA from defective prophages by relaxed homologous recombination: the role of Rad52-like recombinases.* PLoS Genet, 2014. **10**(3): p. e1004181.

126. Koonin, E.V., *The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?* Biol Direct, 2006. **1**: p. 22.

127. Martin, W. and Koonin, E.V., *Introns and the origin of nucleus-cytosol compartmentalization.* Nature, 2006. **440**(7080): p. 41-5.

128. Cavalier-Smith, T., *The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa.* Int J Syst Evol Microbiol, 2002. **52**(Pt 2): p. 297-354.

129. Nobs, S.-J., Macleod, F.I., Wong, H.L., and Burns, B.P., *Eukarya the chimera: eukaryotes, a secondary innovation of the two domains of life?* Trends in Microbiology, 2021.

130. Bell, P.J., *Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus?* J Mol Evol, 2001. **53**(3): p. 251-6.

131. Bell, P.J.L., *Evidence supporting a viral origin of the eukaryotic nucleus.* Virus Res, 2020. **289**: p. 198168.

132. Cavalier-Smith, T., *Intron phylogeny: a new hypothesis.* Trends Genet, 1991. **7**(5): p. 145-8.

133. Martin, W. and Muller, M., *The hydrogen hypothesis for the first eukaryote.* Nature, 1998. **392**(6671): p. 37-41.

134. Rivera, M.C. and Lake, J.A., *The ring of life provides evidence for a genome fusion origin of eukaryotes.* Nature, 2004. **431**(7005): p. 152-5.

135. Eme, L., Spang, A., Lombard, J., Stairs, C.W., and Ettema, T.J.G., *Archaea and the origin of eukaryotes.* Nat Rev Microbiol, 2017. **15**(12): p. 711-723.

136. Cavalier-Smith, T. and Lee, J.J., *Protozoa as Hosts for Endosymbioses and the Conversion of Symbionts into Organelles1,2.* The Journal of Protozoology, 1985. **32**(3): p. 376-379.

137. Cavalier-Smith, T., *The origin of eukaryotic and archaebacterial cells.* Ann N Y Acad Sci, 1987. **503**: p. 17-54.

138. Zillig, W., *Comparative biochemistry of Archaea and Bacteria.* Curr Opin Genet Dev, 1991. **1**(4): p. 544-51.

139. Kyrpides, N., Overbeek, R., and Ouzounis, C., *Universal protein families and the functional content of the last universal common ancestor.* J Mol Evol, 1999. **49**(4): p. 413-23.

140. Ouzounis, C. and Kyrpides, N., *The emergence of major cellular processes in evolution.* FEBS Lett, 1996. **390**(2): p. 119-23.

141. Ouzounis, C.A. and Kyrpides, N.C., *Parallel origins of the nucleosome core and eukaryotic transcription from Archaea.* J Mol Evol, 1996. **42**(2): p. 234-9.

142. Rivera, M.C., Jain, R., Moore, J.E., and Lake, J.A., *Genomic evidence for two functionally distinct gene classes.* Proc Natl Acad Sci U S A, 1998. **95**(11): p. 6239-44.

143. Ribeiro, S. and Golding, G.B., *The mosaic nature of the eukaryotic nucleus.* Mol Biol Evol, 1998. **15**(7): p. 779-88.

144. Doolittle, W.F., *You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes.* Trends Genet, 1998. **14**(8): p. 307-11.

145. Gray, M.W., *The pre-endosymbiont hypothesis: a new perspective on the origin and evolution of mitochondria.* Cold Spring Harb Perspect Biol, 2014. **6**(3).

146. Gray, M.W., Burger, G., and Lang, B.F., *Mitochondrial evolution.* Science, 1999. **283**(5407): p. 1476-81.

147. Lang, B.F., Gray, M.W., and Burger, G., *Mitochondrial genome evolution and the origin of eukaryotes.* Annu Rev Genet, 1999. **33**: p. 351-97.

148. von Dohlen, C.D., Kohler, S., Alsop, S.T., and McManus, W.R., *Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts.* Nature, 2001. **412**(6845): p. 433-6.

149. Tovar, J., Leon-Avila, G., Sanchez, L.B., Sutak, R., Tachezy, J., van der Giezen, M., Hernandez, M., Muller, M., and Lucocq, J.M., *Mitochondrial remnant organelles of Giardia function in iron-sulphur protein maturation.* Nature, 2003. **426**(6963): p. 172-6.

150. Hrdy, I., Hirt, R.P., Dolezal, P., Bardonova, L., Foster, P.G., Tachezy, J., and Embley, T.M., *Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I.* Nature, 2004. **432**(7017): p. 618-22.

151. Ku, C., Nelson-Sathi, S., Roettger, M., Sousa, F.L., Lockhart, P.J., Bryant, D., Hazkani-Covo, E., McInerney, J.O., Landan, G., and Martin, W.F., *Endosymbiotic origin and differential loss of eukaryotic genes.* Nature, 2015. **524**(7566): p. 427-32.

152. Gilbert, W., *Why genes in pieces?* Nature, 1978. **271**(5645): p. 501.

153. Roy, S.W. and Gilbert, W., *Rates of intron loss and gain: implications for early eukaryotic evolution.* Proc Natl Acad Sci U S A, 2005. **102**(16): p. 5773-8.

154. Rogers, J.H., *How were introns inserted into nuclear genes?* Trends Genet, 1989. **5**(7): p. 213-6.

155. Ferat, J.L. and Michel, F., *Group II self-splicing introns in bacteria.* Nature, 1993. **364**(6435): p. 358-61.

156. Shub, D.A., Goodrich-Blair, H., and Eddy, S.R., *Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns.* Trends Biochem Sci, 1994. **19**(10): p. 402-4.

157. Gorbalenya, A.E., *Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family.* Protein Sci, 1994. **3**(7): p. 1117-20.

158. Dai, L. and Zimmerly, S., *Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior.* Nucleic Acids Res, 2002. **30**(5): p. 1091-102.

159. Wise, J.A., *Guides to the heart of the spliceosome.* Science, 1993. **262**(5142): p. 1978-9.

160. Cech, T.R., *The generality of self-splicing RNA: relationship to nuclear mRNA splicing.* Cell, 1986. **44**(2): p. 207-10.

161. Jacquier, A., *Self-splicing group II and nuclear pre-mRNA introns: how similar are they?* Trends Biochem Sci, 1990. **15**(9): p. 351-4.

162. Weiner, A.M., *mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism?* Cell, 1993. **72**(2): p. 161-4.

163. Nilsen, T.W., *RNA-RNA interactions in the spliceosome: unraveling the ties that bind.* Cell, 1994. **78**(1): p. 1-4.

164. Zimmerly, S., Hausner, G., and Wu, X., *Phylogenetic relationships among group II intron ORFs.* Nucleic Acids Res, 2001. **29**(5): p. 1238-50.

165. Lynch, M. and Richardson, A.O., *The evolution of spliceosomal introns.* Curr Opin Genet Dev, 2002. **12**(6): p. 701-10.

166. Toro, N., *Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment.* Environ Microbiol, 2003. **5**(3): p. 143-51.

167. Robart, A.R. and Zimmerly, S., *Group II intron retroelements: function and diversity.* Cytogenet Genome Res, 2005. **110**(1-4): p. 589-97.

168. Vanacova, S., Yan, W., Carlton, J.M., and Johnson, P.J., *Spliceosomal introns in the deep-branching eukaryote Trichomonas vaginalis.* Proc Natl Acad Sci U S A, 2005. **102**(12): p. 4430-5.

169. Toro, N., Jimenez-Zurdo, J.I., and Garcia-Rodriguez, F.M., *Bacterial group II introns: not just splicing.* FEMS Microbiol Rev, 2007. **31**(3): p. 342-58.

170. Chalamcharla, V.R., Curcio, M.J., and Belfort, M., *Nuclear expression of a group II intron is consistent with spliceosomal intron ancestry.* Genes Dev, 2010. **24**(8): p. 827-36.

171. Novikova, O. and Belfort, M., *Mobile Group II Introns as Ancestral Eukaryotic Elements.* Trends Genet, 2017. **33**(11): p. 773-783.

172. Guglielmini, J., Woo, A.C., Krupovic, M., Forterre, P., and Gaia, M., *Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes.* Proc Natl Acad Sci U S A, 2019. **116**(39): p. 19585-19592.

173. Mans, B., Anantharaman, V., Aravind, L., and Koonin, E.V., *Comparative Genomics, Evolution and Origins of the Nuclear Envelope and Nuclear Pore Complex.* Cell Cycle, 2004. **3**(12): p. 1625-1650.

174. Forterre, P., *[The great virus comeback].* Biol Aujourdhui, 2013. **207**(3): p. 153-68.

175. Forterre, P. and Raoult, D., *The transformation of a bacterium into a nucleated virocell reminds the viral eukaryogenesis hypothesis.* Virologie (Montrouge), 2017. **21**(4): p. 28-30.

176. Takemura, M., *Poxviruses and the origin of the eukaryotic nucleus.* J Mol Evol, 2001. **52**(5): p. 419-25.

177. Iyer, L.M., Aravind, L., and Koonin, E.V., *Common origin of four diverse families of large eukaryotic DNA viruses.* J Virol, 2001. **75**(23): p. 11720-34.

178. Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J.M., *The 1.2-megabase genome sequence of Mimivirus.* Science, 2004. **306**(5700): p. 1344-50.

179. Abrahao, J., Silva, L., Silva, L.S., Khalil, J.Y.B., Rodrigues, R., Arantes, T., Assis, F., Boratto, P., Andrade, M., Kroon, E.G., Ribeiro, B., Bergier, I., Seligmann, H., Ghigo, E., Colson, P., Levasseur, A., Kroemer, G., Raoult, D., and La Scola, B., *Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere.* Nat Commun, 2018. **9**(1): p. 749.

180. Schulz, F., Yutin, N., Ivanova, N.N., Ortega, D.R., Lee, T.K., Vierheilig, J., Daims, H., Horn, M., Wagner, M., Jensen, G.J., Kyrpides, N.C., Koonin, E.V., and Woyke, T., *Giant viruses with an expanded complement of translation system components.* Science, 2017. **356**(6333): p. 82-85.

181. Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C.J., Olm, M.R., Bouma-Gregson, K., Amano, Y., He, C., Méheust, R., Brooks, B., Thomas, A., Lavy, A., Matheus-Carnevali, P., Sun, C., Goltsman, D.S.A., Borton, M.A., Sharrar, A., Jaffe, A.L., Nelson, T.C., Kantor, R., Keren, R., Lane, K.R., Farag, I.F., Lei, S., Finstad, K., Amundson, R., Anantharaman, K., Zhou, J., Probst, A.J., Power, M.E., Tringe, S.G., Li, W.-J., Wrighton, K., Harrison, S., Morowitz, M., Relman, D.A., Doudna, J.A., Lehours, A.-C., Warren, L., Cate, J.H.D., Santini, J.M., and Banfield, J.F., *Clades of huge phages from across Earth's ecosystems.* Nature, 2020. **578**(7795): p. 425-431.

182. Doolittle, W.F., *The origins of introns.* Curr Biol, 1991. **1**(3): p. 145-6.

183. Taft, R.J. and Mattick, J.S., Genome Biology, 2003. **5**(1): p. P1.

184. Koonin, E.V., *Energetics and population genetics at the root of eukaryotic cellular and genomic complexity.* Proc Natl Acad Sci U S A, 2015. **112**(52): p. 15777-8.

185. Wolf, Y.I., Katsnelson, M.I., and Koonin, E.V., *Physical foundations of biological complexity.* Proc Natl Acad Sci U S A, 2018. **115**(37): p. E8678-E8687.

186. Gould, K., *Antibiotics: from prehistory to the present day.* J Antimicrob Chemother, 2016. **71**(3): p. 572-5.

187. Zaffiri, L., Gardner, J., and Toledo-Pereyra, L.H., *History of antibiotics. From salvarsan to cephalosporins.* J Invest Surg, 2012. **25**(2): p. 67-77.

188. Wainwright, M. and Swan, H.T., *C.G. Paine and the earliest surviving clinical records of penicillin therapy.* Med Hist, 1986. **30**(1): p. 42-56.

189. Summers, W.C., *The strange history of phage therapy.* Bacteriophage, 2012. **2**(2): p. 130-133.

190. Stent, G.S., *Molecular biology of bacterial viruses*. 1963: San Francisco: W.H. Freeman.

191. Lerminiaux, N.A. and Cameron, A.D.S., *Horizontal transfer of antibiotic resistance genes in clinical environments.* Can J Microbiol, 2019. **65**(1): p. 34-44.

192. Abraham, E.P. and Chain, E., *An enzyme from bacteria able to destroy penicillin. 1940.* Rev Infect Dis, 1988. **10**(4): p. 677-8.

193. Davies, J. and Davies, D., *Origins and evolution of antibiotic resistance.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 417-33.

194. US Department of Health and Human Services, C., *Antibiotic Resistance Threats in the United States*. 2019, Centers for Disease Control and Prevention. p. 1-113.

195. Loc-Carrillo, C. and Abedon, S.T., *Pros and cons of phage therapy.* Bacteriophage, 2011. **1**(2): p. 111-114.

196. Dąbrowska, K., *Phage therapy: What factors shape phage pharmacokinetics and bioavailability? Systematic and critical review.* Medicinal Research Reviews, 2019. **39**(5): p. 2000-2025.

197. Ujmajuridze, A., Chanishvili, N., Goderdzishvili, M., Leitner, L., Mehnert, U., Chkhotua, A., Kessler, T.M., and Sybesma, W., *Adapted Bacteriophages for Treating Urinary Tract Infections.* Front Microbiol, 2018. **9**: p. 1832.

198. Kahn, L.H., Bergeron, G., Bourassa, M.W., Vegt, B., Gill, J., Gomes, F., Malouin, F., Opengart, K., Ritter, G.D., Singer, R.S., Storrs, C., and Topp, E., *From farm management to bacteriophage therapy: strategies to reduce antibiotic use in animal agriculture.* Annals of the New York Academy of Sciences, 2019. **1441**(1): p. 31-39.

199. Moye, Z., Woolston, J., and Sulakvelidze, A., *Bacteriophage Applications for Food Production and Processing.* Viruses, 2018. **10**(4): p. 205.

200. Schooley, R.T., Biswas, B., Gill, J.J., Hernandez-Morales, A., Lancaster, J., Lessor, L., Barr, J.J., Reed, S.L., Rohwer, F., Benler, S., Segall, A.M., Taplitz, R., Smith, D.M., Kerr, K., Kumaraswamy, M., Nizet, V., Lin, L., McCauley, M.D., Strathdee, S.A., Benson, C.A., Pope, R.K., Leroux, B.M., Picel, A.C., Mateczun, A.J., Cilwa, K.E., Regeimbal, J.M., Estrella, L.A., Wolfe, D.M., Henry, M.S., Quinones, J., Salka, S., Bishop-Lilly, K.A., Young, R., and Hamilton, T., *Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant Acinetobacter baumannii Infection.* Antimicrob Agents Chemother, 2017. **61**(10).

201. Dedrick, R.M., Guerrero-Bustamante, C.A., Garlena, R.A., Russell, D.A., Ford, K., Harris, K., Gilmour, K.C., Soothill, J., Jacobs-Sera, D., Schooley, R.T., Hatfull, G.F., and Spencer, H., *Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus.* Nature Medicine, 2019. **25**(5): p. 730-733.

202. LaVergne, S., Hamilton, T., Biswas, B., Kumaraswamy, M., Schooley, R.T., and Wooten, D., *Phage Therapy for a Multidrug-Resistant Acinetobacter baumannii Craniectomy Site Infection.* Open Forum Infect Dis, 2018. **5**(4): p. ofy064.

203. Aslam, S., Lampley, E., Wooten, D., Karris, M., Benson, C., Strathdee, S., and Schooley, R.T., *Lessons Learned From the First 10 Consecutive Cases of Intravenous Bacteriophage Therapy to Treat Multidrug-Resistant Bacterial Infections at a Single Center in the United States.* Open Forum Infectious Diseases, 2020. **7**(9).

204. Furfaro, L.L., Payne, M.S., and Chang, B.J., *Bacteriophage Therapy: Clinical Trials and Regulatory Hurdles.* Front Cell Infect Microbiol, 2018. **8**: p. 376.

205. Stecher, B., *The Roles of Inflammation, Nutrient Availability and the Commensal Microbiota in Enteric Pathogen Infection.* Microbiol Spectr, 2015. **3**(3).

206. Kim, S., Covington, A., and Pamer, E.G., *The intestinal microbiota: Antibiotics, colonization resistance, and enteric pathogens.* Immunol Rev, 2017. **279**(1): p. 90-105.

207. Francino, M.P., *Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances.* Front Microbiol, 2015. **6**: p. 1543.

208. Maier, L., Goemans, C.V., Wirbel, J., Kuhn, M., Eberl, C., Pruteanu, M., Muller, P., Garcia-Santamarina, S., Cacace, E., Zhang, B., Gekeler, C., Banerjee, T., Anderson, E.E., Milanese, A., Lober, U., Forslund, S.K., Patil, K.R., Zimmermann, M., Stecher, B., Zeller, G., Bork, P., and Typas, A., *Unravelling the collateral damage of antibiotics on gut bacteria.* Nature, 2021. **599**(7883): p. 120-124.

209. Blaser, M.J., *Antibiotic use and its consequences for the normal microbiome.* Science, 2016. **352**(6285): p. 544-5.

210. Tanaka, S., Kobayashi, T., Songjinda, P., Tateyama, A., Tsubouchi, M., Kiyohara, C., Shirakawa, T., Sonomoto, K., and Nakayama, J., *Influence of antibiotic exposure in the early postnatal period on the development of intestinal microbiota.* FEMS Immunol Med Microbiol, 2009. **56**(1): p. 80-7.

211. Arrieta, M.C., Stiemsma, L.T., Dimitriu, P.A., Thorson, L., Russell, S., Yurist-Doutsch, S., Kuzeljevic, B., Gold, M.J., Britton, H.M., Lefebvre, D.L., Subbarao, P., Mandhane, P., Becker, A., McNagny, K.M., Sears, M.R., Kollmann, T., Investigators, C.S., Mohn, W.W., Turvey, S.E., and Finlay, B.B., *Early infancy microbial and metabolic alterations affect risk of childhood asthma.* Sci Transl Med, 2015. **7**(307): p. 307ra152.

212. Azad, M.B., Bridgman, S.L., Becker, A.B., and Kozyrskyj, A.L., *Infant antibiotic exposure and the development of childhood overweight and central adiposity.* Int J Obes (Lond), 2014. **38**(10): p. 1290-8.

213. Azad, M.B., Konya, T., Persaud, R.R., Guttman, D.S., Chari, R.S., Field, C.J., Sears, M.R., Mandhane, P.J., Turvey, S.E., Subbarao, P., Becker, A.B., Scott, J.A., Kozyrskyj, A.L., and Investigators, C.S., *Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: a prospective cohort study.* BJOG, 2016. **123**(6): p. 983-93.

214. Neuman, H., Forsythe, P., Uzan, A., Avni, O., and Koren, O., *Antibiotics in early life: dysbiosis and the damage done.* FEMS Microbiol Rev, 2018. **42**(4): p. 489-499.

215. Dethlefsen, L., Huse, S., Sogin, M.L., and Relman, D.A., *The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing.* PLoS Biol, 2008. **6**(11): p. e280.

216. Dethlefsen, L. and Relman, D.A., *Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4554-61.

217. Isaac, S., Scher, J.U., Djukovic, A., Jimenez, N., Littman, D.R., Abramson, S.B., Pamer, E.G., and Ubeda, C., *Short- and long-term effects of oral vancomycin on the human intestinal microbiota.* J Antimicrob Chemother, 2017. **72**(1): p. 128-136.

218. Ramirez, J., Guarner, F., Bustos Fernandez, L., Maruy, A., Sdepanian, V.L., and Cohen, H., *Antibiotics as Major Disruptors of Gut Microbiota.* Front Cell Infect Microbiol, 2020. **10**: p. 572912.

219. Cho, I., Yamanishi, S., Cox, L., Methe, B.A., Zavadil, J., Li, K., Gao, Z., Mahana, D., Raju, K., Teitler, I., Li, H., Alekseyenko, A.V., and Blaser, M.J., *Antibiotics in early life alter the murine colonic microbiome and adiposity.* Nature, 2012. **488**(7413): p. 621-6.

220. Hsiao, E.Y., McBride, S.W., Hsien, S., Sharon, G., Hyde, E.R., McCue, T., Codelli, J.A., Chow, J., Reisman, S.E., Petrosino, J.F., Patterson, P.H., and Mazmanian, S.K., *Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders.* Cell, 2013. **155**(7): p. 1451-63.

221. Volkova, A., Ruggles, K., Schulfer, A., Gao, Z., Ginsberg, S.D., and Blaser, M.J., *Effects of early-life penicillin exposure on the gut microbiome and frontal cortex and amygdala gene expression.* iScience, 2021. **24**(7): p. 102797.

222. Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J., *Meta-analysis of gut microbiome studies identifies disease-specific and shared responses.* Nat Commun, 2017. **8**(1): p. 1784.

223. Abdallah Ismail, N., Ragab, S.H., Abd Elbaky, A., Shoeib, A.R., Alhosary, Y., and Fekry, D., *Frequency of Firmicutes and Bacteroidetes in gut microbiota in obese and normal weight Egyptian children and adults.* Arch Med Sci, 2011. **7**(3): p. 501-7.

224. Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., Sun, X., Li, Z., Tang, A., Zhong, S., Li, X., Chen, W., Xu, R., Wang, M., Feng, Q., Gong, M., Yu, J., Zhang, Y., Zhang, M., Hansen, T., Sanchez, G., Raes, J., Falony, G., Okuda, S., Almeida, M., LeChatelier, E., Renault, P., Pons, N., Batto, J.M., Zhang, Z., Chen, H., Yang, R., Zheng, W., Li, S., Yang, H., Wang, J., Ehrlich, S.D., Nielsen, R., Pedersen, O., Kristiansen, K., and Wang, J., *A metagenome-wide association study of gut microbiota in type 2 diabetes.* Nature, 2012. **490**(7418): p. 55-60.

225. Milosevic, I., Vujovic, A., Barac, A., Djelic, M., Korac, M., Radovanovic Spurnic, A., Gmizic, I., Stevanovic, O., Djordjevic, V., Lekic, N., Russo, E., and Amedei, A., *Gut-Liver Axis, Gut Microbiota, and Its Modulation in the Management of Liver Diseases: A Review of the Literature.* Int J Mol Sci, 2019. **20**(2).

226. Wirbel, J., Pyl, P.T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J.S., Voigt, A.Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L.P., Schrotz-King, P., Vogtmann, E., Habermann, N., Nimeus, E., Thomas, A.M., Manghi, P., Gandini, S., Serrano, D., Mizutani, S., Shiroma, H., Shiba, S., Shibata, T., Yachida, S., Yamada, T., Waldron, L., Naccarati, A., Segata, N., Sinha, R., Ulrich, C.M., Brenner, H., Arumugam, M., Bork, P., and Zeller, G., *Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer.* Nat Med, 2019. **25**(4): p. 679-689.

227. Nguyen, T.T., Hathaway, H., Kosciolek, T., Knight, R., and Jeste, D.V., *Gut microbiome in serious mental illnesses: A systematic review and critical evaluation.* Schizophr Res, 2021. **234**: p. 24-40.

228. Nguyen, T.T., Kosciolek, T., Daly, R.E., Vazquez-Baeza, Y., Swafford, A., Knight, R., and Jeste, D.V., *Gut microbiome in Schizophrenia: Altered functional pathways related to immune modulation and atherosclerotic risk.* Brain Behav Immun, 2021. **91**: p. 245-256.

229. Nguyen, T.T., Kosciolek, T., Maldonado, Y., Daly, R.E., Martin, A.S., McDonald, D., Knight, R., and Jeste, D.V., *Differences in gut microbiome composition between persons with chronic schizophrenia and healthy comparison subjects.* Schizophr Res, 2019. **204**: p. 23-29.

230. Holmes, M., Flaminio, Z., Vardhan, M., Xu, F., Li, X., Devinsky, O., and Saxena, D., *Cross talk between drug-resistant epilepsy and the gut microbiome.* Epilepsia, 2020. **61**(12): p. 2619-2628.

231. Ceccarani, C., Vigano, I., Ottaviano, E., Redaelli, M.G., Severgnini, M., Vignoli, A., and Borghi, E., *Is Gut Microbiota a Key Player in Epilepsy Onset? A Longitudinal Study in Drug-Naive Children.* Front Cell Infect Microbiol, 2021. **11**: p. 749509.

232. Wu, M., Liu, J., Li, F., Huang, S., He, J., Xue, Y., Fu, T., Feng, S., and Li, Z., *Antibiotic-induced dysbiosis of gut microbiota impairs corneal development in postnatal mice by affecting CCR2 negative macrophage distribution.* Mucosal Immunol, 2020. **13**(1): p. 47-63.

233. Strati, F., Pujolassos, M., Burrello, C., Giuffre, M.R., Lattanzi, G., Caprioli, F., Troisi, J., and Facciotti, F., *Antibiotic-associated dysbiosis affects the ability of the gut microbiota to control intestinal inflammation upon fecal microbiota transplantation in experimental colitis models.* Microbiome, 2021. **9**(1): p. 39.

234. Landers, T.F., Cohen, B., Wittum, T.E., and Larson, E.L., *A review of antibiotic use in food animals: perspective, policy, and potential.* Public Health Rep, 2012. **127**(1): p. 4-22.

CHAPTER 2:

A cytoskeletal vortex drives phage nucleus rotation during jumbo phage replication in *Escherichia coli*

2.1 Abstract

Nucleus-forming jumbo phages establish an intricate subcellular organization, enclosing phage genomes within a proteinaceous shell called the phage nucleus. During infection in *Pseudomonas*, *Phikzvirus* phages assemble a bipolar spindle of tubulin-like PhuZ filaments that positions the phage nucleus at midcell, and drives its intracellular rotation, facilitating the distribution of capsids on its surface for genome packaging. Here we show that the *Escherichia coli* jumbo phage Goslar assembles a phage nucleus surrounded by an array of PhuZ filaments resembling a vortex instead of a bipolar spindle. Expression of a mutant PhuZ protein strongly reduces Goslar phage nucleus rotation, demonstrating that the PhuZ cytoskeletal vortex is necessary for rotating the phage nucleus. While vortex-like cytoskeletal arrays are important in eukaryotes for cytoplasmic streaming and nucleus alignment, this work identifies the first known example of a coherent assembly of filaments into a vortex-like structure driving intracellular rotation within the prokaryotic cytoplasm.

2.2 Introduction

Subcellular organization is essential for all domains of life, including viruses [1, 2]. We recently identified a nucleus-like structure formed by *Pseudomonas* jumbo phage 201φ2-1 [3], which was subsequently shown to be conserved in ΦKZ [4, 5], and ΦPA3 [4]. Analogous to a eukaryotic nucleus, the phage nucleus imparts strict separation of transcription from translation by enclosing phage DNA within a proteinaceous shell that excludes ribosomes [3]. This separation requires the export of mRNA to the cytoplasm for translation and the selective import of proteins required for transcription, DNA replication, and DNA repair. The phage nucleus grows as phage DNA is replicated inside and it is positioned in the center of the cell by a bipolar spindle composed of filaments of the tubulin-like PhuZ protein [3, 6, 7]. The phage nucleus protects the phage DNA

by excluding restriction enzymes and DNA-targeting CRISPR-Cas in both *Pseudomonas* and *Serratia* jumbo phages [8-10]. Nucleus-forming *Pseudomonas* jumbo phages rely upon a bipolar spindle to position and rotate the phage nucleus that encloses their DNA and protects it from host defenses [1]. The phage nucleus and spindle are also important for jumbo phage diversification because they contribute to the evolution of new species through subcellular genetic isolation and virogenesis incompatibility [11].

Early in the nucleus-forming jumbo phage infection process, the phage nucleus is maneuvered towards the center of the host cell and oscillates in position once it reaches the midcell due to stochastic growth and shrinkage of PhuZ filaments in the bipolar spindle [3, 6, 7]. PhuZ, which forms a three-stranded polar filament with dynamic (+) and (-) ends, analogous to eukaryotic microtubules [12], was first identified in phage 201φ2-1 infecting *Pseudomonas chlororaphis* and later shown to be conserved, along with the phage nucleus, in the related *Pseudomonas aeruginosa* phages ΦKZ and ΦPA3 [4, 13]. The (-) ends of the spindle are positioned at each cell pole, while the (+) ends point toward the midcell. PhuZ filaments display dynamic instability during which the (+) ends rapidly depolymerize until returning to a growth phase, thereby allowing the spindle to position the phage nucleus in the center of the cell [4, 6, 7].

During the later phases of the nucleus-forming jumbo phage replication cycle, PhuZ filaments exhibit treadmilling when they polymerize at the cell pole at a rate similar to depolymerization at the surface of the nucleus [14]. Treadmilling filaments apply pushing forces to opposing sides of the phage nucleus, causing it to rotate [3, 14]. Capsids, which assemble on the cell membrane, attach to treadmilling filaments of the PhuZ spindle and are trafficked to the surface of the phage nucleus where they dock for packaging of the genome [14]. As the PhuZ spindle rotates the phage nucleus, capsids are distributed to different locations on its surface to

promote efficient DNA packaging. These functions of the PhuZ spindle, including the remarkable capacity for driving intracellular rotation of the phage nucleus, are conserved in all three *Pseudomonas* jumbo phages described above and they are important for full efficiency of the replication cycle [3, 4, 7, 14].

To add to the intricacies of this replication mechanism, we recently described the discovery of phage bouquets [15]. These nearly spherical structures are formed by DNA-filled capsids with tails pointing inward, resembling a bouquet of flowers. Bouquets are established in the final stages of the nucleus-forming jumbo phage replication cycle when fully packaged capsids move from the surface of the phage nucleus to the regions adjacent to the nucleus. The interior of bouquets largely excludes ribosomes and cytoplasmic GFP [15]. Eventually, the cell lyses, releasing the progeny phages into the environment.

It is still unknown how widespread this replication pathway may be among jumbo phages; assembly of a phage nucleus has only been observed in the above-mentioned *Pseudomonas* phages and Serratia phage PCH45, while the PhuZ cytoskeleton has only been studied in *Pseudomonas*. In order to further understand if phage nucleus formation and positioning, bipolar PhuZ spindle assembly, intracellular rotation of the phage nucleus, capsid trafficking, and phage bouquets are conserved among more diverse jumbo phage, we sought out a nucleus-forming jumbo phage that infects *Escherichia coli*. Here we characterize the reproduction pathway of phage vB_EcoM_Goslar (Goslar) [16]. We show that Goslar is a nucleus-forming jumbo phage that assembles a vortex-like cytoskeletal array instead of a bipolar spindle to drive phage nucleus rotation. PhuZ mutations that disrupt vortex assembly also disrupt phage nucleus rotation, linking the cytoskeletal vortex to a key process in the phage replication cycle. Our results show that the

nucleus-forming phage infection pathway is likely widespread and diverse strategies have evolved to achieve intracellular rotation of the phage nucleus.

2.3 Materials and Methods

Growth conditions and bacteriophage preparation

The *Escherichia coli* strain APEC 2248 (APEC, serotype O78++) is an avian pathogen. It is closely related to *E. coli* K-12, approximately 99% identical [17]. *E. coli* strains APEC and MG1655 were grown on Luria-Bertani (LB) plates containing 10 g Bacto-Tryptone, 5 g NaCl, 5 g Bacto-yeast extract, 16 g agar in 1 L ddH_2O and incubated at 37°C overnight. Liquid cultures were obtained by inoculation of LB broth with one colony of *E. coli* from an LB plate. Lysates for Goslar were obtained from Johannes Wittmann at the DSMZ and were amplified by adding 15 µl high titer phage lysate to 300 µl APEC at OD_{600} 0.5, incubating at 37°C for 30 minutes, then adding 300 µl LB broth, plating 200 µl of the suspension onto each of 3 LB plates and incubating at 37°C overnight. 15 ml of Phage Buffer (PB) containing 10 mM Tris (pH 7.5), 10 mM $MgSO_4$, 68 mM NaCl, and 1 mM $CaCl_2$ was chilled on ice before 5 ml was added to each plate and left to soak at room temperature. After 4 hours, 3 ml of PB was added to each plate and after 2 more hours, the buffer was drawn off into a single tube. The lysate was clarified by pelleting the bacteria at 3220 x g for 10 minutes. The supernatant was filtered through 0.45 µm Corning membranes by syringe. 5 drops of chloroform from a pasteur pipette were added to the 10 ml lysate and shaken by hand for 1 minute. The mixture was spun at 3220 x g for 5 minutes and the aqueous phase containing the phage was removed to a clean tube and stored at 4°C.

Plasmid constructions and bacterial transformation

Fluorescent-tagged phage proteins were synthesized and cloned by restriction with SacI and SalI into pDSW206 by Genscript, and delivered as lyophilized plasmid. The plasmids were

hydrated at ~0.2 g/L with Tris-EDTA buffer and diluted 1:10 with ddH_2O. Electroporation competent DH5α and MG1655 cells were prepared by washing with 10% glycerol and stored at -80°C. 30-50 µl of competent cells was combined with 1 µl of diluted plasmid and electroporated with 1.8 kV then incubated at 30°C in SOC media for 30-60 minutes before plating on LB with 100 µg/ml ampicillin and incubating overnight at 37°C.

Single cell infection assay

E. coli strains containing a plasmid with fluorescent-tagged phage protein(s) were grown on 1% agarose pads supplemented with 25% LB and the desired IPTG concentration to induce protein expression. Cells were obtained from overnight incubation at 37°C on an LB plate with 100 µg/ml ampicillin. A colony was resuspended in 25% LB to an OD_{600} of ~0.35 then 8 µl was spotted on the imaging pad and spread with the bottom of an eppendorf tube. Wild-type *E. coli* was grown without ampicillin or IPTG. The imaging pad was then incubated at 37°C for 1.5 hours without a coverslip in a humidor. 6 µl of Goslar lysate was added to the agarose pads and spread as before, then further incubated at 37°C to allow phage infection to proceed. At the desired time point, the slide was placed at room temperature and spotted with 7 µl of dye mix (2 µg/ml DAPI, 4 µg/ml FM4-64, 25% LB). Once dry after ~5 minutes, a coverslip was put on the agarose pad and fluorescent microscopy was initiated. Data of static images and time-lapse imaging were collected and processed as described below.

Live cell static image and time-lapse fluorescence microscopy

The DeltaVision Elite Deconvolution microscope (Applied Precision, Issaquah, WA, USA) was used to visualize the live cells. For static images, the cells were imaged with 12-15 slices in the Z-axis at 0.15 µm increments. For long time-lapse, imaging pads were prepared and infected as above and 30 minutes after the addition of Goslar, pads were coverslipped without

dyes. The environmental control unit surrounding the microscope warmed the imaging space to 35°C. Fields adequate for imaging were marked and time-lapse imaging began within 10 minutes, with Ultimate Focus utilized. For short time-lapse, infections proceeded at 37°C for the indicated time before being coverslipped and imaged at room temperature one field at a time using Ultimate Focus. Images were processed by the aggressive deconvolution algorithm in the DeltaVision SoftWoRx 6.5.2 Image Analysis Program. Further image analysis and processing was performed in FIJI version 2.1.0/1.53c [18]. Figure images were adjusted and layered in Adobe Photoshop.

Cryo-electron tomography sample preparation

Ten agarose pads for infection were prepared as above (1% agarose, 25% LB) and spotted with 10 µl of APEC cells at an OD_{600} of ~0.35 then incubated at 37°C for 1.5 hours in a humidor. 10 µl of Goslar lysate from the DSMZ was added and spread on each pad then incubated for another 1.5 hours until being removed to room temperature for cell collection. Infected cells were removed from the pads by addition of 25 µl of 25% LB and gentle scraping with the bottom of an eppendorf tube. All 10 pads were collected into one tube and after a portion was aliquoted, the remainder was centrifuged at 6000 x g for 45 seconds, resuspended with ¼ volume of the supernatant, and a portion of that was diluted 1:1 in supernatant. Samples were delivered for plunging 20-30 minutes after removal from 37°C which significantly slows infection progression.

Cryo-focused ion beam milling and electron tomography

Infected cells were prepared as described above and at approximately 90 minutes post infection, 4-7 µl of cells were deposited on R2/1 Cu 200 grids (Quantifoil) that had been glow-discharged for 1 min at 0.19 mbar and 20 mA in a PELCO easiGlow device shortly before use. Grids were mounted in a custom-built manual plunging device (Max Planck Institute of Biochemistry, Martinsried) and excess liquid was blotted with filter paper from the backside of

the grid for 5-7 seconds prior to freezing in a 50:50 ethane:propane mixture (Airgas) cooled by liquid nitrogen.

Grids were mounted into modified Autogrids (TFS) compatible with cryo-focused ion beam milling. Samples were loaded into an Aquilos 2 cryo-focused ion beam/scanning electron microscope (TFS) and milled to yield lamellae following published procedures for bacterial samples [19].

Milled specimens were imaged with a Titan Krios G3 transmission electron microscope (TFS) operated at 300 kV and equipped with a K2 directed electron detector (Gatan) mounted post 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-v3.8b11 [20] and all recorded images were collected using the K2 in counting mode.

Tilt-series were acquired at either 4.27 Å and 5.34 Å per pixel. For the higher magnification tilt-series, images were acquired over a nominal range of +/-60^{o} in 2^{o} steps following a dose-symmetric scheme [21] with a per-tilt fluence of 2.6 $e^{-}\cdot Å^{-2}$ and total of ~ 160 $e^{-}\cdot Å^{-2}$ per tilt-series. Lower magnification tilt-series were acquired similarly, but using a 1.5^{o} tilt-step and per-tilt fluence of 1.8 $e^{-}\cdot Å^{-2}$. Target defocus values were set for between -5 and -6 µm.

Image processing and analysis of cryo-electron tomography data

Tilt-movies were corrected for whole-frame motion using Warp-v1.09 [22] and aligned via patch-tracking using Etomo (IMOD-v4.10.28) [23]. Tilt-series CTF parameters were estimated and tomograms reconstructed with exposure-filtering and CTF-correction using Warp-v1.09. For general visualization and membrane segmentation, tomograms were reconstructed using Warp's

deconvolution filter applied at default settings and downsampled to 20 Å and 25 Å per pixel from the original 4.27 Å and 5.34 Å pixel sizes, respectively.

Segmentation of host cell membranes and the phage nucleus perimeter from representative tomograms was performed by first coarsely segmenting using TomoSegMemTV [24] followed by manual patching with Amira-v6.7 (TFS). Ribosomes, capsids, and tails were segmented using subtomogram analysis. For ribosomes, approximately 200 particles were manually selected from the respective tomograms and used to generate an *ab initio* reference using Warp-v1.09 and Relion-v3.1.1 [25] following conventional procedures [22, 26]. The references were used for template-matching with Warp-v1.09 against their respective tomograms. Template-matching results were curated to remove obvious false-positives (e.g., picks outside cell boundaries and cell membranes, etc.). Curated picks were aligned and classified using Relion-v3.1.1 to remove additional false-positives and refine their positions in the tomogram. For capsids, all particles were manually picked. Reference-generation and alignment of capsids was performed while enforcing icosahedral symmetry with Relion-v3.1.1 (despite the capsids possessing C5 symmetry) in order to promote convergence from the low number of particles. For the phage tails, the start and end points along the filament axis were defined manually and used to generate over-sampled filament models in Dynamo-v1.1514 [27, 28]. An initial reference for the tail was generated using Dynamo-v1.1514 from two full-length tails with clearly defined polarity. The resulting reference displayed apparent C6 symmetry, which was enforced for the alignment of all tails from a given tomogram using Dynamo-v1.1514 and Relion-v3.1.1. All interconversions of metadata between Warp/Relion and Dynamo formats were performed using scripts from the dynamo2m-v0.2.2 package [29]. Final averages were placed back in the reference-frame of their respective

tomograms using *dynamo_table_place*. Figures of the segmentation models were prepared using ChimeraX-v1.2.1 [30].

Construction of PhuZ and shell phylogenetic trees

PSI-BLAST was run querying ΦKZ PhuZ (gp039) and shell (gp054). All phage-encoded hits were pulled and only those with both a PhuZ and a shell were aligned by MUSCLE and used to construct trees in MEGA X (10.2.6) [31, 32]. The evolutionary history was inferred using the Neighbor-Joining method and bootstrapped with 1000 replicates. Evolutionary distances were calculated using the Poisson correction method and gaps were eliminated.

Quantification and Statistical Analysis

To quantify filaments of GFP-PhuZ or GFP-PhuZ(D202A), non-deconvolved DeltaVision image files were opened in FIJI version 2.1.0/1.53c using the Bio-formats importer. Images were automatically scaled at 15.456 pixels per μm and filaments were measured in the GFP channel at the slice containing the filament.

For the percentage of the infected population that had a rotating nucleus, an A/B test was performed and a p value < 0.05 was the threshold for statistical significance.

Measurement of the linear velocity of nucleus rotation was performed on differential interference contrast (DIC) images taken every 4 seconds. Dark protrusions on the surface of the phage nucleus were traced by hand over 12 second intervals (3 images) using the segmented line tool and their lengths recorded. Total distance traveled by points on the nuclear surface was divided by total time (120 seconds) and averaged with standard deviation calculated. Data is displayed as a violin plot produced in GraphPad Prism 9.2.0 for Mac OS X. An unpaired t-test was performed to obtain p values ($p < 0.05$, significant).

2.4 Results

E. coli jumbo phage Goslar forms a phage nucleus

E. coli phage vB_EcoM_Goslar (Goslar) was recently discovered and sequenced [16], revealing distantly related homologs of the PhuZ spindle protein and the major phage nucleus protein (Figures S2.1A and B). We studied the progression of Goslar through the lytic cycle in the *E. coli* K-12 lab strain MG1655 and compared basic infection morphology in the isolation host, APEC 2248 (APEC), to ensure our findings in the lab strain represented the replication pathway in the original host. We found that host chromosomal DNA was largely degraded by 60 minutes post-infection (mpi) as a dense ball of phage DNA accumulated, visible by DIC as well as DAPI staining (Figures S2.1C-E), similar to infection of *Pseudomonas* by nucleus-forming phages [6, 7]. To determine whether the Goslar DNA ball is surrounded by a proteinaceous shell, we tagged a homolog of the major phage nucleus shell protein, Goslar gp189, (21.52% amino acid identity with gp105 of 201φ2-1; Figure S2.1A), with GFP on the N-terminus and expressed it from a plasmid by induction with 0.2 mM IPTG. In uninfected cells, fluorescence from the GFP-shell fusion was diffuse (Figure S2.2A). In infected cells imaged without added dyes, GFP-shell formed a distinct halo around the DNA density observed by DIC (Figures 2.1A and S2.2C). We occasionally observed GFP puncta within the halo. When visualized with FM4-64 (cell membrane) and DAPI added to the imaging pad at 30, 60, and 90 mpi, GFP-shell surrounded the DNA (Figure 2.1D), suggesting that Goslar is a nucleus-forming phage.

We next tested whether the Goslar phage nucleus selectively imports proteins, establishing a nuclear compartment that is distinct from the cytoplasm. Using fluorescent protein fusions, we found that the Goslar-encoded protein gp193 of unknown function localizes with the Goslar phage nucleus. We co-expressed GFP-shell with gp193-mCherry and found that gp193-mCherry puncta

were either clearly inside the GFP-shell (28%), inside the shell but also associated with it (43%), or colocalized with the surface (29%) (n = 100; Figure 2.1B). Both fusions were diffuse when co-expressed in uninfected cells (Figure S2.2B). We observed the same association of gp193 with the phage DNA when gp193 was tagged with other fluorescent proteins (GFP, GFPmut1, mEGFP) on either terminus (Figures S2.2D and E). These results demonstrate that in 100% (n = 100) of infections, gp193 was associated with the phage nucleus and never elsewhere in the cytoplasm, showing that it is selectively localized with the phage DNA.

The phage nuclei of the *Pseudomonas* jumbo phages exclude ribosomes and metabolic proteins [3] so we examined the localization of the *E. coli* 50S ribosomal protein L20 (RplT) and thymidylate kinase (TMK) fused to GFP, as well as GFP alone as a generally soluble protein. RplT-GFP, TMK-GFPmut1, and GFP were all diffuse in uninfected cells (Figure S2.2F) and they were largely excluded from the phage nucleus during infection (Figure 2.1C), demonstrating that the phage nucleus forms a compartment that is distinct from the cell cytoplasm. These data suggest that the Goslar phage nucleus successfully segregates DNA from ribosomes and metabolic enzymes, thereby uncoupling transcription from translation.

In order to examine the ultrastructure and molecular organization of the phage nucleus, we performed cryo-focused ion beam milling coupled with cryo-electron tomography (cryo-FIB-ET) on Goslar-infected APEC cells at late stages of infections (~90 mpi). Despite the distant sequence homology of the shell (Figure S2.1A), the ultrastructure of the Goslar phage nucleus is strikingly similar to phage nuclei observed in the previously characterized *Pseudomonas* jumbo phages. As observed for those phage nuclei, the perimeter of the Goslar nucleus appears to be composed of a single layer of protein, though the estimation of the Goslar shell thickness is ~6 nm. In agreement with our fluorescence microscopy results, the Goslar nucleus separates phage DNA from the

cytosol and is devoid of host ribosomes (Figures 2.1E and F). Capsids containing various amounts of DNA, based on the density within the capsids, were docked on the surface of the phage nucleus shell, representing various stages of Goslar DNA packaging.

Goslar forms a vortex-like array of PhuZ filaments

Goslar encodes a divergent homolog (gp201) of the phage cytoskeletal protein PhuZ (26.96% amino acid identity with 201φ2-1 gp59; Figure S2.1B). To determine whether this divergent homolog forms a bipolar spindle that organizes phage replication similarly to the nucleus-forming *Pseudomonas* jumbo phages, we tagged the N-terminus of Goslar gp201 with GFP and visualized it with fluorescence microscopy *in vivo*. Since the polymerization of tubulins like PhuZ occurs spontaneously above a critical concentration of monomers, we chose a concentration of IPTG (0.2 mM) for induction of GFP-PhuZ that resulted in spontaneous filament formation in less than 0.5% of cells (Figures S2.3B and C), but strongly labeled PhuZ filaments in infected cells (Figures 2.2A and B). During infection by Goslar, GFP-PhuZ was incorporated into filaments that organized into a vortex-like array around the phage nucleus (Figures 2.2A, B, and D). The percentage of cells containing GFP-PhuZ filaments increased from 19% (n = 75) at 30 mpi to 97% (n = 114) at 60 mpi (Figure 2.2C). At 30 mpi, one or two long filaments (yellow) could be observed in some cells (Figure 2.2A). By 60 mpi, GFP-PhuZ filaments were arranged into a vortex wrapping around the phage nucleus (visualized by DAPI-staining and as a DIC density) and terminating at the membrane with some reaching the cell poles (Figure 2.2A). This vortex-like cytoskeletal structure remained assembled at both 90 mpi and 120 mpi (Figure 2.2B). In some cells where the nucleus was positioned in the center of the cell, the longest filaments were observed on either side of the nucleus resembling a bipolar spindle, however, they were accompanied by smaller filaments around the phage nucleus forming the vortex. When GFP-PhuZ

was expressed at lower induction levels, including 20 μM IPTG and 100 μM IPTG, the cytoskeletal vortex was still observed (Figure S2.3A). Even without any IPTG present, GFP signal from the fusion protein was most concentrated around the periphery of the phage DNA visualized by DIC, although the signal was very low. These results suggest that GFP-PhuZ filaments wrap around the phage nucleus and create a vortex-like cytoskeletal structure (Figure 2.2D) and that the vortex is not due to an artifact of overexpression.

To better visualize the spatial relationship between the PhuZ vortex and the phage nucleus, we simultaneously visualized the Goslar shell and PhuZ filaments by expressing GFP-shell and mCherry-PhuZ from the same plasmid with 0.2 mM IPTG. At 90 mpi, the mCherry-PhuZ filaments wrapped around the GFP-shell and protruded towards the membrane in all directions (Figure 2.3A). To determine whether the choice of fluorescent protein fusion affected the co-localization, we also imaged mCherry-shell with GFP-PhuZ and found a similar organization (Figure 2.3B). In time course experiments, we found that at 30 mpi the shell of the phage nucleus had grown to the point that it was clearly distinguishable as a compartment containing DAPI-stained DNA before PhuZ filaments assembled a visible vortex around it (Figure 2.3C). By 60 mpi and beyond, the filaments colocalized with the phage nucleus surface and assembled a cytoskeletal vortex. This contrasts with the bipolar spindle assembled by nucleus-forming *Pseudomonas* phages, where filaments emanate from each pole of the cell and extend toward the phage nucleus at the cell center, providing the pushing forces necessary for both positioning and rotation (see model in Figure 2.5D) [3, 4, 6, 7]. Thus, the vortex-like organization of PhuZ filaments during Goslar infections represents a novel type of cytoskeletal structure found in prokaryotes.

The Goslar nucleus is not positioned at midcell

The bipolar PhuZ spindle of *Pseudomonas* jumbo phages uses dynamic instability to center the phage nucleus [6]. However, we measured the position of the Goslar phage nucleus and found that, unlike nucleus-forming *Pseudomonas* phages, the Goslar nucleus was not specifically positioned at any one location along the cell length, in neither MG1655 nor APEC (Figures 2.4A and C). In fact, the likelihood of finding a Goslar nucleus near the midcell was not significantly greater than finding it elsewhere in the cell. The Goslar nucleus grew in size over time as DNA was replicated inside, ultimately reaching an average 2D area of ~4 μm^2 (Figures 2.4B and D). These data demonstrate that the Goslar-encoded distantly related PhuZ protein does not position the growing phage nucleus at midcell.

Goslar nucleus rotation is dependent on vortex orientation and PhuZ function

Midway through the infection cycle, the phage nucleus of *Pseudomonas* phages 201φ2-1, ΦKZ, and ΦPA3 begins to rotate due to the opposing forces of the PhuZ spindle (see model in Figure 2.5D) [3, 14]. Rotation is coupled to the delivery of capsids to the surface of the phage nucleus and disrupting PhuZ filament dynamics results in the production of fewer phage virions [7]. Rotation is therefore an important and conserved aspect of the nucleus-forming jumbo phage replication cycle. We collected time-lapse images of the GFP-tagged shell every 4 seconds and found that the Goslar phage nucleus also rotates (Figure 2.5A). We also used DIC imaging to visualize and quantify nucleus rotation without the potential interference of any fluorescent stains or protein fusions (Figure 2.5B). The direction of phage nucleus rotation occurred along any plane relative to the long axis of the cell and sometimes appeared to reverse direction. We measured the rate of rotation using the distance a point traveled along the perimeter of the phage nucleus, when that point appeared to be traveling along the focal plane. Reporting this rate in linear velocity

allows for comparison to the rate of PhuZ polymerization and nucleus rotation reported for the *Pseudomonas* phages [14]. We found an average linear velocity of 50 ± 13 nm/s (measured at the nucleus periphery; n = 20; Figure 2.5C). This average rate of movement is nearly identical to the rates of nucleus rotation and PhuZ filament growth for the nucleus-forming *Pseudomonas* phages [14].

We hypothesized a model of nucleus rotation for Goslar in which the PhuZ cytoskeletal vortex provides pushing forces tangential to the surface of the phage nucleus (Figure 2.5D). This model predicts that the direction of nucleus rotation should be correlated with the orientation of filaments within the vortex, and that mutations that disrupt PhuZ filament dynamics by inactivating GTP hydrolysis will reduce the rate of rotation. To test the first prediction, we examined the direction of nucleus rotation relative to the orientation of the filaments. Nucleus rotation was observed using DIC time-lapse at 60 mpi and 90 mpi in cells induced with 0.2 mM IPTG to express GFP-PhuZ, and the PhuZ cytoskeleton was observed at a single time point just prior to the time-lapse. As shown in Figure 2.5E, the nucleus in the top panels rotated counterclockwise while the nucleus in the bottom panels rotated clockwise, and in both cases the cytoskeletal filaments were arranged in such a way that filament elongation would drive rotation of the phage nucleus in the corresponding direction. These results are consistent with a model in which the vortex of filaments drives rotation of the nucleus by pushing against the cell membrane (Figure 2.5D).

In order to further test this model, we mutated a conserved PhuZ aspartic acid residue, D202, to an alanine to disrupt the putative site of GTP hydrolysis. Expression of this mutant protein PhuZ(D202A) in uninfected cells yielded filaments in 56% of cells during induction with only 0.02 mM IPTG (Figures S2.4A and B), whereas filaments are only observed for wild-type PhuZ above 0.3 mM IPTG (Figure S2.3C). The ability of the mutant to assemble visible filaments

when expressed at a much lower level compared to wild-type PhuZ suggests that the D202A mutation disrupts PhuZ polymer dynamics by inhibiting GTP hydrolysis. Based on our prior studies creating similar mutations in phage and plasmid-encoded tubulins [6, 7, 33], we expected GFP-PhuZ(D202A) to behave as a dominant-negative mutant that could co-assemble with wild-type PhuZ to form inactive polymers. During Goslar infection in the presence of GFP-PhuZ(D202A) induced with 0.2 mM IPTG, the mutant protein failed to form a vortex (Figure 2.6A), yet phage nuclei positioning and growth appeared to be unaffected, compared to GFP-PhuZ (Figures S2.5A-E). However, in time-lapse microscopy, the intracellular rotation of phage nuclei was greatly reduced (Figure 2.6B). We measured the rates of rotation in cells expressing GFP-tagged PhuZ for comparison to the previous measurements in cells with no fusion proteins, to ensure there are no drastic effects on rotation due to GFP-tagging. In the presence of wild-type GFP-PhuZ, 92% (n = 265; Figure 2.6C) of nuclei rotated visibly by DIC with an average linear velocity of 44 ± 9 nm/s (n = 20; Figures 2.6D and S2.4C and D, black dotted lines), whereas with the expression of GFP-PhuZ(D202A), only 29% (n = 101; Figure 2.6C) of nuclei appeared to rotate and for those 29%, they had an average linear velocity of 18 ± 9 nm/s (n = 23; Figure 2.6D; Figure S2.4D, magenta dotted line), significantly slower than the wild-type. As a control, nucleus rotation in the presence of wild-type GFP-PhuZ (92% rotated, 44 ± 9 nm/s) behaved similarly in the absence of any fusion proteins (97% rotated, n = 105, Figure 2.6C; 50 ± 13 nm/s, n = 20, Figure 2.6D; Figure S2.4C, red dotted line). The disruption of function by expression of PhuZ(D202A) is likely not complete in some cells due to the presence of wild-type PhuZ expressed by the phage. These results are consistent with a model in which the PhuZ vortex is required for Goslar nucleus rotation.

Goslar forms bouquets of mature virions

As the PhuZ cytoskeleton of nucleus-forming *Pseudomonas* phages rotates the phage nucleus, it simultaneously delivers capsids to the surface of the nucleus for DNA packaging. Once filled with DNA, maturing phage particles assemble structures we recently termed "phage bouquets" [15]. We therefore examined if Goslar also forms phage bouquets late in the infection cycle. We tagged the putative Goslar capsid protein, gp41, with GFP at its C-terminus to determine whether Goslar virions also assemble bouquets. At 90 mpi, several stages of capsid organization were visible. In some cells, capsid-GFP was found around the nucleus, where the capsids were likely docked and in the process of packaging phage DNA (Figures 2.7A and S2.6A, left panels). We also observed the formation of spherical structures of capsid protein adjacent to the phage nucleus, or "bouquets", with capsids also appearing within the centers of the bouquets (Figures 2.7A and S2.6A, middle panels, white arrowheads). Since *Pseudomonas* phage bouquets have not been shown to have capsids localized to the interior of the bouquet [15], this could be a newly observed type of bouquet organization. Finally, in very swollen cells, capsids accumulated in the cytoplasm without any obvious organization (Figures 2.7A and S2.6A, right panels).

To demonstrate whether capsids within these spherical structures are filled with DNA, we stained Goslar infections with 10 μg/ml DAPI at 90 mpi because the addition of DAPI prior to infection halted phage replication in MG1655 (Figure 2.7B). We observed a spherical pattern of faint DAPI staining, similar to the capsid-GFP localization and consistent with DNA-filled capsids arranged in a bouquet (white arrows). Curiously, no DAPI staining was visible in the interior of the bouquets, despite observing capsid-GFP localized in this region. We also visualized bouquet formation during Goslar infections of the APEC strain, which could be grown in 200 ng/ml DAPI without inhibiting phage replication. The presence of DAPI throughout the entire infection

allowed capsids to be stained prior to organizing into the bouquets. This resulted in very bright staining of spherically shaped DAPI-stained phage bouquets that also contained DAPI staining within the interior of the bouquet (Figure 2.7C), as predicted by capsid protein localization (Figure 2.7A). Cells with a single phage nucleus typically contained one or two bouquets that could be found on either one side of the phage nucleus or opposite sides of the phage nucleus (Figure 2.7C). An example of a cell with two phage nuclei and three phage bouquets is shown in Figure 2.7C with more examples in Figure S2.6B. The length and width of 105 bouquets at a median focal plane is displayed as a scatter plot in Figure S2.6C. Phage bouquets could become larger than the nucleus, reaching over 2 μm in width and over 5 μm in length, with a median maximum dimension of 1.5 ± 0.75 μm (Figure S2.6D).

These observations suggest that Goslar capsids are organized into a sphere with the tails packed together as observed in the *Pseudomonas* phages [15], but also deforming to fill the cell and containing capsids within the bouquet, not previously seen in the *Pseudomonas* phages (Figure 2.7D). To further investigate this model and better understand the stages of capsid localization, we simultaneously visualized capsid-GFP and DAPI staining at various times post infection (30, 40, 50, 70, and 90 minutes) (Figure 2.7E). We were not able to express capsid-GFP in APEC so we colocalized capsid-GFP with DAPI in MG1655. At early time points (30 mpi), the capsid fusion was devoid of DAPI staining and was found near the membrane and in the cytosol, or concentrated around the nucleus (Figure 2.7E). By 40 mpi, small circular bouquets stained with DAPI were found next to the DAPI-filled nucleus (Figure 2.7E, white arrow). Between 40 and 50 mpi, the small bouquets expanded and acquired internal capsids but only the exterior ring of capsids was stained with DAPI in MG1655. This time course resembles the trafficking of capsids from the membrane to the nuclear shell as described for the nucleus-forming *Pseudomonas* phages

[14] and the assembly of "bouquets" of DNA-filled capsids in those phages [15]. Time-lapse microscopy of the capsid fusion without any dyes present confirmed the temporal progression of capsid localization from membrane or cytosol, to the periphery of the nucleus, and then to the adjacent bouquet(s) (Figure S2.6E).

Cryo-FIB-ET confirms that fully packaged capsids are arranged in a circular shape, forming bouquets (Figures 2.7F and G). Consistent with our fluorescence microscopy data, the interior of the observed bouquet is greatly depleted of host ribosomes compared to the surrounding cytosol. In the observed bouquet the majority of the tails point to the inside of the bouquet and the capsids to the exterior, but a capsid localized inside the bouquet is also observed (Figures 2.7F and G). Given the sizes of the bouquets (> 1.5 μm) compared to the thickness of FIB-milled lamellae (< 200 nm), we were unable to capture a bouquet in its entirety and ascertain its complete ultrastructure from our current cryo-FIB-ET dataset. Nevertheless, when combined with our fluorescence microscopy results, our cryo-FIB-ET data support the assembly of Goslar virions into double-layered bouquets at late stages of infection. Taken together, these data reveal that Goslar forms a distinctive type of prominent bouquet containing DNA-packed capsids within its center.

2.5 Discussion

Nucleus-forming jumbo phages rely on a tubulin-based cytoskeleton for their complex subcellular organization. In *Pseudomonas* jumbo phages 201φ2-1, ΦPA3, and ΦKZ, PhuZ forms a dynamic bipolar spindle that positions the phage nucleus at midcell and delivers capsids to the shell of the phage nucleus while rotating it [3, 4, 6, 7, 14]. Phage nucleus rotation occurs concomitantly with the delivery of newly assembled capsids to the surface of the nucleus [3, 14]. We have proposed that this conserved feature of nucleus-forming jumbo phage replication allows capsids to be distributed uniformly around the nucleus surface for efficient DNA packaging into

capsids [14]. Intracellular rotation in prokaryotes has only been reported for *Pseudomonas* infected by nucleus-forming jumbo phages, where a bipolar array of PhuZ filaments provides the forces necessary for rotation. Here we characterize Goslar infecting *E. coli* and demonstrate that this phage also has a complex replication cycle with both a nucleus-like compartment and a tubulin-based cytoskeleton that rotates it (Figure 2.8).

Surprisingly, Goslar PhuZ filaments assemble a vortex-like cytoskeletal structure in which filaments wrap around the entire phage nucleus and project radially toward the membrane (Figure 2.8). The orientation of the cytoskeletal vortex relative to the direction of rotation is consistent with the hypothesis that it generates force against the membrane to cause nucleus rotation, and expression of the PhuZ(D202A) mutant impairs both vortex formation and nucleus rotation. How can radially arranged filaments lead to intracellular rotation? If filaments emanate from the surface of the phage nucleus and are arranged radially around it, presumably no motion would occur if the forces they generated were perfectly balanced. A stochastic imbalance of forces is more likely to occur and this would lead to rotation in some direction. The process of rotation and consequential flow of cytoplasm can impose the alignment of filaments with the current direction of rotation, reinforcing the vortex array, as demonstrated *in vitro* [34, 35].

Taken together, these results suggest that the function of the vortex-like cytoskeletal network of PhuZ filaments is to drive phage nucleus rotation. Previous work on phage 201φ2-1 has demonstrated that inactivating the PhuZ cytoskeleton eliminates both nucleus rotation and capsid trafficking thereby reducing the number of phage progeny produced by approximately 50% [7, 14]. In Goslar, the PhuZ cytoskeleton drives intracellular phage nucleus rotation, which is expected to facilitate capsid distribution. The PhuZ vortex may also participate in capsid trafficking, and radially projecting filaments would be ideal conduits for capsid migration from

the cell membrane to the surface of the phage nucleus, but further investigation must be done to determine whether Goslar PhuZ plays a role in capsid trafficking.

A cytoskeletal vortex has not previously been described within any prokaryotes as far as we know. In eukaryotes, a vortex of cytoplasmic motion dependent on microtubules and actin has been shown to be critical for development by driving cytoplasmic streaming in the eggs of several species as well as in plant cells [36-39]. The eukaryotic cytoskeleton is responsible for the rotation of the nucleus to align with cell polarity in motile cells, particularly fibroblasts [40-46] and has been observed in a vortex-like arrangement during misalignment [43]. A self-organizing vortex array of microtubules was demonstrated to spontaneously arise *in vitro* among purified microtubules with *Chlamydomonas* dynein c [34] and in spatially confined droplets of *Xenopus* oocyte extract [35]. The vortex of PhuZ filaments that organizes during Goslar infection visually resembles these *in vitro* microtubule vortex arrays.

While a cytoskeletal vortex is capable of providing intracellular rotation, it did not appear to play a role in phage nucleus positioning. The Goslar phage nucleus is not positioned at the midcell as it is for the nucleus-forming *Pseudomonas* phages. Instead, it is positioned along the entire lateral axis of the cell. Expression of the PhuZ(D202A) mutant, which significantly reduces nucleus rotation, does not affect nucleus positioning. Thus, the main function of the Goslar PhuZ cytoskeleton may be to provide the driving forces for nuclear rotation and perhaps capsid delivery, rather than midcell positioning. It is not presently clear why some phages position the nucleus at the midcell and Goslar doesn't. The midcell positioning in *Pseudomonas* phages may be a consequence of a bipolar spindle that facilitates rotation and it would have the added benefit of maintaining space between the surface of the phage nucleus and the cell membrane to allow the

docking of capsids. In Goslar, with a vortex array instead of a bipolar spindle, the benefit of keeping space between the nucleus and the membrane is provided without midcell positioning.

At late stages of infection, Goslar virus particles accumulate in large spherical structures similar to the ones recently described for the nucleus-forming *Pseudomonas* phages and termed bouquets [15]. In phage bouquets, virions are arranged in a sphere with capsids on the outside and tails facing inward. Goslar bouquets appeared similar to ΦPA3 bouquets except Goslar's contained capsids located in the center of the phage bouquets rather than only around the outside, suggesting that Goslar bouquets are organized with internal capsids oriented inversely relative to the larger outer layer, with tails packed together (Figures 2.8D and 2.7D). The role of phage bouquets is currently unknown, and the fact that they are not always detected in phages ΦPA3 and ΦKZ, and are only rarely detected in phage 201φ2-1, suggests that they are not essential for phage replication [15]. However, the discovery of prominent phage bouquets in Goslar, which is distantly related to the nucleus-forming *Pseudomonas* jumbo phages, suggests that bouquets likely offer an advantage to the phage.

The detailed characterization of the replication cycle of Goslar has brought to light a cytoskeletal vortex that drives phage nucleus rotation which is a key process in the nucleus-forming jumbo phage replication cycle. This work also demonstrates that the replication mechanism involving a phage nucleus and spindle is widespread across diverse hosts and that it likely confers a selective advantage since divergent strategies for rotation of the phage nucleus have evolved.

2.6 Acknowledgements

This work was supported by the National Institutes of Health R01-GM129245 (to JP and EV) and R21-AI148814 (to KDC) and the National Science Foundation MRI grant NSF DBI

1920374 (to EV). EV is an investigator of the Howard Hughes Medical Institute. We acknowledge the use of the UC San Diego cryo-EM facility, which was built and equipped with funds from UC San Diego and an initial gift from the Agouron Institute. We thank Arshad Desai, Justin R. Meyer, and Amy M. Prichard for their helpful suggestions and comments on the manuscript.

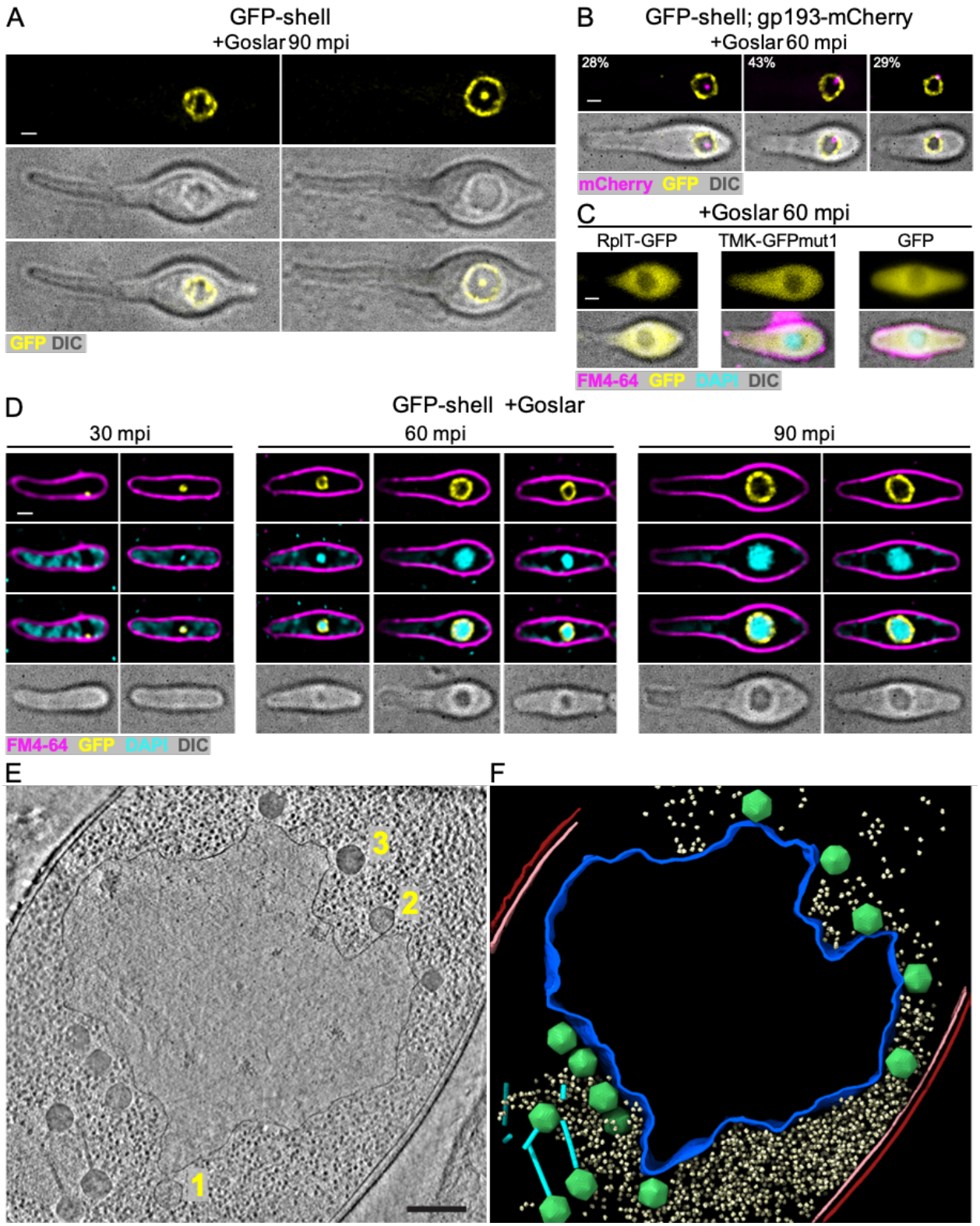
Chapter 2, in entirety, is a reprint of the material in review as Birkholz, EA., Laughlin, TG., Suslov, S., Armbruster, E., Lee, J., Wittmann, J., Corbett, KD., Villa, E., & Pogliano, J., 2022, A cytoskeletal vortex drives phage nucleus rotation during jumbo phage replication in *E. coli* in review with Cell Reports. The dissertation author was the primary investigator and author of this material.

Author Contributions

E.A.B., T.G.L., S.S., E.A., and J.L. conducted experiments. E.A.B. and T.G.L. analyzed data. E.A.B. and J.P. conceptualized the original manuscript. J.W. provided the phage for study. E.A.B., T.G.L., E.A., J.W., K.D.C., E.V., and J.P. contributed to editing the manuscript.

2.7 Figures

Figure 2.1 Goslar builds a phage nucleus separating DNA from translation and metabolism. (A) Goslar infecting *E. coli* (MG1655) expressing GFP-shell (gp189, yellow) at 0.2 mM IPTG and infected by Goslar for 90 minutes (mpi). White scale bars are 1 μm. (B) *E. coli* co-expressing GFP-shell and gp193-mCherry at 0.2 mM IPTG, infected by Goslar for 60 minutes. Displayed in the upper left is the percentage of the infected population that has gp193-mCherry localized clearly to the interior of the GFP-shell halo (left), localized to the interior of the shell but on the periphery (center), and localized to the shell (right) (n = 100). 100% of the time, gp193-mCherry is colocalized with the phage nucleus. (C) *E. coli* expressing GFP fusions (yellow) with 50S ribosomal protein L20 (RplT), thymidylate kinase (TMK), and soluble GFP, and infected with Goslar for 60 minutes before being dyed with FM4-64 (4 μg/ml, membrane, magenta) and DAPI (2 μg/ml, DNA, cyan). (D) *E. coli* expressing GFP-shell, infected by Goslar for 30, 60, or 90 minutes then dyed with FM4-64 and DAPI. (E) Slice through a deconvolved tomogram of a phage nucleus in a 90 mpi Goslar-infected APEC cell. Inset scale bar is 250 nm (1 - empty capsid, 2 - partially filled capsid, 3 - nearly full capsid). (F) Annotation of the tomogram shown in (E). Outer and inner host cell membranes are colored red and pink, respectively. The phage nucleus shell is colored blue. Goslar capsids and tails are colored green and cyan, respectively. Host 70S ribosomes are colored pale yellow.


A
GFP-shell
+Goslar 90 mpi
GFP DIC
B
GFP-shell; gp193-mCherry
+Goslar 60 mpi
28%
43%
29%
mCherry GFP DIC
C
+Goslar 60 mpi
RplT-GFP
TMK-GFPmut1
GFP
FM4-64 GFP DAPI DIC
D
GFP-shell +Goslar
30 mpi
60 mpi
90 mpi
FM4-64 GFP DAPI DIC
E
1
2
3
F

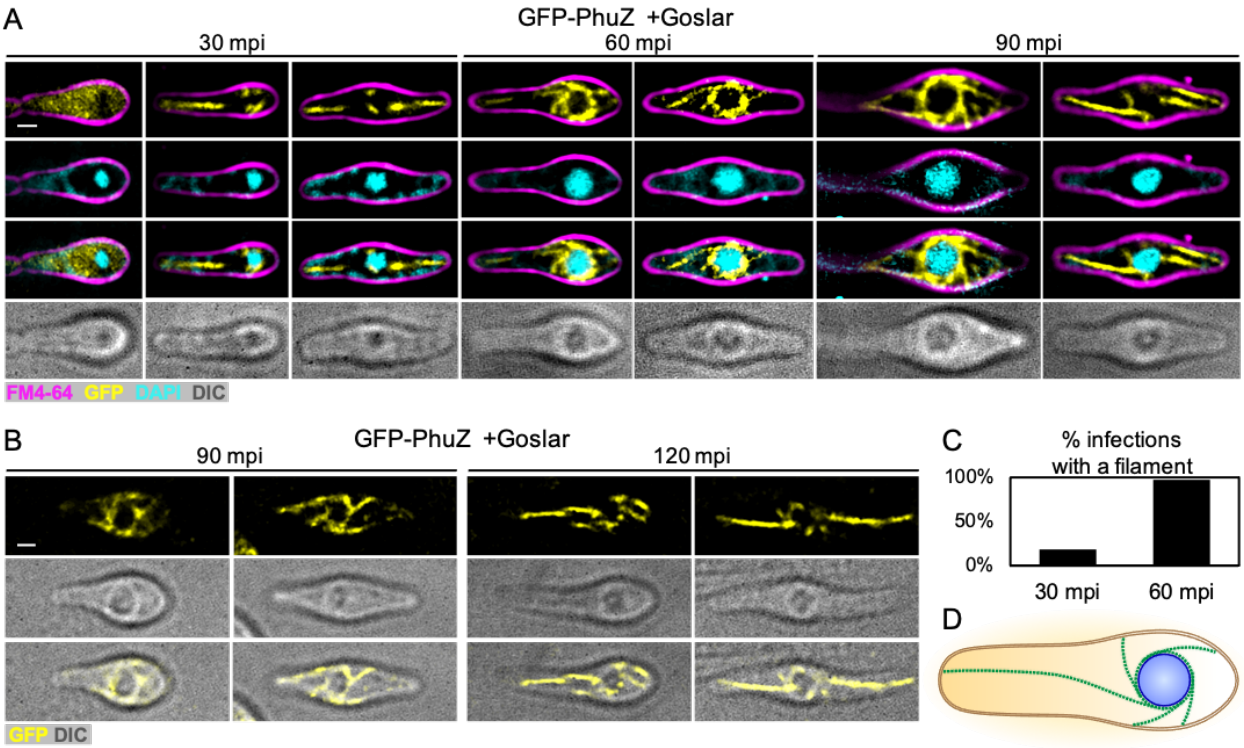


Figure 2.2 Goslar PhuZ forms a vortex-like cytoskeletal array. (A) *E. coli* (MG1655) expressing GFP-PhuZ at 0.2 mM IPTG and infected with Goslar for 30, 60, or 90 minutes prior to being dyed with FM4-64 and DAPI for 5-10 minutes. All scale bars are 1 μm. (B) *E. coli* (MG1655) expressing GFP-PhuZ and infected by Goslar for 90 and 120 minutes. (C) Percentage of 30 mpi (n = 75) or 60 mpi (n = 114) infections with a GFP-PhuZ filament over 0.3 μm. (D) Model of the PhuZ cytoskeletal vortex. PhuZ filaments (green) extend radially from the phage nucleus (blue) to the cell membrane (gold).

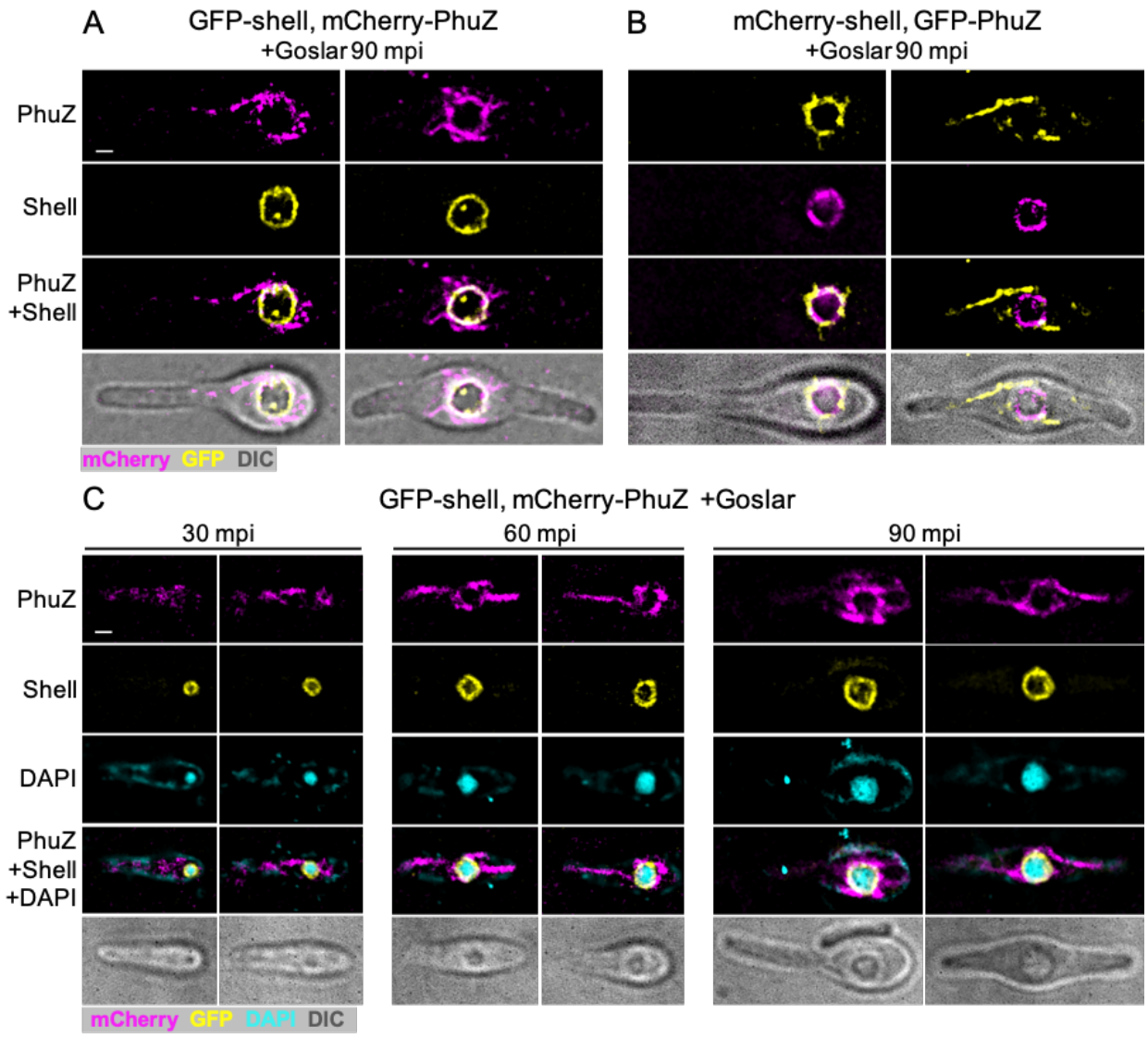


Figure 2.3 Colocalization experiments show the PhuZ cytoskeletal vortex wraps around the proteinaceous phage nucleus. (A) *E. coli* (MG1655) co-expressing GFP-shell (yellow) and mCherry-PhuZ (magenta) and infected with Goslar for 90 minutes. All scale bars are 1 μm. (B) *E. coli* co-expressing mCherry-shell and GFP-PhuZ and infected with Goslar for 90 minutes. (C) *E. coli* co-expressing GFP-shell and mCherry-PhuZ and infected with Goslar for 30, 60, or 90 minutes then dyed with DAPI (cyan).

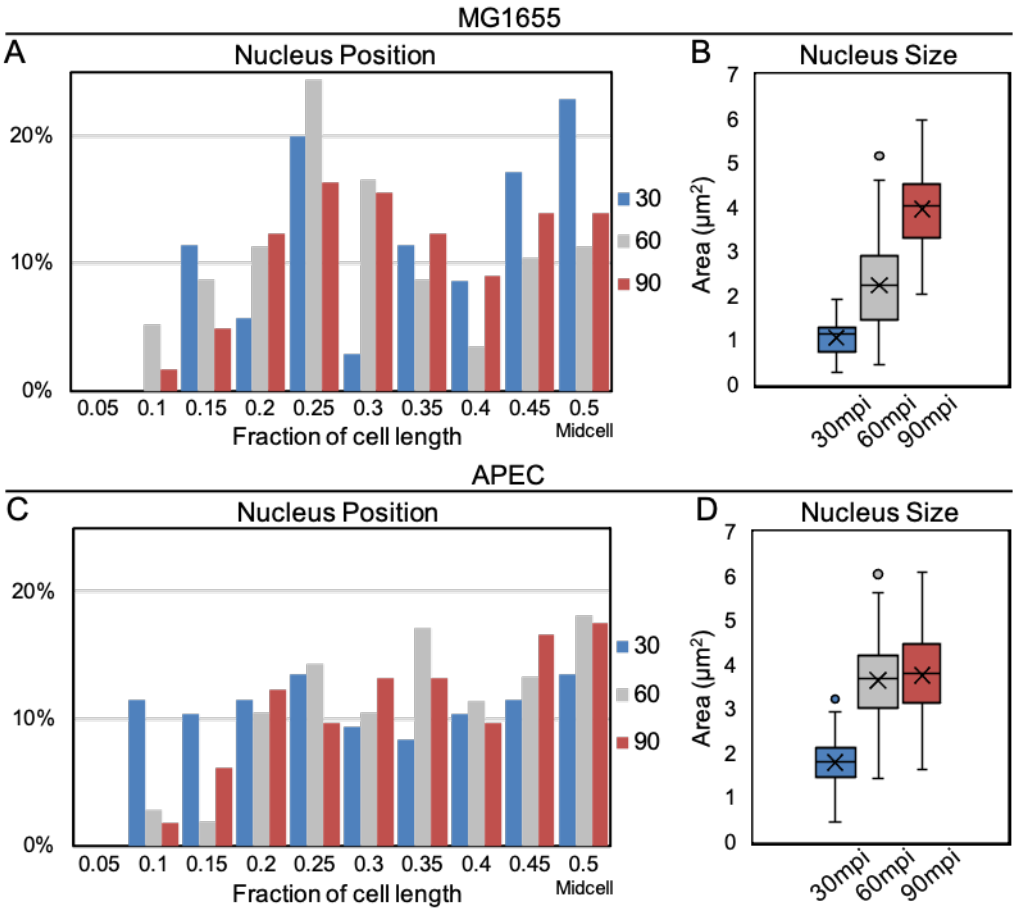
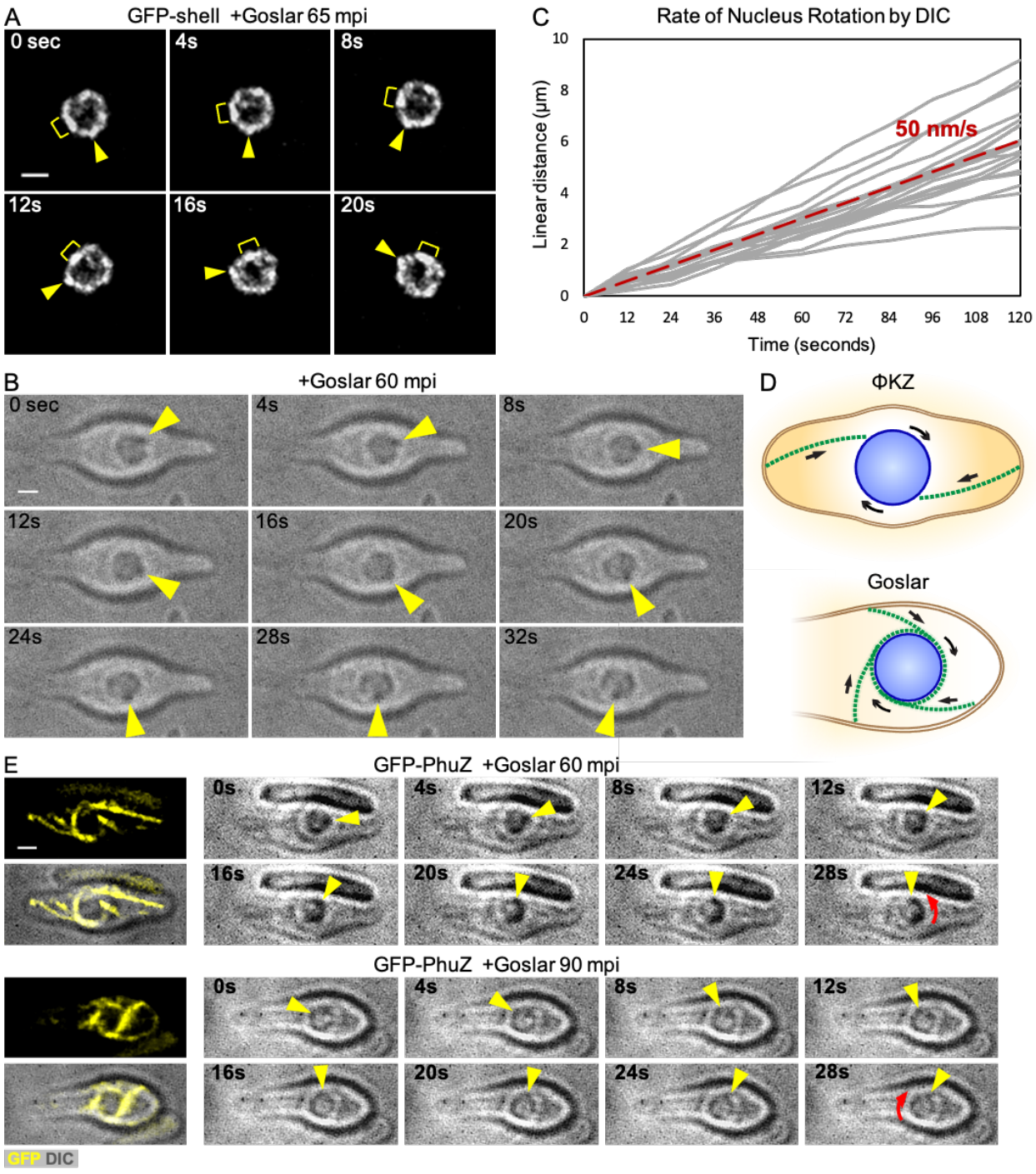


Figure 2.4 The Goslar nucleus is not positioned at midcell and is excluded from the cell pole. (A) Distribution of DIC phage nuclei positions along the lateral length of the cell (MG1655), in 0.05 fraction of cell length bins. For each time point, there is no significantly greater chance of finding a phage nucleus near midcell (0.5) or quarter cell (0.25) than in the neighboring bins (30 mpi, n = 35; 60 mpi, n = 115; 90 mpi, n = 122). (B) 2D area of the DAPI-stained phage nucleus in MG1655 at 30, 60, and 90 mpi (30 mpi; n = 50, 60 mpi; n = 114, 90 mpi; n = 121). (C) Distribution of DIC phage nuclei positions along the lateral length of the APEC cell, in 0.05 fraction of cell length bins. No significant enrichment occurs at midcell or any other bins (30 mpi; n = 96, 60 mpi; n = 105, 90 mpi; n = 114). (D) 2D area of the DAPI-stained phage nucleus in APEC at 30, 60, and 90 mpi (30 mpi; n = 115, 60 mpi; n = 120, 90 mpi; n = 145).

Figure 2.5 The Goslar phage nucleus rotates and the PhuZ vortex pushes against the cell membrane. (A) Time-lapse of the phage nucleus every 4 seconds for 20 seconds in *E. coli* (MG1655) expressing GFP-shell (white) infected with Goslar for 65 minutes. Yellow arrowhead and bracket indicate distinctions for following rotation. All scale bars are 1 μm. (B) DIC time-lapse every 4 seconds for 20 seconds in *E. coli* infected with Goslar for 60 minutes. Yellow arrowhead indicates dark spot to follow rotation. (C) Linear velocity of nucleus rotation measured from DIC time-lapse, averaging 50 nm/s (n = 20), red dotted line, individual measurements shown as gray lines. (D) Model of ΦKZ phage nucleus rotation by bipolar PhuZ spindle (top) and Goslar phage nucleus rotation by PhuZ cytoskeletal vortex (bottom). Arrows indicate the direction of forces applied to the phage nucleus and direction of rotation. (E) GFP imaging coupled with DIC time-lapse every 4 seconds on *E. coli* expressing GFP-PhuZ; yellow arrowhead indicates DIC-dense spot to follow for rotation, red curved arrows in final panels indicate direction of rotation (CCW top cell, CW bottom cell).


A
GFP-shell +Goslar 65 mpi
0 sec
4s
8s
12s
16s
20s
C
Rate of Nucleus Rotation by DIC
Linear distance (µm)
Time (seconds)
50 nm/s
B
+Goslar 60 mpi
0 sec
4s
8s
12s
16s
20s
24s
28s
32s
D
ΦKZ
Goslar
E
GFP-PhuZ +Goslar 60 mpi
0s
4s
8s
12s
16s
20s
24s
28s
GFP-PhuZ +Goslar 90 mpi
GFP DIC

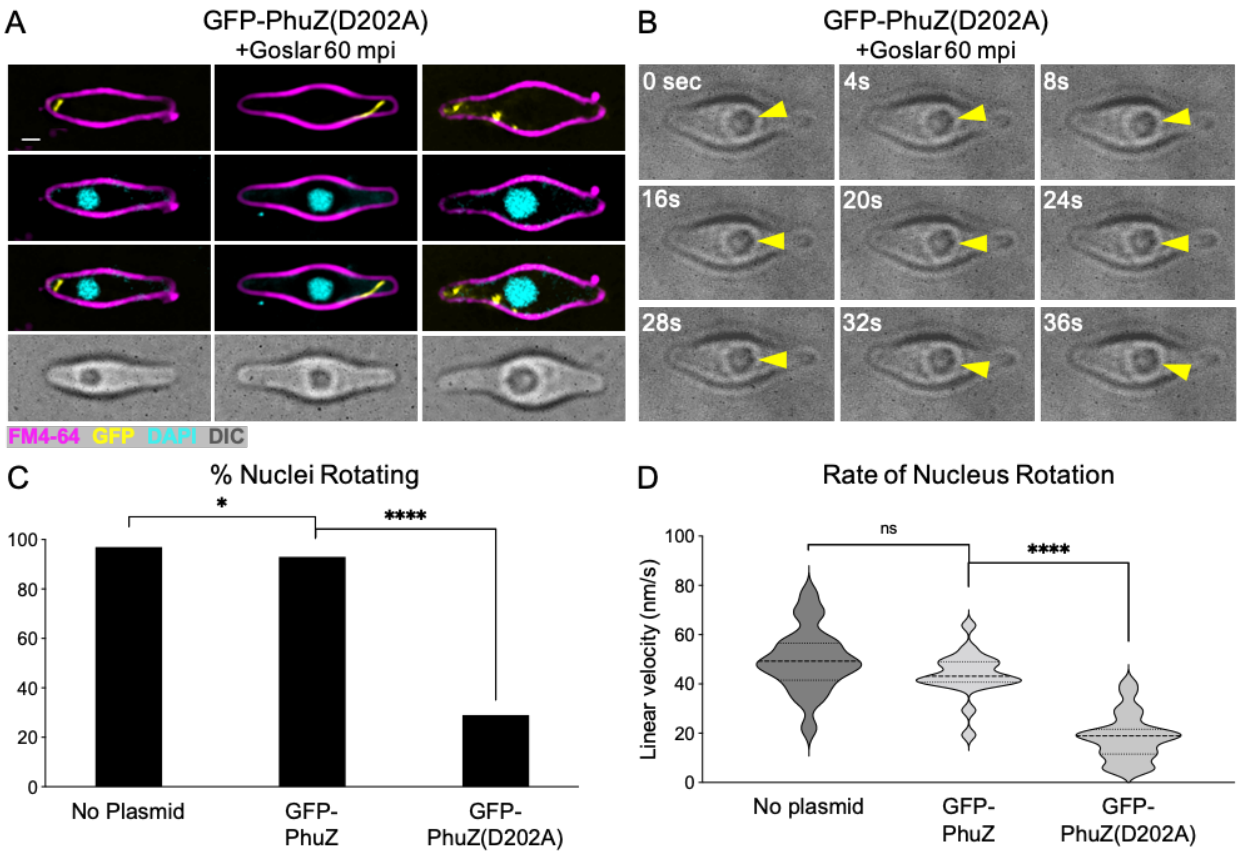
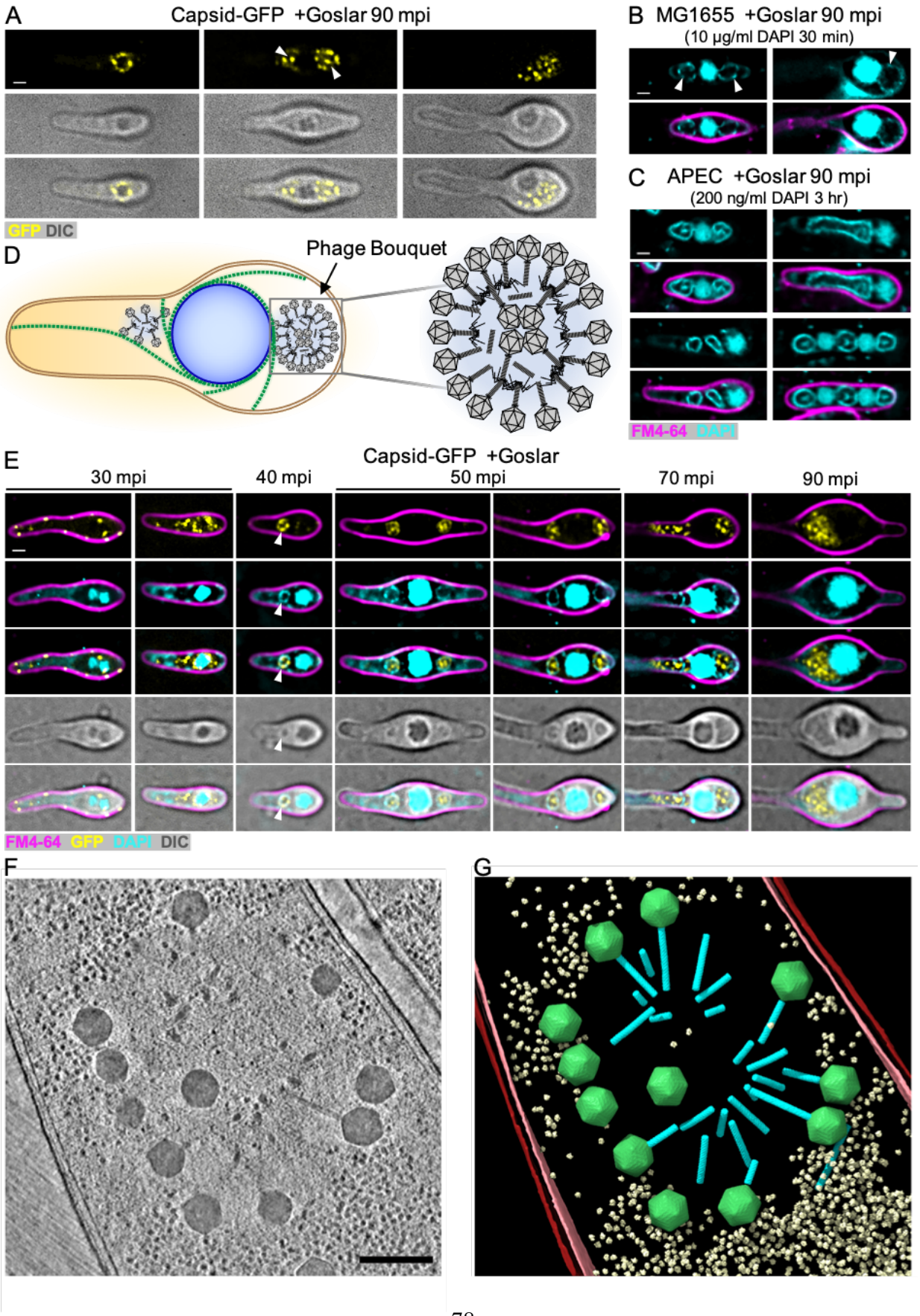


Figure 2.6 Mutant PhuZ(D202A) disrupts filament formation and nucleus rotation. (A) *E. coli* expressing GFP-PhuZ(D202A) at 0.2 mM IPTG and infected by Goslar for 60 minutes before being stained with FM4-64 and DAPI. All scale bars are 1 μm. (B) DIC time-lapse every 4 seconds for 36 seconds on *E. coli* expressing GFP-PhuZ(D202A) at 0.2 mM IPTG and infected by Goslar for 60 minutes. Yellow arrowhead tracks dark spot for rotation. (C) Percentage of the infected cells at 60 mpi that had a rotating nucleus with any amount of progressive movement imaged by DIC time-lapse for *E. coli* with no plasmid (MG1655, n = 105) or cells expressing GFP-PhuZ (n = 164) or GFP-PhuZ(D202A) (n = 101) (*unpaired t-test*, * p value = 0.04, **** p value < 0.0001). (D) Linear velocity of the most progressively rotating nuclei at 60 mpi by DIC time-lapse for *E. coli* with no plasmid (MG1655, n = 20) or cells expressing GFP-PhuZ (n = 20) or GFP-PhuZ(D202A) (n = 23). Violin plot generated and unpaired t tests performed using GraphPad Prism 9 (*unpaired t-test*, ns - p value > 0.05, **** p value < 0.0001).

Figure 2.7 Goslar capsids migrate from the cytoplasm, surround the phage nucleus, and form phage bouquets. (A) *E. coli* (MG1655) expressing the putative capsid protein (gp41) fused to GFP (yellow) induced with 0.2 mM IPTG and infected with Goslar for 90 minutes. Three predominant localizations can be found; around the phage nucleus (left panels), in adjacent bouquets (middle panels), and filling more of the cytoplasm (right panels). White scale bars are 1 μm. (B) MG1655 infected with Goslar for 90 minutes then stained with 10 μg/ml DAPI for 30 minutes at room temperature. White arrowheads indicate faint bouquets. (C) APEC grown with 200 ng/ml DAPI for 90 minutes and then infected with Goslar for 90 minutes and imaged. Large, brightly fluorescent phage bouquets are formed. (D) Model of Goslar phage bouquet organization with tails packed together. (E) MG1655 expressing capsid-GFP (yellow) and infected with Goslar for 30, 40, 50, 60, 70, or 90 minutes before being stained with FM4-64 (magenta) and DAPI (cyan). (F) Slice through a deconvolved tomogram of a phage bouquet in a Goslar-infected APEC cell. Inset scale bar is 250 nm. (G) Annotation of the tomogram shown in (F). Outer and inner host cell membranes are colored red and pink, respectively. The phage nucleus shell is colored blue. Goslar capsids and tails are colored green and cyan, respectively. Host 70S ribosomes are colored pale yellow.


A
Capsid-GFP +Goslar 90 mpi
GFP DIC
B MG1655 +Goslar 90 mpi
(10 µg/ml DAPI 30 min)
C APEC +Goslar 90 mpi
(200 ng/ml DAPI 3 hr)
FM4-64 DAPI
D
Phage Bouquet
E
Capsid-GFP +Goslar
30 mpi
40 mpi
50 mpi
70 mpi
90 mpi
FM4-64 GFP DAPI DIC
F
G

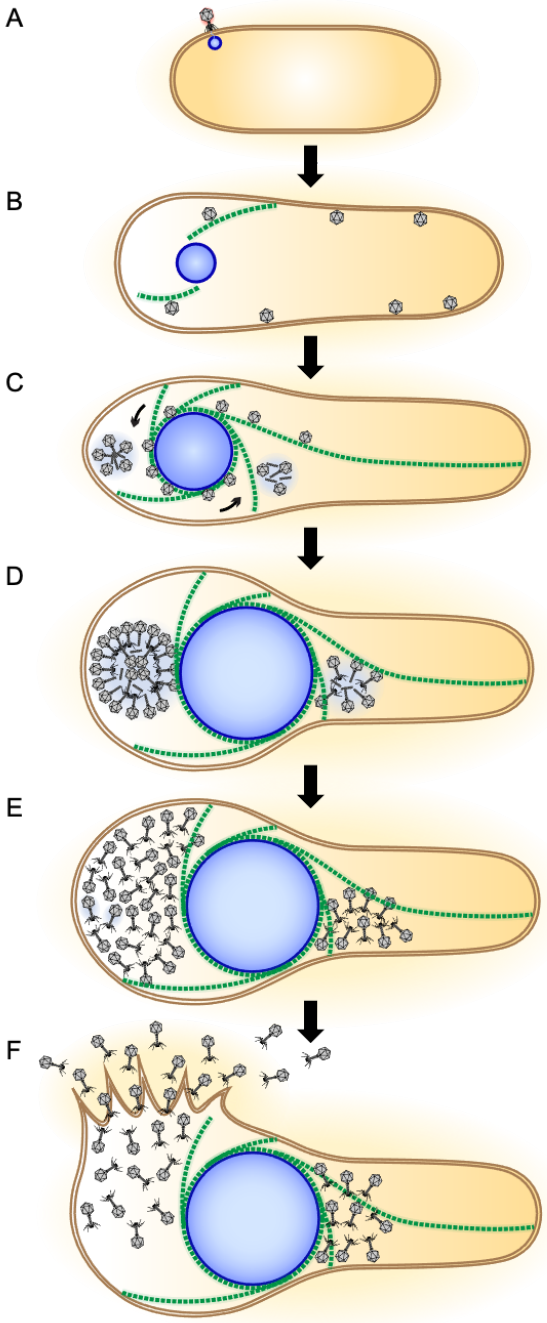


Figure 2.8 Model of the Goslar infection cycle. (A) The Goslar phage injects its DNA into an *E. coli* cell and the formation of a shell begins. (B) The shell grows in size as DNA replicates inside and the PhuZ vortex begins to form. Capsids form near the periphery of the cell and migrate towards the phage nucleus, possibly by trafficking along PhuZ filaments as we have demonstrated for *Pseudomonas* phages [14]. (C) The PhuZ vortex is fully formed, wrapping around and rotating the phage nucleus. Capsids dock on the nuclear shell to be filled with DNA prior to localizing to the adjacent bouquets. (D) Large bouquets form with internally localized capsids on either one side or both sides of the phage nucleus. (E) Final assembly of the progeny virions is completed as they fill the cell in a more disordered fashion. (F) Lysis of the *E. coli* cell is achieved, releasing the progeny virions to find the next host.

Figure S2.1 Goslar encodes distant homologs of the major phage nucleus shell protein and tubulin-like PhuZ protein and forms a dynamic DNA density during infection. (A & B) Unrooted phylogenetic trees of phage nucleus shell proteins (A) or phage tubulin PhuZ proteins (B) with bootstrap values (1000 replicates). Red asterisks: Goslar, gold asterisks: characterized nucleus-forming *Pseudomonas* phages, blue asterisks: nucleus-forming Serratia phage. (C & D) Two examples of MG1655 cells (C) and APEC cells (D) either uninfected (uninf) or infected by Goslar for 30, 60, or 90 minutes (mpi) then stained with FM4-64 (membrane, magenta) and DAPI (DNA, cyan). A concentrated mass of DNA appears only after addition of phage lysate and this mass can be observed without any fluorescent stains using DIC. All scale bars are 1 μm. Goslar replicates equally well in both strains. (E) DIC time-lapse every 12 minutes for 94 minutes of MG1655 infected with Goslar for 30 minutes prior to imaging. The DIC density first visible at the yellow arrowhead shows that the nucleoid grows in size and moves around over time as the cell bulges.

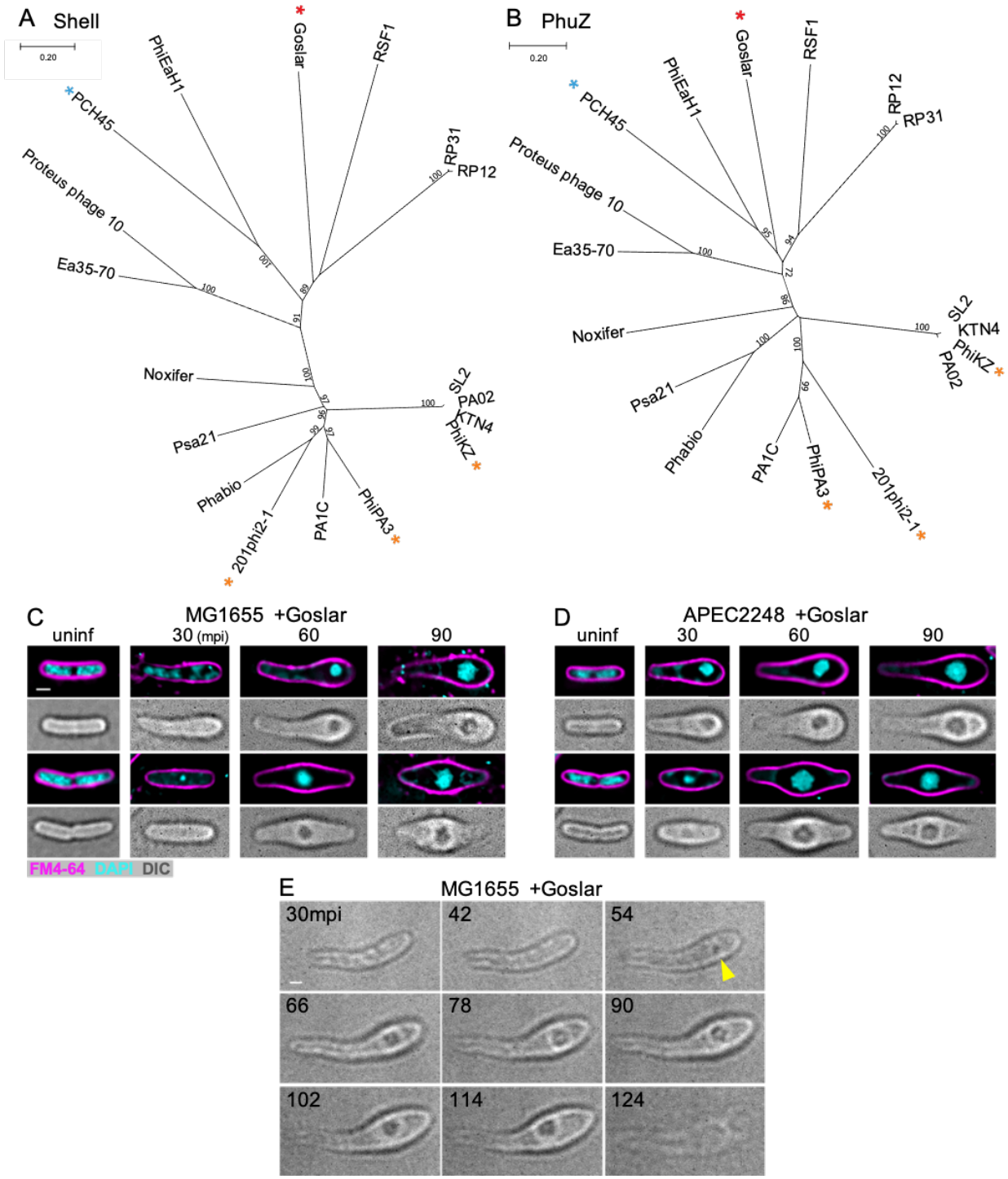
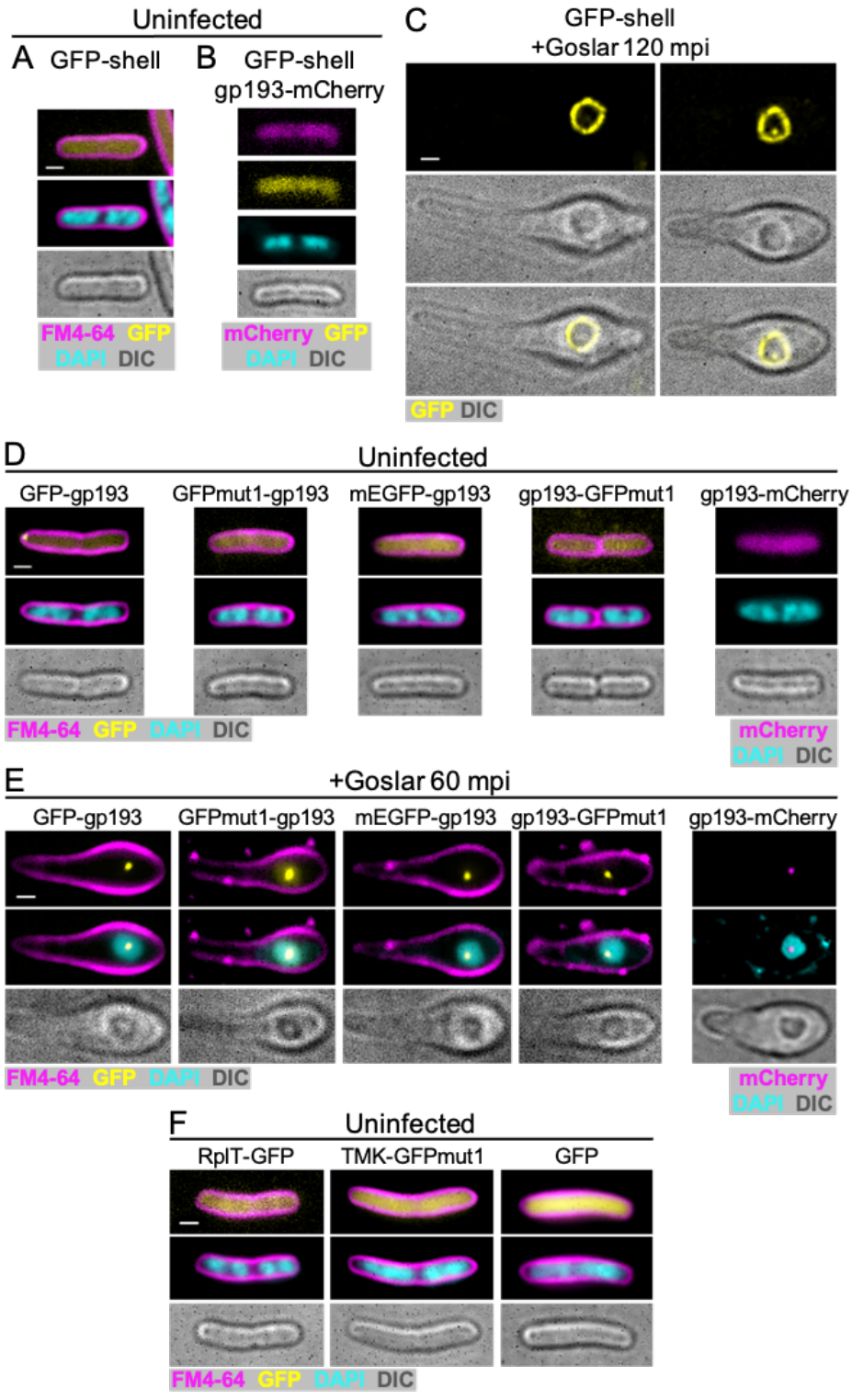

A Shell
0.20
PhiEaH1
Goslar
RSF1
PCH45
RP31
RP12
Proteus phage 10
Ea35-70
Noxifer
SL2
PA02
KTN4
PhiKZ
Psa21
Phabio
201phi2-1
PA1C
PhiPA3
B PhuZ
0.20
PhiEaH1
Goslar
RSF1
PCH45
RP12
RP31
Proteus phage 10
Ea35-70
Noxifer
SL2
KTN4
PhiKZ
PA02
Psa21
Phabio
PA1C
PhiPA3
201phi2-1
C MG1655 +Goslar
uninf
30 (mpi)
60
90
FM4-64 DAPI DIC
D APEC2248 +Goslar
uninf
30
60
90
E MG1655 +Goslar
30mpi
42
54
66
78
90
102
114
124

Figure S2.2 Control experiments show that the fusion proteins do not form specific structures in uninfected cells and that the choice of fusion does not alter the results. (A) *E. coli* (MG1655) expressing GFP-shell induced with 0.2 mM IPTG for 1.5 hours (uninfected). All scale bars are 1 μm. (B) *E. coli* co-expressing GFP-shell with gp193-mCherry and induced at 0.2 mM IPTG for 1.5 hours then stained with DAPI (uninfected). (C) *E. coli* expressing GFP-shell induced with 0.2 mM IPTG and infected by Goslar for 120 minutes. (D) *E. coli* expressing each gp193 fusion indicated and induced at 0.2 mM IPTG for 1.5 hours then stained with FM4-64 and DAPI, or just DAPI for gp193-mCherry (uninfected). Fusions to different fluorescent proteins (mCherry, mEGFP, GFPmut1) are generally uniformly distributed throughout the cell, only GFP-gp193 forms a polar punctum. (E) *E. coli* expressing the listed gp193 fusion protein at 0.2 mM IPTG, infected by Goslar for 60 minutes then stained with FM4-64 and DAPI. Fusing different fluorescent proteins (mCherry, GFP, mEGFP, GFPmut1) to gp193 at either the N- or C-terminus produces the same results. (F) *E. coli* expressing 50S ribosomal protein L20 (RplT-GFP), thymidylate kinase (TMK-GFP), or GFP alone, induced at 0.2 mM IPTG for 1.5 hours then stained with FM4-64 and DAPI. These fusions are generally uniformly distributed throughout the cell.


Uninfected
A GFP-shell
B GFP-shell
gp193-mCherry
FM4-64 GFP
DAPI DIC
mCherry GFP
DAPI DIC
C
GFP-shell
+Goslar 120 mpi
GFP DIC
D
Uninfected
GFP-gp193
GFPmut1-gp193
mEGFP-gp193
gp193-GFPmut1
gp193-mCherry
FM4-64 GFP DAPI DIC
mCherry
DAPI DIC
E
+Goslar 60 mpi
GFP-gp193
GFPmut1-gp193
mEGFP-gp193
gp193-GFPmut1
gp193-mCherry
FM4-64 GFP DAPI DIC
mCherry
DAPI DIC
F
Uninfected
RplT-GFP
TMK-GFPmut1
GFP
FM4-64 GFP DAPI DIC

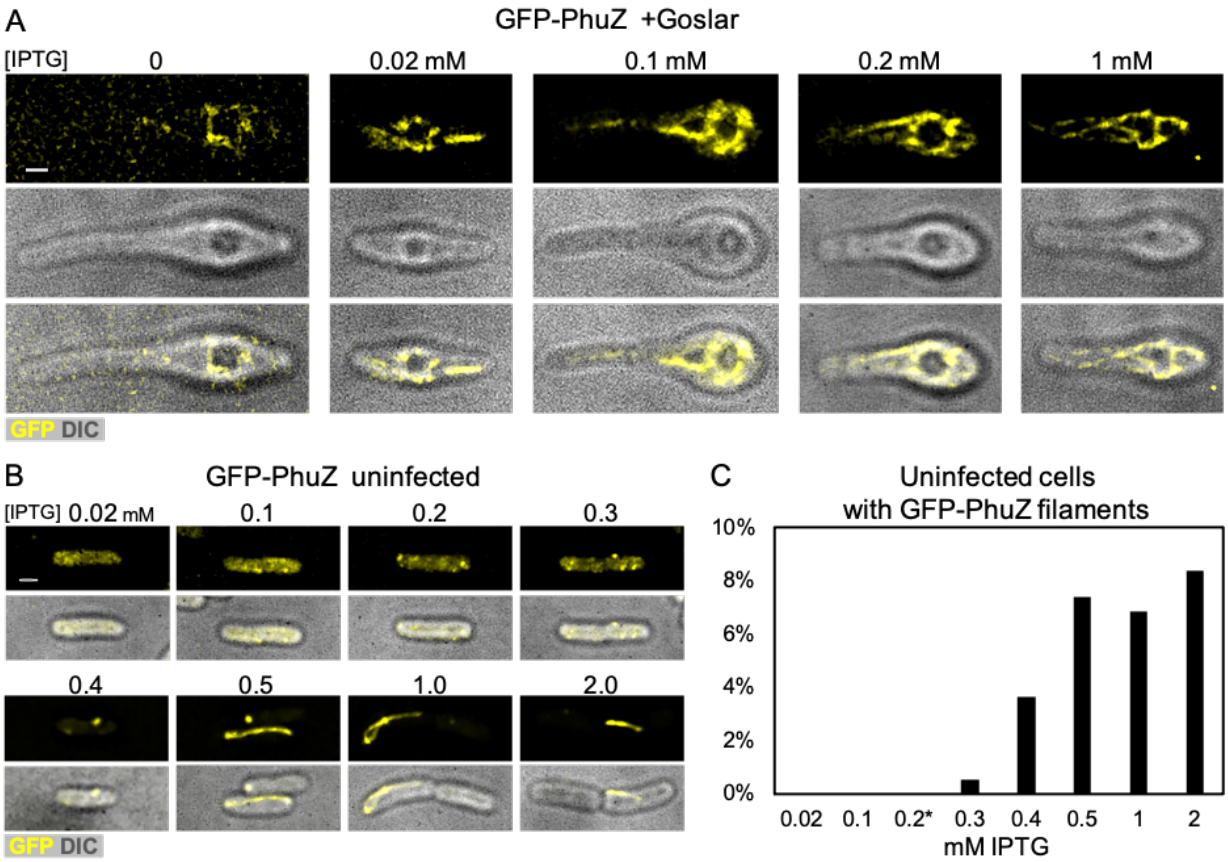


Figure S2.3 Assembly of GFP-PhuZ in uninfected and infected cells. (A) GFP-PhuZ accumulates around the phage nucleus even when expressed at very low levels. All levels of induction by IPTG show a similar vortex-like phenotype. All scale bars are 1 μm. (B) Uninfected *E. coli* (MG1655) expressing GFP-PhuZ (yellow) does not form filaments over 0.3 μm in length until 0.3 mM IPTG. (C) Percent of uninfected cells with a PhuZ filament over 0.3 μm at IPTG concentrations from 0.02 - 2.0 mM (0.02, n = 78; others, n > 100).

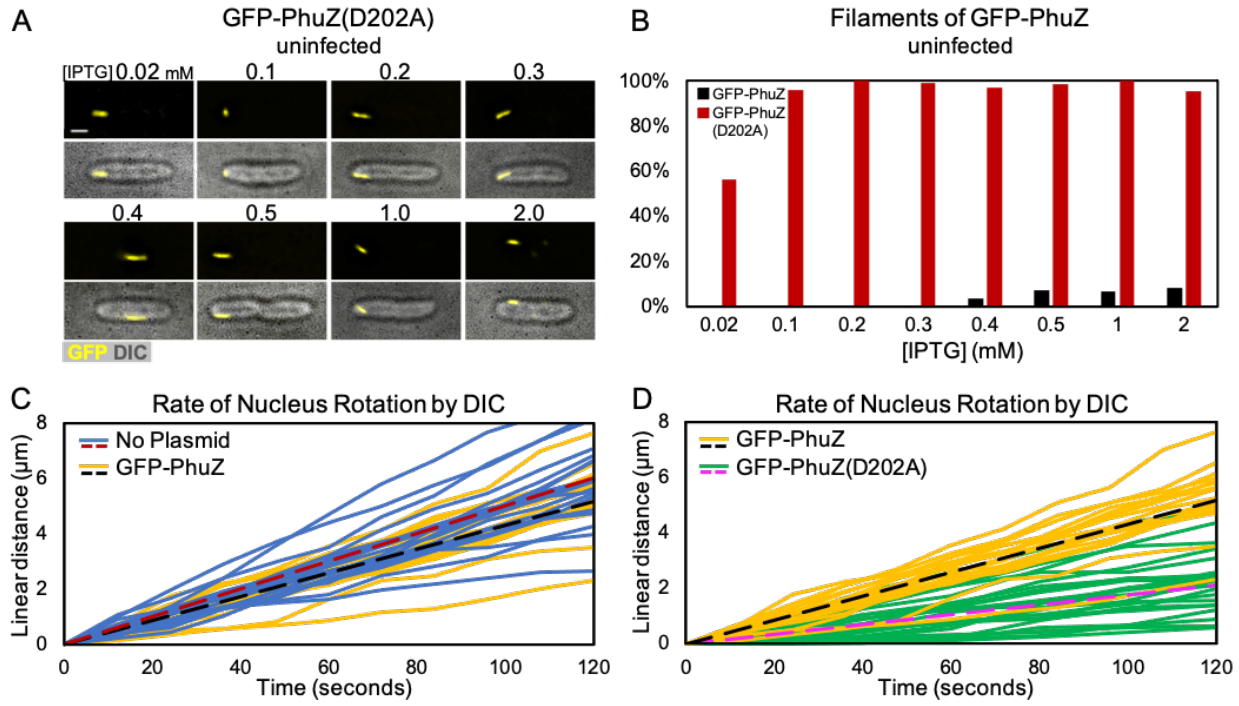


Figure S2.4 Expression of a catalytically defective PhuZ, GFP-PhuZ(D202A), alters PhuZ assembly properties and phage nucleus rotation. (A) Uninfected *E. coli* (MG1655) expressing GFP-PhuZ(D202A) induced with 0.02-2.0 mM IPTG and grown on 50 μg/ml ampicillin for 90 minutes. Scale bar 1 μm. (B) Percentage of cells expressing either GFP-PhuZ(D202A) (red) or GFP-PhuZ (black) that contained filaments over 0.3 μm at each IPTG concentration (0.2, n = 99; 0.5, n = 85; others, n > 100). (C) Linear velocity of nuclear rotation of 20 measured nuclei for *E. coli* (blue solid lines; average, red dotted line) compared to those in *E. coli* expressing GFP-PhuZ (yellow solid lines; average, black dotted line), 60 minutes after the addition of Goslar. (D) Linear velocity of nuclear rotation of 20 measured nuclei in *E. coli* expressing GFP-PhuZ (yellow solid lines; average, black dotted line) compared to the 23 nuclei in *E. coli* expressing GFP-PhuZ(D202A) (green solid lines; average, magenta dotted line), 60 minutes after the addition of Goslar.

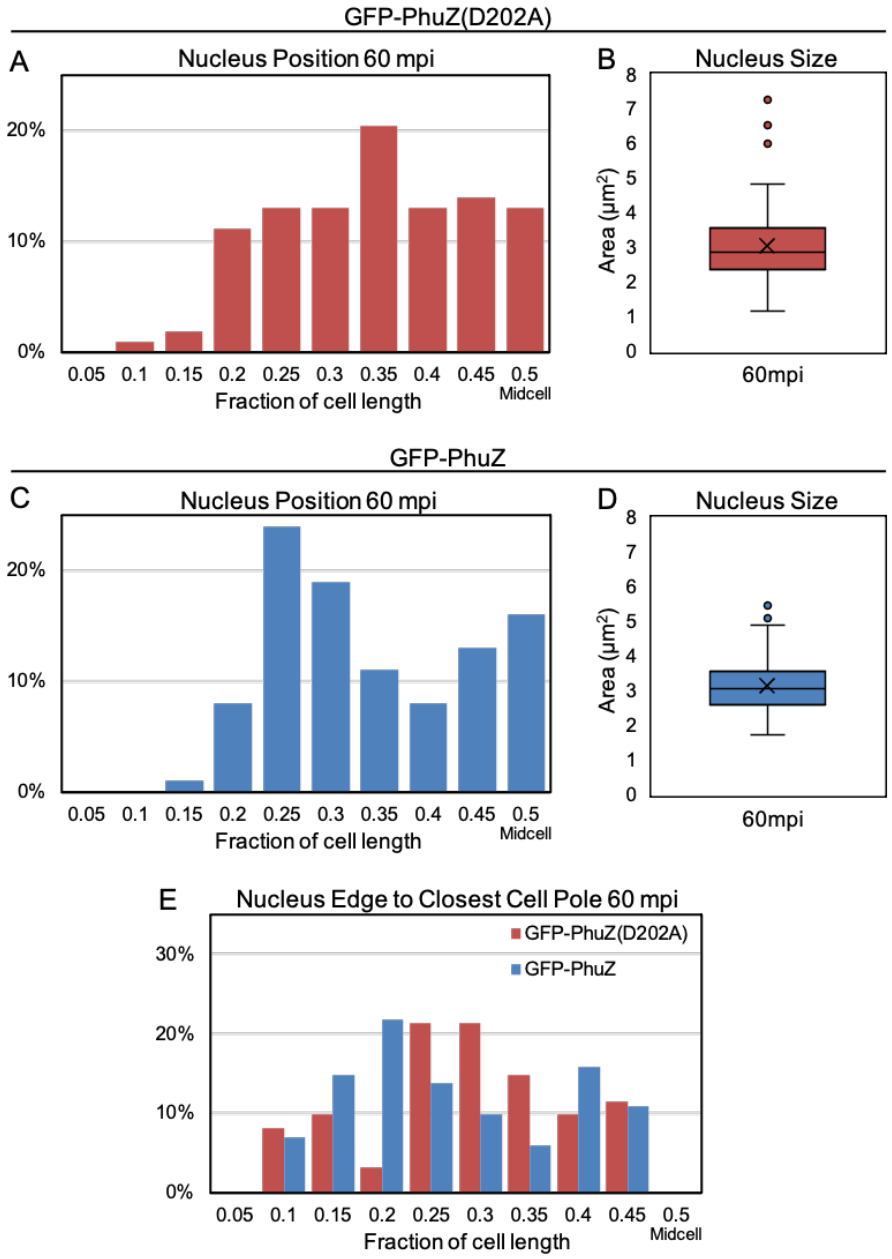
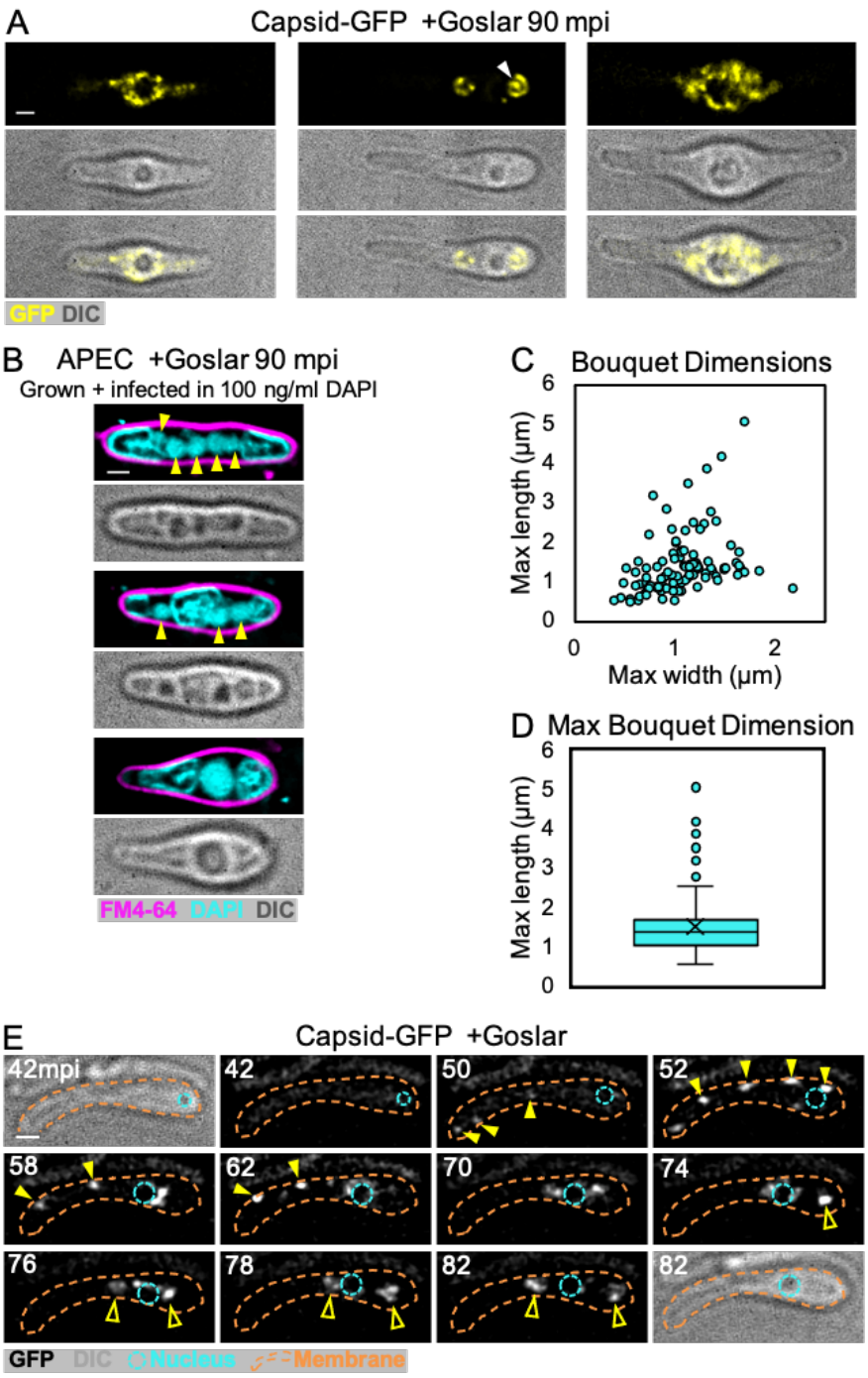


Figure S2.5 PhuZ(D202A) expression does not affect Goslar nucleus positioning or size. (A) Distribution of 60 mpi phage nuclei (measured by DIC) along the lateral length of *E. coli* (MG1655) expressing GFP-PhuZ(D202A) induced with 0.2 mM IPTG. For each time point, there is no significantly greater chance of finding a phage nucleus in one certain bin than in the neighboring bins (n = 108). (B) 2D area of the DAPI-stained 60 mpi phage nucleus in the presence of GFP-PhuZ(D202A) (n = 108). (C) Distribution of 60 mpi DIC phage nuclei along the lateral length of *E. coli* expressing GFP-PhuZ induced with 0.2 mM IPTG. For each time point, there is no significantly greater chance of finding a phage nucleus in one certain bin than in the neighboring bins (n = 100). (D) 2D area of the DAPI-stained 60 mpi phage nucleus in the presence of GFP-PhuZ (n = 111). (E) Distribution of distances between the edge of the phage nucleus to the closest cell pole, normalized to cell length. Either GFP-PhuZ(D202A) or GFP-PhuZ was infected for 60 minutes and stained with FM4-64 and DAPI. DIC was used for measurement (n = 88).

Figure S2.6 Capsid-GFP localization and bouquet formation during Goslar infections. (A) More examples of capsid-GFP localization as in Figure 2.7A. *E. coli* (MG1655) expressing the putative capsid protein (gp41) fused to GFP (yellow) induced with 0.2 mM IPTG and infected with Goslar for 90 minutes. Capsids localized around the phage nucleus (left panels), in adjacent bouquets (middle panels), or filling more of the cytoplasm (right panels). All scale bars are 1 μm. (B) Additional examples of multiple nuclei and bouquets (top four panels) and a single nucleus (bottom two panels) imaged by DAPI and FM4-64 after 90 mpi in APEC (yellow arrowheads indicate nuclei). Top 2 panels show 5 nuclei observed in one cell, middle 2 panels show 3 nuclei with a large central bouquet, bottom panels show a single nucleus with the most common bouquet phenotype. (C) DAPI-stained bouquets at 90 mpi in APEC were measured along the maximum x and y dimensions. Each point represents one bouquet (n = 105). (D) Box and whisker plot of maximum bouquet dimension for each bouquet in (C). Quartiles are calculated with an excluded median and a mean of 1.5 μm (n = 105). (E) Time-lapse of capsid-GFP (white) at various intervals over 40 minutes starting at 42 mpi in MG1655. Capsids migrate from the cell periphery or cytoplasm (yellow solid arrowheads) to the phage nucleus exterior and then to adjacent phage bouquets (yellow open arrowheads) (cyan dashed line, nucleus; orange dashed line, cell membrane).


A
Capsid-GFP +Goslar 90 mpi
GFP DIC
B APEC +Goslar 90 mpi
Grown + infected in 100 ng/ml DAPI
FM4-64 DAPI DIC
C Bouquet Dimensions
Max length (µm)
Max width (µm)
D Max Bouquet Dimension
Max length (µm)
E
Capsid-GFP +Goslar
42mpi
42
50
52
58
62
70
74
76
78
82
82
GFP DIC Nucleus Membrane

2.8 References

1. Chaikeeratisak, V., Birkholz, E.A., and Pogliano, J., *The Phage Nucleus and PhuZ Spindle: Defining Features of the Subcellular Organization and Speciation of Nucleus-Forming Jumbo Phages.* Front Microbiol, 2021. **12**: p. 641317.

2. Charman, M. and Weitzman, M.D., *Replication Compartments of DNA Viruses in the Nucleus: Location, Location, Location.* Viruses, 2020. **12**(2).

3. Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., Villa, E., Agard, D.A., and Pogliano, J., *Assembly of a nucleus-like structure during viral replication in bacteria.* Science, 2017. **355**(6321): p. 194-197.

4. Chaikeeratisak, V., Nguyen, K., Egan, M.E., Erb, M.L., Vavilina, A., and Pogliano, J., *The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages.* Cell Reports, 2017. **20**(7): p. 1563-1571.

5. Danilova, Y.A., Belousova, V.V., Moiseenko, A.V., Vishnyakov, I.E., Yakunina, M.V., and Sokolova, O.S., *Maturation of Pseudo-Nucleus Compartment in P. aeruginosa, Infected with Giant phiKZ Phage.* Viruses, 2020. **12**(10).

6. Erb, M.L., Kraemer, J.A., Coker, J.K., Chaikeeratisak, V., Nonejuie, P., Agard, D.A., and Pogliano, J., *A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells.* Elife, 2014. **3**.

7. Kraemer, J.A., Erb, M.L., Waddling, C.A., Montabana, E.A., Zehr, E.A., Wang, H., Nguyen, K., Pham, D.S., Agard, D.A., and Pogliano, J., *A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell.* Cell, 2012. **149**(7): p. 1488-99.

8. Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gumy, L.F., and Fineran, P.C., *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, 2020. **5**(1): p. 48-55.

9. Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J., *A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases.* Nature, 2020. **577**(7789): p. 244-248.

10. Nguyen, K.T., Sugie, J., Khanna, K., Egan, M.E., Birkholz, E.A., Lee, J., Beierschmitt, C., Villa, E., and Pogliano, J., *Selective transport of fluorescent proteins into the phage nucleus.* PLoS One, 2021. **16**(6): p. e0251429.

11. Chaikeeratisak, V., Birkholz, E.A., Prichard, A.M., Egan, M.E., Mylvara, A., Nonejuie, P., Nguyen, K.T., Sugie, J., Meyer, J.R., and Pogliano, J., *Viral speciation through subcellular genetic isolation and virogenesis incompatibility.* Nat Commun, 2021. **12**(1): p. 342.

12. Zehr, E.A., Kraemer, J.A., Erb, M.L., Coker, J.K., Montabana, E.A., Pogliano, J., and Agard, D.A., *The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA.* Structure, 2014. **22**(4): p. 539-48.

13. Aylett, C.H., Izore, T., Amos, L.A., and Lowe, J., *Structure of the tubulin/FtsZ-like protein TubZ from Pseudomonas bacteriophage PhiKZ.* J Mol Biol, 2013. **425**(12): p. 2164-73.

14. Chaikeeratisak, V., Khanna, K., Nguyen, K.T., Sugie, J., Egan, M.E., Erb, M.L., Vavilina, A., Nonejuie, P., Nieweglowska, E., Pogliano, K., Agard, D.A., Villa, E., and Pogliano, J., *Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria.* Cell, 2019. **177**(7): p. 1771-1780.e12.

15. Chaikeeratisak, V., Khanna, K., Nguyen, K. T., Egan, M. E., Enustun, E., Armbruster, E., Pogliano, K., Villa, E., & Pogliano, J. , *Subcellular Organization of Viral Particles During Maturation of Nucleus-Forming Jumbo Phage*. 2021: BioRxiv.

16. Korf, I.H.E., Meier-Kolthoff, J.P., Adriaenssens, E.M., Kropinski, A.M., Nimtz, M., Rohde, M., Van Raaij, M.J., and Wittmann, J., *Still Something to Discover: Novel Insights into Escherichia coli Phage Diversity and Taxonomy.* Viruses, 2019. **11**(5): p. 454.

17. Dziva, F., Hauser, H., Connor, T.R., van Diemen, P.M., Prescott, G., Langridge, G.C., Eckert, S., Chaudhuri, R.R., Ewers, C., Mellata, M., Mukhopadhyay, S., Curtiss, R., 3rd, Dougan, G., Wieler, L.H., Thomson, N.R., Pickard, D.J., and Stevens, M.P., *Sequencing and functional annotation of avian pathogenic Escherichia coli serogroup O78 strains reveal the evolution of E. coli lineages pathogenic for poultry via distinct mechanisms.* Infect Immun, 2013. **81**(3): p. 838-49.

18. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A., *Fiji: an open-source platform for biological-image analysis.* Nature Methods, 2012. **9**(7): p. 676-682.

19. Lam, V. and Villa, E., *Practical Approaches for Cryo-FIB Milling and Applications for Cellular Cryo-Electron Tomography*, in *cryoEM*. 2021, Springer US. p. 49-82.

20. Mastronarde, D.N., *Automated electron microscope tomography using robust prediction of specimen movements.* J Struct Biol, 2005. **152**(1): p. 36-51.

21. Hagen, W.J.H., Wan, W., and Briggs, J.A.G., *Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging.* J Struct Biol, 2017. **197**(2): p. 191-198.

22. Tegunov, D. and Cramer, P., *Real-time cryo-electron microscopy data preprocessing with Warp.* Nature Methods, 2019. **16**(11): p. 1146-1152.

23. Mastronarde, D.N. and Held, S.R., *Automated tilt series alignment and tomographic reconstruction in IMOD.* Journal of Structural Biology, 2017. **197**(2): p. 102-113.

24. Martinez-Sanchez, A., Garcia, I., Asano, S., Lucic, V., and Fernandez, J.J., *Robust membrane detection based on tensor voting for electron tomography.* J Struct Biol, 2014. **186**(1): p. 49-61.

25. Scheres, S.H., *RELION: implementation of a Bayesian approach to cryo-EM structure determination.* J Struct Biol, 2012. **180**(3): p. 519-30.

26. Bharat, T.A. and Scheres, S.H., *Resolving macromolecular structures from electron cryo-tomography data using subtomogram averaging in RELION.* Nat Protoc, 2016. **11**(11): p. 2054-65.

27. Castano-Diez, D., Kudryashev, M., Arheit, M., and Stahlberg, H., *Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments.* J Struct Biol, 2012. **178**(2): p. 139-51.

28. Castano-Diez, D., Kudryashev, M., and Stahlberg, H., *Dynamo Catalogue: Geometrical tools and data management for particle picking in subtomogram averaging of cryo-electron tomograms.* J Struct Biol, 2017. **197**(2): p. 135-144.

29. Burt, A., Gaifas, L., Dendooven, T., and Gutsche, I., *Tools enabling flexible approaches to high-resolution subtomogram averaging*. 2021, Cold Spring Harbor Laboratory.

30. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E., *UCSF ChimeraX:Structure visualization for researchers, educators, and developers.* Protein Science, 2021. **30**(1): p. 70-82.

31. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K., *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Molecular Biology and Evolution, 2018. **35**(6): p. 1547-1549.

32. Stecher, G., Tamura, K., and Kumar, S., *Molecular Evolutionary Genetics Analysis (MEGA) for macOS.* Molecular Biology and Evolution, 2020. **37**(4): p. 1237-1239.

33. Larsen, R.A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P., and Pogliano, J., *Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in Bacillus thuringiensis.* Genes Dev, 2007. **21**(11): p. 1340-52.

34. Sumino, Y., Nagai, K.H., Shitaka, Y., Tanaka, D., Yoshikawa, K., Chaté, H., and Oiwa, K., *Large-scale vortex lattice emerging from collectively moving microtubules.* Nature, 2012. **483**(7390): p. 448-452.

35. Suzuki, K., Miyazaki, M., Takagi, J., Itabashi, T., and Ishiwata, S.I., *Spatial confinement of active microtubule networks induces large-scale rotational cytoplasmic flow.* Proceedings of the National Academy of Sciences, 2017. **114**(11): p. 2922-2927.

36. Schroeder, T.E. and Battaglia, D.E., *"Spiral asters" and cytoplasmic rotation in sea urchin eggs: induction in Strongylocentrotus purpuratus eggs by elevated temperature.* The Journal of Cell Biology, 1985. **100**(4): p. 1056-1062.

37. Serbus, L.R., Cha, B.-J., Theurkauf, W.E., and Saxton, W.M., *Dynein and the actin cytoskeleton control kinesin-driven cytoplasmic streaming inDrosophilaoocytes.* Development, 2005. **132**(16): p. 3743-3752.

38. Woodhouse, F.G. and Goldstein, R.E., *Cytoplasmic streaming in plant cells emerges naturally by microfilament self-organization.* Proceedings of the National Academy of Sciences, 2013. **110**(35): p. 14132-14137.

39. Stein, D.B., De Canio, G., Lauga, E., Shelley, M.J., and Goldstein, R.E., *Swirling Instability of the Microtubule Cytoskeleton.* Physical Review Letters, 2021. **126**(2).

40. Gerashchenko, M.V., Chernoivanenko, I.S., Moldaver, M.V., and Minin, A.A., *Dynein is a motor for nuclear rotation while vimentin IFs is a "brake".* Cell Biology International, 2009. **33**(10): p. 1057-1064.

41. Levy, J.R. and Holzbaur, E.L.F., *Dynein drives nuclear rotation during forward progression of motile fibroblasts.* Journal of Cell Science, 2008. **121**(19): p. 3187-3195.

42. Kim, D.-H., Cho, S., and Wirtz, D., *Tight coupling between nucleus and cell migration through the perinuclear actin cap.* Journal of Cell Science, 2014. **127**(11): p. 2528-2541.

43. Kumar, A., Maitra, A., Sumit, M., Ramaswamy, S., and Shivashankar, G.V., *Actomyosin contractility rotates the cell nucleus.* Scientific Reports, 2015. **4**(1).

44. Maninová, M., Iwanicki, M.P., and Vomastek, T., *Emerging role for nuclear rotation and orientation in cell migration.* Cell Adhesion & Migration, 2014. **8**(1): p. 42-48.

45. Wu, J., Lee, K.C., Dickinson, R.B., and Lele, T.P., *How dynein and microtubules rotate the nucleus.* Journal of Cellular Physiology, 2011. **226**(10): p. 2666-2674.

46. Fruleux, A. and Hawkins, R.J., *Physical role for the nucleus in cell migration.* J Phys Condens Matter, 2016. **28**(36): p. 363002.

CHAPTER 3:

Viral speciation through subcellular genetic isolation and virogenesis incompatibility

3.1 Abstract

Understanding how biological species arise is critical for understanding the evolution of life on Earth. Bioinformatic analyses have recently revealed that viruses, like multicellular life, form reproductively isolated biological species. Viruses are known to share high rates of genetic exchange, so how do they evolve genetic isolation? Here, we evaluate two related bacteriophages and describe three factors that limit genetic exchange between them: 1) A nucleus-like compartment that physically separates replicating phage genomes, thereby limiting inter-phage recombination during coinfection; 2) A tubulin-based spindle that orchestrates phage replication and forms nonfunctional hybrid polymers; and 3) A nuclear incompatibility factor that reduces phage fitness. Together, these traits maintain species differences through Subcellular Genetic Isolation where viral genomes are physically separated during coinfection, and Virogenesis Incompatibility in which the interaction of cross-species components interferes with viral production.

3.2 Introduction

Bacteriophages are the most abundant microbes on Earth and arguably harbor the most genetic diversity of any taxonomic group [1, 2]. Viruses are known to evolve quickly due to their large population sizes, short generation times, and frequently high mutation rates. Mutations can become fixed in the population by strong selective forces or by drift. Until recently, it was thought that this genetic diversity was relatively unstructured with phages freely exchanging genes [3-5]. Through analyses of a growing number of full phage genome sequences, it has become clear that the diversity coalesces into biological species clusters where more genetic information is exchanged within, rather than between clusters [6]. This observation leads to the question of how phages evolve barriers to genetic exchange. Adaptation to different hosts has been shown to cause

viral speciation since viruses recombine when multiple particles infect the same cell. Evolving divergent host specificities therefore has the side effect of developing barriers to genetic exchange [7-9]. However, the evolution of host specificities cannot fully explain viral speciation because viruses that infect the same hosts sometimes form genetically isolated species [6, 10]. Viral traits must have evolved that form genetic barriers during coinfection of the same cell and could ultimately lead to viral speciation. In the study of two closely related jumbo phages that infect *Pseudomonas aeruginosa*, we hypothesized that specific viral mechanisms must exist that contribute to genetic isolation. Here we sought to identify potential barriers to viral genetic exchange and quantitate the extent to which they contribute to reproductive isolation. Our approach is analogous to those historically performed on eukaryotes where the reproductive isolation caused by different phenotypic and genetic characteristics is quantified [11]. During infection of *Pseudomonas*, jumbo phages 201φ2-1, ΦPA3, and ΦKZ establish an intricate subcellular organization that we reasoned may impact gene flow [12-16]. These jumbo phages replicate by enclosing their DNA within a proteinaceous shell that forms a nucleus-like compartment, separating enzymes involved in DNA replication, transcription, and repair from ribosomes and metabolic enzymes in the cytoplasm [15, 16]. A tubulin-like protein, PhuZ, assembles a spindle early during lytic growth that plays multiple roles during the life cycle of the phage [12-17]. At early stages of infection, the PhuZ spindle uses dynamic instability to position the phage nucleus at midcell [12, 13, 15, 16]. Later during infection, the spindle uses treadmilling to traffic newly assembled capsids through the cell and to distribute them around the nucleus [14]. Nonfunctional spindles result in severe nucleus mispositioning, loss of capsid trafficking, and a 50% decrease in phage progeny [12-16]. Given the organizational complexity of these phages, it was unclear how they might interact with one another during coinfection of a single cell when

there is an opportunity for genetic exchange. Here we characterize coinfecting phages and identify two types of barriers to gene flow. We show that Subcellular Genetic Isolation occurs when a nucleus-like compartment physically separates viral genomes during coinfection. We also show that coinfections by different viruses can be less productive because of incompatibility between divergent viral components. This Virogenesis Incompatibility blocks gene flow by reducing the chance of producing recombinant progeny.

3.3 Materials and Methods

General methods

Pseudomonas aeruginosa strain PAO1 was grown on LB agar containing 10 g Bacto-Tryptone, 5 g yeast extract, 5 g NaCl, and 15 g agar in 1 L ddH_2O and incubated at 30°C. Lysates of phages ΦPA3 and ΦKZ were prepared by infecting 109 cells of *P. aeruginosa* strain PAO1 with phage (~10^8 pfu/ml) at room temperature for 15 minutes and then adding 5 ml of 0.35% LB top agar onto an LB plate. Following overnight incubation at 30°C, plates were flooded with 5 ml SM phage buffer and incubated at room temperature for 4 to 6 hours. Phage lysates were harvested, centrifuged at 21,130 × g for 10 minutes and stored at 4°C.

Plasmid and strain constructions

Pseudomonas aeruginosa strain PAO1-K2733 was used for all experiments. Plasmids were introduced into strain PAO1-K2733 by electroporation and selecting on LB containing 15 μg/mL gentamycin. The vector pHERD-30T [14-16] was used to express all GFP fusions. Plasmids expressing sfGFP-ΦPA3PhuZ, sfGFP-ΦKZPhuZ, sfGFP-ΦPA3PhuZ(D190A), sfGFP-ΦPA3shell and mCherry-ΦKZshell were previously described [14-16]. To construct a colocalization construct of ΦPA3 shell and ΦKZ shell, the pHERD-30T backbone with mCherry-ΦKZshell was first amplified using primers oVC701 and oVC406 by PCR from plasmid

pMAC011 [15]. The GFPmut1 gene tagged at the N-terminus of ΦPA3shell gene in plasmid pVC077 was replaced by sfGFP [15]. The insert of sfGFP-ΦPA3shell was then made by PCR amplification using primers oVC403 and oVC702 from the resulting plasmid. These 2 amplicons were assembled together by NEBuilder® HiFi DNA Assembly Cloning Kit (New England Biolabs) resulting in a colocalization construct of sfGFP-ΦPA3shell and mCherry-ΦKZshell (pMAC082). We then performed the same cloning strategy to construct a colocalization plasmid of ΦPA3 PhuZ and ΦKZ PhuZ. The primers oVC701 and oVC406 were used to amplify the pHERD-30T backbone with mCherry-ΦKZPhuZ from plasmid pVC116 while the primers oVC403 and oVC702 were used to amplify the insert of sfGFP-ΦPA3PhuZ from plasmid pVC028. These 2 amplicons were then assembled using NEBuilder® HiFi DNA Assembly Cloning Kit (New England Biolabs) resulting in a colocalization construct of sfGFP-ΦPA3PhuZ and mCherry-ΦKZPhuZ (pVC120). An untagged ΦKZPhuZ plasmid was made using oVC704 and oVC705 to amplify the linear amplicon ofΦKZ PhuZ linked with pHERD-30T from pVC029 [15]. The linear amplicon was then circularized by ligase resulting in plasmid pVC121. Plasmids containing GFPmut1 and sfGFP were previously described [14-16]. gp210 was amplified from ΦPA3 lysate using primers oEB003 and oEB004. Plasmid pKN053 and pKN069 were linearized using oEB005 and oEB006. gp210 was assembled with each linearized plasmid using NEBuilder® HiFi DNA Assembly Cloning Kit (New England Biolabs). See Tables 1 and 2 for lists of primers, plasmids, and strains. All plasmids were confirmed by DNA sequencing.

Live cell fluorescence microscopy

Single cell infections of *P. aeruginosa* were visualized using fluorescence microscopy [14-16]. Briefly, 1% agarose pads containing 25% LB, 2 μg/ml FM 4-64, and 1 μg/ml DAPI were inoculated with 5 μL of cells (~1×10^8 cfu/ml) allowed to grow at 30°C in a humidor for 3 hours

and then infected with 10 µl of a high titer (10^{12} pfu/ml) phage lysate. For dual infections, lysates of ΦPA3 and ΦKZ were mixed at an equal ratio prior to infections. At desired time point after phage infection, a coverslip was put on the slide and cells were imaged using a DeltaVision Spectris Deconvolution microscope (Applied Precision, Issaquah, WA, USA). Images were further processed by the deconvolution algorithm in DeltaVision SoftWoRx 6.5.2 Image Analysis Program, and analyzed using Fiji 1.53c software. For time-lapse microscopy, infections were established on agarose pads (25% LB, 1% agarose, 0.1% arabinose) without stains. At 30 mpi, locations for taking time-lapse images were identified, and at 45 mpi, images were captured every 10 minutes for 4 hours. The cells from the resulting images were counted and the number of infected cells was divided by the total number of starting cells to determine the percentage of cells that were lysing over the course of the 4 hours. The percentages were averaged and plotted for both conditions.

Fluorescence recovery after photobleaching (FRAP)

Cells containing sfGFP-ΦPA3PhuZ filaments were prepared as indicated for fluorescence microscopy and photobleach using a laser (QLM module, API) for 0.05 seconds at 31% power. Images were collected every 2 seconds for 1 minute by an Applied Precision/GE OMX V2.2 Microscope. Images were deconvolved with DeltaVision SoftWoRx 6.5.2 and analyzed by Fiji 1.53c.

IC50 growth curves

We determined the concentration of phage required to inhibit 50% cell growth over 6 hours as a measure of phage fitness. 5 mL cultures of *P. aeruginosa* in liquid LB containing 15 µg/mL gentamycin were grown to an OD_{600} of approximately 0.6 to 0.8 and then diluted to an OD_{600} of 0.1. 5 µL of 10-fold serial dilutions of a ΦKZ lysate with a titer of 10^{12} pfu/ml were added to each

well of a 96-well plate containing 100 μL of diluted bacterial culture. The 96-well plates were shaken for an initial 40 minutes at 30°C in a microplate reader after which OD_{600} measurements were taken every 10 minutes for 6 hours, with continuous shaking at 30°C between the timepoints. The resulting OD_{600} values were averaged and plotted as a growth curve. Fusions were expressed from a plasmid at very low levels without arabinose induction.

Efficiency of plating (EOP) titers

To compare the efficiency of phage plaque formation on different strains, 10^8 cells were infected with 10-fold serial dilutions of phage lysate at room temperature for 15 minutes and then 5 ml of 0.35% LB top agar was added and poured onto an LB plate. Plaques were counted after 24 hours of incubation at 37°C and the efficiency of plating calculated as the ratio of the test sample to the control. Protein expression was achieved with the indicated fusions by inducing with either 1.0% or 0.1% arabinose [14-16].

Coinfection shell analysis

Fiji image analysis program was used for image analysis. A total of 89 coinfected cells containing both shells were included in the analysis. For size and DAPI intensity analysis of the shell, a polygon was drawn over the shell to measure the area and mean DAPI intensity of the shell. Then, the mean DAPI intensity was subtracted by the background DAPI intensity of the corresponding image yielding the absolute DAPI intensity of the shell for the analysis. For shell position analysis, the distance between the center of the shell and the mid-position of the cell length was measured. Then, the distance was normalized by the corresponding cell length creating the fraction of the cell length for data visualization.

Statistics and reproducibility

All experiments, including microscopy, were repeated at least three times. Representative microscopy images are shown. Standard deviations were calculated for phage titer experiments.

3.4 Results

We expected that two identical phages infecting one cell could either form a single nucleus that would facilitate inter-phage recombination or two separate nuclei, greatly reducing gene exchange. We first used DAPI staining to visualize phage DNA and soluble GFP to visualize the cytoplasm, and found that cells infected with either ΦPA3 or ΦKZ frequently (24%, n = 156 and 19%, n = 368 at 20 minutes post-infection (mpi); 21%, n = 142 and 18%, n = 307 at 50 mpi, respectively; Figure S3.1A) formed two separate phage nucleoids in a single cell (Figure 3.1A). The similar percentage of cells with more than one nucleoid at 20 mpi and 50 mpi suggests that the nucleoids remained separated throughout the infection. To confirm that these nucleoids were in separate compartments, we fluorescently tagged the ΦPA3 shell protein with GFP and the ΦKZ shell with mCherry and followed nucleus assembly during infection (Figure 3.1B). Each fluorescent protein-shell fusion formed diffuse fluorescence and small foci when uninfected (Figure S3.1B) and assembled a nuclear shell during infection (Figure 3.1B). Direct visualization of each nucleus confirmed that a significant fraction of infected host cells contained two nuclei (ΦPA3, 18%, n = 104; ΦKZ, 14%, n = 117; Figure 3.1B). In cells containing two distinct nuclei, both were adjacent to one another and positioned at midcell by the PhuZ spindle (Figure 3.1B). These results demonstrate that during single-species coinfections, separate replication compartments for identical genomes can be established within the same cell. These data also suggest that the phage nucleus forms a physical barrier that separates co-replicating viral genomes, thereby reducing the potential for genetic exchange even between two identical genomes. We

refer to the separation of co-replicating viral genomes by physical or spatial barriers, such as the phage nucleus, as Subcellular Genetic Isolation. This phenomenon has already been observed in *herpesviruses* and *poxviruses* which form spatially separated replication factories [18, 19]. Recombination does not occur unless the replication factories coalesce. Traits that cause Subcellular Genetic Isolation create a barrier to genetic exchange which allows for evolutionary divergence and the accumulation of reproductive incompatibilities that arise by neutral drift or other processes. Subcellular Genetic Isolation is therefore a potentially common mechanism of viral speciation. To further test the theory that the phage nucleus results in Subcellular Genetic Isolation, we visualized coinfections of ΦPA3 and ΦKZ, two related phages whose shells share ~45% amino acid identity. We expressed both fluorescently tagged shell proteins from a single plasmid and determined if assembly into a nucleus was promiscuous or specific to the cognate phage from which the shell originated (Figure 3.1C). We found that each shell protein only assembled into a nuclear shell containing DNA during infection with the cognate phage (Figure 3.1C) and therefore can be used to track individual phage nuclei during coinfection of a single cell. Next, we followed the fate of the two different phages during coinfection of *P. aeruginosa* expressing both fluorescently tagged shell proteins, by infecting it simultaneously with ΦPA3 and ΦKZ (Figure 3.1D). Approximately 25% (n = 295) of doubly infected cells contained shell structures that incorporated components from each species (Figure 3.1D, half red and half green shell), whereas the majority (75%; n = 295) of coinfections resulted in separate red and green shells that physically separated the genomes (Figure 3.1D). Thus, during coinfections resulting from either a single species (Figures 3.1A and B) or two species (Figure 3.1D), phage can assemble separate nuclei, supporting the idea that the phage nucleus is a speciation factor that creates a physical barrier which separates co-replicating viral genomes and establishes Subcellular Genetic

Isolation. A second mechanism that is likely universal among viruses and that can contribute to genetic isolation, even without the formation of nuclear shells, is intracellular competition between speciating strains. If one strain of virus gains an advantage over the other, then it will reproduce more within cells and it will be more likely to recombine with conspecific genomes. The probability of cross-species recombination can be predicted by the rate at which genomes encounter each other, which, without subcellular organization or other factors that influence recombination rates, is equal to their relative genome frequencies (Figure S3.2). When each strain is equally competitive, then half of all recombinations will occur within the same strain and the other half between strains. As competitive differences increase between strains, their relative frequencies will shift, and the amount of cross-species recombination will equal two times the product of each strains' frequency (Figure S3.2). Phages ΦKZ and ΦPA3 were added simultaneously to cells, yet after 1 hour, one phage nucleus was almost always (97%, n = 89) larger than the other (Figure 3.1E), demonstrating competition between ΦKZ and ΦPA3. DAPI staining was used to estimate relative DNA content by integrated intensity within individual phage nuclei. We found a direct correlation between nucleus size and amount of phage DNA (Figures S3.1E and F). ΦKZ outcompetes ΦPA3 in 78% (n = 89) of coinfections based on DNA content (Figure 3.1F). For comparison, we quantitated DAPI staining during coinfection by a single species and found that one phage usually (~88%; n = 50) out-replicated the identical phage, suggesting that competition is inherent during coinfections (Figure 3.1G and H). We measured the relative genome frequencies of ΦPA3 (q) and ΦKZ (p) based on DAPI staining for each coinfected cell, and compared the expected probability of cross-species recombination, 2pq, to the probability that two genomes present at equal frequencies would recombine (when p = q, 2pq = 0.5) (Figure S3.2). For ΦPA3 and ΦKZ coinfections, the average 2pq value was 0.47, which corresponds to a 6

percent reduction of cross-species recombination (Figure S3.2). Intracellular competition does not have much of an impact on limiting interspecific recombination for these species; however, the effect is predicted to magnify as competitive differences increase (Figure S3.2). Competition between viruses may have an even more important long-term effect on speciation than calculated above since inter-strain competition can trigger a coevolutionary arms race that will drive further genetic divergence and help reinforce species separation. In eukaryotes, reproductive isolation can be driven by pre- or post-zygotic incompatibilities between gametes, limiting the ability to form offspring. While viruses do not form gametes or zygotes, our coinfection experiments revealed that the PhuZ spindle is an incompatibility factor that limits the production of phage offspring due to deleterious interactions between cross-species alleles (Figures 3.2 and 3.3). Normally, the phage nucleus is positioned at midcell by the spindle within 1 hour post-infection for single or double infections by one species of phage (Figures 3.1A-C and 3.2A, B, and D) [12, 13, 15, 16]. However, in coinfections with both ΦPA3 and ΦKZ, the two nuclei were always adjacent to one another and usually (92%, n = 89) mispositioned (Figures 3.1D and 3.2A-C). Normally, > 95% of nuclei occur within 20% of the cell midpoint at 60 mpi [12, 13, 15, 16]. Mispositioning results in a large percentage of nuclei outside of the 20% boundary. In time-lapse microscopy (Figure 3.2C), two small nuclei at the cell pole grew in size over time but failed to be moved to the center. This loss of positioning suggests that spindles are nonfunctional during coinfections. One possibility is that each phage establishes a separate bipolar spindle that cannot properly position two adjacent nuclei. Alternatively, a single nonfunctional spindle forms by co-assembly of the divergent PhuZ monomers which share only 37% sequence identity [15]. To understand the molecular basis of nuclear mispositioning, we expressed GFP fusions to either ΦPA3 PhuZ or ΦKZ PhuZ at low levels below the critical threshold for filament assembly, and

followed spindle assembly and dynamics during cognate or cross-infection with ΦPA3. When ΦPA3 infected cells expressing the cognate sfGFP-ΦPA3PhuZ, a properly functioning bipolar spindle formed, flanking a centrally located phage nucleus as expected (Figures 3.2D and G, green). However, when ΦPA3 infected cells expressing sfGFP-ΦKZPhuZ (Figures 3.2F and H) or untagged ΦKZPhuZ (Figure S3.3), the spindles were malformed, usually resulting in one long static filament within the cell (Figure 3.2F), and the nuclei were as severely mispositioned (Figure 3.2H, pink) as in cells expressing a ΦPA3PhuZD190A catalytic mutant [15] that renders the spindle inactive (Figures 3.2E and G, purple).

This suggests that the two divergent PhuZ monomers interfere with one another by co-assembling in a non-productive manner. To directly observe if these proteins can co-assemble, we simultaneously expressed both mCherry-ΦKZPhuZ and sfGFP-ΦPA3PhuZ proteins in *P. aeruginosa*. Upon infection with ΦPA3, instead of forming a dynamic bipolar spindle as occurs when expressing sfGFP-ΦPA3PhuZ alone (Figure 3.2D; [15]), hybrid spindles composed of both ΦPA3 PhuZ and ΦKZ PhuZ monomers were formed, and the nuclei were mispositioned (Figure 3.3A). Fluorescence recovery after photobleaching (FRAP) demonstrated that normal PhuZ spindles treadmill (Figure 3.3B, left; [14]), while the hybrid spindles were not dynamic (Figure 3.3B, right). Nuclei are mispositioned during coinfections of cells that do not express GFP-tagged PhuZ proteins (Figure S3.3). Taken together, these results demonstrate that PhuZ monomers from the two phages co-assemble nonproductively, disrupting the dynamic properties of the spindle. Given the central roles of the PhuZ spindle in nucleus centering [12, 13, 15, 16], nucleus rotation [14, 16], and capsid transport for DNA packaging [14], the nonfunctional hybrid spindles suggest that PhuZ is a speciation factor that limits the ability of these two phages to reproduce while coinfecting a single host cell. Analogous to pre-zygotic incompatibilities experienced by

eukaryotes where combinations of alleles prevent the production of hybrid offspring, hybrid spindles present an incompatibility that limits the production of progeny during the process of viral replication. We term this Virogenesis Incompatibility, in which viral encoded factors create barriers to the successful production of two different viruses within the same cell. Since loss of spindle function causes a 50% decrease in phage production [13], we estimate that the Virogenesis Incompatibility caused by hybrid spindles results in a 50% reduction in progeny phage during coinfection. Since we found an incompatibility factor in the cytoplasm, we reasoned that nuclear incompatibility factors might be important when hybrid nuclei between two distinct phages are formed, which occurs in ~25% of ΦKZ/ΦPA3 coinfections (Figure 3.1D). We hypothesized that nucleases normally present in the nucleus may diverge between species and, if introduced into the opposing phage's nucleus, would reduce its fitness. We identified five potential nucleases from ΦPA3 and imported them into the ΦKZ nucleus by taking advantage of our fortuitous finding that GFPmut1 and any proteins fused to it, are imported into the nucleus of ΦKZ but not ΦPA3 (Figure 3.4C) [20]. In contrast, sfGFP does not alter protein localization and naturally resides in the cytoplasm (Figure 3.4C), serving as a negative control. Of the five ΦPA3 proteins tested, only the putative endonuclease gp210 strongly inhibited ΦKZ reproduction when imported into its nucleus. Bioinformatically, gp210 has an HNH nuclease domain, but its specific biochemical activities are presently uncharacterized. When tagged with either GFPmut1 or sfGFP, and expressed from a plasmid with 1% arabinose, gp210 localized inside the nucleus of ΦPA3 without affecting the efficiency of plating (EOP) (Figures 3.4A and C). In contrast, importing gp210-GFPmut1 into the ΦKZ nucleus decreased EOP by 99.4% compared to GFPmut1 alone (Figures 3.4B and D). The cytoplasmically localized gp210-sfGFP control had a relatively small effect (~60% reduction) on ΦKZ EOP (Figure 3.4B). When gp210-GFPmut1 was expressed at low levels with 0.1%

arabinose, the EOP of ΦKZ was reduced by ~93% while cytoplasmic gp210-sfGFP was only reduced by 20% (Figures S3.5A and B). This suggests a reduction in fitness resulting from importing an endonuclease from a closely related phage, representing another type of Virogenesis Incompatibility. Surprisingly, when we examined ΦKZ replication in cells expressing gp210-GFPmut1, we saw no obvious effect on the first round of phage replication as judged by nucleus size or lysis time (Figures S3.5C–E). We used bacterial growth curves to further quantify the effect of gp210 on phage fitness over multiple generations in liquid cultures. Ten-fold serial dilutions of phage were added to exponentially growing cells expressing either gp210-sfGFP or gp210-GFPmut1 at a low level from an uninduced arabinose promoter and then cell growth (OD_{600}) was monitored over time. If the imported gp210-GFPmut1 impairs ΦKZ fitness, we expect an increase in the concentration of ΦKZ required to inhibit 50% of cell growth (IC50) after 6.5 hours. We found the IC50 of ΦKZ for cells expressing control proteins sfGFP or GFPmut1 was a multiplicity of infection (MOI) of 5×10^{-3} (Figures 3.4E and F). In comparison, the IC50 of ΦKZ for cells expressing gp210-GFPmut1 was an MOI of 500 (Figure 3.4G). This shows that 100,000-fold more ΦKZ phage were required to inhibit the growth of cells by 50% when gp210-GFPmut1 was imported into the nucleus compared to the GFPmut1 control. As another control, the IC50 of ΦKZ for cells expressing gp210-sfGFP (which is localized in the cytoplasm) was an MOI of 5×10^{-3} (Figure 3.4H) which is similar to the IC50 in the cells expressing control proteins sfGFP or GFPmut1. These experiments demonstrate that gp210 is a native nuclear protein of ΦPA3 that is naturally excluded by the ΦKZ nucleus and is incompatible with ΦKZ replication when imported. Such nuclear incompatibility factors can provide continued selection for the nucleus, thereby reinforcing Subcellular Genetic Isolation and consequent divergence of these two species.

3.5 Discussion

The Biological Species Concept defines species as interbreeding populations that are reproductively isolated from other populations [21]. Speciation factors in eukaryotes establish reproductive barriers that allow genetic divergence and the formation of new species. Here we describe three viral speciation factors that we categorize into two distinct reproductive isolation mechanisms. Subcellular Genetic Isolation can act by reducing genetic exchange between viruses coinfecting a single host cell, as in the case of the phage nucleus that physically separates phage genomes. Virogenesis Incompatibility can act by limiting the production of viable viral particles and reducing viral fitness, as occurs with incompatible hybrid spindles or the presence of a nuclear incompatibility factor such as gp210. These general principles of Subcellular Genetic Isolation and Virogenesis Incompatibility provide an understanding of the underlying mechanisms of viral speciation. We calculated that these mechanisms are sufficient for nearly complete genetic isolation between these two phages. The probability of a successful hybrid was estimated by multiplying the effects of Subcellular Genetic Isolation due to the phage nucleus (25%), relative genome frequency due to competition (94%), and Virogenesis Incompatibility due to hybrid spindles (50%) and a nuclear incompatibility factor (0.59%). While we cannot discern from these data if any of these factors caused the speciation of the two phages, we have demonstrated that they significantly reduce the chance of obtaining a successful hybrid to 0.07% under ideal circumstances where two diverging phages simultaneously infect the same cell. These results show Subcellular Genetic Isolation and Virogenesis Incompatibility lead to a limitation of gene flow, and together with previous work, directly support the application of the Biological Species Concept to viral speciation.

In line with other speciation mechanisms like Dobzhansky-Muller genetic incompatibilities, we predict that each of these mechanisms will have a greater impact on isolation with increased evolutionary divergence [22]. As the nucleus and spindles diverge, they are likely to accumulate functional and structural differences that result in increasingly defective configurations when they co-assemble. When mutations accumulate between diverging strains, evolving nucleases that are beneficial to one phage can become detrimental to a divergent phage. As shown in Figure S3.2, when strains diverge in their competitive fitness they are less likely to recombine. Taken together, we expect that these mechanisms will favor genetic exchange between closely related strains, and create increasingly reinforced barriers between divergent strains.

These evolutionary principles likely apply to many viruses and help to explain their rapid evolution and diversity. *Herpesvirus* has been shown to form spatially separated replication compartments (Subcellular Genetic Isolation) that limit genetic exchange but recombination occurs when these compartments coalesce [18]. In comparison, the phage nucleus also limits genetic exchange of viral genomes regardless of intracellular proximity. The phage nucleus, which may have evolved to provide protection against host defenses [15, 16, 23, 24], is likely widespread since it has been observed during infection of *Serratia* with phage PCH45 [23], and we have identified shell homologs in many distinct phage families infecting a diverse range of hosts including *Salmonella, Ralstonia, Cronobacter, Erwinia, Vibrio, and E. coli*. The PhuZ tubulin that makes up the spindle is also conserved and the genetic divergence that has occurred between the PhuZ proteins of ΦPA3 and ΦKZ results in Virogenesis Incompatibility that further limits exchange and drives speciation. Virogenesis Incompatibility could conceivably occur for any virus through the co-assembly of divergent proteins that make up any macromolecular structure (including virion structural proteins) required for phage propagation. These are the first described

examples of intracellular speciation factors for any virus, demonstrating that even viruses evolve traits that facilitate reproductive isolation. This discovery suggests that all domains of life, including viruses, are challenged to optimize genetic exchange by way of speciation in order to enhance the supply of adaptive variation while minimizing the influx of incompatible genes

3.6 Acknowledgements

This research was supported by National Institutes of Health grants GM104556 (J.P.) and GM129245 (J.P.) and by National Science Foundation grant DEB1934515 (J.R.M). We thank Eray Enüstun and Chase Morgan for carefully reading the manuscript.

Chapter 3, in entirety, is a reprint of the material as it appears in Chaikeeratisak, V., Birkholz, EA., Prichard, A., Egan, ME., Mylvara, A., Nonejuie, N., Nguyen, KT., Sugie, J., Meyer, JR., & Pogliano, J., 2021, Viral speciation through subcellular genetic isolation and virogenesis incompatibility in Nature Communications. The dissertation author was a primary author and the secondary investigator of this material.

Author contributions

V.C., E.A.B., A.P., A.M., and M.E.E. conducted experiments and contributed to analysis and writing of the paper. E.A.B., A.P., and J.P. conceptualized the original manuscript. K.N., P.N., J.S., J.R.M., and J.P. contributed to experimental design, analysis, and writing of the paper.

3.7 Figures

Figure 3.1 Subcellular Genetic Isolation occurs between identical and divergent coinfecting phages. (A) *P. aeruginosa* infected for 50 minutes with either ΦPA3 or ΦKZ frequently harbors more than one phage nucleoid. Cell membranes (magenta), DNA (cyan), GFP (green). (B) GFP-tagged ΦPA3 shell (green) and mCherry-tagged ΦKZ shell (red) reveals two nucleoids separated by nuclear shells at 50 mpi. (C, D) Cells expressing both GFP-ΦPA3shell and mCherry-ΦKZshell and were infected with either ΦPA3 or ΦKZ (C) or both phages (D). (C) GFP-ΦPA3shell only forms a shell (green) when infected with ΦPA3 and mCherry-ΦKZshell only forms a shell (red) when infected with ΦKZ. Phage DNA is blue. (D) Coinfected cells formed two separate red and green nuclei (~75%) or one nucleus with both red and green shell components (~25%). (E-H) n = 89 coinfected cells, (E) Percentage of coinfected cells with a larger ΦPA3 or ΦKZ shell. (F) Distribution of coinfected cells based on DAPI intensity inside shells. (G, H) Distribution of single-species coinfecting nuclei by DAPI intensity for ΦPA3 or ΦKZ, respectively. Using a cut off of at least 5% difference between the two values to establish if they are different from one another shows that 88% of the nuclei have unequal DNA content. Scale bars in panels (A-D) are 1 μm.

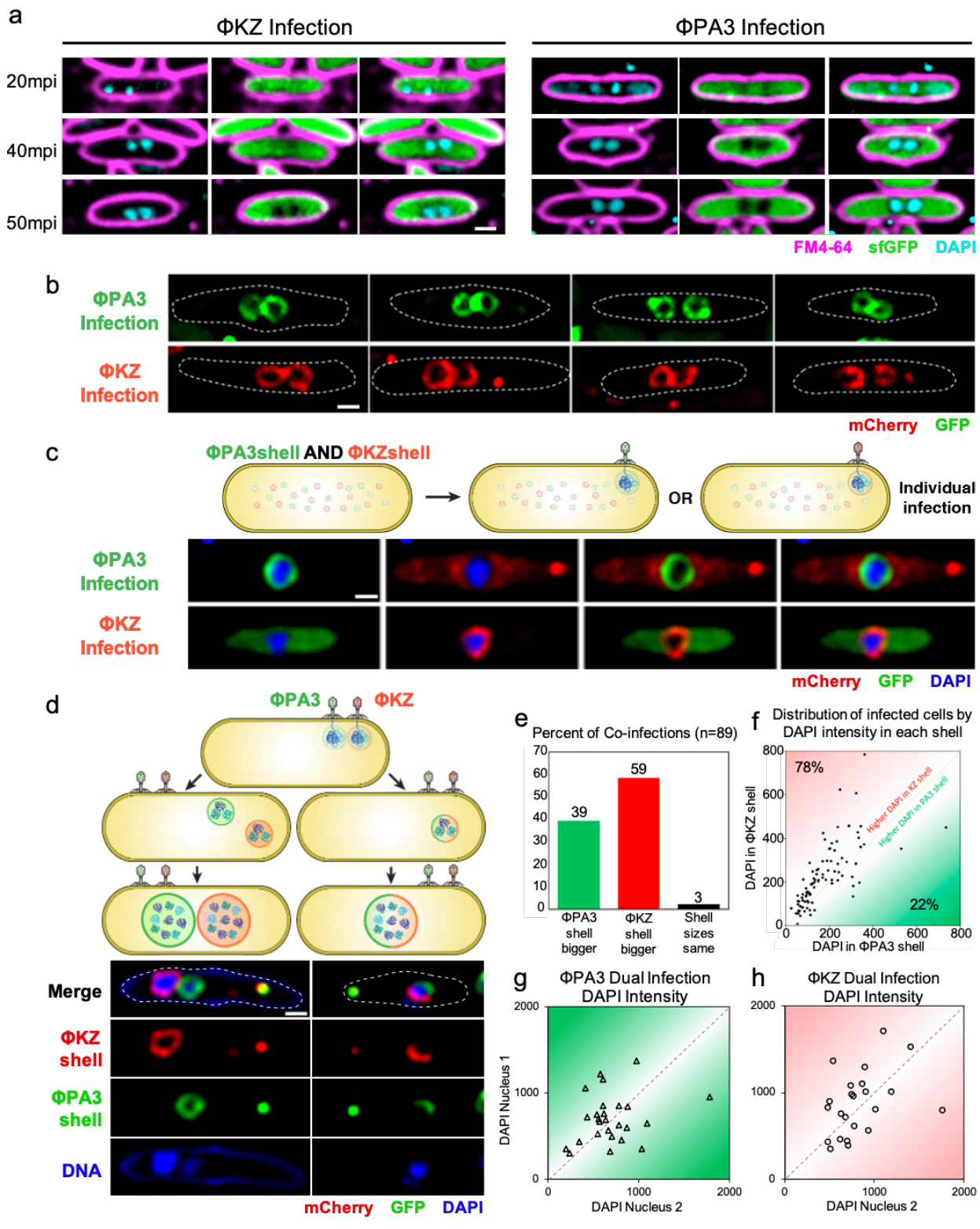
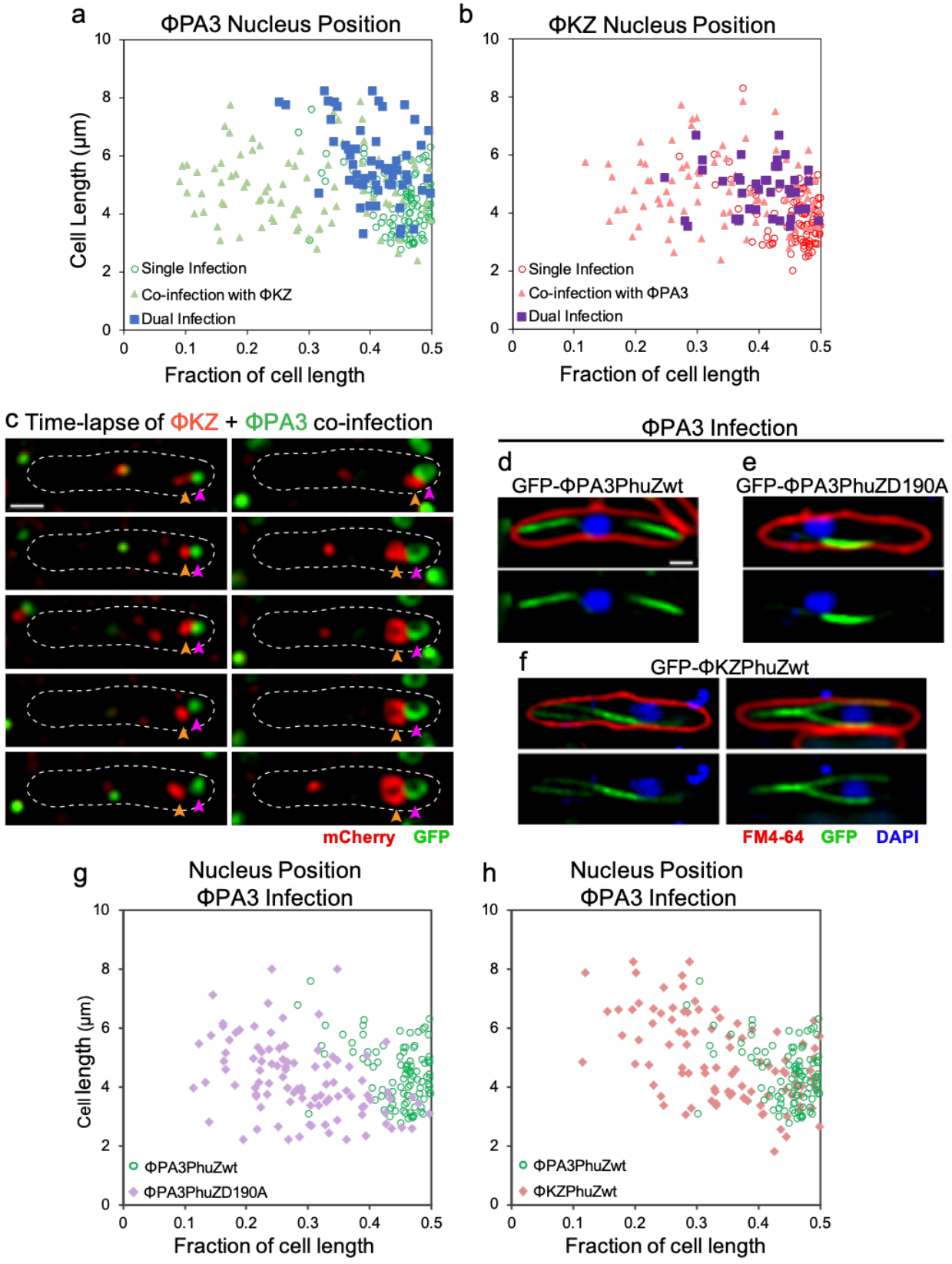

a
ΦKZ Infection
ΦPA3 Infection
20mpi
40mpi
50mpi
FM4-64 sfGFP DAPI
b
ΦPA3 Infection
ΦKZ Infection
mCherry GFP
c
ΦPA3shell AND ΦKZshell
OR
Individual infection
ΦPA3 Infection
ΦKZ Infection
mCherry GFP DAPI
d
ΦPA3
ΦKZ
Merge
ΦKZ shell
ΦPA3 shell
DNA
mCherry GFP DAPI
e
Percent of Co-infections (n=89)
70
60
50
40
30
20
10
0
39
59
3
ΦPA3 shell bigger
ΦKZ shell bigger
Shell sizes same
f
Distribution of infected cells by DAPI intensity in each shell
800
600
400
200
0
78%
22%
Higher DAPI in KZ shell
Higher DAPI in PA3 shell
DAPI in ΦKZ shell
0
200
400
600
800
DAPI in ΦPA3 shell
g
ΦPA3 Dual Infection DAPI Intensity
2000
1000
0
DAPI Nucleus 1
0
1000
2000
DAPI Nucleus 2
h
ΦKZ Dual Infection DAPI Intensity
2000
1000
0
0
1000
2000
DAPI Nucleus 2

Figure 3.2 Nuclei are mispositioned during cross-species coinfections. (A-C) Cells expressing both GFP-ΦPA3shell (green) and mCherry-ΦKZshell (red) were infected with either ΦPA3 or ΦKZ or both. (A) ΦPA3 nucleus position when singly infected with ΦPA3 (open circles, n = 100), when dual infected with ΦPA3 (solid squares, n = 56) resulting in two ΦPA3 nuclei in a single cell, or when coinfected with ΦKZ (solid triangles, n = 89). (B) ΦKZ nucleus position when singly infected with ΦKZ (open circles, n = 100), when dual infected with ΦKZ (solid squares, n = 40) resulting in two ΦKZ nuclei in a single cell, or when coinfected with ΦPA3 (solid triangles, n = 89). (C) Time-lapse (sec, seconds) of coinfecting nuclei failing to migrate to midcell (mpi, minutes post-infection). (D-F) Cells expressing wildtype sfGFP-ΦPA3PhuZ (D), mutant sfGFP-ΦPA3PhuZD190A (E), or wildtype sfGFP-ΦKZPhuZ (F) showing the nucleus (blue, DAPI) and spindle (green, GFP), cell membrane (red, FM4-64) when infected with ΦPA3. (G, H) ΦPA3 nucleus position displayed as fraction of cell length for cells expressing wildtype sfGFP-ΦPA3PhuZ (G, H, green, n = 100), mutant sfGFP-ΦPA3PhuZD190A (G, purple, n = 100), or wildtype sfGFP- ΦKZPhuZ (H, pink, n = 100). Scale bar in panel (D) equals 1 μm and all panels in (C-F) are at the same scale. Data for ΦPA3 nucleus position is repeated in graphs (A, G & H) for reference.


a
ΦPA3 Nucleus Position
Cell Length (μm)
Fraction of cell length
Single Infection
Co-infection with ΦKZ
Dual Infection
b
ΦKZ Nucleus Position
Fraction of cell length
Single Infection
Co-infection with ΦPA3
Dual Infection
c Time-lapse of ΦKZ + ΦPA3 co-infection
mCherry GFP
ΦPA3 Infection
d
GFP-ΦPA3PhuZwt
e
GFP-ΦPA3PhuZD190A
f
GFP-ΦKZPhuZwt
FM4-64 GFP DAPI
g
Nucleus Position
ΦPA3 Infection
Cell length (μm)
Fraction of cell length
ΦPA3PhuZwt
ΦPA3PhuZD190A
h
Nucleus Position
ΦPA3 Infection
Fraction of cell length
ΦPA3PhuZwt
ΦKZPhuZwt

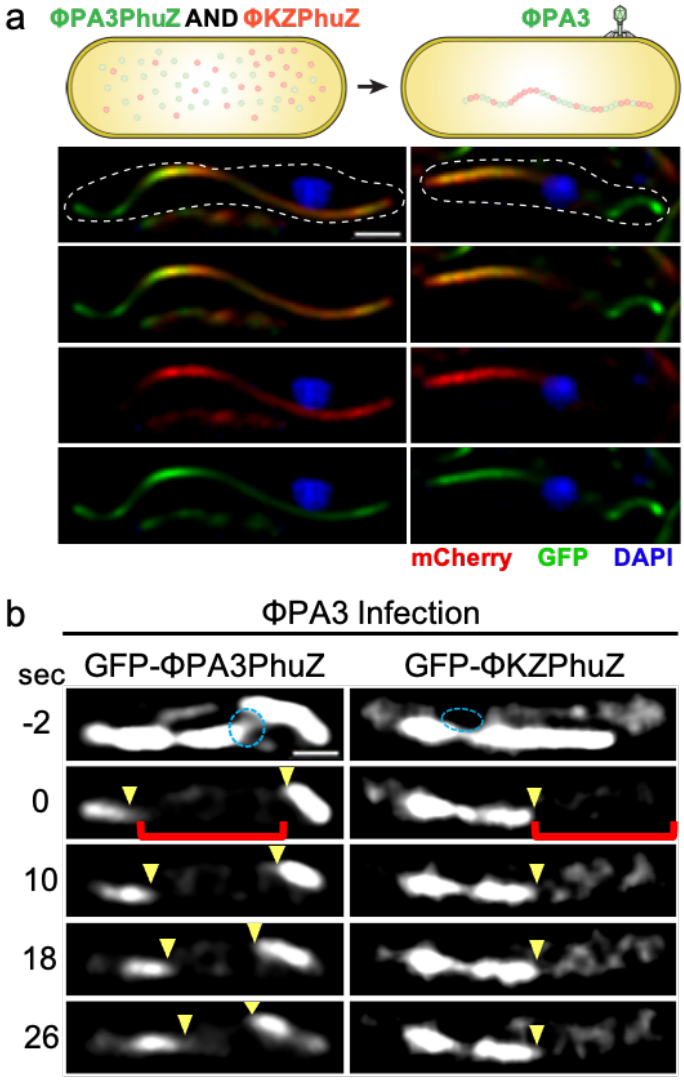
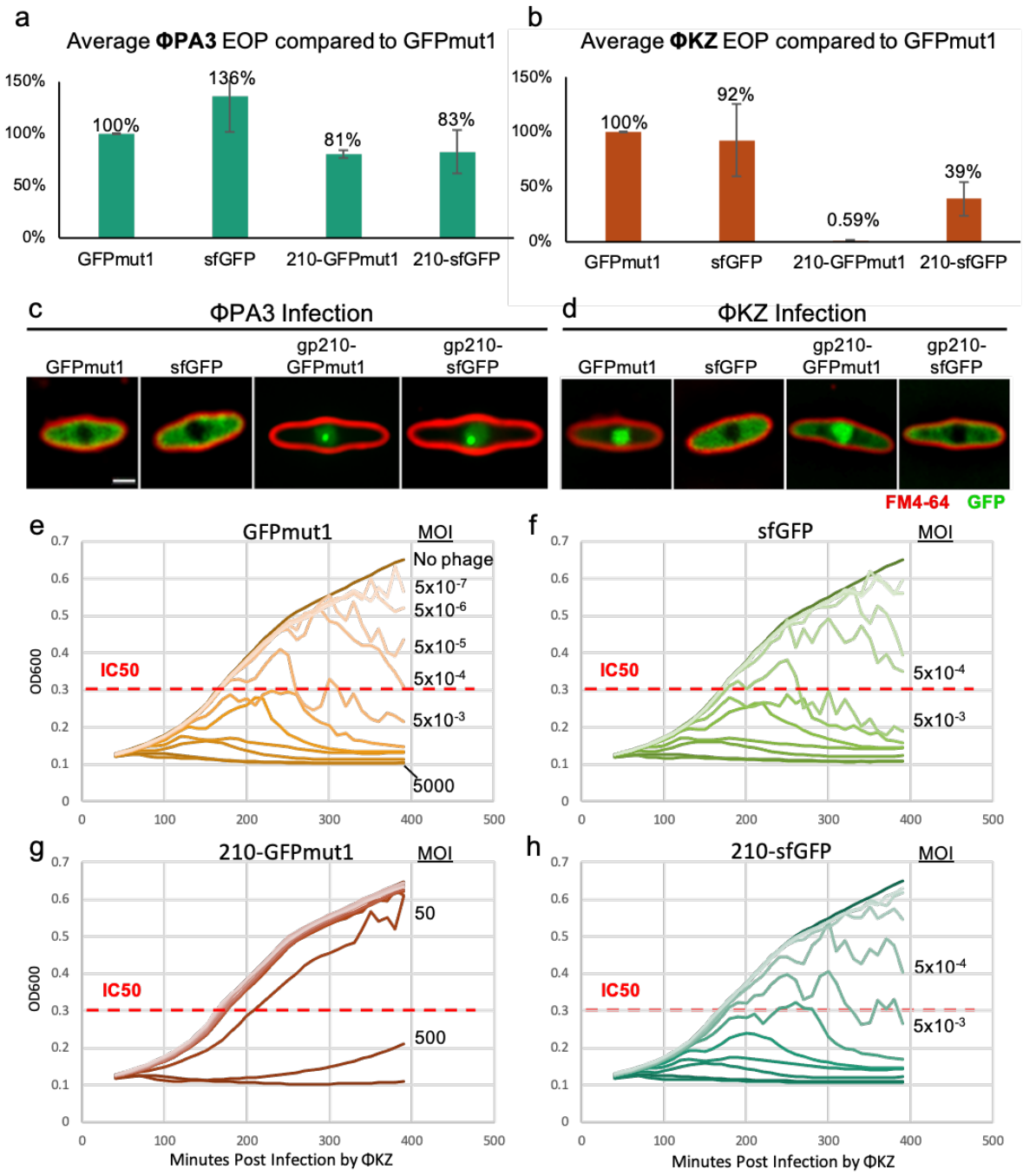


Figure 3.3 Nonfunctional hybrid spindles form through cross-species mixing of PhuZ monomers. (A) Simultaneous expression of sfGFP-ΦPA3PhuZ (green) and mCherry-ΦKZPhuZ (red) results in hybrid filaments (yellow) and mispositioning of the ΦPA3 nucleus (blue). (B) Photobleaching of sfGFP-ΦPA3PhuZ (left) shows movement of bleached zones indicating treadmilling (yellow arrows) during ΦPA3 infection. Photobleaching of hybrid spindles formed by expressing sfGFP-ΦKZPhuZ and cross infecting with ΦPA3 demonstrates filaments are not dynamic. Scale bars equal 1 μm.

Figure 3.4 A nuclear incompatibility determinant impairs phage fitness. (A) Efficiency of plating (EOP) relative to GFPmut1 for ΦPA3 in cells expressing the indicated fusions with 1% arabinose. The number of independent replicates, n, equals 4 (GFPmut1), 3 (sfGFP), 4 (gp210-GFPmut1), and 4 (gp210-sfGFP). Error bars show standard deviation. (B) EOP relative to GFPmut1 for ΦKZ in cells expressing the indicated fusions with 1% arabinose demonstrate a 99.4% decrease in viable ΦKZ with gp210-GFPmut1 in the nucleus. The number of independent replicates, n, equals 7 (GFPmut1), 6 (sfGFP), 6 (gp210-GFPmut1), 6 (gp210-sfGFP). Error bars equal standard deviation. (C, D) GFP fusions (green) and FM4-64 stained cell membranes (red). Individual data points are shown as circles. (C) GFPmut1 and sfGFP localize to the cytoplasm during ΦPA3 infections. gp210-GFPmut1 and gp210-sfGFP localize to the nucleus and form puncta during ΦPA3 infections. (D) sfGFP and gp210-sfGFP localize to the cytoplasm of ΦKZ infected cells. GFPmut1 and gp210-GFPmut1 localize to the nucleus of ΦKZ infected cells. (E-H) Determination of ΦKZ IC50 (red dotted line) for cells expressing control proteins GFPmut1 or sfGFP (E, F) or fusions to gp210 (G, H) as indicated by measuring cell growth (OD_{600}) over 6.5 hours of infection. Ten-fold serial dilutions of phage were added to cells resulting in a multiplicity of infection (MOI) ranging from 5000 to 5×10^{-7} as shown in (D). All growth curves represent an average of eight independent trials. For the sake of clarity, only the MOIs neighboring the IC50 threshold are shown for (F-H).


a
Average ΦPA3 EOP compared to GFPmut1
150%
100%
50%
0%
100%
136%
81%
83%
GFPmut1
sfGFP
210-GFPmut1
210-sfGFP
b
Average ΦKZ EOP compared to GFPmut1
150%
100%
50%
0%
100%
92%
0.59%
39%
GFPmut1
sfGFP
210-GFPmut1
210-sfGFP
c
ΦPA3 Infection
GFPmut1
sfGFP
gp210-GFPmut1
gp210-sfGFP
d
ΦKZ Infection
GFPmut1
sfGFP
gp210-GFPmut1
gp210-sfGFP
FM4-64
GFP
e
GFPmut1
MOI
No phage
5x10-7
5x10-6
5x10-5
5x10-4
5x10-3
5000
IC50
OD600
f
sfGFP
MOI
5x10-4
5x10-3
IC50
g
210-GFPmut1
MOI
50
500
IC50
OD600
Minutes Post Infection by ΦKZ
h
210-sfGFP
MOI
5x10-4
5x10-3
IC50
Minutes Post Infection by ΦKZ

a

mpi	ΦPA3	ΦKZ
20	24%	19%
50	21%	18%

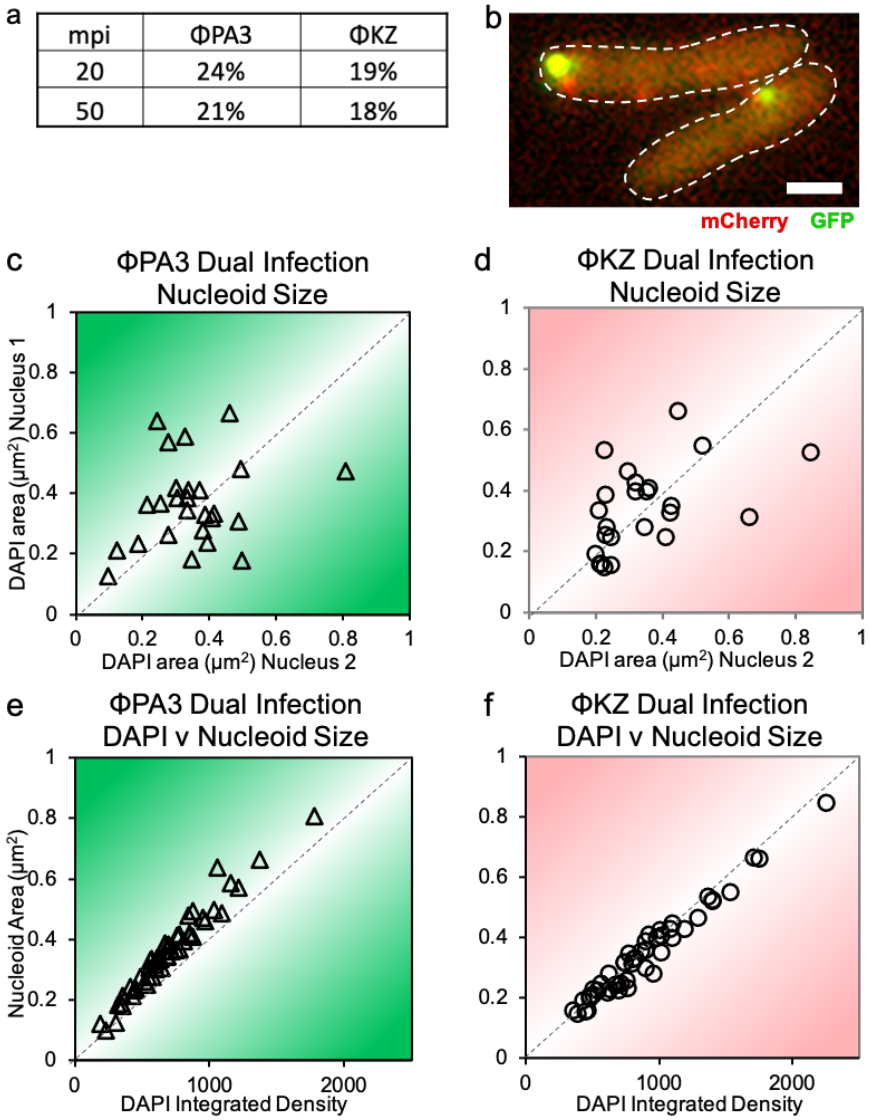


Figure S3.1 Viral nucleoid size and shell size of single species dual infections. (A) Table of single species dual infection percent per total infected cells, observed at 20 or 50 minutes post infection (mpi). For ΦKZ, n = 368 (20 mpi) and n = 307 (50 mpi). For ΦPA3, n = 156 (20 mpi) and n = 142 (50 mpi). (B) Uninfected *P. aeruginosa* expressing GFP-ΦPA3shell (green) and mCherry-ΦKZshell (red) form diffuse fluorescence and foci. (C, D) Nucleoid area as measured by area of DAPI staining and plotted for two nucleoids in single cell for ΦPA3 (C, n = 26 cells) and ΦKZ (D, n = 25 cells) dual infections. (E, F) The relative amount of nucleoid DNA was estimated by measuring integrated DAPI intensity and found to be proportional to nucleoid area for ΦPA3 (E, n = 52) and ΦKZ (F, n = 50). (G) Average size of each shell during a coinfection (n = 89).

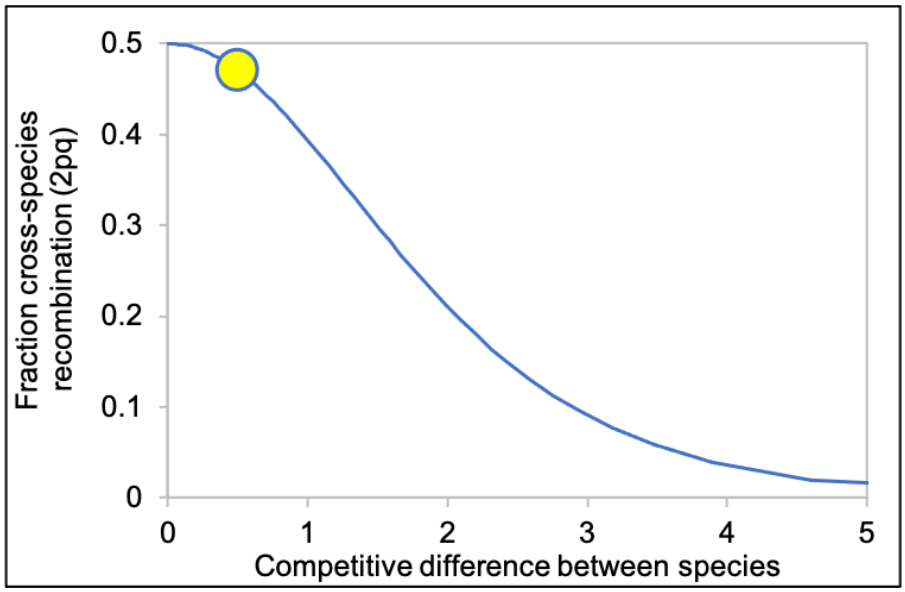


Figure S3.2 Fraction of cross-species recombination varies with competitive imbalance between species. Competitive differences were calculated using a modified selection rate equation, which computes the difference in Malthusian growth parameters over a period of time (t). Here, t will equal one round of infection and thus factors out of the equation. Competitive difference = ln(p/0.5) – ln(q/0.5). The yellow circle indicates the value calculated for ΦPA3 and ΦKZ based on their relative DNA concentrations.

Figure S3.2 Fraction of cross-species recombination varies with competitive imbalance between species (continued). Calculation of assortative mating frequency. One mechanism that can reduce the chance of cross-species phage recombination, even in the absence of protected shells, is through intracellular competition. Imagine a scenario where phage species infect a cell simultaneously and their genomes begin to replicate diffusely throughout the cytoplasm. If the two phages reproduce equally well, then they will achieve equal frequencies intracellularly and each phage species will have an equal probability of recombining with conspecifics or the other species. Alternatively, if one phage is able to allocate more resources and replicate more DNA, then the probability of cross-species recombination will be reduced since the majority of the recombinations will happen between conspecific genomes of the dominant species. This can be shown mathematically with calculations that are similar to the Hardy Weinberg equation from population genetics. If a cell were infected by two phages and these phages vary in their growth rates, then their frequencies would shift from 0.5 and 0.5 to p and q. Additionally, if it is assumed that most recombination between genomes occurs near the end of DNA replication when phage DNA is at its highest concentration, then p and q can be used to predict the relative frequencies of intraspecies versus interspecies recombination. The probability that recombination will occur between two P genomes is equal to the likelihood that two P genomes encounter each other, which is the product of their frequency in the cell, p^2. The probability of recombination between two Q genomes is q^2. The total amount of intraspecies recombination will be $p^2 + q^2$. The probability of P recombining with Q is pq, and the probability of Q recombining with P is qp. Therefore, 2pq is the probability of cross-species recombination. When p = q, then 50% of the recombination is interspecies and 50% is intraspecies. As the difference between the phages' ability to reproduce is magnified, their final frequencies will diverge, and interspecies recombination will become increasingly rare (Figure 3.2S). We calculated 2pq for each individual coinfection based on DAPI staining intensity within each nucleus at 60 mpi and found the average 2pq value equaled 0.47. Therefore, the expected impact on the rate of recombination between ΦPA3 and ΦKZ based on genome frequency was a 6 percent reduction (Figure S3.2). This calculation makes many simplifying assumptions and additional factors could also contribute to genetic isolation. Among the additional factors we could conceive, all of them would reduce the potential for cross-strain recombination, making our estimations conservative. For example, our calculation assumes that the genetic differences between the strains does not impact the potential for them to recombine. However, it is well known that as DNA sequences accumulate more genetic mutations, they will be less likely to recombine [25]. Therefore, the probability of recombination will be the product of the potential for two genomes to encounter each other within a cell, as we have computed here, and how similar the genomes are to one another. Another factor that would also reduce the potential for cross-strain recombination is the timing of infection. The greater the time lag between infections, the less likely the different strains will recombine. There are many other mechanisms that could reduce recombination. Our goal was to show that a simple ecological mechanism, such as resource competition, can impact inter-strain recombination and thus reproductive isolation. To the best of our knowledge, the effects of intracellular viral competition on recombination had not been previously considered. Given that intracellular competition should arise between any two strains infecting the same cell, we predict that this mechanism could have a significant impact on establishing genetic isolation in the virosphere.

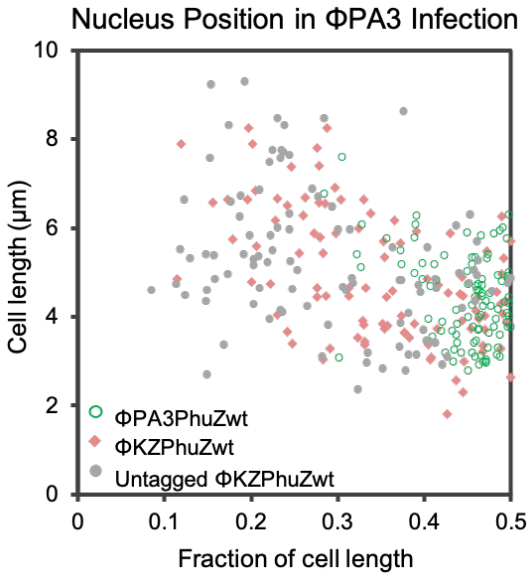


Figure S3.3 Nucleus positioning of ΦPA3-infected cells showing that the nuclei are mispositioned during cross-infections. ΦPA3 nucleus position is displayed as fraction of cell length for cells expressing wild type sfGFP-ΦPA3PhuZ (green, n = 100), or wild type sfGFP-ΦKZPhuZ (red, n = 100), or untagged ΦKZPhuZ (grey, n = 100).

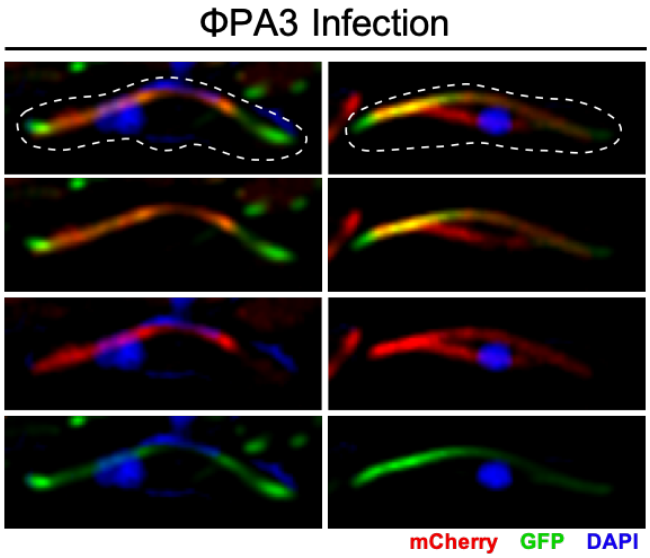


Figure S3.4 Two examples of nonfunctional hybrid spindles formed through cross species mixing of PhuZ monomers. Simultaneous expression of sfGFP-ΦPA3PhuZ (green) and mCherry-ΦKZPhuZ (red) results in hybrid filaments (yellow) and mispositioning of the ΦPA3 nucleus (blue).

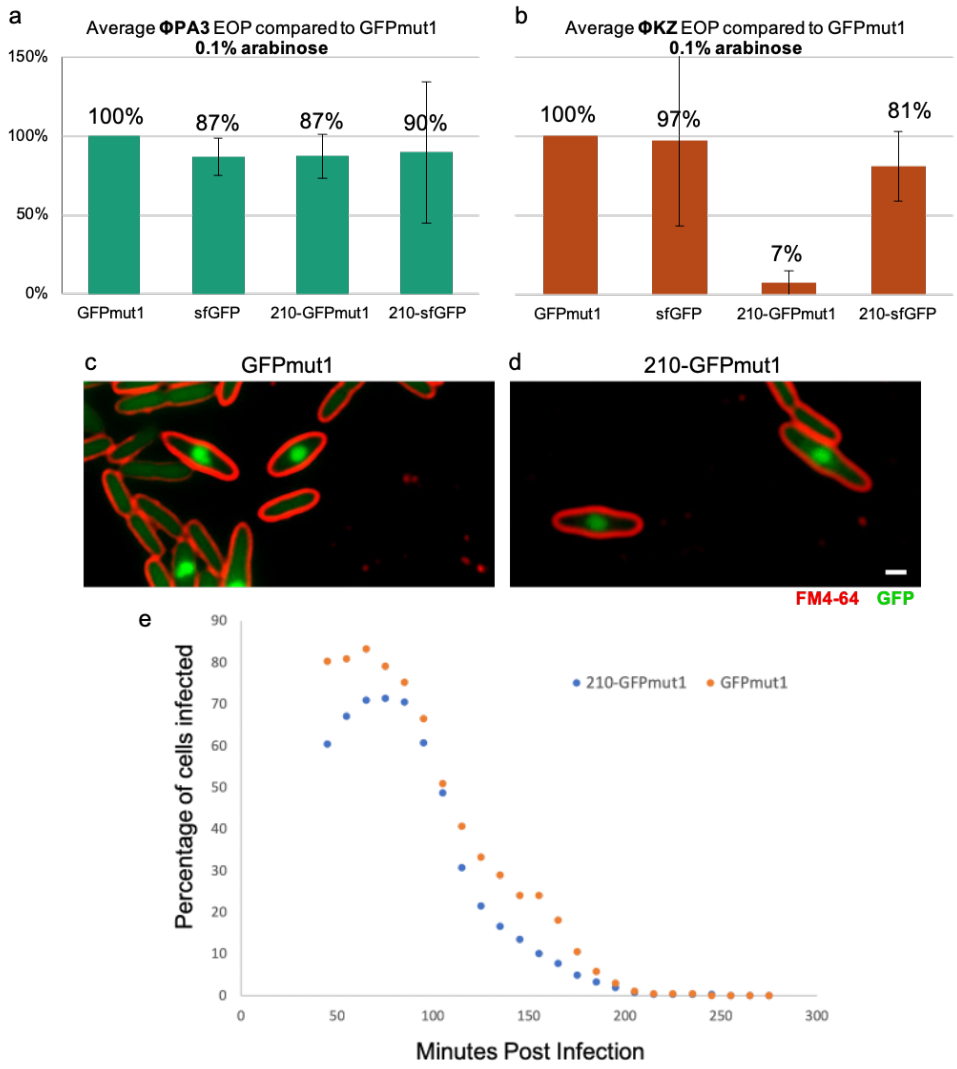


Figure S3.5 Low expression of gp210-GFPmut1 still incurs a large knockdown of ΦKZ titer but infection morphology and lysis not affected. (A) Efficiency of plating (EOP) relative to GFPmut1 for ΦPA3 in cells expressing the indicated fusions with 0.1% arabinose. The number of independent replicates, n, equals 7 (GFPmut1), 3 (sfGFP), 7 (gp210-GFPmut1), and 6 (gp210-sfGFP). Error bars equal standard deviation. (B) EOP relative to GFPmut1 for ΦKZ in cells expressing the indicated fusions with 0.1% arabinose demonstrate a 93% decrease in viable ΦKZ with gp210-GFPmut1 in the nucleus. The number of independent replicates, n, equals 19 (GFPmut1), 6 (sfGFP), 19 (gp210-GFPmut1), 18 (gp210-sfGFP). Error bars represent standard deviation. (C, D) Fields of cells at 70 mpi showing that *P. aeruginosa* cells expressing gp210-GFPmut1 (C, n = 3 replicates with 177 total cells) or GFPmut1 alone (D, n = 5 replicates with 172 total cells) both form large, centrally positioned nuclei. GFP (green), cell membranes stained with FM4-64 (red). ΦKZ replication in cells expressing gp210-GFPmut1 appears normal during a single round of infection. (E) The percentage of phage cells infected, as indicated by the presence of a nucleus, decreases over time as cells lyse at the end of the infection cycle. The timing of cell lysis after phage infection was measured in time-lapse microscopy. Cells expressing gp210-GFPmut1 (blue circles) or GFPmut1 alone (orange circles) were infected and followed for 4.5 hours.

3.8 References

1. Rohwer, F. and Barott, K., *Viral information.* Biology & Philosophy, 2013. **28**(2): p. 283-297.

2. Kristensen, D.M., Mushegian, A.R., Dolja, V.V., and Koonin, E.V., *New dimensions of the virus world discovered through metagenomics.* Trends in Microbiology, 2010. **18**(1): p. 11-19.

3. Juhala, R.J., Ford, M.E., Duda, R.L., Youlton, A., Hatfull, G.F., and Hendrix, R.W., *Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages.* J Mol Biol, 2000. **299**(1): p. 27-51.

4. Casjens, S.R., *Diversity among the tailed-bacteriophages that infect the Enterobacteriaceae.* Res Microbiol, 2008. **159**(5): p. 340-8.

5. Mavrich, T.N. and Hatfull, G.F., *Bacteriophage evolution differs by host, lifestyle and genome.* Nat Microbiol, 2017. **2**: p. 17112.

6. Bobay, L.M. and Ochman, H., *Biological species in the viral world.* Proc Natl Acad Sci U S A, 2018. **115**(23): p. 6040-6045.

7. Duffy, S., Burch, C.L., and Turner, P.E., *Evolution of host specificity drives reproductive isolation among RNA viruses.* Evolution, 2007. **61**(11): p. 2614-22.

8. Meyer, J.R., Dobias, D.T., Medina, S.J., Servilio, L., Gupta, A., and Lenski, R.E., *Ecological speciation of bacteriophage lambda in allopatry and sympatry.* Science, 2016. **354**(6317): p. 1301-1304.

9. Saxenhofer, M., Schmidt, S., Ulrich, R.G., and Heckel, G., *Secondary contact between diverged host lineages entails ecological speciation in a European hantavirus.* PLoS Biol, 2019. **17**(2): p. e3000142.

10. Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B., *Viral tagging reveals discrete populations in Synechococcus viral genome sequence space.* Nature, 2014. **513**(7517): p. 242-5.

11. Ramsey, J., Bradshaw, H.D., Jr., and Schemske, D.W., *Components of reproductive isolation between the monkeyflowers Mimulus lewisii and M. cardinalis (Phrymaceae).* Evolution, 2003. **57**(7): p. 1520-34.

12. Erb, M.L., Kraemer, J.A., Coker, J.K., Chaikeeratisak, V., Nonejuie, P., Agard, D.A., and Pogliano, J., *A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells.* Elife, 2014. **3**.

13. Kraemer, J.A., Erb, M.L., Waddling, C.A., Montabana, E.A., Zehr, E.A., Wang, H., Nguyen, K., Pham, D.S., Agard, D.A., and Pogliano, J., *A phage tubulin assembles*

dynamic filaments by an atypical mechanism to center viral DNA within the host cell. Cell, 2012. **149**(7): p. 1488-99.

14. Chaikeeratisak, V., Khanna, K., Nguyen, K.T., Sugie, J., Egan, M.E., Erb, M.L., Vavilina, A., Nonejuie, P., Nieweglowska, E., Pogliano, K., Agard, D.A., Villa, E., and Pogliano, J., *Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria.* Cell, 2019. **177**(7): p. 1771-1780.e12.

15. Chaikeeratisak, V., Nguyen, K., Egan, M.E., Erb, M.L., Vavilina, A., and Pogliano, J., *The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages.* Cell Reports, 2017. **20**(7): p. 1563-1571.

16. Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., Villa, E., Agard, D.A., and Pogliano, J., *Assembly of a nucleus-like structure during viral replication in bacteria.* Science, 2017. **355**(6321): p. 194-197.

17. Zehr, E.A., Kraemer, J.A., Erb, M.L., Coker, J.K., Montabana, E.A., Pogliano, J., and Agard, D.A., *The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA.* Structure, 2014. **22**(4): p. 539-48.

18. Tomer, E., Cohen, E.M., Drayman, N., Afriat, A., Weitzman, M.D., Zaritsky, A., and Kobiler, O., *Coalescing replication compartments provide the opportunity for recombination between coinfecting herpesviruses.* FASEB J, 2019. **33**(8): p. 9388-9403.

19. Kieser, Q., Noyce, R.S., Shenouda, M., Lin, Y.J., and Evans, D.H., *Cytoplasmic factories, virus assembly, and DNA replication kinetics collectively constrain the formation of poxvirus recombinants.* PLoS One, 2020. **15**(1): p. e0228028.

20. Nguyen, K.T., Sugie, J., Khanna, K., Egan, M.E., Birkholz, E.A., Lee, J., Beierschmitt, C., Villa, E., and Pogliano, J., *Selective transport of fluorescent proteins into the phage nucleus.* PLoS One, 2021. **16**(6): p. e0251429.

21. Mayr, E., *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. 1942: New York: Columbia Univ. Press.

22. Matute, D.R., Butler, I.A., Turissini, D.A., and Coyne, J.A., *A test of the snowball theory for the rate of evolution of hybrid incompatibilities.* Science, 2010. **329**(5998): p. 1518-21.

23. Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gumy, L.F., and Fineran, P.C., *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, 2020. **5**(1): p. 48-55.

24. Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J., *A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases.* Nature, 2020. **577**(7789): p. 244-248.

25. Watt, V.M., Ingles, C.J., Urdea, M.S., and Rutter, W.J., *Homology requirements for recombination in Escherichia coli.* Proc Natl Acad Sci U S A, 1985. **82**(14): p. 4768-72.

CHAPTER 4:

A homing endonuclease of ΦPA3 inhibits virogenesis of coinfecting phage ΦKZ

4.1 Abstract

Viruses constitute a large portion of the genetic diversity on Earth yet their taxonomy relies largely on sequence similarity. The boundaries between species of viruses can be biologically defined by the Biological Species Concept through the identification of speciation factors limiting the potential for recombination between viruses in different gene pools. The first viral intracellular speciation factors were discovered in nucleus-forming jumbo phage. The potential for gene flow between coinfecting phages was restricted by subcellular genetic isolation, the physical segregation of DNA, and virogenesis incompatibilities, the interference between machinery required for virion production. Here we characterize the virogenesis incompatibility factor gp210, encoded by ΦPA3, and show that it is an intronic endonuclease that targets ΦKZ DNA at a site homologous to the intron insertion site in the ΦPA3 genome. The intron containing gp210 interrupts a subunit of the ΦPA3 RNA polymerase and targets the allele of coinfecting phage ΦKZ which lacks an intron, acting as a virogenesis incompatibility factor by severely limiting virion production. Our studies indicate that homing endonucleases can reinforce the subcellular genetic isolation that maintains separation of phage genomes and their respective endonucleases. This presents an explanation for the disparities in selective protein import that has been observed between closely related phages. The prevalence of homing endonucleases in phage genomes infecting diverse hosts suggests that this is a widespread competitive mechanism that drives viral speciation.

4.2 Introduction

The mechanisms of viral speciation are not fully understood despite viruses constituting the majority of genetic information and diversity on our planet [1, 2]. Viruses are officially classified into taxa mainly by comparative sequence analyses with arbitrary thresholds [3-

6]. Virus genomes have been found to cluster into biological species where there is more genetic exchange within than between the clusters [7, 8]. This definition of a biological species as a gene pool comes from The Biological Species Concept [9, 10]. Gene pools can be separated spatially, temporally, or by mechanisms that prevent mating or recombination. Gene pools can also be isolated by incompatibilities, such as Dobzhansky-Muller genetic incompatibilities [9, 11-13], that reduce the production or virulence of the progeny generated from the mixing of two different gene pools. The divergence resulting from these reproductive isolating mechanisms can enhance or introduce speciation factors that provide a selection against the mixing of the gene pools.

In viruses, the identification of speciation factors in relation to the Biological Species Concept has only been reported for spatially separated populations [14-16]. Duffy et al. demonstrated that when RNA virus Φ6 was propagated on a new host, it lost the ability to infect the host of the ancestral virus, separating the gene pools [14]. Meyer et al. showed that two λ populations infecting two different hosts accumulated enough genetic divergence to result in genetic incompatibilities [15]. Saxenhofer et al. reported genetic isolation between two clades of a virus infecting a speciating rodent population and interestingly the amount of isolation between the virus clades exceeded that between the emerging rodent species [16]. Since viral genomes are only accessible to recombination when replicating inside a cell, viruses that are geographically separated or do not share host tropism are largely isolated from recombining their genes, making these the first barriers to be considered when defining viral gene pools.

The first characterized intracellular factors that separate the gene pools of coinfecting phages were observed in the nucleus-forming jumbo phages ΦKZ and ΦPA3 of *Pseudomonas aeruginosa* [17]. These phages encode proteins that interfere with the recombination of their genetic traits by two different mechanisms. The major nuclear shell proteins of each phage form

separate phage nuclei for 75% of coinfections. The shell protein even separates coinfecting genomes from the same phage stock for about 20% of infections and does not show evidence of fusing during infection [17]. This physical separation of coinfecting genomes was termed Subcellular Genetic Isolation, a speciation factor that has also been observed between the replication compartments of HSV-1 and vaccinia virus [18-20]. The bipolar spindle that centers the phage nucleus is composed of filaments of tubulin-like monomers called PhuZ [21-23], and the divergent PhuZ proteins from ΦKZ and ΦPA3 can co-assemble to create dysfunctional static polymers [17]. These filaments of incompatible monomers reduce the efficiency of the intracellular production of virions, or virogenesis. Their genomes have diverged enough that the mutations in the PhuZ gene have created a reinforcement in the limitation of genetic exchange, or speciation, by making coinfections less productive than single species infections. This type of speciation factor was termed Virogenesis Incompatibility [17]. Another virogenesis incompatibility found between these two phages involves a resident protein of the ΦPA3 nucleus, gp210. In nature, ~25% of the coinfections between these two phages would be subject to this speciation factor, when the shells co-assemble and potentially contain both genomes and their associated proteins [17].

The inhibition of ΦKZ by gp210 was discovered using our recent discovery that GFPmut1, along with a fused protein, is imported into the ΦKZ nucleus [24]. GFPmut1 is the only fluorophore tested that gets imported into ΦKZ but no fluorophores used were imported into the ΦPA3 nucleus. ΦPA3 does import tagged gp210 demonstrating it is naturally a resident of the nucleus while ΦKZ excludes gp210 without the GFPmut1 tag. This fascinating disparity in which proteins are imported into the nuclei of such closely related phages as ΦKZ and ΦPA3 has yet to be explained. By fusing gp210 to GFPmut1, we forced the import of this ΦPA3 endonuclease into

the ΦKZ nucleus and saw a great reduction in the fitness of ΦKZ (99.4% reduction by spot titer, 100,000-fold reduction in ability to lyse cells in liquid culture over 6 hours) [17]. Since gp210 is a nuclear protein of one phage that is drastically detrimental when in the nucleus of a coinfecting phage, it is a virogenesis incompatibility reinforcing the separation of these gene pools. It may also offer an explanation for the evolution of phage nucleus protein import mechanisms that differ between closely related phages. No details about the nature of gp210 and its activity against ΦKZ have been published. Here we present evidence supporting that gp210 is a homing endonuclease from a group I intron, and that it can target the ΦKZ genome, interfering with the production of virions.

Homing endonucleases are widespread in all kingdoms of life [25]. They target the DNA surrounding their open reading frames (ORFs), causing a DNA break in a conserved site of a homologous chromosome that does not contain the endonuclease gene, triggering recombination which results in unilateral gene conversion the intron(+) allele with the loss of the homing site [26-32]. While these homing endonucleases can be freestanding, they are also often found within group I or group II introns as well as inteins. Those self-splicing elements allow the homing endonuclease to invade highly conserved open reading frames by ensuring the essential genes of the host remain functional. The introns invaded essential genes at conserved residues such as enzyme active sites, so that any imperfect attempt to delete the intron could lead to a nonfunctional essential gene [27, 28, 33]. It has been proposed that once an intron was in place, a homing endonuclease, which also targeted highly conserved sequences to optimize spread to related but diverging populations, invaded the intron that offered protection against disruption of the host gene [27, 28, 34-38]. This creates a composite mobile genetic element consisting of an intron and a protein that promotes the intron's mobility, a mobile intron.

In phage, the most common introns are group I introns with a homing endonuclease interrupting genes involved in DNA metabolism [28]. Group I introns are also found interrupting mRNA, rRNA, and tRNA genes in every domain of life [29, 33, 39]. A stop codon often occupies the intron at the 5' splice site, to reduce RNA folding interference by ribosomes, which on the other hand, can chaperone RNA folding [28, 29]. The ribozyme formed by group I intron RNA is highly conserved in structure [33, 40] with a terminal loop that can tolerate large insertions, ripe for invasion by homing endonuclease ORFs [29, 33, 41].

Homing endonucleases in phage instigate intron mobility by employing the recombination machinery of phage replication [26, 28, 29]. They stimulate homologous recombination by creating either a double strand break (DSB) [42-44] or a single strand nick [41, 45-48]. Nicking produces a less toxic effect [49], suggesting an advantage of nicking to protect the host. When these introns invade a related genome, the dependence on homologous recombination will lead to less efficient repair between more divergent genomes [50, 51]. The sequence tolerant recombinase of phage λ, Redβ$_{\lambda}$, loses 100-fold efficiency when the recombining sequences are 22% diverged and recombination does not occur at 52% divergence [52, 53]. This feature of the homing process sets it up to reinforce speciation if the genomes diverge too much to efficiently recombine but the DNA targeting of the homing endonuclease is preserved through the most conserved amino acids.

These nucleases housed in the introns of phage are often elongated proteins that enable recognition of a DNA target region over 30 bp long, despite the small number of residues making up the protein [26, 28, 46, 54, 55]. Their loosely connected DNA-binding (NUMOD [56]) and catalytic domains allow non-uniform contacts in the major and minor groove of the target region, generating mismatch tolerant endonucleases that retain high specificity at only the most conserved residues [26, 28, 29, 54]. Homing endonucleases with an H-N-H/N domain are common in viruses

and they utilize a leading histidine to activate a water molecule as a nucleophile and an asparagine or histidine to coordinate the metal ion (magnesium or zinc), stabilizing the transition state [26, 46, 57]. The phage-encoded nicking homing endonucleases referenced above all belong to the HNH family.

Group I introns containing a homing endonuclease have been studied in the context of coinfecting phage but only at the genome level and not for the entire organism. Nicking HNH endonuclease I-PfoP3I from the group I intron in the DNA polymerase (DNAP) of cyanophage Pf-WMP3 is an example of a typical homing endonuclease [45]. It prefers the intron(-) allele of the homologous phage over related phage Pf-WMP4 and cuts both genomes 4 nucleotides upstream of the intron insertion site with the recognition region spanning only 14 bp. *Bacillus subtilis* phages SPO1 and SP82 also contain group I introns with nicking HNH homing endonucleases, I-HmuI and I-HmuII respectively, in their DNAPs [38, 58]. However, unlike many homing endonucleases, these two are able to cut both the intron(+) and intron(-) alleles of the DNAP and they each prefer the allele of the heterologous phage. Interestingly, I-HmuII cuts SPO1 DNA 52 nucleotides from the intron insertion site and excludes the I-HmuI intron while causing the co-conversion of the flanking genes [41] as seen similarly for the SegF and SegG homing endonucleases [30, 59]. This activity is termed ‘marker exclusion’ or ‘partial exclusion’ [60-64] and has been implicated as a selective advantage that only acts on the specific genes that benefit from co-conversion, resulting in a localized genetic isolation for a portion of the genome [60, 62].

These mobile intron systems are “selfish” genetic elements but whether or not their nuclease confers a benefit to the host organism has remained unclear [65-68]. When these mobile introns inhabit a virus, they have the potential to influence the fitness of certain genes, the virus itself, or the bacterial host. The selective advantage of mobile introns on certain genes is the

marker exclusion referenced above. We could not find any report of a selective advantage conferred by a viral intron to the host of the virus. For benefits to the entire virus from the intron, only stress-induced ectopic insertion has been posited, but not demonstrated, to provide an advantage by introducing genetic diversity, which could also be destructive [29, 69, 70]. Other host benefits that have been suggested merely mitigate the negative impact of the intron itself and do not provide any advantage over intron(-) competitors. There has not been any reported competitive advantage for an entire virus that is conferred by a homing endonuclease, though it has been hypothesized [67].

Here we characterize the virogenesis incompatibility factor gp210, a homing endonuclease encoded by ΦPA3, that was previously identified through its ability to significantly decrease the fitness of related coinfecting phage ΦKZ [17]. We show that ΦPA3 gp210 targets ΦKZ DNA at the unoccupied locus of the ΦPA3 intron. This nucleolytic activity was supported by an *in vitro* nuclease assay. Genetic and biochemical studies demonstrate that gp210 cuts a specific sequence in the ΦKZ DNA within a gene encoding a subunit of the virion-packaged RNA polymerase, gp178. Using cryo-FIB-ET to visualize the consequence of this cleavage on phage replication we demonstrated that the import of gp210 into the ΦKZ nucleus inhibited the production of new viral capsids. These studies allow us to propose a model in which gp210 acts as a virogenesis incompatibility factor by preventing the production of ΦKZ virions through targeting of gp178 which is required for viral capsid assembly.

4.3 Materials and Methods

Bacterial strains, phage, and growth conditions

P. aeruginosa PAO1 derivative K2733 was used as the host in all experiments. It was

grown at 30°C in LB and 15 μg/ml gentamycin was used for plasmid selection. Phage stocks of ΦKZ and ΦPA3 were collected from medium titer plate lysates using phage buffer (10 mM Tris (pH 7.5), 10 mM $MgSO_4$, 68 mM NaCl, and 1 mM $CaCl_2$) and stored at 4°C with titers of 10^{12} and 10^{10}, respectively.

Plasmid construction and transformation

Plasmids were either constructed by Gibson assembly [17] or synthesized and cloned by Genscript. The vector for all plasmids was pHERD-30T [17, 71-73] and Genscript cloned the inserts using SacI and SalI. Plasmid transformation was accomplished by 2 kV electroporation of K2733 cells washed with 300 mM sucrose and stored at -80°C.

Phylogenetic tree construction and alignments

All trees are unrooted neighbor-joining trees [74] constructed from a neighbor-joining MUSCLE alignment using MEGA X [75, 76]. Trees are bootstrapped [77] by 1000 replicates. Tree scale is given in units of amino acid substitutions per site with branch lengths equal to the evolutionary distances computed using the Poisson correction method [78]. Positions containing gaps were eliminated.

Protein expression, purification, and characterization

Full length wild-type gp210 was cloned into UC Berkeley Macrolab vector 1B (Addgene #29653) to generate N-terminal fusions to a TEV protease-cleavable His_6-tag. Proteins were expressed in *E. coli* strain Rosetta2 pLysS (EMD Millipore) by induction with 0.25 mM IPTG at 20°C for 16 hours. For protein purification, cells were harvested by centrifugation, suspended in resuspension buffer (20 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM imidazole, 5 mM 2-mercaptoethanol and 10% glycerol) and lysed by sonication. Lysates were clarified by centrifugation (16,000 rpm 30 minutes), then supernatant was loaded onto a 5 mL Ni^{2+} affinity

column (HisTrap HP, GE Life Sciences) pre-equilibrated with resuspension buffer. The column was washed with buffer containing 20 mM imidazole and 100 mM NaCl, and eluted with a buffer containing 250 mM imidazole and 100 mM NaCl. The elution was loaded onto an anion-exchange column (Hitrap Q HP, GE Life Sciences) and eluted using a 100-600 mM NaCl gradient. Fractions containing the protein were pooled and mixed with TEV protease (1:20 protease:gp210 by weight), then incubated 48 hours at 4°C for tag cleavage. Cleavage reactions were passed over a Ni^{2+} affinity column, and the flow-through containing cleaved protein was collected and concentrated to 2 mL by ultrafiltration (Amicon Ultra-15, EMD Millipore), then passed over a size exclusion column (HiLoad Superdex 200 PG, GE Life Sciences) in a buffer containing 20 mM Tris-HCl pH 7.5, 300 mM NaCl, and 1 mM dithiothreitol (DTT). Purified proteins were concentrated by ultrafiltration and aliquoted and frozen at -80°C for biochemical assays.

Nuclease activity assay

Plasmids were miniprepped from 5 mL of overnight culture of NovaBlue *E. coli* cells. Purified gp210 was mixed with 500 ng plasmid DNA in a buffer containing 10 mM Tris-HCl (pH 7.5), 25 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT (50 μL reaction volume), incubated 1 hour at 37°C, then separated on a 1.0% agarose gel. Gels were stained with ethidium bromide and imaged by UV illumination.

Efficiency of plating (EOP) by spot titers

To determine the efficiency of phage plaque formation on bacteria expressing different fusion proteins, previously published methods were followed [17]. Protein expression was achieved with the indicated fusions by inducing with 1.0% arabinose. For the EOP of progeny, the indicated expression protein was expressed during the replication of the progeny and the titers were performed on the *P. aeruginosa* host, K2733, without any plasmid present.

IC50 growth curves

Growth curves to determine the amount of phage required for suppression of the bacterial culture to 50% of maximum growth (IC50) were performed as described previously [17]. 95 µl of bacterial host at OD_{600} of 0.1 was combined with 5 µl of ten-fold serial dilutions of a ΦKZ lysate across a 96-well plate. The plates were shaken for an initial 40 minutes at 30°C in a microplate reader (Tecan Infinite MPlex and Tecan Sunrise) after which OD_{600} measurements were taken every 10 minutes, with continuous shaking at 30°C between the timepoints. The OD_{600} values were averaged and plotted as a growth curve. IC50 is annotated at half of the OD_{600} reached by the strain when grown with only phage buffer. Fusion proteins were expressed by 0.1% arabinose.

Single cell infection assay

Single cell infections of *P. aeruginosa* were visualized using fluorescence microscopy [17, 71-73]. Briefly, 1% agarose pads containing 25% LB, 2 µg/ml FM 4-64, and 1 µg/ml DAPI were inoculated with 5 µl of cells (OD_{600} 0.4) and incubated for 3 hours at 30°C in a humidor and then infected with 10 µl of a high titer phage lysate. Imaging was performed with a DeltaVision Elite Deconvolution microscope (Applied Precision, Issaquah, WA, USA). Images were further processed by the aggressive deconvolution algorithm in DeltaVision SoftWoRx 6.5.2 Image Analysis Program, and analyzed using Fiji 1.53c software [79].

Lysis time-lapse

Time-lapse to measure lysis was performed as previously described [17]. Infections were established on agarose pads (described above) without any stains. Beginning after 45 mpi, images were captured every 10 minutes for 4 hours using UltimateFocus. The cells that lysed during the imaging were counted and reported as a percentage of the total number of starting cells. The percentages were averaged and plotted for both bacterial host strains.

Progeny collection

ΦKZ progeny were generated in the presence or absence of functional or nonfunctional gp210. Homogeneous liquid cultures of *P. aeruginosa* K2733 were grown in LB for 3 hours at 30°C. The OD_{600} was measured and cultures were back-diluted to an OD_{600} of 0.05 and induced with 1% arabinose for 1 hour to express the protein of interest (GFPmut1, gp210-GFPmut1, gp210-sfGFP, or gp210H82A-GFPmut1). Cells were grown to an OD_{600} of about 0.1 and infected with ΦKZ at an MOI of 30 to ensure high infection rates. During early infection (30mpi, after cells would be infected but before lysis begins), cells were centrifuged at 3220 rcf for 10 minutes to pellet infected cells and the supernatant was removed, filtered through a 0.45 μm filter to remove any residual bacterial cells, and saved as the "parent" phage sample. Cells were resuspended in fresh LB and washed twice more by centrifugation and resuspension to remove any residual parent phage from the supernatant. When the infections would be entering later stages (80 mpi), centrifugation was halted to prevent premature lysis of infected cells and resuspended cells were concentrated 5-fold in LB to ensure a high titer of progeny phage and incubated for 2-4 hours at 30°C to allow for complete lysis of infected cells. After incubation, the cells were centrifuged for 1 minute at 21,130 rcf to pellet the cells and cell debris, and the supernatant containing the progeny phage lysate was collected and filtered through a 0.45 μm filter to remove any residual bacterial cells. Progeny and parent phages were stored at 4°C.

Cryo-FIB-ET

Cells expressing gp210-GFPmut1 were grown for 3 hours in a humidor at 30°C on 10 agarose pads (1% arabinose, 1% agarose, 25% LB). 10 μl of high titer ΦKZ lysate was added to the pads and incubated for another hour before 25 μl of 25% LB was added at RT and the cells

were gently scraped from the pads using the bottom of an eppendorf tube and collected for plunging.

A custom-built manual vitrification device (Max Planck Institute for Biochemistry, Munich) was filled with liquid nitrogen (LN2) to maintain a and an ethane/propane gas mixture was condensed to a liquid in a LN2-cooled cup in the device. Room humidity was kept around 30% to mitigate water contamination. Holey carbon-coated QUANTIFOIL® R2/1 copper grids were glow discharged (0.2mbar, 20mA, 60sec) 30 minutes prior to plunging using a Pelco easiGlow™ system. 7µL of concentrated infected cells were put on the carbon side of each grid. Samples were blotted with Whatman filter paper No. 1 to remove excess liquid and plunge-frozen in the liquid ethane/propane to be vitrified.

Grids were mounted into cryo-FIB AutoGrids (TFS) and milled as previously described using an Aquilos (TFS) dual-beam scanning electron microscope under cryogenic conditions [80]. Briefly, areas of interest were rough-milled with an ion-beam current of 0.1-0.50 nA, followed by fine milling and polishing at 30-50 pA. One grid was prepared for each condition to yield two lamellae for the wild-type infection and four lamellae for infected host cells expressing the gp210-GFPmut1 fusion.

Lamellae were imaged under cryogenic conditions using a Titan Krios (TFS) transmission electron microscope operated at 300 kV and equipped with a K2 direct electron detector (Gatan) mounted post a Quantum 968 LS imaging filer (Gatan). The microscope was operated in EFTEM using a 20 eV slit-width and the detector operated in counting mode. Data collection was performed semi-automatically using SerialEM [81]. Tilt-series were collected following a dose-symmetric scheme starting at the specimen pre-tilt and targeting a tilt-range of +/- 60^{o} in increments of either 2^{o}, 2.5^{o}, or 3^{o}, using a pixel size of either 0.3457 nm or 0.4265 nm. Exposure

times for each tilt-movie were adjusted to maintain constant counts on the detector throughout the tilt-series. The cumulative dose for each tilt-series was usually between ~150-200 e/Å^2 and defocus between 5 to 7 microns.

Tilt-movies were motion-corrected and dose-weighted using MotionCor2 [82]. Tilt-series alignments of motion-corrected stacks were performed by patch-tracking using Etomo from IMOD [83]. Tomograms were then reconstructed by Fourier inversion on the aligned stack from Etomo at a binning level of 4 using the implementation from EMAN2 [84].

Isolation and sequencing of ΦKZ mutants resistant to gp210

To obtain ΦKZ mutants that were able to plaque successfully when the host was expressing gp210-GFPmut1, whole plate infections were generated from 3 separate aliquots of 100 µl of cells (OD_{600} ~0.4) induced for over an hour with 0.1% arabinose and combined with 10 µl of ΦKZ lysate (10^{12}). Phage adsorption was allowed to proceed while stationary for 10 minutes at room temperature (RT) prior to the addition of 5 ml warm 0.35% LB top agar. The mixture was poured onto an LB plate with 15 µg/ml gentamycin and incubated overnight at 30°C in a humidor. Isolated plaques from each plate were streaked onto a new plate and overlaid with top agar containing 0.1% arabinose and cells expressing gp210-GFPmut1 then incubated overnight. Plaque isolation was repeated twice more and the final isolate was streaked to achieve web lysis. Those plates were soaked with 5 ml chilled phage buffer for 5 hours then collected and centrifuged at 3,220 rcf for 10 minutes at 4°C. The supernatant was filtered through a 0.45 µm Corning membrane filter by syringe, 5 drops of chloroform were added, it was shaken by hand for 2 minutes, then centrifuged again to separate the aqueous phase. Lysate was drawn off the top for standard titers on cells expressing GFPmut1 or gp210-GFPmut1. For mutants displaying resistance to gp210, 2 whole plate infections for each isolate were generated and the lysates collected as described above.

Phage genomic DNA was isolated from each mutant using 10 ml lysate incubated with 5 µl of each RNaseA (100 mg/mL) and DNaseI (20 mg/mL) at 37°C for 30 minutes. Then 4 ml of phage precipitant solution (30% PEG 8000, 19.3% NaCl in ddH_2O) was added and incubated overnight at 4°C. Samples were centrifuged at 10,000 rcf 4°C for 20 minutes and the pellets were resuspended in 0.5 ml sterile water for incubation at 5 minutes at RT. Then 2.5 ml of Qiagen Buffer PB was added and incubated at RT for 10 minutes with occasional swirling. The resuspensions were filtered through Qiagen PCR Purification columns and washed twice with Qiagen Buffer PE. The columns were dried with an extra 2-minute spin before 100 µl of 37°C Tris-EDTA buffer was added and allowed to soak for 5 minutes before elution.

Samples were sequenced by the Microbial Genome Sequencing Center (MiGS) in Pittsburgh. Whole genome sequencing was performed using the Illumina NextSeq 2000 platform at a depth of 200Mbp. MiGS provided paired end reads (2x151bp) and reported variations from the Genbank entry for ΦKZ (NC_004629.1).

Image quantitation

To measure differences in DNA content of the phage nuclei using DAPI concentration, cells expressing either GFPmut1 or gp210-GFPmut1 were grown with 1% arabinose and 1 µg/ml DAPI for 3 hours then infected by ΦKZ for 45 minutes. Raw images were analyzed in Fiji 1.53c [79]. The raw integrated density of each phage nucleus was measured by an inscribed circle on the Z slice closest to the middle of the nucleus. Background raw intensity was measured from empty space next to each infected cell that was measured. Intensity was normalized to the area of the region measured and the background subtracted. A violin plot of these values was generated using Prism 9.2.0 by GraphPad.

4.4 Results

Gp210 requires HNH endonuclease domain to inhibit ΦKZ

The ΦPA3 nucleus-residing protein that inhibits ΦKZ fitness [17], gp210, is predicted to contain an HNHc endonuclease domain (accession cl00083, e-value: 2.05 e^{-5}). A ClustalO alignment of 20 jumbo phage proteins with similarity to gp210 reveals high conservation of the H-N-N motif (Figure 4.1A, red asterisks). One of those proteins is I-HmuI of *Bacillus* jumbo phage SPO1 (Figure 4.1A, blue asterisk) which has been characterized in detail as an HNH homing endonuclease (PDB 1u3e) [46] and was the template used by Phyre2 to produce a 100% confidence match with the H-N-N motif conserved (Figure 4.1B), supporting the prediction that gp210 is an HNH endonuclease. These analyses predict that H82 of gp210 is necessary for nucleolytic activity so we mutated it to an arginine and produced gp210(H82R) from a plasmid in *P. aeruginosa*. It was previously established that ΦKZ plaque formation is significantly inhibited on a host that is expressing gp210-GFPmut1 [17] which is forcibly imported into the ΦKZ nucleus by GFPmut1 [24]. This result is reaffirmed here in Figure 4.1C by a 0.0017% efficiency of plating compared to cells expressing only GFPmut1, while the expression of gp210(H82R)-GFPmut1 results in a full rescue of ΦKZ. Gp210 without any tags is presumed to remain in the cytoplasm based on previous localization of gp210-sfGFP [17] and does not have a significant effect on ΦKZ titer. These proteins do not significantly affect ΦPA3 EOP (Figure S4.1A). In growth curves, ΦKZ required a 1,000-fold higher MOI to suppress 50% of the growth of cells expressing gp210-GFPmut1 and this defect was again fully rescued by the H82R mutation of gp210 (Figure 4.1E). ΦPA3 growth curves are not significantly affected by the expression of these proteins (Figure S4.1C). Gp210(H82R) is still imported into the phage nucleus, so the rescue is not due to physical exclusion of the mutant protein (Figure 4.1D). These data support the hypothesis that the

mechanism of ΦKZ inhibition by gp210 requires H82 of the H-N-N motif and access to the ΦKZ DNA within the phage nucleus.

Despite this drastic inhibition of ΦKZ, there are no apparent morphological defects of the ΦKZ nucleus when visualized with fluorescence microscopy, however, by 70 mpi, expression of gp210-GFPmut1 resulted in a lack of stained capsids in phage bouquets compared to GFPmut1 and gp210(H82R)-GFPmut1 (Figure 4.1D). ΦPA3 infection morphology is undisturbed by the expression of these proteins (Figure S4.1B). To determine whether there was an overall loss of ΦKZ DNA in a nucleus containing gp210-GFPmut1 which might indicate nonspecific large-scale cleavage by gp210, we quantified the amount of DAPI signal in the nuclei. At 45 mpi, the mean DAPI intensity for ΦKZ nuclei containing gp210-GFPmut1 was 491 arbitrary units and in nuclei containing GFPmut1 it was significantly lower at 395 arbitrary units, possibly indicating a retention of DNA in the nucleus due to a lack of genome packaging into capsids (Figure 4.1F, *unpaired t-test, $p = 0.0019$, $n = 53$*). Since gp210 may affect the final stages of ΦKZ infection related to the formation of phage bouquets (Figure 4.1D and F), we looked for an effect in the subsequent process of cell lysis. Lysis delay could allow the host to outgrow ΦKZ, producing the effect seen in growth curves (Figure 4.1E) and premature lysis could end the infection before the maximal number of virions are produced, reducing the effectiveness of ΦKZ infection, and again resulting in the growth curves obtained. We performed time-lapse DIC microscopy on ΦKZ infections beginning at 45 minutes post-infection (mpi) with images taken every 10 minutes and the results were averaged from five replicates. 50% of the infected cells expressing either GFPmut1 or gp210-GFPmut1 were lysed by ΦKZ after 115 mpi and 95% lysis was achieved by ΦKZ after 195 mpi in both strains (Figure 4.1G). These results suggest that gp210-GFPmut1 does

not affect the ability of ΦKZ to lyse its host nor does it instigate general destruction of the ΦKZ genome as visualized by DAPI, but it may play a role in the packaging of DNA into capsids.

Gp210 is encoded within a group I intron interrupting an RNAP subunit

To deduce the target of nucleolysis by gp210, we analyzed its genomic locus in ΦPA3. From Phyre2, the best match for gp210 was I-HmuI (Figure 4.1B), which is a homing endonuclease that can target its own DNA as well as the DNA of a coinfecting phage [46, 58, 63]. Homing endonucleases are either contained within an intron, fused to a host protein, or freestanding. I-HmuI is nested within an intron interrupting the DNA polymerase gene of SPO1, so we inquired whether gp210 is also housed within an intron. We used Rfam to predict the nature of the entire ΦPA3 sequence between the start of gp211 and the stop of gp209 (coding strand is transcribed), which both contain homology to the ΦKZ RNA polymerase (RNAP) subunit gp178, and the bacterial RNAP subunit RpoB [85] (Figure 4.2A). This resulted in a single high confidence hit for a group I intron (accession RF00028, 6.2 e^{-16}) which is also the type of intron that contains the I-HmuI homing endonuclease [58].

Analysis of the related RNAP genes in ΦKZ and nucleus-forming jumbo phage 201φ2-1 that infects *P. chlororaphis* [73], revealed that they are also interrupted by group I introns. The RNAP gene corresponding to bacterial RpoB is interrupted in ΦPA3 and 201φ2-1 but not in ΦKZ where the locus is unoccupied, while the gene corresponding to bacterial RpoC is interrupted in ΦKZ but not in ΦPA3 or 201φ2-1 (Figure 4.2A). Each intron interrupts the gene in a critical region, which is expected for mobile introns that have evolved to evade deletion and target divergent genomes at sites of high conservation [26]. The ΦKZ intron interrupts the catalytic DFDGD motif of the RpoC homolog while the ΦPA3 and 201φ2-1 introns are inserted in the RpoB homolog at the same site, immediately following two highly conserved aspartic acid residues

(Figure 4.2D), the second of which has been implicated in polymerase fidelity in *E. coli* (D675, [86]). The ΦKZ and ΦPA3 introns both contain putative endonucleases but the 201φ2-1 intron does not contain any open reading frame (ORF). Pairwise alignment of the entire ΦPA3 intron DNA sequence with each of the other two displays high conservation at the 3' end (Figures 4.2B and C), allowing their identification as group I introns [47, 58, 87, 88]. Interestingly, the first 109 nucleotides of the 201φ2-1 intron are 45% identical to the last 108 nucleotides of the gp210 ORF, and in the middle there is a 64-nucleotide gap in the 201φ2-1 sequence compared to the ΦPA3 sequence. The entire 201φ2-1 intron (336 nt) is significantly shorter than the ΦPA3 intron (1352 nt) and shares 63.4% identity with the ΦPA3 intron, excluding gaps. This is indicative of a shared origin of the ΦPA3 and 201φ2-1 introns, and a subsequent degeneration of the 201φ2-1 intron that deleted the majority of the homing endonuclease ORF, as well as another 64 nucleotides, reducing the burden of the intron on the host but retaining the ribozyme activity that effectively splices the intronic sequence from the 201φ2-1 RNAP subunit which was confirmed by mass spectrometry to be intact and properly spliced [89]. The phylogenetic relationship of RNAPs from phages infecting 9 different genera of hosts, along with the RpoB genes of *E. coli* (MG1655) and *P. aeruginosa* (PAO1) is presented as an unrooted neighbor-joining tree in Figure 4.2E.

Alignment of the phage-encoded RNAP genes with RpoB of *E. coli* strain MG1655 and *P. aeruginosa* strain PAO1 uncovered an extra 15 residues annotated as part of the C-terminal end of ΦPA3 gp211 (Figure S4.1D). However, these 15 amino acids following the conserved D-D are likely excluded from the exon in the splicing process to reconstitute the site as presented in Figure 4.2D. We hypothesized that the region in ΦKZ that aligns to the intron insertion site in ΦPA3 is the target site of gp210. This gene of ΦKZ, gp178, is part of the polymerase that transcribes only early genes [85] and it is also packaged into the virion of newly produced ΦKZ progeny, making

it a "vRNAP" [90]. If this gene is disrupted by gp210 and the protein cannot be made, that may prevent the proper assembly of virions as well as the transcription of early genes in the second round of infections being carried out by the progeny that were produced when gp210-GFPmut1 was in the ΦKZ nucleus. This hypothesis is consistent with all of the observations of Figure 4.1.

Gp210 targets the unoccupied locus in the ΦKZ vRNAP subunit

To test this hypothesis, we performed a nuclease assay *in vitro* using purified gp210 protein and a plasmid containing the potential target gene ΦKZ gp178. Incubation of increasing concentrations of gp210 (0-5 μM) with ΦKZ gp178 plasmid resulted in cutting of the plasmid that presented as a decrease of supercoiled plasmid and an increase of nicked and linearized plasmid (Figure 4.3A). Nicked and linear species were identified based on the control reactions using enzymes Nt.BsaI (nicking) and NcoI (linearizing). Quantitation of the bands and comparison to the reaction performed with the empty vector (Figure 4.3B) demonstrated an elimination of supercoiled plasmid (Figure 4.3C) and an accumulation of nicked plasmid (Figure 4.3D) that was more specific than for the empty vector. The 5 μM condition does not continue the trend for the gp178 plasmid which may be an artifact of high nuclease concentration. This assay will be performed again to confirm the activity and to compare with purified gp210(H82R) to determine whether this nucleolysis is from gp210 and not a contaminant. To narrow down the target site of gp210 *in vitro*, we are working to gel purify and sequence the linearized band of ΦKZ gp178 digested by gp210. We expect that the sequencing will drop off at the site of gp210 cutting. We are also testing whether gp210 can cut the intron(+) and intron(-) alleles of the ΦPA3 RNAP gene that is interrupted by it. This could help to fully define gp210 as a homing endonuclease.

Despite a strong knockdown of ΦKZ titer in the presence of gp210-GFPmut1, there are always a few plaques that are able to form. To identify mutations that allow ΦKZ to evade the

activity of gp210, we isolated clones from three independent infections in cells expressing gp210-GFPmut1 by induction with 0.1% arabinose. We confirmed the resistance of these ΦKZ mutants to gp210-GFPmut1 using titers expressed as EOP (Figure 4.3E, *ratio paired t test, p = 0.0004, n = 4*) and sequenced their genomes. We found that these gp210-resistant ΦKZ phages all had a mutation of the adenosine at position 3215 in the gp178 vRNAP gene (Figure 4.3F). Two mutants replaced the adenosine with a cytosine and the third mutated it to a guanine. These changes resulted in a missense mutation of the highly conserved aspartic acid residue at position 1072 to alanine or glycine, respectively. D1072 of gp178 aligns with the aspartic acid residue neighboring the gp210 intron insertion site in ΦPA3 (Figure 4.3G). The mutated nucleotide is 2 bases upstream of the intron insertion site so it is well within range to be required for gp210 recognition of the ΦKZ gene. To further support this target site of gp210, we attempted to complement ΦKZ infections inhibited by gp210-GFPmut1 with the mutant gp178(D1072A) found in resistant isolates. Coexpression of gp178(D1072A) with gp210-GFPmut1 resulted in an average 10-fold increase in titer suggesting a partial but significant recovery of EOP (Figure 4.3H, *ratio paired t test, p = 0.0078, n = 4*). We are currently performing further controls to ensure the rescue is not due to decreased production of gp210 which is downstream of gp178(D1072A). It was expected that the rescue of ΦKZ would not be complete due to the missense mutation at a highly conserved residue which possibly plays a role in polymerase fidelity. These mutants and the resistance provided by gp178(D1072A) support the hypothesis that gp210 of ΦPA3 is able to hydrolyze ΦKZ DNA within vRNAP gene gp178. The proximity of the mutation to the native intron insertion site of gp210 also indicates classic homing activity at a conserved site in a divergent genome.

Gp210 disrupts ΦKZ capsid assembly when imported into the ΦKZ nucleus

The biological consequences of gp210 targeting of the ΦKZ vRNAP gene were investigated next. Gp178 is a subunit of the ΦKZ vRNAP that is injected into the cell along with the phage DNA where it transcribes early genes [85, 90]. If this gene is made nonfunctional by gp210, the early transcripts produced by gp178 will not be affected in the first round of infections which originate from the ΦKZ virions that have not yet been exposed to gp210. However, the ΦKZ progeny produced in the presence of gp210-GFPmut1 will not have functional gp178 and therefore may not be able to transcribe the early genes in the second round of infections. To test the virulence of ΦKZ progeny produced from infections containing gp210-GFPmut1, parent phage were washed out from a liquid infection and the host cells were allowed to lyse and release the progeny for collection. The ΦKZ progeny lysate produced with gp210-GFPmut1 plated with an efficiency of 0.000014% (*paired t test, $p < 0.0001$, $n = 5$*) compared to the progeny produced with only GFPmut1 and this is rescued by the H82R mutation of gp210 (Figure 4.4A). These data support the idea that the production of infective virions is inhibited by gp210 targeting of gp178.

This loss of virulence of the ΦKZ progeny could potentially be due to a structural defect that prevents the assembly of virions, since gp178 is incorporated into the capsid [85]. To visualize the macromolecular organization of a ΦKZ infection proceeding in the presence of gp210-GFPmut1, we performed cryo-focused ion beam milling coupled with cryo-electron tomography (cryo-FIB-ET), presented in Figures 4.4B, D, and E as a 2D slice of a 3D tomogram. The ΦKZ infections in *P. aeruginosa* without any plasmid formed many complete capsids full of DNA by 90 mpi. They were grouped together in a non-spherical bouquet and also observed were empty capsids and nascent capsids assembling on the cell membrane (Figure 4.4B "No Plasmid"). However, when gp210-GFPmut1 was expressed, there was a noticeable loss of

completed capsids (cyan arrowheads) and the observation of unidentified structures (orange and red arrowheads, Figure 4.4B "210-GFPmut1"). In 5 tomograms of infections without plasmid, 113 filled capsids were observed (average: 22.6 ± 3.3 SEM). In 17 tomograms of infections with gp210-GFPmut1, only 13 filled capsids were observed (average: 0.76 ± 0.34 SEM) demonstrating a 96.6% decrease in virion production (Figure 4.4C, *unpaired t test, $p < 0.0001$*). Along with the significant deficit of completed capsids, we discovered many unusual structures never before observed in these cells, shown in Figure 4.4B. In all of the tomograms of wild-type ΦKZ infecting a cell expressing gp210-GFPmut1, large geometric and sometimes tubular structures filled the cell. A phage bouquet has not previously been observed without capsids and tails, but in Figure 4.4B, the yellow arrowheads highlight the boundaries of ribosome-free regions that we hypothesize to be nascent phage bouquets due to the location and size, as well as the association of capsid-like assemblies near the perimeter. DNA-filled capsids that can be observed with gp210-GFPmut1 do not appear to have an obvious defect compared to those observed in cells without a plasmid (Figure 4.4D). This is likely due to the production of a few copies of vRNAP subunit gp178 before the gene was cut by gp210. Those limited gp178 proteins could then be incorporated into the capsids that assemble occasionally to form a wild-type virion capable of packaging DNA, but leaving many other assembling capsids without the integral gp178 protein. Some of the unusual structures visualized with gp210-GFPmut1 closely resemble partially assembled capsids and this is exemplified in Figure 4.4E.

This work demonstrates that gp210 is a homing endonuclease housed within an intron interrupting a ΦPA3 RNAP gene and it is able to cut the vRNAP gene of coinfecting phage ΦKZ at the site homologous to the intron insertion site in ΦPA3. This results in the inhibition of ΦKZ

virion assembly and prevents subsequent rounds of infection, significantly reducing ΦKZ fitness (Figure 4.5A).

4.5 Discussion

Gp210 is one of the first bacteriophage-encoded intracellular speciation factors to be identified [17], but its mechanism was not previously understood. Our data allow us to propose a model to explain how this factor could influence speciation (Figure 4.5B). In a natural coinfection of phages ΦPA3 and ΦKZ, a hybrid nuclear compartment forms in 25% of coinfections [17]. Hybrid compartments are expected to import gp210, which will cleave ΦKZ DNA, disrupting the expression of the vRNAP, and decreasing the production of ΦKZ capsids. DNA cleavage by gp210 can be repaired by homologous recombination with either ΦKZ DNA or ΦPA3 DNA as the donor. Repairs made with ΦKZ DNA will simply result in the reconstitution of the gp210 target site so recurrent nucleolysis will occur until the insertion of the intron from the ΦPA3 template disrupts the site. However, rescuing the ΦKZ DNA by conversion to the intron(+) ΦPA3 allele becomes less efficient as the genomes diverge. These two alleles currently share only 60.4% sequence identity, which is likely poorly recombinogenic given that lambda recombinase Redβ_{λ} loses 100-fold efficiency at 78% sequence identity and no longer recombines at 48% identity [52, 53]. The resulting damage to ΦKZ DNA by this nuclease will reinforce the reproductive isolation between these two sister phages by acting as a virogenesis incompatibility factor that limits the number of ΦKZ progeny produced from a hybrid phage nucleus. This nuclease natively resides in the nucleus of ΦPA3 and therefore is a potent virogenesis incompatibility speciation factor that would bias the replication outcome in favor of ΦPA3 for the 25% of coinfections that form a hybrid nucleus (Figure 4.5B). Under these conditions, the intron provides a selective advantage for ΦPA3 over ΦKZ.

Do mobile introns generally provide a selective advantage for the organism they inhabit? How selfish genetic elements like introns and inteins affect the evolution of viral populations, and whether they confer selective advantages for the virus itself has been unclear [65-68]. For coinfections of two nearly identical viruses, one with and one without the intronic endonuclease, the cleaved DNA can be repaired efficiently by homologous recombination, resulting in the spread of the intron (Figure 4.5C). Rescue of the intron(-) virus by homologous recombination would be diminished as the divergence between the viruses increases (Figure 4.5D). Since homing endonucleases generally target highly conserved sites in essential genes, the recognition sites diverge more slowly than the rest of the genome. This can result in sustained targeting by the homing endonuclease yet inefficient recombination to rescue the intron(-) virus. Hence, over time, the phage carrying the homing endonuclease would possess an increasing advantage during coinfections as the two phage genomes diverge. For viruses that establish distinct replication factories and display subcellular genetic isolation, the rates of repair are expected to be greatly diminished, even if closely related, due to physical sequestration of the intron(+) repair template (Figure 4.5E).

In a broader sense, speciation can be considered a process during which two populations of organisms diverge from recognizing each other as self, the same species, to defining each other as non-self, two different species. In this context, homing endonucleases can spread between organisms of the same gene pool without causing significant harm, but can destroy related organisms that have diverged enough to prevent rescue by homologous recombination. This makes the homing endonuclease toxic to organisms that are related enough to be a competitor but divergent enough to be non-self. Along the continuum of sequence divergence between two organisms, there exist conditions where the genomes recombine at low frequencies but the homing

endonuclease still targets the insertion site. These conditions define where the two organisms behave as non-self and are therefore a different species biologically. We can imagine a relationship between sequence divergence, recombination frequency, and competitive fitness for two phages coinfecting the same cell, one containing and one lacking a mobile intron. As sequence divergence increases, the rate of recombination decreases, and the competitive fitness of the intron-containing phage increases. This model provides evidence of a competitive mechanism that defines self from non-self and therefore determines which individuals are in competition and which are cooperating. This model suggests that the intron-containing organisms have a strong fitness advantage during speciation and provides a possible explanation for intron pervasiveness throughout life.

The effect of mobile introns on competition has profound implications for diverging nucleus-forming jumbo phages. Since two identical phages infecting the same cell form separate phage nuclei [17] yet import the same components, the acquisition of a homing endonuclease by one of those genomes would lead to the production of the nuclease and its import into the nuclei of both the intron(+) and the intron(-) phages, but the intron(-) genome would not have access to the intron(+) DNA template for recombination (Figure 4.6C). We can imagine a scenario in which the ancestor to ΦPA3 and ΦKZ evolved a phage nucleus to protect its DNA from host defenses (Figure 4.6A) and at some point, it acquired a mobile intron, the ancestor of the gp210 intron (Figure 4.6B). The phage that acquired the mobile intron (Phage 2) would be able to coinfect with its relative that did not acquire the intron (Phage 1). As Phage 1 and Phage 2 competed during coinfections, the endonuclease from Phage 2 would be imported into the nucleus of both Phage 1 and Phage 2 (Figure 4.6C). Phage 1 DNA would be cut by the endonuclease without repair by the Phage 2 DNA, which is fully enclosed within the phage nucleus. Phage 1 would therefore be

unable to produce progeny phage, except when these phages formed mixed nuclei, which happens at a relatively low frequency [17]. Phage 2 would therefore have a significant advantage over Phage 1. This selective pressure might promote more frequent mixing through formation of hybrid nuclei, but the fusion of different nuclear compartments was not previously observed [17], and is even less likely when the phages inject their DNA near opposing cell poles. Mutations in the target site of Phage 2 will also arise, but these are expected to occur at a low frequency due to its location within a highly conserved region of an essential gene. The other possible outcome of this selective pressure would be the evolution of Phage 1 mutants that exclude the homing endonuclease from the nucleus (Figure 4.6D).

While we don't yet understand the mechanisms underlying selective protein import, we hypothesize that they are likely shaped by the evolutionary pressure to exclude enzymes harmful to the phage DNA [17, 24, 91, 92]. This predicts that phages ΦPA3 and ΦKZ should have evolved different specificities for protein import to avoid importing toxic endonucleases native to the other phage. In support of this hypothesis, we know of two proteins, GFPmut1 and gp210, that are differentially imported between the two phages [17, 24]. ΦPA3 imports gp210 and excludes GFPmut1, while ΦKZ excludes gp210 and imports GFPmut1. While initially puzzling why these two closely related phages would display such disparate protein import selectivity, this homing endonuclease model provides an explanation and predicts such disparities must evolve for both phages to co-exist.

These studies reveal a complex evolutionary relationship between the phage nucleus and mobile introns. The competition of homing endonucleases selects for the phage nucleus in two different ways. The intron(-) phage is pressured to retain the nucleus so that competitive endonucleases are excluded from its genome, while the intron(+) phage receives an advantage

from the nucleus since it sequesters the repair template away from the intron(-) phage, increasing the effectiveness of the competitive homing endonuclease. Taken together, our results present a clear example of how a mobile intron can provide a selective advantage for a virus and drive viral speciation. Given the ubiquity of introns, this selective advantage is likely widespread in nature.

4.6 Acknowledgements

This work was supported by the National Institutes of Health R01-GM129245 (to JP and EV) and R21-AI148814 (to KDC) and the National Science Foundation MRI grant NSF DBI 1920374 (to EV). EV is an investigator of the Howard Hughes Medical Institute. We acknowledge the use of the UC San Diego cryo-EM facility, which was built and equipped with funds from UC San Diego and an initial gift from the Agouron Institute. We would like to thank Avani Mylvara, Jina Lee, Emily Armbruster, and Livia Songster for their participation in experiments.

This chapter is being formulated into a manuscript for publication. Erica A. Birkholz, Chase Morgan, Rebecca K. Lau, Amy Prichard, Sergey Suslov, Justin R. Meyer, Elizabeth Villa, Kevin D. Corbett, Joe Pogliano, 2022. The dissertation author was the primary investigator and author of this material.

Author Contributions

E.A.B., C.M., R.K.L., A.P., and S.S. conducted experiments. E.A.B. analyzed data and created figures. E.A.B. and J.P. conceptualized the original manuscript. K.D.C. and J.R.M. provided feedback on the manuscript. J.R.M., K.D.C., E.V., and J.P. secured funding.

4.7 Figures

Figure 4.1 Gp210 of ΦPA3 requires the HNH endonuclease domain to inhibit ΦKZ. (A) Protein alignment of phage-encoded endonucleases related to gp210 (green asterisk), including well-characterized homing endonuclease I-HmuI (blue asterisk), shows high conservation in the predicted HNH nuclease domain, including critical H-N-N residues (red asterisks). Residue numbers based on gp210. ClustalO alignment annotated by ESPript3. (B) Phyre2 alignment of gp210 with I-HmuI (PDB: 1u3e). H-N-N is conserved including H82 of gp210 indicated above. (C) Efficiency of plating (EOP) of ΦKZ as measured by spot titer and normalized to the paired titer on GFPmut1. gp210-GFPmut1 causes a 99.9983% decrease in ΦKZ titer ($p = 0.0011$, n = 3) while untagged 210, not imported into the nucleus, causes an insignificant 27% decrease ($p = 0.1371$, n = 3). H82R mutation (gp210(H82R)-GFPmut1) fully rescues ΦKZ compared to 210-GFPmut1 ($p = 0.0026$, n = 3). Error bars represent SEM, p values calculated by ratio paired t test. (D) Live fluorescent microscopy of ΦKZ infections stained with FM4-64 (magenta: membranes) and DAPI (blue: DNA). Infections proceeded in the presence of either GFPmut1, gp210-GFPmut1, or gp210(H82R)-GFPmut1 (green: GFP). Nuclei formed, grew in size, and centered at midcell with no apparent deficit, but stained capsids in phage bouquets did not appear with gp210-GFPmut1. Scale bar is 1 μm. (E) ΦKZ growth curves measuring OD_{600} of bacteria in liquid culture showed an MOI of $7e^{-3}$ was required to achieve 50% inhibition of cell growth (IC50: red dotted lines) when cells were expressing GFPmut1 (n = 10). gp210-GFPmut1 increased the IC50 MOI to 7, a 1,000-fold decrease in ΦKZ fitness (n = 10). 210H82R-GFPmut1 rescued the IC50 back to $7e^{-3}$ (n = 10), while untagged 210 in the cytoplasm caused only a 10-fold increase in required MOI (n = 5). (F) Violin plot of the total DAPI signal counts from the middle image slice, divided by the area of the circle measured, and normalized to a matching background measurement taken outside the cell. Gp210 significantly increases (*unpaired t test,* $p = 0.0012$, n = 53) the total amount of ΦKZ DNA contained within the nucleus at 45 mpi. This may indicate the retention of DNA in the nucleus instead of packaging into capsids. Dashed lines indicate quartiles with the center line at the median. (G) Lysis kinetics for ΦKZ infecting cells expressing GFPmut1 (blue) or gp210-GFPmut1 (orange) was measured by DIC time lapse microscopy. For both strains, 50% lysis is reached by 115 mpi, 95% lysis by 195 mpi. Line plots are an average of biological replicates (GFPmut1 n = 6, gp210-GFPmut1 n = 7).

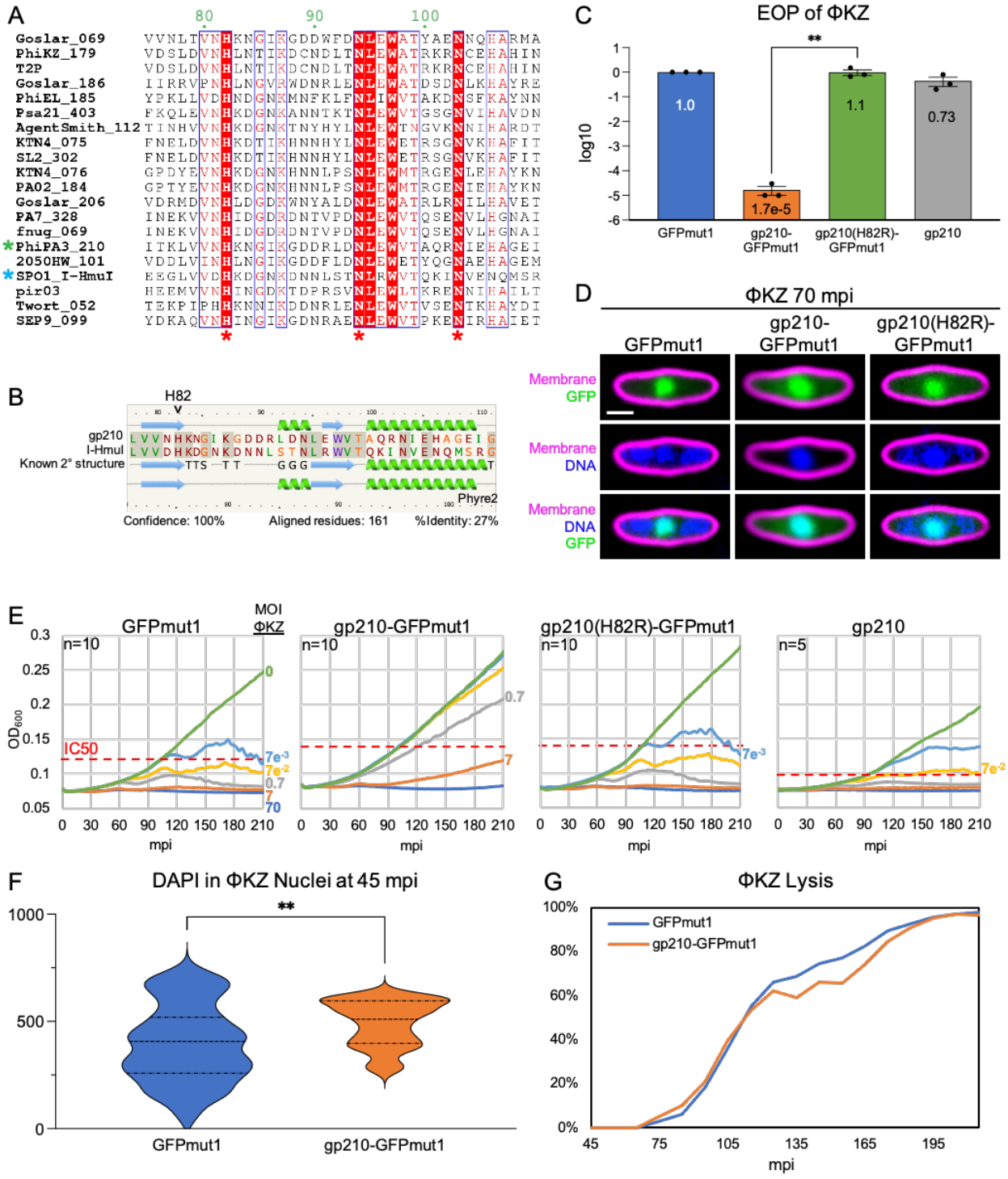

A
80
90
100
Goslar_069
PhiKZ_179
T2P
Goslar_186
PhiEL_185
Psa21_403
AgentSmith_112
KTN4_075
SL2_302
KTN4_076
PA02_184
Goslar_206
PA7_328
fnug_069
PhiPA3_210
2050HW_101
SPO1_I-HmuI
pir03
Twort_052
SEP9_099
B
H82
gp210
I-Hmul
Known 2° structure
Phyre2
Confidence: 100%
Aligned residues: 161
%Identity: 27%
C
EOP of ΦKZ
log10
1.0
1.7e-5
1.1
0.73
GFPmut1
gp210-GFPmut1
gp210(H82R)-GFPmut1
gp210
D
ΦKZ 70 mpi
GFPmut1
gp210-GFPmut1
gp210(H82R)-GFPmut1
Membrane
GFP
Membrane
DNA
Membrane
DNA
GFP
E
MOI
ΦKZ
GFPmut1
gp210-GFPmut1
gp210(H82R)-GFPmut1
gp210
n=10
n=10
n=10
n=5
OD600
IC50
mpi
F
DAPI in ΦKZ Nuclei at 45 mpi
GFPmut1
gp210-GFPmut1
G
ΦKZ Lysis
GFPmut1
gp210-GFPmut1
mpi

Figure 4.2 Gp210 is within a group I intron interrupting an RNAP subunit with homology to ΦKZ gp178, a virion-packaged RNAP subunit. (A) Diagram of 2 RNAP subunit genes of nucleus-forming *Pseudomonas* jumbo phages ΦPA3, ΦKZ, and 201φ2-1. ΦPA3 and 201φ2-1 contain introns at the same conserved position in the subunit with homology to bacterial RpoB, while the corresponding gene in ΦKZ, gp178, does not contain any introns. ΦKZ does contain an intron in the subunit with homology to RpoC, interrupting the catalytic DFDGD motif, while ΦPA3 and 201φ2-1 do not. All three introns are putative group I introns but only the ΦPA3 and ΦKZ introns harbor open reading frames, which are predicted to encode HNH endonucleases. The leading 109 nt of the 201φ2-1 intron contains homology to the 3' end of gp210. Another 64 nt deletion leaves the 201φ2-1 intron 27% of the length of the ΦPA3 intron. (B) Nucleotide alignment between the introns of ΦPA3 and ΦKZ, based on ClustalO alignment. Nucleotide identities, including gaps, are reported as a percentage for each 60 nt block and color coded. The ORFs of each HNH endonuclease are annotated by red lines and the catalytic histidines are noted in their respective positions. (C) Nucleotide alignment between the introns of ΦPA3 and 201φ2-1 as represented in (B). The 201φ2-1 intron does not contain an ORF but begins with homology to the gp210 ORF annotated by the red line above. It is also lacking 64 nt compared to the ΦPA3 intron, represented by the gap that is followed by 60 nt with 70% nucleotide identity. (D) Protein alignment of the intron(-) version of the ΦPA3 RNAP, after editing from the annotation in Genbank which can be found in Figure S4.1D. It is aligned with RpoB of MG1655 (*E. coli*) and PAO1 (*P. aeruginosa*), 8 RNAPs encoded by jumbo phages infecting different genera of hosts, and 4 RNAPs encoded by other *Pseudomonas* jumbo phages including the intron(-) allele of the 201φ2-1 intron, as annotated in Genbank. Alignment was performed by ClustalO and annotated by ESPript3. The intron insertion site (IIS) occupied in ΦPA3 and 201φ2-1 is indicated by a green carrot and text. The IIS is in a region of high conservation, a hallmark of mobile introns, following the *E. coli* D675 which is important for polymerase fidelity. Residue numeration in green above is based on the ΦPA3 sequence. (E) Unrooted neighbor-joining tree of the homologous phage-encoded RNAP subunits, showing an expected clustering of *Pseudomonas* phages.

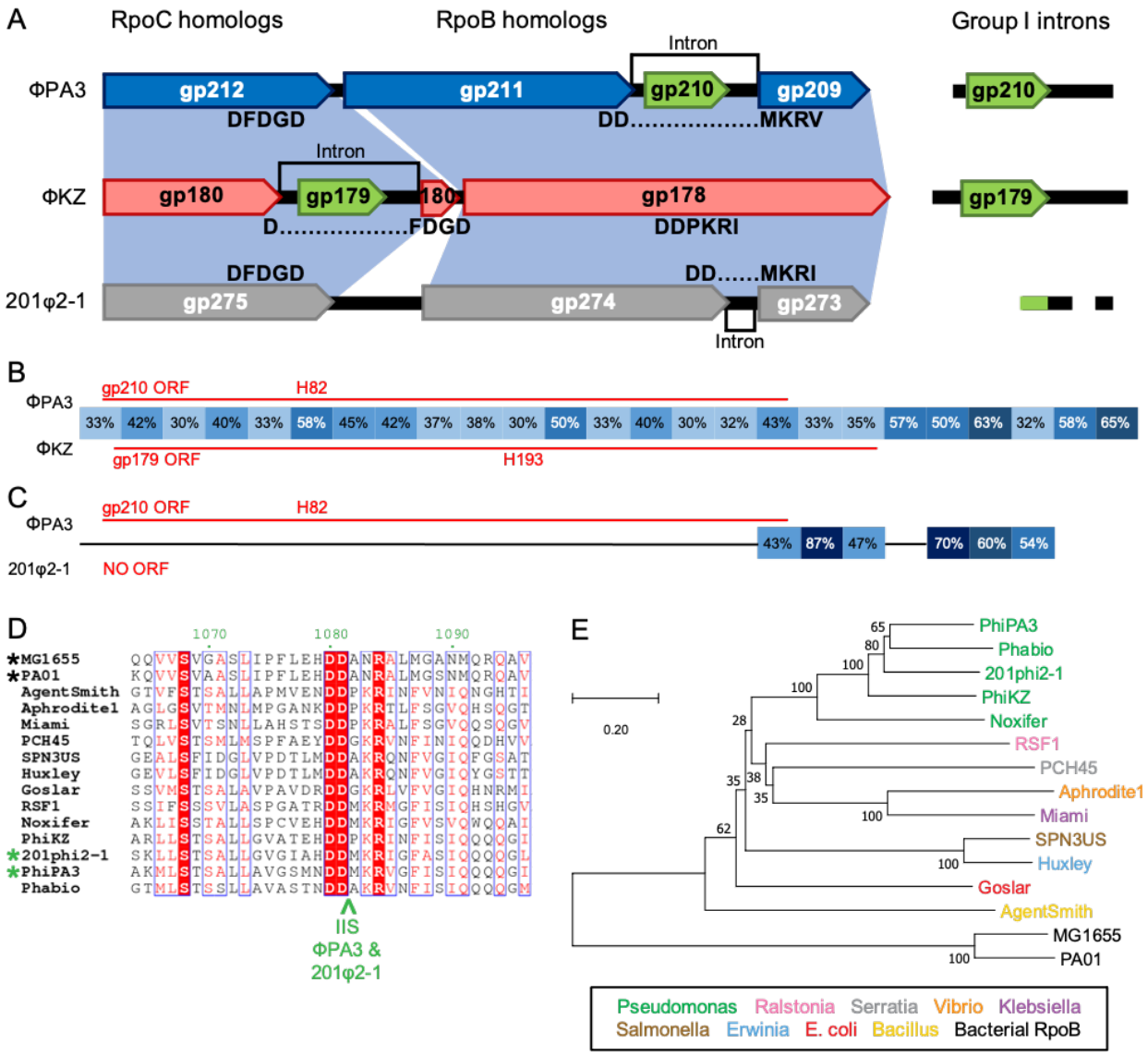

A
RpoC homologs
RpoB homologs
Group I introns
ΦPA3
gp212
gp211
gp210
gp209
DFDGD
DD.....MKRV
Intron
ΦKZ
gp180
gp179
gp178
D.....FDGD
DDPKRI
201φ2-1
gp275
gp274
gp273
DD.....MKRI
B
gp210 ORF
H82
gp179 ORF
H193
33% 42% 30% 40% 33% 58% 45% 42% 37% 38% 30% 50% 33% 40% 30% 32% 43% 33% 35% 57% 50% 63% 32% 58% 65%
C
NO ORF
43% 87% 47% 70% 60% 54%
D
IIS
ΦPA3 & 201φ2-1
E
0.20
PhiPA3
Phabio
201phi2-1
PhiKZ
Noxifer
RSF1
PCH45
Aphrodite1
Miami
SPN3US
Huxley
Goslar
AgentSmith
MG1655
PA01
Pseudomonas Ralstonia Serratia Vibrio Klebsiella
Salmonella Erwinia E. coli Bacillus Bacterial RpoB

Figure 4.3 Gp210 cuts ΦKZ gp178 *in vitro*, resistant ΦKZ mutated gp178(D1072A) which provides some protection against gp210. (A, B) Nuclease assay of purified gp210 incubated with plasmid DNA containing KZgp178 (A) or only the empty vector (B). Gp210 concentration from left to right was 5.0, 2.5, 1.25, and 0 μM. Nt.BsaI is a reference digest for nicked plasmid (red asterisks) and NcoI is a reference for linearized plasmid (blue asterisks). Supercoiled plasmid is indicated by black asterisks. (C) Quantitation of bands in (A) and (B) revealed a more drastic decrease of supercoiled plasmid when KZgp178 was present (orange bars) compared to the empty vector (blue bars). Analysis was performed in Fiji. (D) Quantitation of bands in (A) and (B) shows a specific accumulation of nicked plasmid when the KZgp178 sequence is present (orange bars). (E) EOP of ΦKZ on cells expressing gp210-GFPmut1. Mutants 1-3 (ΦKZ isolated from infections with gp210-GFPmut1 expressed by 0.1% arabinose) displayed an average 266,907-fold increase in EOP compared to wild-type ΦKZ (WT). Mutant 2 had the lowest EOP which was still a significant increase over WT (*ratio paired t test, p = 0.0004,* error bars ± SEM, n = 4). (F) All 3 mutants contained a single point mutation at adenosine 3215 in ΦKZ vRNAP gene gp178. The adenosine was mutated to either cytosine (mutants 1 and 2) or guanine (mutant 3), and resulted in a coding change of D1072A or D1072G, respectively. (G) The point mutation in the ΦKZ vRNAP gp178 at position 3215 (red box) is 2 bases upstream of the site that aligns with the intron insertion site (IIS, black dotted line) in ΦPA3 and 201φ2-1 (asterisks). Red sequences code for the conserved aspartic acid residues. The exons of ΦPA3 and 201φ2-1 are noted above. (H) The EOP of ΦKZ on cells containing no plasmid (blue) or expressing gp210-GFPmut1 (orange) or the mutant ΦKZ vRNAP gp178(D1072A) co-expressed with 210-GFPmut1 (light green). Complementation of gp210-GFPmut1 with gp178(D1072A) resulted in an average 10-fold (maximum 40-fold) rescue of ΦKZ titer (*ratio paired t test, p = 0.0078*, error bars ± SEM, n = 4).

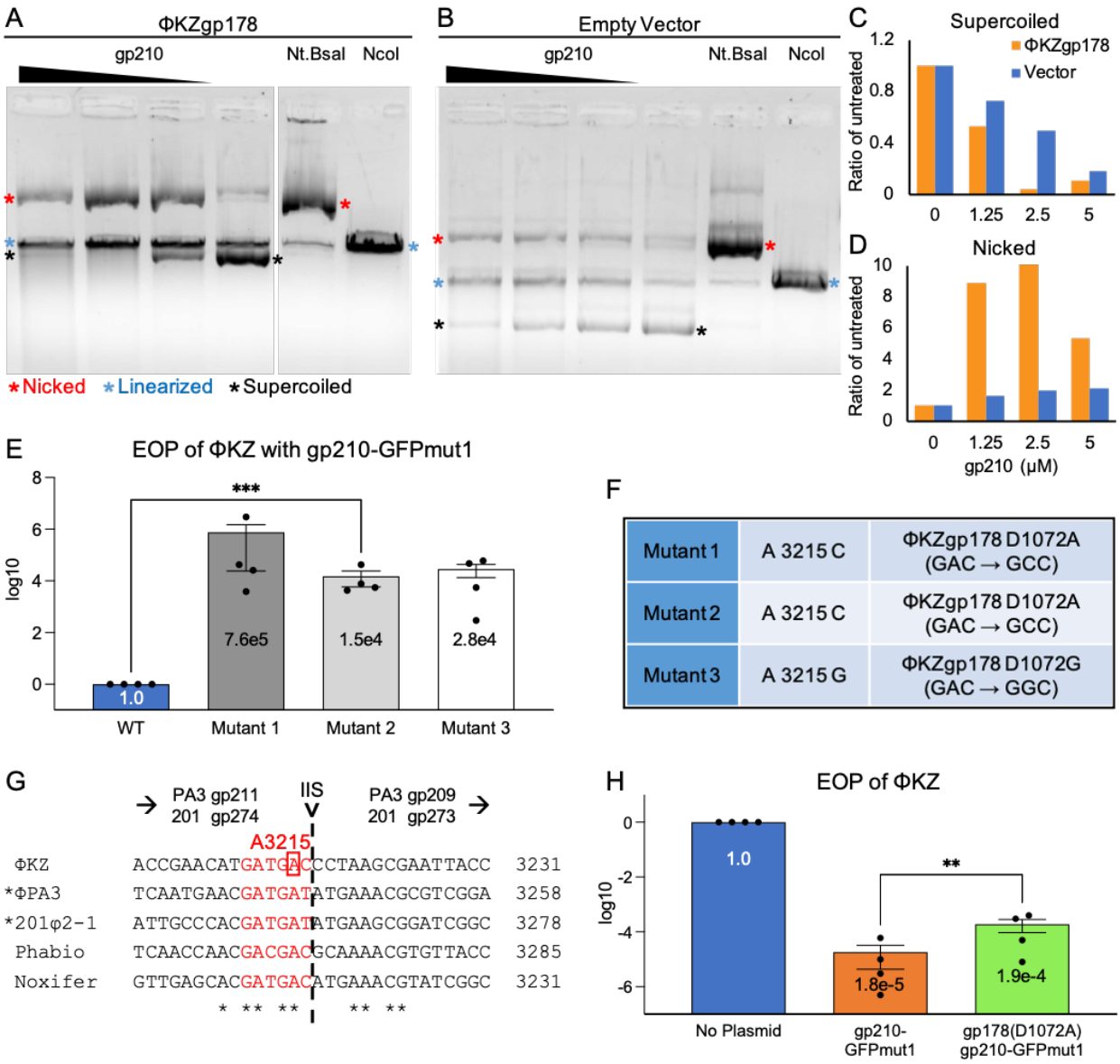

A
ΦKZgp178
gp210
Nt.BsaI
NcoI
B
Empty Vector
gp210
Nt.BsaI
NcoI
*Nicked *Linearized *Supercoiled
C
Supercoiled
ΦKZgp178
Vector
Ratio of untreated
1.2
0.8
0.4
0
0
1.25
2.5
5
D
Nicked
Ratio of untreated
10
8
6
4
2
0
0
1.25
2.5
5
gp210 (μM)
E
EOP of ΦKZ with gp210-GFPmut1
log10
8
6
4
2
0
***
1.0
7.6e5
1.5e4
2.8e4
WT
Mutant 1
Mutant 2
Mutant 3
F
Mutant 1 | A 3215 C | ΦKZgp178 D1072A (GAC → GCC)
Mutant 2 | A 3215 C | ΦKZgp178 D1072A (GAC → GCC)
Mutant 3 | A 3215 G | ΦKZgp178 D1072G (GAC → GGC)
G
PA3 gp211
201 gp274
IIS
PA3 gp209
201 gp273
A3215
ΦKZ ACCGAACATGATGACCCTAAGCGAATTACC 3231
*ΦPA3 TCAATGAACGATGATATGAAACGCGTCGGA 3258
*201φ2-1 ATTGCCCACGATGATATGAAGCGGATCGGC 3278
Phabio TCAACCAACGACGACGCAAAACGTGTTACC 3285
Noxifer GTTGAGCACGATGACATGAAACGTATCGGC 3231
* ** ** ** **
H
EOP of ΦKZ
log10
0
-2
-4
-6
1.0
**
1.8e-5
1.9e-4
No Plasmid
gp210-GFPmut1
gp178(D1072A) gp210-GFPmut1

Figure 4.4 Gp210 causes a loss of virulent ΦKZ progeny and capsid production is disrupted. (A) EOP of ΦKZ progeny collected from washed infections of cells expressing the indicated protein, measured by spot titer on cells without plasmid. Progeny grown in the presence of gp210-GFPmut1 plaque with an efficiency of 0.00001% of the progeny grown with GFPmut1 ($p < 0.0001$, n = 5). Progeny produced with gp210(H82R)-GFPmut1 have a relative EOP of 190% (n = 5). Error bars represent SEM and a paired t test was used. (B) 2D slices of 3D tomograms of 90 mpi ΦKZ infections of *P. aeruginosa* containing either no plasmid (top left panel) or gp210-GFPmut1 (other 3 panels). A decrease of completed capsids full of DNA (cyan arrowheads) and the bouquets they normally organize into can be observed with gp210-GFPmut1. Unidentified geometric (orange arrowheads) or tubular protein structures (red arrowhead) occurred in all gp210-GFPmut1 tomograms analyzed (n = 17). Black arrowheads indicate tails, purple arrowheads highlight fully assembled capsids that have not yet been filled with DNA, and the pink arrowhead shows a capsid partially assembled on the cell membrane. The shell of the phage nucleus is delineated by blue arrowheads and the compartment is labeled when in view. Scale bars are 250 nm. (C) Counts of assembled capsids containing DNA that were observed in infections of cells without plasmid (22.6 ± *SEM* 3.28, n = 5) were significantly higher ($p < 0.0001$, *unpaired t test*) than counts of completed capsids observed in infections with gp210-GFPmut1 (0.76 ± *SEM* 0.34, n = 17). Error bars represent 95% CI. (D, E) Expanded portions of tomograms highlighting the morphology of full capsids (D) which appear similar between samples, and of empty capsids (E) from the control host (left panel) compared to what may be improperly assembled empty capsids in the host expressing gp210-GFPmut1 (right panel).

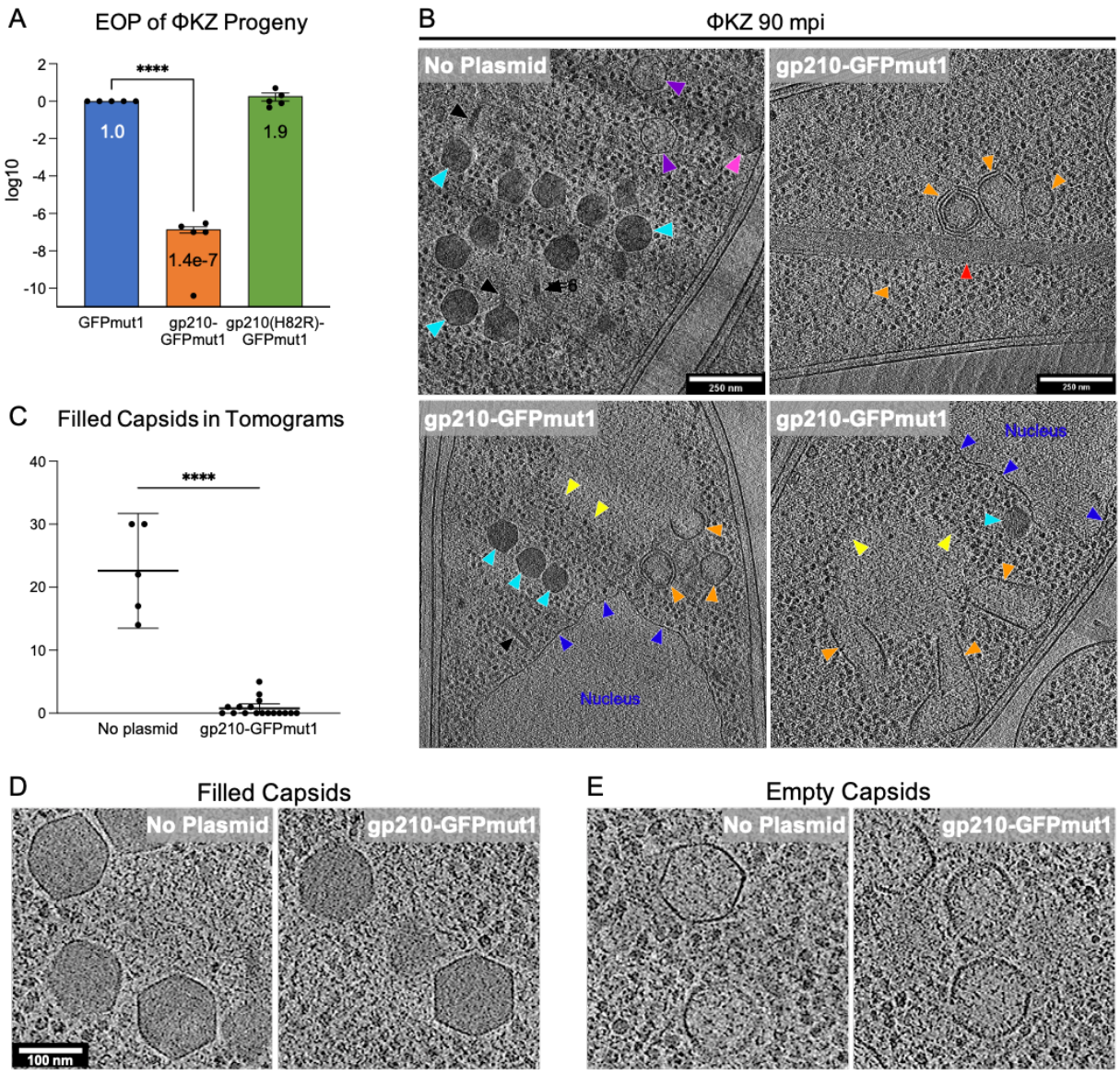

A
EOP of ΦKZ Progeny
log10
2
0
-2
-4
-6
-8
-10
****
1.0
1.4e-7
1.9
GFPmut1
gp210-GFPmut1
gp210(H82R)-GFPmut1
B
ΦKZ 90 mpi
No Plasmid
gp210-GFPmut1
250 nm
gp210-GFPmut1
gp210-GFPmut1
Nucleus
Nucleus
C
Filled Capsids in Tomograms
40
30
20
10
0
****
No plasmid
gp210-GFPmut1
D
Filled Capsids
No Plasmid
gp210-GFPmut1
100 nm
E
Empty Capsids
No Plasmid
gp210-GFPmut1

Figure 4.5 Models of gp210 mechanism of ΦKZ knockdown and general mobile intron competition between viruses. (A) gp210 (green pacman) tagged with GFPmut1 is artificially imported into the ΦKZ nucleus (red shell, blue fill) where it cuts ΦKZ DNA in the vRNAP gene of gp178, dependent upon adenosine 3215, and this results in a stark reduction of properly assembled virions, thereby inhibiting ΦKZ bouquet formation and subsequent infections. (B) Hypothesis of the effects of gp210 on a hybrid nucleus containing ΦPA3 and ΦKZ which occurs for 25% of coinfections. In a hybrid nucleus, gp210 would be imported and cut the ΦKZ vRNAP gene gp178. Some ΦKZ may be repaired by homologous recombination with the ΦPA3 allele that no longer contains the gp210 target site, but with only 60.4% identity between the vRNAP alleles, recombination efficiency is reduced and that may lead to the formation of more virions with the ΦPA3 genome. (C) Model of mobile intron spread during coinfection between any related viruses that replicate their genomes freely in the cytoplasm of the host. (D) Theoretical model of mobile intron competition between divergent viruses that can physically mix their genomes but sequence divergence has significantly reduced the efficiency of homologous recombination while the highly conserved site in the essential gene is still targeted by the homing endonuclease. This results in a selection for the intron(+) virus which treats the divergent coinfecting virus as non-self. (E) Theoretical model of mobile intron competition between closely related viruses that perform subcellular genetic isolation. Whether spatially separated in distinct replication factories or physically isolated by a barrier such as the phage nucleus, the homing endonuclease can target the essential gene in the intron(-) virus but the repair template containing the intron that interrupts the target site is sequestered. This results in the destruction of the intron(-) virus and a strong selective advantage for the intron(+) virus.

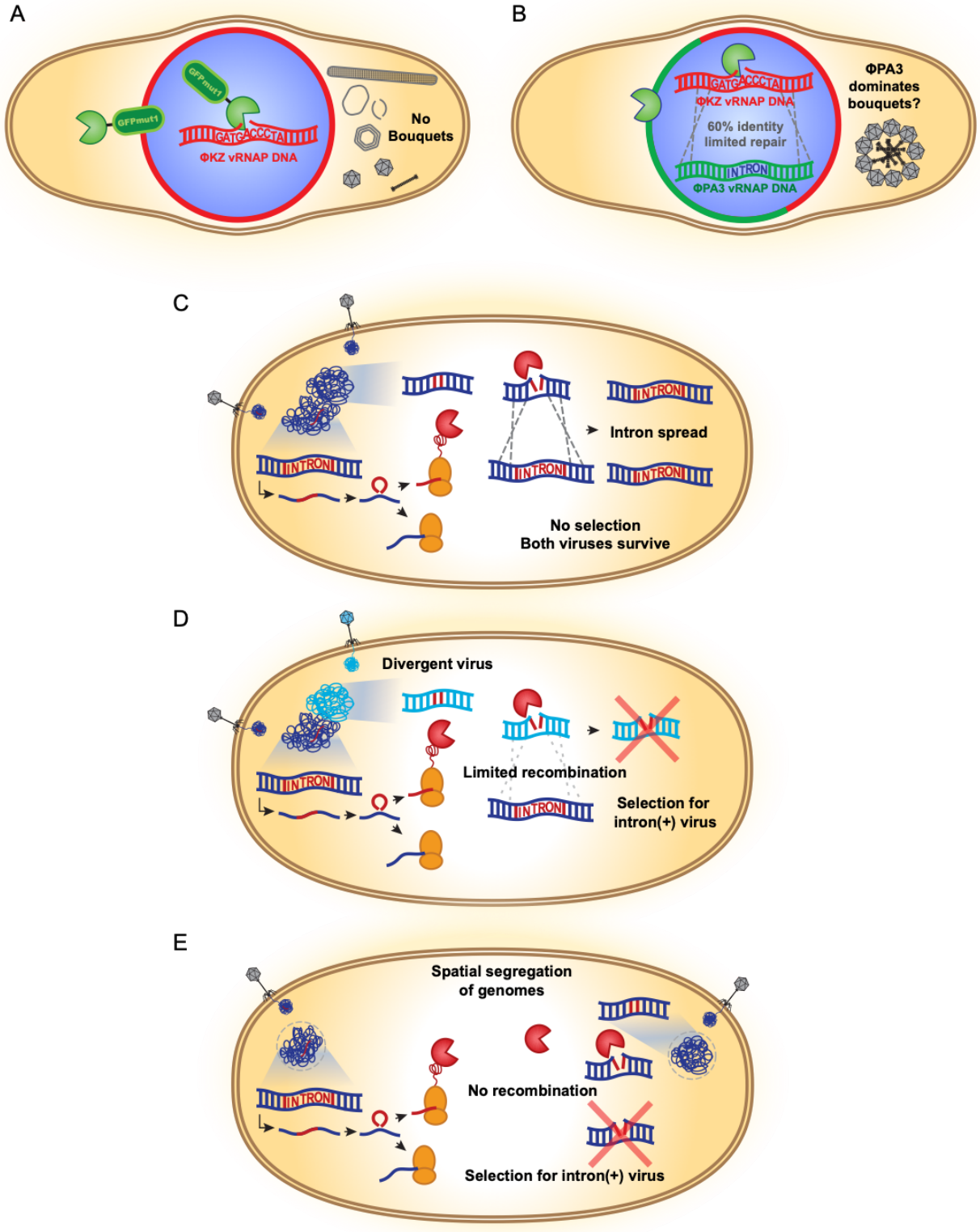
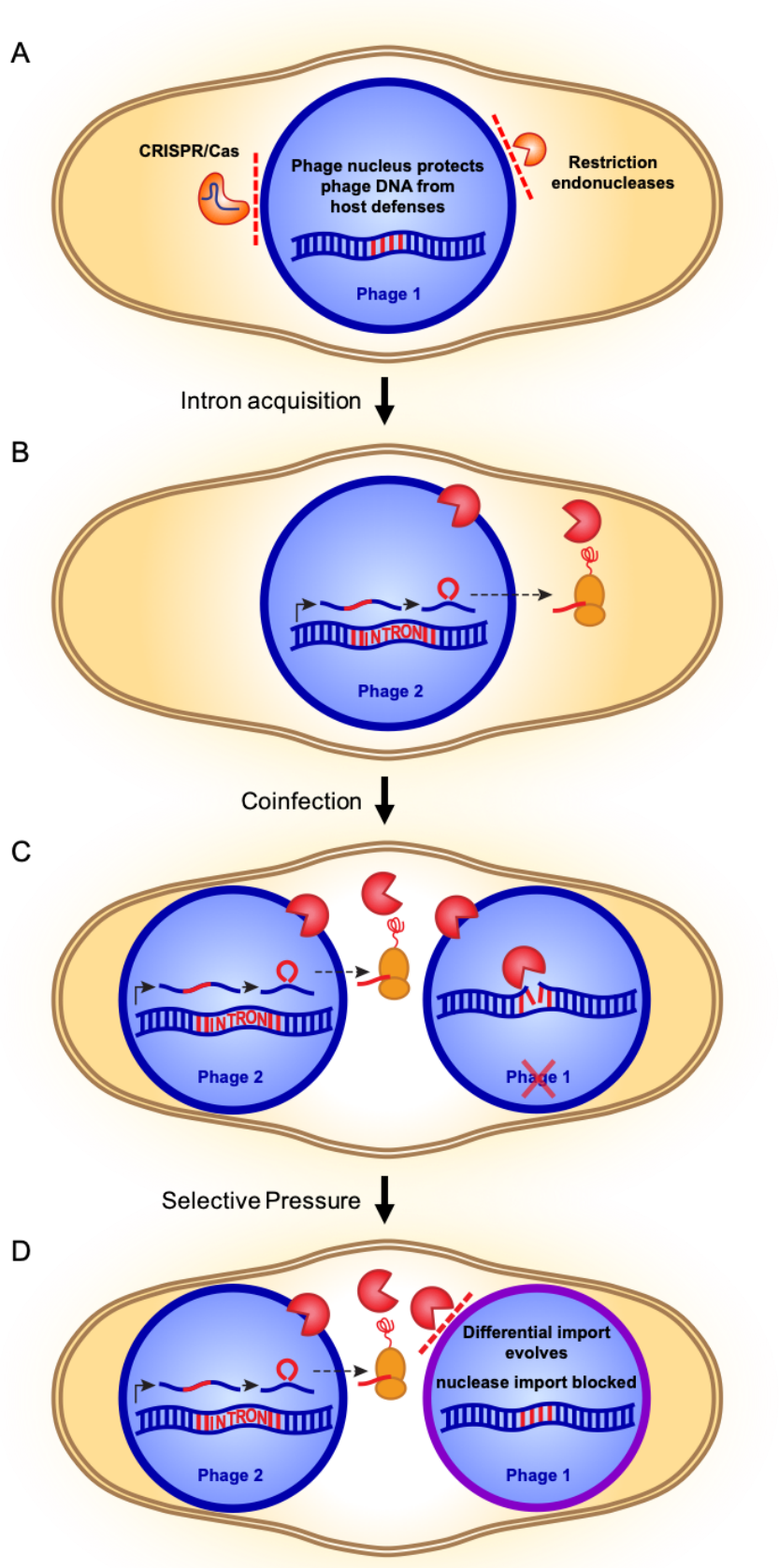

A
GFPmut1
GFPmut1
GATCACCCTA
ΦKZ vRNAP DNA
No
Bouquets
B
GATCACCCTA
ΦKZ vRNAP DNA
60% identity
limited repair
INTRON
ΦPA3 vRNAP DNA
ΦPA3
dominates
bouquets?
C
INTRON
INTRON
Intron spread
INTRON
INTRON
No selection
Both viruses survive
D
Divergent virus
INTRON
Limited recombination
INTRON
Selection for
intron(+) virus
E
Spatial segregation
of genomes
INTRON
No recombination
Selection for intron(+) virus

Figure 4.6 Model for the evolution of differential selective protein import by the phage nucleus driven by mobile intron competition. (A) The phage nucleus likely evolved in the ancestor to ΦPA3 and ΦKZ (Phage 1) to protect replicating phage DNA from host defenses such as restriction endonucleases and CRISPR/Cas. (B) One of the ancestral phages (Phage 2) then acquired the intron with the homing endonuclease that is ancestral to gp210. (C) Coinfection between Phage 1 and Phage 2 would result in the import of the homing endonuclease into both nuclei since they are identical except for the intronic sequence that Phage 2 acquired. The endonuclease would be able to cut the conserved sequence in Phage 1 but there would be no repair of Phage 1 since the template with the disrupted target site is sequestered in the nucleus of Phage 2. (D) The nuclease attack in (C) would provide a selective pressure on Phage 1 to evolve a different mechanism of protein import into the phage nucleus that would exclude the homing endonuclease that Phage 2 can import without issue. This model of mobile intron competition offers an explanation for the natural import of GFPmut1 into the nucleus of ΦKZ but not into the nucleus of ΦPA3 or 201φ2-1. Similarly, it explains the ability of the ΦKZ nucleus to naturally exclude gp210 while closely related phage ΦPA3 imports gp210.


A
CRISPR/Cas
Phage nucleus protects phage DNA from host defenses
Restriction endonucleases
Phage 1
Intron acquisition
B
INTRON
Phage 2
Coinfection
C
INTRON
Phage 2
Phage 1
Selective Pressure
D
INTRON
Phage 2
Differential import evolves
nuclease import blocked
Phage 1

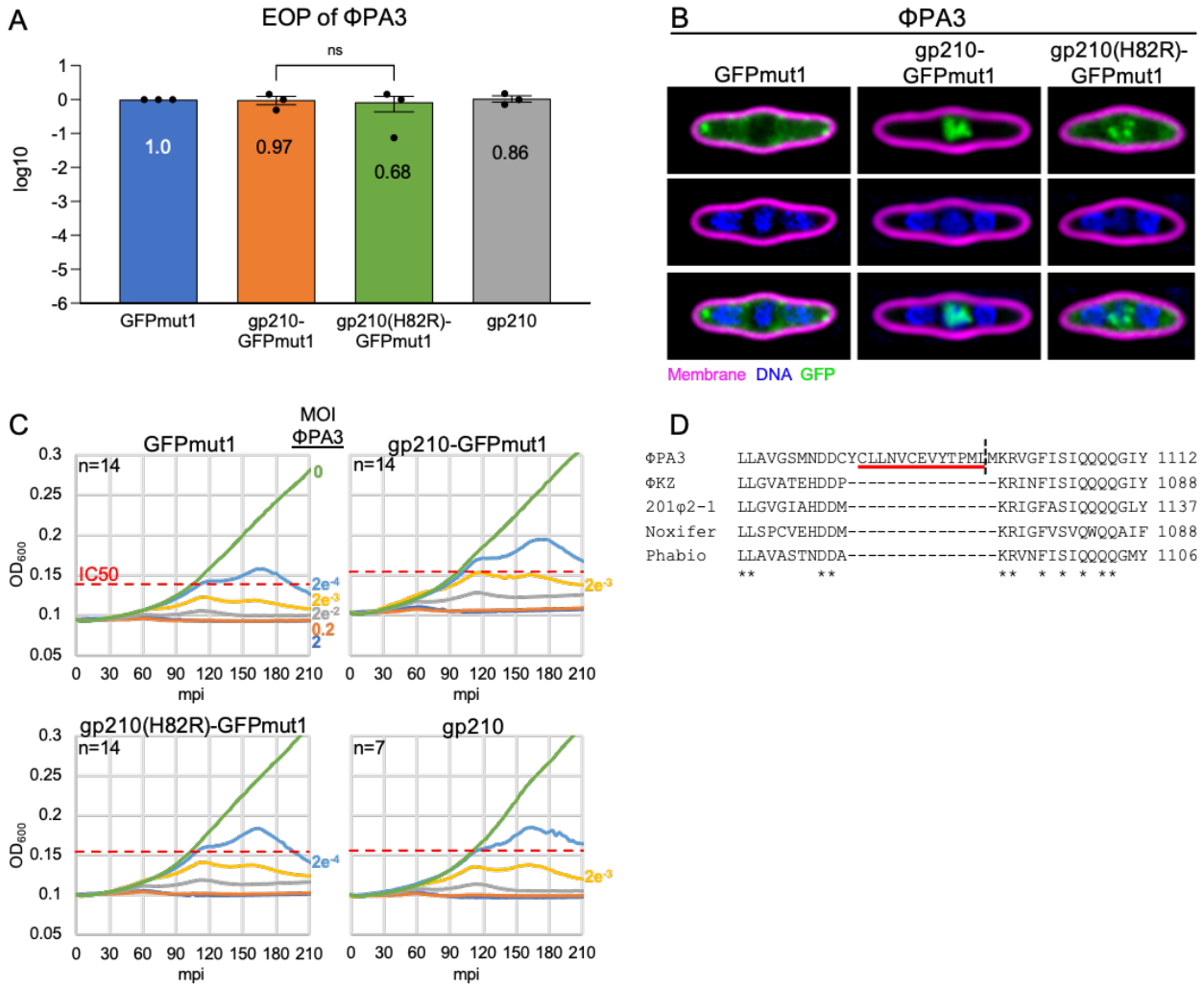


Figure S4.1 ΦPA3 is unaffected by gp210 and its RNAP is annotated with superfluous residues in Genbank. (A) ΦPA3 is not significantly affected by expression of gp210-GFPmut1 ($p = 0.74$, n = 3), 210H82R-GFPmut1 ($p = 0.51$, n = 3), or untagged 210 ($p = 0.93$, n = 3). The difference between gp210-GFPmut1 and gp210(H82R)-GFPmut1 is also insignificant ($p = 0.42$). Ratio paired t tests were used to determine p values. (B) Live fluorescent microscopy of stained ΦPA3 infections in the presence of either GFPmut1, gp210-GFPmut1, or 210H82R-GFPmut1, with no obvious morphological differences between them. (C) ΦPA3 growth curves demonstrating only a 10-fold decrease in ΦPA3 fitness with the expression of gp210-GFPmut1 (n = 14) or untagged 210 (n = 7) but no change with gp210(H82R)-GFPmut1. This is not surprising since overexpression of an active nuclease can be toxic, even to its original host. (D) Protein alignment of *Pseudomonas* jumbo phage RNAPs reveals an extra 15 residues are included in the Genbank annotation of ΦPA3 gp211. They are likely spliced out to reconstitute the protein as seen in homologous genes and as presented in Figure 4.2D.

4.8 References

1. Rohwer, F. and Barott, K., *Viral information.* Biology & Philosophy, 2013. **28**(2): p. 283-297.

2. Kristensen, D.M., Mushegian, A.R., Dolja, V.V., and Koonin, E.V., *New dimensions of the virus world discovered through metagenomics.* Trends in Microbiology, 2010. **18**(1): p. 11-19.

3. Adriaenssens, E. and Brister, J.R., *How to Name and Classify Your Phage: An Informal Guide.* Viruses, 2017. **9**(4).

4. Calisher, C.H., Briese, T., Brister, J.R., Charrel, R.N., Durrwald, R., Ebihara, H., Fulhorst, C.F., Gao, G.F., Groschup, M.H., Haddow, A.D., Hyndman, T.H., Junglen, S., Klempa, B., Klingstrom, J., Kropinski, A.M., Krupovic, M., LaBeaud, A.D., Maes, P., Nowotny, N., Nunes, M.R.T., Payne, S.L., Radoshitzky, S.R., Rubbenstroth, D., Sabanadzovic, S., Sasaya, T., Stenglein, M.D., Varsani, A., Wahl, V., Weaver, S.C., Zerbini, F.M., Vasilakis, N., and Kuhn, J.H., *Strengthening the Interaction of the Virology Community with the International Committee on Taxonomy of Viruses (ICTV) by Linking Virus Names and Their Abbreviations to Virus Species.* Syst Biol, 2019. **68**(5): p. 828-839.

5. International Committee on Taxonomy of Viruses Executive, C., *The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks.* Nat Microbiol, 2020. **5**(5): p. 668-674.

6. Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G., and Smith, D.B., *Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV).* Nucleic Acids Res, 2018. **46**(D1): p. D708-D717.

7. Bobay, L.M. and Ochman, H., *Biological species in the viral world.* Proc Natl Acad Sci U S A, 2018. **115**(23): p. 6040-6045.

8. Deng, L., Ignacio-Espinoza, J.C., Gregory, A.C., Poulos, B.T., Weitz, J.S., Hugenholtz, P., and Sullivan, M.B., *Viral tagging reveals discrete populations in Synechococcus viral genome sequence space.* Nature, 2014. **513**(7517): p. 242-5.

9. Dobzhansky, T., *Genetics and the Origin of Species*. 1937: Columbia University Press.

10. Mayr, E., *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. 1942: New York: Columbia Univ. Press.

11. Muller, H.J., *Isolating Mechanisms, Evolution, and Temperature.* Biology Symposium, 1942. **6**: p. 71–125.

12. Muller, H.J., *Reversibility in evolution considered from the standpoint of genetics.* Biological Reviews of the Cambridge Philosophical Society, 1939. **14**: p. 261-280.

13. Muller, H.J. and Pontecorvo, G., *Recombinants between Drosophila Species the F1 Hybrids of which are Sterile.* Nature, 1940. **146**(3693): p. 199-200.

14. Duffy, S., Burch, C.L., and Turner, P.E., *Evolution of host specificity drives reproductive isolation among RNA viruses.* Evolution, 2007. **61**(11): p. 2614-22.

15. Meyer, J.R., Dobias, D.T., Medina, S.J., Servilio, L., Gupta, A., and Lenski, R.E., *Ecological speciation of bacteriophage lambda in allopatry and sympatry.* Science, 2016. **354**(6317): p. 1301-1304.

16. Saxenhofer, M., Schmidt, S., Ulrich, R.G., and Heckel, G., *Secondary contact between diverged host lineages entails ecological speciation in a European hantavirus.* PLoS Biol, 2019. **17**(2): p. e3000142.

17. Chaikeeratisak, V., Birkholz, E.A., Prichard, A.M., Egan, M.E., Mylvara, A., Nonejuie, P., Nguyen, K.T., Sugie, J., Meyer, J.R., and Pogliano, J., *Viral speciation through subcellular genetic isolation and virogenesis incompatibility.* Nat Commun, 2021. **12**(1): p. 342.

18. Tomer, E., Cohen, E.M., Drayman, N., Afriat, A., Weitzman, M.D., Zaritsky, A., and Kobiler, O., *Coalescing replication compartments provide the opportunity for recombination between coinfecting herpesviruses.* FASEB J, 2019. **33**(8): p. 9388-9403.

19. Kieser, Q., Noyce, R.S., Shenouda, M., Lin, Y.J., and Evans, D.H., *Cytoplasmic factories, virus assembly, and DNA replication kinetics collectively constrain the formation of poxvirus recombinants.* PLoS One, 2020. **15**(1): p. e0228028.

20. Paszkowski, P., Noyce, R.S., and Evans, D.H., *Live-Cell Imaging of Vaccinia Virus Recombination.* PLoS Pathog, 2016. **12**(8): p. e1005824.

21. Erb, M.L., Kraemer, J.A., Coker, J.K., Chaikeeratisak, V., Nonejuie, P., Agard, D.A., and Pogliano, J., *A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells.* Elife, 2014. **3**.

22. Kraemer, J.A., Erb, M.L., Waddling, C.A., Montabana, E.A., Zehr, E.A., Wang, H., Nguyen, K., Pham, D.S., Agard, D.A., and Pogliano, J., *A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell.* Cell, 2012. **149**(7): p. 1488-99.

23. Zehr, E.A., Kraemer, J.A., Erb, M.L., Coker, J.K., Montabana, E.A., Pogliano, J., and Agard, D.A., *The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA.* Structure, 2014. **22**(4): p. 539-48.

24. Nguyen, K.T., Sugie, J., Khanna, K., Egan, M.E., Birkholz, E.A., Lee, J., Beierschmitt, C., Villa, E., and Pogliano, J., *Selective transport of fluorescent proteins into the phage nucleus.* PLoS One, 2021. **16**(6): p. e0251429.

25. Stoddard, B.L., *Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification.* Structure, 2011. **19**(1): p. 7-15.

26. Stoddard, B.L., *Homing endonucleases from mobile group I introns: discovery to genome engineering.* Mob DNA, 2014. **5**(1): p. 7.

27. Edgell, D.R., *Selfish DNA: homing endonucleases find a home.* Curr Biol, 2009. **19**(3): p. R115-7.

28. Edgell, D.R., Belfort, M., and Shub, D.A., *Barriers to intron promiscuity in bacteria.* J Bacteriol, 2000. **182**(19): p. 5281-9.

29. Edgell, D.R., Chalamcharla, V.R., and Belfort, M., *Learning to live together: mutualism between self-splicing introns and their hosts.* BMC Biol, 2011. **9**: p. 22.

30. Belle, A., Landthaler, M., and Shub, D.A., *Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns.* Genes Dev, 2002. **16**(3): p. 351-62.

31. Belfort, M., *Bacteriophage introns: parasites within parasites?* Trends Genet, 1989. **5**(7): p. 209-13.

32. Belfort, M. and Roberts, R.J., *Homing endonucleases: keeping the house in order.* Nucleic Acids Res, 1997. **25**(17): p. 3379-88.

33. Nielsen, H., *Ribozymes: Methods and Protocols*. Group I Intron Ribozymes, ed. J.S. Hartig. 2012, Totowa, NJ: Humana Press. 73–89.

34. Zeng, Q., Bonocora, R.P., and Shub, D.A., *A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages.* Curr Biol, 2009. **19**(3): p. 218-22.

35. Bonocora, R.P. and Shub, D.A., *A likely pathway for formation of mobile group I introns.* Curr Biol, 2009. **19**(3): p. 223-8.

36. Loizos, N., Tillier, E.R., and Belfort, M., *Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease.* Proc Natl Acad Sci U S A, 1994. **91**(25): p. 11983-7.

37. Bechhofer, D.H., Hue, K.K., and Shub, D.A., *An intron in the thymidylate synthase gene of Bacillus bacteriophage beta 22: evidence for independent evolution of a gene, its group I intron, and the intron open reading frame.* Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11669-73.

38. Goodrich-Blair, H. and Shub, D.A., *The DNA polymerase genes of several HMU-bacteriophages have similar group I introns with highly divergent open reading frames.* Nucleic Acids Res, 1994. **22**(18): p. 3715-21.

39. Katharios, P., Kalatzis, P.G., Kokkari, C., Sarropoulou, E., and Middelboe, M., *Isolation and characterization of a N4-like lytic bacteriophage infecting Vibrio splendidus, a pathogen of fish and bivalves.* PLoS One, 2017. **12**(12): p. e0190083.

40. Cech, T.R., *Conserved sequences and structures of group I introns: building an active site for RNA catalysis--a review.* Gene, 1988. **73**(2): p. 259-71.

41. Goodrich-Blair, H. and Shub, D.A., *Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers.* Cell, 1996. **84**(2): p. 211-21.

42. Robbins, J.B., Stapleton, M., Stanger, M.J., Smith, D., Dansereau, J.T., Derbyshire, V., and Belfort, M., *Homing endonuclease I-TevIII: dimerization as a means to a double-strand break.* Nucleic Acids Res, 2007. **35**(5): p. 1589-600.

43. Bell-Pedersen, D., Quirk, S., Clyman, J., and Belfort, M., *Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications.* Nucleic Acids Res, 1990. **18**(13): p. 3763-70.

44. Chu, F.K., Maley, G., Pedersen-Lane, J., Wang, A.M., and Maley, F., *Characterization of the restriction site of a prokaryotic intron-encoded endonuclease.* Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3574-8.

45. Kong, S., Liu, X., Fu, L., Yu, X., and An, C., *I-PfoP3I: a novel nicking HNH homing endonuclease encoded in the group I intron of the DNA polymerase gene in Phormidium foveolarum phage Pf-WMP3.* PLoS One, 2012. **7**(8): p. e43738.

46. Shen, B.W., Landthaler, M., Shub, D.A., and Stoddard, B.L., *DNA binding and cleavage by the HNH homing endonuclease I-HmuI.* J Mol Biol, 2004. **342**(1): p. 43-56.

47. Landthaler, M., Begley, U., Lau, N.C., and Shub, D.A., *Two self-splicing group I introns in the ribonucleotide reductase large subunit gene of Staphylococcus aureus phage Twort.* Nucleic Acids Res, 2002. **30**(9): p. 1935-43.

48. Landthaler, M. and Shub, D.A., *The nicking homing endonuclease I-BasI is encoded by a group I intron in the DNA polymerase gene of the Bacillus thuringiensis phage Bastille.* Nucleic Acids Res, 2003. **31**(12): p. 3071-7.

49. Metzger, M.J., McConnell-Smith, A., Stoddard, B.L., and Miller, A.D., *Single-strand nicks induce homologous recombination with less toxicity than double-strand breaks using an AAV vector template.* Nucleic Acids Res, 2011. **39**(3): p. 926-35.

50. Thaler, D.S., Stahl, M.M., and Stahl, F.W., *Tests of the double-strand-break repair model for red-mediated recombination of phage lambda and plasmid lambda dv.* Genetics, 1987. **116**(4): p. 501-11.

51. Quirk, S.M., Bell-Pedersen, D., and Belfort, M., *Intron mobility in the T-even phages: high frequency inheritance of group I introns promoted by intron open reading frames.* Cell, 1989. **56**(3): p. 455-65.

52. Martinsohn, J.T., Radman, M., and Petit, M.A., *The lambda red proteins promote efficient recombination between diverged sequences: implications for bacteriophage genome mosaicism.* PLoS Genet, 2008. **4**(5): p. e1000065.

53. Li, X.T., Thomason, L.C., Sawitzke, J.A., Costantino, N., and Court, D.L., *Bacterial DNA polymerases participate in oligonucleotide recombination.* Mol Microbiol, 2013. **88**(5): p. 906-20.

54. Landthaler, M., Shen, B.W., Stoddard, B.L., and Shub, D.A., *I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities.* J Mol Biol, 2006. **358**(4): p. 1137-51.

55. Bell-Pedersen, D., Quirk, S.M., Bryk, M., and Belfort, M., *I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target.* Proc Natl Acad Sci U S A, 1991. **88**(17): p. 7719-23.

56. Sitbon, E. and Pietrokovski, S., *New types of conserved sequence domains in DNA-binding regions of homing endonucleases.* Trends Biochem Sci, 2003. **28**(9): p. 473-7.

57. Mate, M.J. and Kleanthous, C., *Structure-based analysis of the metal-dependent mechanism of H-N-H endonucleases.* J Biol Chem, 2004. **279**(33): p. 34763-9.

58. Goodrich-Blair, H., Scarlato, V., Gott, J.M., Xu, M.Q., and Shub, D.A., *A self-splicing group I intron in the DNA polymerase gene of Bacillus subtilis bacteriophage SPO1.* Cell, 1990. **63**(2): p. 417-24.

59. Liu, Q., Belle, A., Shub, D.A., Belfort, M., and Edgell, D.R., *SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site.* J Mol Biol, 2003. **334**(1): p. 13-23.

60. Russell, R.L. and Huskey, R.J., *Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism.* Genetics, 1974. **78**(4): p. 989-1014.

61. Kadyrov, F.A., Shlyapnikov, M.G., and Kryukov, V.M., *A phage T4 site-specific endonuclease, SegE, is responsible for a non-reciprocal genetic exchange between T-even-related phages.* FEBS Lett, 1997. **415**(1): p. 75-80.

62. Gary, T.P., Colowick, N.E., and Mosig, G., *A species barrier between bacteriophages T2 and T4: exclusion, join-copy and join-cut-copy recombination and mutagenesis in the dCTPase genes.* Genetics, 1998. **148**(4): p. 1461-73.

63. Landthaler, M., Lau, N.C., and Shub, D.A., *Group I intron homing in Bacillus phages SPO1 and SP82: a gene conversion event initiated by a nicking homing endonuclease.* J Bacteriol, 2004. **186**(13): p. 4307-14.

64. Sandegren, L., Nord, D., and Sjoberg, B.M., *SegH and Hef: two novel homing endonucleases whose genes replace the mobC and mobE genes in several T4-related phages.* Nucleic Acids Res, 2005. **33**(19): p. 6203-13.

65. Edgell, D.R., Fast, N.M., and Doolittle, W.F., *Selfish DNA: the best defense is a good offense.* Curr Biol, 1996. **6**(4): p. 385-8.

66. Landthaler, M. and Shub, D.A., *Unexpected abundance of self-splicing introns in the genome of bacteriophage Twort: introns in multiple genes, a single gene with three introns, and exon skipping by group I ribozymes.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 7005-10.

67. Deeg, C.M., Chow, C.T., and Suttle, C.A., *The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea.* Elife, 2018. **7**.

68. Douglas, G.M. and Shapiro, B.J., *Genic Selection Within Prokaryotic Pangenomes.* Genome Biol Evol, 2021. **13**(11).

69. Robbins, J.B., Smith, D., and Belfort, M., *Redox-responsive zinc finger fidelity switch in homing endonuclease and intron promiscuity in oxidative stress.* Curr Biol, 2011. **21**(3): p. 243-8.

70. Coros, C.J., Piazza, C.L., Chalamcharla, V.R., Smith, D., and Belfort, M., *Global regulators orchestrate group II intron retromobility.* Mol Cell, 2009. **34**(2): p. 250-6.

71. Chaikeeratisak, V., Khanna, K., Nguyen, K.T., Sugie, J., Egan, M.E., Erb, M.L., Vavilina, A., Nonejuie, P., Nieweglowska, E., Pogliano, K., Agard, D.A., Villa, E., and Pogliano, J., *Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria.* Cell, 2019. **177**(7): p. 1771-1780.e12.

72. Chaikeeratisak, V., Nguyen, K., Egan, M.E., Erb, M.L., Vavilina, A., and Pogliano, J., *The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages.* Cell Reports, 2017. **20**(7): p. 1563-1571.

73. Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., Villa, E., Agard, D.A., and Pogliano, J., *Assembly of a nucleus-like structure during viral replication in bacteria.* Science, 2017. **355**(6321): p. 194-197.

74. Saitou, N. and Nei, M., *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Mol Biol Evol, 1987. **4**(4): p. 406-25.

75. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K., *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Molecular Biology and Evolution, 2018. **35**(6): p. 1547-1549.

76. Stecher, G., Tamura, K., and Kumar, S., *Molecular Evolutionary Genetics Analysis (MEGA) for macOS.* Molecular Biology and Evolution, 2020. **37**(4): p. 1237-1239.

77. Felsenstein, J., *Confidence Limits on Phylogenies: An Approach Using the Bootstrap.* Evolution, 1985. **39**(4): p. 783.

78. Zuckerkandl, E.a.P., L., *Evolutionary divergence and convergence in proteins.* Evolving Genes and Proteins, ed. V.a.V. Bryson, H.J. 1965, New York: Academic Press.

79. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A., *Fiji: an open-source platform for biological-image analysis.* Nature Methods, 2012. **9**(7): p. 676-682.

80. Lam, V. and Villa, E., *Practical Approaches for Cryo-FIB Milling and Applications for Cellular Cryo-Electron Tomography*, in *cryoEM*. 2021, Springer US. p. 49-82.

81. Mastronarde, D.N., *Automated electron microscope tomography using robust prediction of specimen movements.* J Struct Biol, 2005. **152**(1): p. 36-51.

82. Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., and Agard, D.A., *MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy.* Nature Methods, 2017. **14**(4): p. 331-332.

83. Kremer, J.R., Mastronarde, D.N., and McIntosh, J.R., *Computer visualization of three-dimensional image data using IMOD.* J Struct Biol, 1996. **116**(1): p. 71-6.

84. Chen, M., Bell, J.M., Shi, X., Sun, S.Y., Wang, Z., and Ludtke, S.J., *A complete data processing workflow for cryo-ET and subtomogram averaging.* Nature Methods, 2019. **16**(11): p. 1161-1168.

85. Ceyssens, P.J., Minakhin, L., Van den Bossche, A., Yakunina, M., Klimuk, E., Blasdel, B., De Smet, J., Noben, J.P., Blasi, U., Severinov, K., and Lavigne, R., *Development of giant bacteriophage varphiKZ is independent of the host transcription apparatus.* J Virol, 2014. **88**(18): p. 10501-10.

86. Holmes, S.F., Santangelo, T.J., Cunningham, C.K., Roberts, J.W., and Erie, D.A., *Kinetic investigation of Escherichia coli RNA polymerase mutants that influence nucleotide discrimination and transcription fidelity.* J Biol Chem, 2006. **281**(27): p. 18677-83.

87. Zaug, A.J. and Cech, T.R., *The intervening sequence excised from the ribosomal RNA precursor of Tetrahymena contains a 5-terminal guanosine residue not encoded by the DNA.* Nucleic Acids Res, 1982. **10**(9): p. 2823-38.

88. Cech, T.R., *The generality of self-splicing RNA: relationship to nuclear mRNA splicing.* Cell, 1986. **44**(2): p. 207-10.

89. Thomas, J.A., Rolando, M.R., Carroll, C.A., Shen, P.S., Belnap, D.M., Weintraub, S.T., Serwer, P., and Hardies, S.C., *Characterization of Pseudomonas chlororaphis myovirus 201varphi2-1 via genomic sequencing, mass spectrometry, and electron microscopy.* Virology, 2008. **376**(2): p. 330-8.

90. Lecoutere, E., Ceyssens, P.J., Miroshnikov, K.A., Mesyanzhinov, V.V., Krylov, V.N., Noben, J.P., Robben, J., Hertveldt, K., Volckaert, G., and Lavigne, R., *Identification and*

comparative analysis of the structural proteomes of phiKZ and EL, two giant Pseudomonas aeruginosa bacteriophages. Proteomics, 2009. **9**(11): p. 3215-9.

91. Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gumy, L.F., and Fineran, P.C., *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, 2020. **5**(1): p. 48-55.

92. Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J., *A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases.* Nature, 2020. **577**(7789): p. 244-248.

CHAPTER 5:

Concluding remarks and discussion

Bacteriophage are the most abundant and diverse entities on Earth [1, 2], yet we have only recently discovered that they have the ability to build the incredibly complex structures of the phage nucleus and PhuZ spindle within host cells [3-8]. The characterization of this intricate replication system established by three *Phikzviruses* infecting *Pseudomonas chlororaphis* and *P. aeruginosa* begs the question of how widespread this astonishing replication mechanism could be and how conserved or divergent it is in more distantly related phage. To investigate this, I characterized the major stages of nucleus-forming jumbo phage replication in the tractable model bacterium *Escherichia coli*. Phage vB_EcoM_Goslar (Goslar) [9] was isolated as a potential therapeutic to fight avian pathogenic *E. coli* and we discovered that its genome encodes homologs to the *Phikzvirus* proteins that form the phage nucleus and tubulin-like PhuZ spindle.

Several key processes are conserved in Goslar; a proteinaceous phage nucleus forms around the phage genome and grows as the DNA replicates, excluding metabolic proteins and ribosomes to maintain uncoupled transcription and translation while the entire structure rotates, a cytoskeleton of PhuZ filaments is assembled, and phage bouquets [10] are formed by DNA-filled capsids. Despite this general conservation of the three major stages of nucleus-forming jumbo phage replication, I found intriguing divergences as well. Instead of the midcell positioning of the phage nucleus performed by the *Phikzviruses*, Goslar could position the nucleus anywhere along the cell length. I observed that positioning relied on the location of cell bulging which is a stage of this replication cycle that is currently under investigation in our lab. Another difference I detected was that the phage bouquets of Goslar are more prominent than those of the *Phikzviruses* and they are more organized with capsids localizing to the inside of the spherical bouquet at later time points. This contrasts the *Phikzvirus* bouquets that range from disorganized to spherical without internal capsids [10]. Finally, the most fascinating divergent feature of the Goslar

replication cycle is the assembly of the PhuZ cytoskeleton into a vortex-like array. As opposed to the bipolar spindle of filaments formed by the *Phikzviruses* [3-8], Goslar PhuZ filaments wrap around the phage nucleus and project radially towards the cell membrane in every direction, resembling a vortex. This cytoskeletal vortex composed of the tubulin-like PhuZ protein functions to rotate the phage nucleus inside the *E. coli* cell.

This is the first time a cytoskeletal vortex has been observed within a prokaryote while in eukaryotes, a vortex-like array of microtubules drives cytoplasmic streaming [11-14] and has been shown to spontaneously self-organize *in vitro* when spatially confined [15]. It is also remarkable that the function of the Goslar cytoskeletal vortex is to perform intracellular rotation, an activity only witnessed thus far in the *Phikzviruses* and in eukaryotic motile cells that must realign the nucleus with cell polarity [16-22]. However, the nuclear rotation observed in eukaryotes is not a continually progressive rotation as occurs in the nucleus-forming phage jumbo phage. These noteworthy features of the Goslar replication cycle are not only reminiscent of eukaryotic mechanisms but their divergence between nucleus-forming jumbo phage implies their evolutionary significance. The evolution of diverse methods for achieving this style of viral replication suggests that it is established in evolutionary history and that it offers advantages to surviving evolutionary pressures. The diversity of this replication cycle revealed by the characterization of Goslar in *E. coli* points to the significance of nucleus-forming jumbo phage on the evolutionary stage.

Competition has been fierce on this evolutionary stage and since phage far outnumber their host bacteria, they have been locked in a millennia-long competition for host resources. This competition has resulted in a milieu of mechanisms, both antagonistic and cooperative, that govern the interactions of these intracellular parasites, as well as the viruses that infect all kingdoms of

life. The way that viruses interact with each other shapes their speciation as they either recombine their genomes or they isolate and diverge from each other, both outcomes contributing more genetic diversity to our biome. This wealth of viral genetic diversity is currently organized by species delineations that rely largely on sequence similarity [23], with no consideration of the biological capacity for different viruses to mate, or recombine their genetic information. A greater understanding of viral diversity is being reached by the application of the Biological Species Concept [24-26] which defines species as gene pools that are separated by reproductive isolating mechanisms.

Mechanisms that isolate different viral gene pools from each other have been observed for decades, with only recent acknowledgement that traditional concepts of speciation and evolution also apply to viruses [24]. Since viruses in general can only exchange genetic material while replicating inside of a host, coinfection is the major opportunity for viral "mating" [27, 28] and it happens commonly in nature [29-33]. If two viruses occupy physically distinct ecosystems [30, 34, 35], or they do not share host tropism, they cannot achieve coinfection, and therefore experience a strong reproductive isolation that can lead to their divergence into separate species [36-38]. If two viruses have the capacity to infect the same cell, superinfection exclusion may still prevent coinfection [39, 40]. Once a coinfection is successfully established, the viruses may recombine to produce hybrids with either increased or decreased virulence [41-44] which will either reinforce genetic mixing or it will reinforce genetic isolation, respectively.

Interestingly, the intracellular mechanisms of reproductive isolation between viruses with the capacity to coinfect a single host cell have been understudied but appear to exist based on the clustering of viral genomes into biological species within overlapping host tropisms [24]. The intracellular aspect of viral interaction has been observed in vaccinia virus [45-48] and HSV-1 [49]

where viral factories occupy discrete regions of the host and only recombine when they spatially overlap. We have since then investigated the intricate intracellular interactions between two nucleus-forming jumbo phage that infect the human pathogen *Pseudomonas aeruginosa* and discovered two new universally applicable viral speciation factors that we have termed Subcellular Genetic Isolation, which can be applied to HSV-1 and vaccinia virus, and Virogenesis Incompatibility [50].

ΦKZ or ΦPA3 infecting *P. aeruginosa* separately were observed to form 2 or more distinct phage nuclei in ~20% of single species infections [50]. This number did not significantly decrease as the infection progressed, suggesting that the nuclei do not regularly fuse together. This same spatial separation of single species coinfections was seen when the major phage nucleus protein was fluorescently tagged, demonstrating that the shell of the phage nucleus formed a physical barrier around each nucleus. This is expected to essentially eliminate genetic exchange when compared to the amount of recombination possible when the genomes can physically interact with one another [46, 49]. Since the major protein of the phage nucleus appears to require no genetic divergence to reduce the capacity for genetic exchange, the phage nucleus is a potent facilitator of subcellular genetic isolation. ΦKZ and ΦPA3 (~80% sequence identity) coinfections were then studied for speciation factors acting between diverging phage. Each major phage nucleus protein, tagged with two different fluorophores, formed strictly around its own genome in a single infection of either phage. Of the coinfections of ΦKZ and ΦPA3, 75% formed two separate nuclei, demonstrating subcellular genetic isolation between the majority of coinfections.

Coinfections between ΦKZ and ΦPA3, but not between the same phage, displayed a defect in the centering of the nucleus [50], which is important for the full efficiency of the replication cycle [3]. The PhuZ proteins from the two different phages were found to colocalize into hybrid

spindles that lost their dynamic nature, explaining the defect in nucleus positioning, and introducing another new speciation factor for viruses, Virogenesis Incompatibility [50]. This term regards the incompatibility between divergent replication components that interferes with the generation of virions. A second virogenesis incompatibility factor, gp210, was discovered in the nucleus of ΦPA3. Since this protein is normally imported into the ΦPA3 nucleus, it would come into contact with ΦKZ DNA in the 25% of coinfections that form a hybrid nuclear compartment. Artificial import of gp210 into the ΦKZ nucleus resulted in a drastic decrease in ΦKZ titer (99.41%) and in ΦKZ growth curves (100,000-fold decrease in IC50 of host), while displaying no obvious morphological defects of the single cell infections [50]. This ΦPA3 nucleus-residing protein is expected to affect the outcome of the 25% of coinfections that form a hybrid nucleus, possibly by reducing the proportion of ΦKZ virions produced from that coinfection. This incompatibility between ΦKZ DNA and gp210 in the ΦPA3 nucleus which interferes with the success of ΦKZ replication, makes gp210 another virogenesis incompatibility factor. The nature and mechanism of ΦKZ interference caused by this ΦPA3 protein gp210 was therefore investigated in more detail.

Bioinformatic analyses predicted that gp210 is an HNH homing endonuclease and it was experimentally confirmed that histidine 82 of the HNH endonuclease motif is required for gp210 inhibition of ΦKZ. However, gp210 did not cause an overall decrease in the amount of DNA produced by ΦKZ, nor did it affect host lysis by ΦKZ. Further bioinformatic analysis revealed that gp210 is housed within a group I intron interrupting a subunit of the ΦPA3 RNA polymerase (RNAP). That genomic locus is unoccupied by any intron in ΦKZ but at that locus in 201φ2-1, which cannot coinfect with ΦPA3, there is a much smaller intron that contains no ORF yet displays high nucleotide conservation to the gp210 intron. Interestingly, the intron of 201φ2-1 appears to

be a degenerate version beginning with strong homology to the last 108 nucleotides of the gp210 coding sequence. The rest of the intron is highly conserved with the gp210 3' intronic RNA except for a 64-nucleotide deletion in the middle. It is possible that this degeneration of the gp210 intron in ΦPA3 has not occurred because there is a competitive advantage conferred by the ability of gp210 to interfere with coinfecting ΦKZ DNA. Gp210 would also be protected from deletion in ΦPA3 if it plays a role in regulating the intron as a folding and splicing chaperone or as a transcriptional regulator, consistent with its native import into the ΦPA3 nucleus. Targeting of a coinfecting phage at the unoccupied genomic locus of the mobile intron is a common characteristic of phage-encoded homing endonucleases [51, 52], so I hypothesized that gp210 is able to recognize and cut the homologous ΦKZ RNAP gene, gp178. Gp178 is a subunit of the ΦKZ RNAP that transcribes only early genes [53] so a loss of this gene product would not affect the first round of infection but instead would leave the progeny unable to transcribe early genes in the next round of infection, consistent with prior data. It is important to note that gp178 is packaged into the virion [54] and therefore it may also be structurally necessary for the proper assembly of virions.

In vitro, purified gp210 appeared to preferentially nick a plasmid containing ΦKZ gp178, supporting my hypothesis. To look for genetic mutations that allowed ΦKZ to escape inhibition by gp210, I isolated 3 clones that could plaque on a host expressing gp210. Three independent ΦKZ isolates with verified resistance to gp210 were sequenced and found to all have a single nucleotide substitution in the same position, adenosine 3215 of ΦKZ gp178, 2 nucleotides upstream of the gp210 intron insertion site. Two isolates mutated adenosine 3215 to cytosine while the third mutated it to guanine, resulting in a coding change of D1072A or D1072G, respectively. That aspartic acid residue is highly conserved among phage RNAPs and is also

conserved in the RpoB gene of *P. aeruginosa* and in *E. coli* where it has been shown to control RNAP fidelity (D675; [55]). These results are consistent with homing endonuclease targeting of a conserved region in an essential gene, matching the genomic locus where the intron is inserted in its host genome.

To elucidate the biological consequences of gp210 targeting of ΦKZ DNA within gp178, we isolated the progeny produced from ΦKZ infections proceeding with gp210 synthetically imported into the nucleus and found that they had severely reduced virulence (0.00001% plaquing efficiency relative to progeny produced without gp210). Cryo-FIB-ET of ΦKZ infections with gp210 compared to ΦKZ infections without any artificially expressed proteins revealed a significant loss of DNA-filled capsids and an abundance of strange structures never before observed in this system, some of which resemble empty capsids. Due to a lack of DNA-filled capsids, infections with gp210 also lack phage bouquets where final virion maturation occurs. However, we did observe ribosome-free regions of approximately the same size and location as bouquets, lacking a discernible protein barrier which differentiated them from phage nuclei. This is an exciting avenue for future investigations into the composition and assembly mechanisms of phage bouquets. For the investigation of gp210, these data revealed that gp210 prevents ΦKZ capsid assembly.

Introns have been shown to influence speciation in yeast [56, 57] and archaea [58, 59] while in viruses, they have only been shown to provide an advantage to certain neighboring genes, rather than to the entire virus [60-62]. Introns have been implicated as a driver of evolutionary transitions [63] and their ubiquity among viruses [64, 65] supports the magnitude of their influence on viral speciation. The prevalence of homologs to the major phage nucleus protein and PhuZ in phage infecting diverse hosts including *Salmonella, Ralstonia, Cronobacter, Erwinia, Vibrio,*

Serratia, and *E. coli* also suggests that these reproductive isolating mechanisms of the phage nucleus and the PhuZ spindle, along with mobile introns, are widespread among phage and therefore greatly influence the genetic landscape of the earth's virome.

The human microbiome also consists of an entire ecosystem of viruses and microbes that are constantly evolving and speciating in response to competition and cooperation with each other and with the human host. This magnificent biome of complex relationships plays vital roles in human health [66-69], yet it is vulnerable to perturbation that results in dysbiosis. The most common disruption of this delicate balance of intricate interactions is the use of broad-spectrum antibiotics [70-73]. These chemical compounds are most often discovered from natural sources because bacteria have been producing them to compete with each other for millenia [74]. However, when used as a therapeutic to eliminate a pathogenic bacterial overgrowth, the pathogens are not the only microbes that suffer. Many species of our resident bacteria that positively contribute to the ecosystem of the human body are obliterated along with the ones causing trouble. This slash-and-burn approach to bacterial pathologies has saved countless lives but we are overdue for a makeover that utilizes a point-and-shoot strategy with more finesse. Our (over)use of broad-spectrum antibiotics and the saturation of the biosphere with them [74] has led us into an era of antibiotic resistance where many people are again losing their lives to bacterial infections that have gained the ability to evade every antibiotic we have [75]. Resistance to antibiotics arises quickly, which is to be expected considering that bacteria have been using them to compete with each other for millions of years.

To remain in harmony with our bacterial counterparts, we will need to enlist the help of the entities which have been evolving in harmony with these microbes since the beginning of life as we know it [76, 77]. Bacteriophage shape bacterial populations with more finesse than we could

ever achieve with chemical antibiotics [78]. They are more discerning about which microbes are eliminated and they continually evolve to overcome resistance tactics developed by the bacterial hosts [79]. Phages have been used as a therapeutic for over 100 years, particularly in the countries of Georgia [80] and Poland [81], since their discovery by Felix d'Herelle [82] and Frederick Twort [83]. In Poland, phage therapy remains a second line of defense only to be used once antibiotics fail, but in Georgia, phages are part of the standard of care. The commercial arm of the Eliava Consortium, Eliava BioPreparations LTD, currently produces 6 phage cocktails which can be picked up at a local pharmacy to combat the most common bacterial infections, without disturbing the majority of the beneficial microbes of the human microbiome. One of those cocktails, PyoPhage, has traditionally included our good friend, the nucleus-forming jumbo phage ΦKZ [84]. ΦPA3 was also demonstrated to lyse clinical isolates of *P. aeruginosa* from cystic fibrosis patients [85]. Other jumbo phages have made it to the spotlight of growing phage therapy in the West [86-89] and giant phages with genomes up to 552 kb have been discovered to be residents of the human gut biome [90]. Until 2017 [6], it was completely unknown that a jumbo phage could construct a proteinaceous nucleus around its DNA and protect its genome from bacterial defenses [91-93]. This is a glimpse into the realm of possibility presented by giant bacteriophages that we have yet to fully explore. These are the phages that contribute to who we are as human biomes and the phages that have been saving us from a fatal imbalance with our microbes [84, 85, 94].

Not only are nucleus-forming jumbo phages vital to who we are today, but they may hold the key to how we came to be. Two billion years ago the Last Eukaryotic Common Ancestor (LECA) [95-97] was born, but how that occurred remains a mystery. The first eukaryote is predicted to have harbored all of the hallmarks of a eukaryote since there is not yet any evidence of a gradual acquisition of each trait [98]. All eukaryotes contain a nucleus that encloses the

genome and uncouples transcription from translation as well as membrane-bound organelles such as the mitochondria and endoplasmic reticulum. One of the theories for the seemingly sudden emergence of the first eukaryote is the Viral Eukaryogenesis hypothesis first published by Philip Bell in 2001 [99, 100]. He proposed that the eukaryotic nucleus evolved from a complex DNA virus that persisted in an archaeal host [96] where viral membrane fusion mechanisms developed into phagocytosis which allowed the predation of bacterial syntrophs that would become the mitochondria [101]. As the viral compartment acquired a set of essential genes from the archaeal host, and metabolism was carried out by the bacterial hostages, the archaeal chromosome became redundant and was lost, and this tripartite organism became the first eukaryote [99, 100]. Bell provides arguments based on characteristics that are shared between the eukaryotic nucleus and viruses but not by prokaryotes including the closest archaeal relatives; mRNA capping, linear chromosomes, and uncoupled transcription and translation, all of which were fully developed in the LECA. Another theory pertaining to the advent of the nucleus during eukaryogenesis is the hypothesis spearheaded by Eugene Koonin which posits that intron invasion provided the selective pressure for a genome that is compartmentalized by a nucleus [63, 102, 103]. Koonin and colleagues make a case for the evolution of a nuclear membrane by an unknown mechanism to protect the archaeal genome that was being disrupted by selfish group II mobile introns harbored in the ancestral mitochondria [104-106]. I propose that these two theories can be combined to create a simplified version, satisfying Occam's razor. I hypothesize that the intron invasion from the ancestral mitochondria was the selective force that championed the protected viral DNA compartment over the increasingly redundant archaeal chromosome, sealing its fate and establishing the first eukaryotic cell composed of an archaeal cell boundary, a bacterial/archaeal

hybrid cytoplasm with membrane-bound organelles, and a viral genetic compartment – the nucleus.

When investigating theories of ancient evolutionary transitions, we tend to look for extant organisms that resemble a stepping stone between the two stages of development, the "missing link". Nucleus-forming jumbo phage are a promising link between simple prokaryotes and the complex nucleated eukaryotes. The phage nucleus displays several characteristics that support its relation to the ancestor of the eukaryotic nucleus [100, 107]. While the term "phage nucleus" originally received criticism due to the lack of a double membrane surrounding it, it is important to remember that the eukaryotic nucleus is scaffolded by a protein shell called the lamina [108]. Therefore, it is reasonable to hypothesize that the nucleus first consisted of a barrier made of protein only, akin to the phage nucleus, later becoming surrounded by layers of lipid membranes. Also important to the assignment of the phage nucleus as the "missing link" between prokaryotes and eukaryotes, is the essential function of uncoupling transcription and translation. This is achieved by the phage nucleus and necessitates an intricate system to determine which proteins are to be imported into the nucleus, which must be excluded, and how to export mRNA for translation. It may be revealed with further investigation that nucleus-forming phage process mRNA with mechanisms resembling those of eukaryotes. Alternative splicing and exon skipping made possible by introns and uncoupled transcription and translation drastically increases the coding potential of a single gene while noncoding DNA has been positively correlated with biological complexity [109-111]. This suggests that the ancient intron invasion that may have selected the viral nucleus over the archaeal chromosome, could also hold the key to the explosion of organismal complexity that eventually brought humans into existence.

Bacteriophage are much more than mere “bacteria-eaters”. This dissertation demonstrates that nucleus-forming jumbo phage are enlightening us to the vast potential of intracellular complexity encoded by phage, the diverse mechanisms influencing the speciation of viruses as they compete inside a single host cell, and the answer to our calls for a weapon against antibiotic resistance and the big question of how we came to be.

5.1 References

1. Rohwer, F. and Barott, K., *Viral information.* Biology & Philosophy, 2013. **28**(2): p. 283-297.

2. Kristensen, D.M., Mushegian, A.R., Dolja, V.V., and Koonin, E.V., *New dimensions of the virus world discovered through metagenomics.* Trends in Microbiology, 2010. **18**(1): p. 11-19.

3. Kraemer, J.A., Erb, M.L., Waddling, C.A., Montabana, E.A., Zehr, E.A., Wang, H., Nguyen, K., Pham, D.S., Agard, D.A., and Pogliano, J., *A phage tubulin assembles dynamic filaments by an atypical mechanism to center viral DNA within the host cell.* Cell, 2012. **149**(7): p. 1488-99.

4. Erb, M.L., Kraemer, J.A., Coker, J.K., Chaikeeratisak, V., Nonejuie, P., Agard, D.A., and Pogliano, J., *A bacteriophage tubulin harnesses dynamic instability to center DNA in infected cells.* Elife, 2014. **3**.

5. Zehr, E.A., Kraemer, J.A., Erb, M.L., Coker, J.K., Montabana, E.A., Pogliano, J., and Agard, D.A., *The structure and assembly mechanism of a novel three-stranded tubulin filament that centers phage DNA.* Structure, 2014. **22**(4): p. 539-48.

6. Chaikeeratisak, V., Nguyen, K., Khanna, K., Brilot, A.F., Erb, M.L., Coker, J.K., Vavilina, A., Newton, G.L., Buschauer, R., Pogliano, K., Villa, E., Agard, D.A., and Pogliano, J., *Assembly of a nucleus-like structure during viral replication in bacteria.* Science, 2017. **355**(6321): p. 194-197.

7. Chaikeeratisak, V., Nguyen, K., Egan, M.E., Erb, M.L., Vavilina, A., and Pogliano, J., *The Phage Nucleus and Tubulin Spindle Are Conserved among Large Pseudomonas Phages.* Cell Reports, 2017. **20**(7): p. 1563-1571.

8. Chaikeeratisak, V., Khanna, K., Nguyen, K.T., Sugie, J., Egan, M.E., Erb, M.L., Vavilina, A., Nonejuie, P., Nieweglowska, E., Pogliano, K., Agard, D.A., Villa, E., and Pogliano, J., *Viral Capsid Trafficking along Treadmilling Tubulin Filaments in Bacteria.* Cell, 2019. **177**(7): p. 1771-1780.e12.

9. Korf, I.H.E., Meier-Kolthoff, J.P., Adriaenssens, E.M., Kropinski, A.M., Nimtz, M., Rohde, M., Van Raaij, M.J., and Wittmann, J., *Still Something to Discover: Novel Insights into Escherichia coli Phage Diversity and Taxonomy.* Viruses, 2019. **11**(5): p. 454.

10. Chaikeeratisak, V., Khanna, K., Nguyen, K. T., Egan, M. E., Enustun, E., Armbruster, E., Pogliano, K., Villa, E., & Pogliano, J. , *Subcellular Organization of Viral Particles During Maturation of Nucleus-Forming Jumbo Phage*. 2021: BioRxiv.

11. Schroeder, T.E. and Battaglia, D.E., *"Spiral asters" and cytoplasmic rotation in sea urchin eggs: induction in Strongylocentrotus purpuratus eggs by elevated temperature.* The Journal of Cell Biology, 1985. **100**(4): p. 1056-1062.

12. Serbus, L.R., Cha, B.-J., Theurkauf, W.E., and Saxton, W.M., *Dynein and the actin cytoskeleton control kinesin-driven cytoplasmic streaming inDrosophilaoocytes.* Development, 2005. **132**(16): p. 3743-3752.

13. Woodhouse, F.G. and Goldstein, R.E., *Cytoplasmic streaming in plant cells emerges naturally by microfilament self-organization.* Proceedings of the National Academy of Sciences, 2013. **110**(35): p. 14132-14137.

14. Stein, D.B., De Canio, G., Lauga, E., Shelley, M.J., and Goldstein, R.E., *Swirling Instability of the Microtubule Cytoskeleton.* Physical Review Letters, 2021. **126**(2).

15. Suzuki, K., Miyazaki, M., Takagi, J., Itabashi, T., and Ishiwata, S.I., *Spatial confinement of active microtubule networks induces large-scale rotational cytoplasmic flow.* Proceedings of the National Academy of Sciences, 2017. **114**(11): p. 2922-2927.

16. Gerashchenko, M.V., Chernoivanenko, I.S., Moldaver, M.V., and Minin, A.A., *Dynein is a motor for nuclear rotation while vimentin IFs is a "brake".* Cell Biology International, 2009. **33**(10): p. 1057-1064.

17. Levy, J.R. and Holzbaur, E.L.F., *Dynein drives nuclear rotation during forward progression of motile fibroblasts.* Journal of Cell Science, 2008. **121**(19): p. 3187-3195.

18. Kim, D.-H., Cho, S., and Wirtz, D., *Tight coupling between nucleus and cell migration through the perinuclear actin cap.* Journal of Cell Science, 2014. **127**(11): p. 2528-2541.

19. Kumar, A., Maitra, A., Sumit, M., Ramaswamy, S., and Shivashankar, G.V., *Actomyosin contractility rotates the cell nucleus.* Scientific Reports, 2015. **4**(1).

20. Maninová, M., Iwanicki, M.P., and Vomastek, T., *Emerging role for nuclear rotation and orientation in cell migration.* Cell Adhesion & Migration, 2014. **8**(1): p. 42-48.

21. Wu, J., Lee, K.C., Dickinson, R.B., and Lele, T.P., *How dynein and microtubules rotate the nucleus.* Journal of Cellular Physiology, 2011. **226**(10): p. 2666-2674.

22. Fruleux, A. and Hawkins, R.J., *Physical role for the nucleus in cell migration.* J Phys Condens Matter, 2016. **28**(36): p. 363002.

23. Calisher, C.H., Briese, T., Brister, J.R., Charrel, R.N., Durrwald, R., Ebihara, H., Fulhorst, C.F., Gao, G.F., Groschup, M.H., Haddow, A.D., Hyndman, T.H., Junglen, S., Klempa, B., Klingstrom, J., Kropinski, A.M., Krupovic, M., LaBeaud, A.D., Maes, P., Nowotny, N., Nunes, M.R.T., Payne, S.L., Radoshitzky, S.R., Rubbenstroth, D., Sabanadzovic, S., Sasaya, T., Stenglein, M.D., Varsani, A., Wahl, V., Weaver, S.C., Zerbini, F.M., Vasilakis, N., and Kuhn, J.H., *Strengthening the Interaction of the Virology Community with the International Committee on Taxonomy of Viruses (ICTV) by Linking Virus Names and Their Abbreviations to Virus Species.* Syst Biol, 2019. **68**(5): p. 828-839.

24. Bobay, L.M. and Ochman, H., *Biological species in the viral world.* Proc Natl Acad Sci U S A, 2018. **115**(23): p. 6040-6045.

25. Dobzhansky, T., *Genetics and the Origin of Species*. 1937: Columbia University Press.

26. Mayr, E., *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. 1942: New York: Columbia Univ. Press.

27. Refardt, D., *Within-host competition determines reproductive success of temperate bacteriophages.* ISME J, 2011. **5**(9): p. 1451-60.

28. Worobey, M. and Holmes, E.C., *Evolutionary aspects of recombination in RNA viruses.* J Gen Virol, 1999. **80 (Pt 10)**: p. 2535-2543.

29. Diaz-Munoz, S.L. and Koskella, B., *Bacteria-phage interactions in natural environments.* Adv Appl Microbiol, 2014. **89**: p. 135-83.

30. Diaz-Munoz, S.L., Tenaillon, O., Goldhill, D., Brao, K., Turner, P.E., and Chao, L., *Electrophoretic mobility confirms reassortment bias among geographic isolates of segmented RNA phages.* BMC Evol Biol, 2013. **13**: p. 206.

31. Silander, O.K., Weinreich, D.M., Wright, K.M., O'Keefe, K.J., Rang, C.U., Turner, P.E., and Chao, L., *Widespread genetic exchange among terrestrial bacteriophages.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 19009-14.

32. DaPalma, T., Doonan, B.P., Trager, N.M., and Kasman, L.M., *A systematic approach to virus-virus interactions.* Virus Res, 2010. **149**(1): p. 1-9.

33. Luque, A. and Silveira, C.B., *Quantification of Lysogeny Caused by Phage Coinfections in Microbial Communities from Biophysical Principles.* mSystems, 2020. **5**(5).

34. Diaz-Munoz, S.L., *Viral coinfection is shaped by host ecology and virus-virus interactions across diverse microbial taxa and environments.* Virus Evol, 2017. **3**(1): p. vex011.

35. O'Keefe, K.J., Silander, O.K., McCreery, H., Weinreich, D.M., Wright, K.M., Chao, L., Edwards, S.V., Remold, S.K., and Turner, P.E., *Geographic differences in sexual reassortment in RNA phage.* Evolution, 2010. **64**(10): p. 3010-23.

36. Duffy, S., Burch, C.L., and Turner, P.E., *Evolution of host specificity drives reproductive isolation among RNA viruses.* Evolution, 2007. **61**(11): p. 2614-22.

37. Meyer, J.R., Dobias, D.T., Medina, S.J., Servilio, L., Gupta, A., and Lenski, R.E., *Ecological speciation of bacteriophage lambda in allopatry and sympatry.* Science, 2016. **354**(6317): p. 1301-1304.

38. Saxenhofer, M., Schmidt, S., Ulrich, R.G., and Heckel, G., *Secondary contact between diverged host lineages entails ecological speciation in a European hantavirus.* PLoS Biol, 2019. **17**(2): p. e3000142.

39. Dulbecco, R., *Mutual exclusion between related phages.* J Bacteriol, 1952. **63**(2): p. 209-17.

40. Abedon, S.T., *Bacteriophage secondary infection.* Virol Sin, 2015. **30**(1): p. 3-10.

41. Turner, P.E., Burch, C.L., Hanley, K.A., and Chao, L., *Hybrid frequencies confirm limit to coinfection in the RNA bacteriophage phi6.* J Virol, 1999. **73**(3): p. 2420-4.

42. Li, C., Hatta, M., Nidom, C.A., Muramoto, Y., Watanabe, S., Neumann, G., and Kawaoka, Y., *Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence.* Proc Natl Acad Sci U S A, 2010. **107**(10): p. 4687-92.

43. Li, C., Hatta, M., Watanabe, S., Neumann, G., and Kawaoka, Y., *Compatibility among polymerase subunit proteins is a restricting factor in reassortment between equine H7N7 and human H3N2 influenza viruses.* J Virol, 2008. **82**(23): p. 11880-8.

44. Song, M.S., Pascua, P.N., Lee, J.H., Baek, Y.H., Park, K.J., Kwon, H.I., Park, S.J., Kim, C.J., Kim, H., Webby, R.J., Webster, R.G., and Choi, Y.K., *Virulence and genetic compatibility of polymerase reassortant viruses derived from the pandemic (H1N1) 2009 influenza virus and circulating influenza A viruses.* J Virol, 2011. **85**(13): p. 6275-86.

45. Kieser, Q., Noyce, R.S., Shenouda, M., Lin, Y.J., and Evans, D.H., *Cytoplasmic factories, virus assembly, and DNA replication kinetics collectively constrain the formation of poxvirus recombinants.* PLoS One, 2020. **15**(1): p. e0228028.

46. Paszkowski, P., Noyce, R.S., and Evans, D.H., *Live-Cell Imaging of Vaccinia Virus Recombination.* PLoS Pathog, 2016. **12**(8): p. e1005824.

47. Lin, Y.C. and Evans, D.H., *Vaccinia virus particles mix inefficiently, and in a way that would restrict viral recombination, in coinfected cells.* J Virol, 2010. **84**(5): p. 2432-43.

48. Cairns, J., *The initiation of vaccinia infection.* Virology, 1960. **11**: p. 603-23.

49. Tomer, E., Cohen, E.M., Drayman, N., Afriat, A., Weitzman, M.D., Zaritsky, A., and Kobiler, O., *Coalescing replication compartments provide the opportunity for recombination between coinfecting herpesviruses.* FASEB J, 2019. **33**(8): p. 9388-9403.

50. Chaikeeratisak, V., Birkholz, E.A., Prichard, A.M., Egan, M.E., Mylvara, A., Nonejuie, P., Nguyen, K.T., Sugie, J., Meyer, J.R., and Pogliano, J., *Viral speciation through subcellular genetic isolation and virogenesis incompatibility.* Nat Commun, 2021. **12**(1): p. 342.

51. Landthaler, M., Shen, B.W., Stoddard, B.L., and Shub, D.A., *I-BasI and I-HmuI: two phage intron-encoded endonucleases with homologous DNA recognition sequences but distinct DNA specificities.* J Mol Biol, 2006. **358**(4): p. 1137-51.

52. Landthaler, M., Lau, N.C., and Shub, D.A., *Group I intron homing in Bacillus phages SPO1 and SP82: a gene conversion event initiated by a nicking homing endonuclease.* J Bacteriol, 2004. **186**(13): p. 4307-14.

53. Ceyssens, P.J., Minakhin, L., Van den Bossche, A., Yakunina, M., Klimuk, E., Blasdel, B., De Smet, J., Noben, J.P., Blasi, U., Severinov, K., and Lavigne, R., *Development of giant*

bacteriophage varphiKZ is independent of the host transcription apparatus. J Virol, 2014. **88**(18): p. 10501-10.

54. Lecoutere, E., Ceyssens, P.J., Miroshnikov, K.A., Mesyanzhinov, V.V., Krylov, V.N., Noben, J.P., Robben, J., Hertveldt, K., Volckaert, G., and Lavigne, R., *Identification and comparative analysis of the structural proteomes of phiKZ and EL, two giant Pseudomonas aeruginosa bacteriophages.* Proteomics, 2009. **9**(11): p. 3215-9.

55. Holmes, S.F., Santangelo, T.J., Cunningham, C.K., Roberts, J.W., and Erie, D.A., *Kinetic investigation of Escherichia coli RNA polymerase mutants that influence nucleotide discrimination and transcription fidelity.* J Biol Chem, 2006. **281**(27): p. 18677-83.

56. Carvunis, A.R., Rolland, T., Wapinski, I., Calderwood, M.A., Yildirim, M.A., Simonis, N., Charloteaux, B., Hidalgo, C.A., Barbette, J., Santhanam, B., Brar, G.A., Weissman, J.S., Regev, A., Thierry-Mieg, N., Cusick, M.E., and Vidal, M., *Proto-genes and de novo gene birth.* Nature, 2012. **487**(7407): p. 370-4.

57. Repar, J. and Warnecke, T., *Mobile Introns Shape the Genetic Diversity of Their Host Genes.* Genetics, 2017. **205**(4): p. 1641-1648.

58. Aagaard, C., Dalgaard, J.Z., and Garrett, R.A., *Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron- cells of Sulfolobus acidocaldarius.* Proc Natl Acad Sci U S A, 1995. **92**(26): p. 12285-9.

59. Naor, A., Altman-Price, N., Soucy, S.M., Green, A.G., Mitiagin, Y., Turgeman-Grott, I., Davidovich, N., Gogarten, J.P., and Gophna, U., *Impact of a homing intein on recombination frequency and organismal fitness.* Proc Natl Acad Sci U S A, 2016. **113**(32): p. E4654-61.

60. Goodrich-Blair, H. and Shub, D.A., *Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers.* Cell, 1996. **84**(2): p. 211-21.

61. Edgell, D.R., Fast, N.M., and Doolittle, W.F., *Selfish DNA: the best defense is a good offense.* Curr Biol, 1996. **6**(4): p. 385-8.

62. Russell, R.L. and Huskey, R.J., *Partial exclusion between T-even bacteriophages: an incipient genetic isolation mechanism.* Genetics, 1974. **78**(4): p. 989-1014.

63. Koonin, E.V., *Viruses and mobile elements as drivers of evolutionary transitions.* Philos Trans R Soc Lond B Biol Sci, 2016. **371**(1701).

64. Lambowitz, A.M. and Belfort, M., *INTRONS AS MOBILE GENETIC ELEMENTS.* Annual Review of Biochemistry, 1993. **62**(1): p. 587-622.

65. Grivell, L.A., *Transposition: Mobile introns get into line.* Current Biology, 1996. **6**(1): p. 48-51.

66. Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., and Alm, E.J., *Meta-analysis of gut microbiome studies identifies disease-specific and shared responses.* Nat Commun, 2017. **8**(1): p. 1784.

67. Wirbel, J., Pyl, P.T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J.S., Voigt, A.Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L.P., Schrotz-King, P., Vogtmann, E., Habermann, N., Nimeus, E., Thomas, A.M., Manghi, P., Gandini, S., Serrano, D., Mizutani, S., Shiroma, H., Shiba, S., Shibata, T., Yachida, S., Yamada, T., Waldron, L., Naccarati, A., Segata, N., Sinha, R., Ulrich, C.M., Brenner, H., Arumugam, M., Bork, P., and Zeller, G., *Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer.* Nat Med, 2019. **25**(4): p. 679-689.

68. Holmes, M., Flaminio, Z., Vardhan, M., Xu, F., Li, X., Devinsky, O., and Saxena, D., *Cross talk between drug-resistant epilepsy and the gut microbiome.* Epilepsia, 2020. **61**(12): p. 2619-2628.

69. Nguyen, T.T., Hathaway, H., Kosciolek, T., Knight, R., and Jeste, D.V., *Gut microbiome in serious mental illnesses: A systematic review and critical evaluation.* Schizophr Res, 2021. **234**: p. 24-40.

70. Dethlefsen, L. and Relman, D.A., *Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4554-61.

71. Francino, M.P., *Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances.* Front Microbiol, 2015. **6**: p. 1543.

72. Blaser, M.J., *Antibiotic use and its consequences for the normal microbiome.* Science, 2016. **352**(6285): p. 544-5.

73. Maier, L., Goemans, C.V., Wirbel, J., Kuhn, M., Eberl, C., Pruteanu, M., Muller, P., Garcia-Santamarina, S., Cacace, E., Zhang, B., Gekeler, C., Banerjee, T., Anderson, E.E., Milanese, A., Lober, U., Forslund, S.K., Patil, K.R., Zimmermann, M., Stecher, B., Zeller, G., Bork, P., and Typas, A., *Unravelling the collateral damage of antibiotics on gut bacteria.* Nature, 2021. **599**(7883): p. 120-124.

74. Davies, J. and Davies, D., *Origins and evolution of antibiotic resistance.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 417-33.

75. US Department of Health and Human Services, C., *Antibiotic Resistance Threats in the United States*. 2019, Centers for Disease Control and Prevention. p. 1-113.

76. Duzgunes, N., Sessevmez, M., and Yildirim, M., *Bacteriophage Therapy of Bacterial Infections: The Rediscovered Frontier.* Pharmaceuticals (Basel), 2021. **14**(1).

77. Dissanayake, U., Ukhanova, M., Moye, Z.D., Sulakvelidze, A., and Mai, V., *Bacteriophages Reduce Pathogenic Escherichia coli Counts in Mice Without Distorting Gut Microbiota.* Front Microbiol, 2019. **10**: p. 1984.

78. Gurney, J., Pradier, L., Griffin, J.S., Gougat-Barbera, C., Chan, B.K., Turner, P.E., Kaltz, O., and Hochberg, M.E., *Phage steering of antibiotic-resistance evolution in the bacterial pathogen, Pseudomonas aeruginosa.* Evol Med Public Health, 2020. **2020**(1): p. 148-157.

79. Divya Ganeshan, S. and Hosseinidoust, Z., *Phage Therapy with a Focus on the Human Microbiota.* Antibiotics (Basel), 2019. **8**(3).

80. Kutateladze, M. and Adamia, R., *Phage therapy experience at the Eliava Institute.* Med Mal Infect, 2008. **38**(8): p. 426-30.

81. Zaczek, M., Weber-Dabrowska, B., Miedzybrodzki, R., Lusiak-Szelachowska, M., and Gorski, A., *Phage Therapy in Poland - a Centennial Journey to the First Ethically Approved Treatment Facility in Europe.* Front Microbiol, 2020. **11**: p. 1056.

82. D'Herelle, F., *On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917.* Res Microbiol, 2007. **158**(7): p. 553-4.

83. Twort, F.W., *An Investigation on the Nature of Ultra-Microscopic Viruses.* The Lancet, 1915. **186**(4814): p. 1241-1243.

84. Villarroel, J., Larsen, M., Kilstrup, M., and Nielsen, M., *Metagenomic Analysis of Therapeutic PYO Phage Cocktails from 1997 to 2014.* Viruses, 2017. **9**(11): p. 328.

85. Monson, R., Foulds, I., Foweraker, J., Welch, M., and Salmond, G.P.C., *The Pseudomonas aeruginosa generalized transducing phage phiPA3 is a new member of the phiKZ-like group of 'jumbo' phages, and infects model laboratory strains and clinical isolates from cystic fibrosis patients.* Microbiology (Reading), 2011. **157**(Pt 3): p. 859-867.

86. Chan, B.K., Sistrom, M., Wertz, J.E., Kortright, K.E., Narayan, D., and Turner, P.E., *Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa.* Scientific Reports, 2016. **6**(1): p. 26717.

87. Chan, B.K., Turner, P.E., Kim, S., Mojibian, H.R., Elefteriades, J.A., and Narayan, D., *Phage treatment of an aortic graft infected with Pseudomonas aeruginosa.* Evol Med Public Health, 2018. **2018**(1): p. 60-66.

88. Lewis, R., Clooney, A.G., Stockdale, S.R., Buttimer, C., Draper, L.A., Ross, R.P., and Hill, C., *Isolation of a Novel Jumbo Bacteriophage Effective Against Klebsiella aerogenes.* Front Med (Lausanne), 2020. **7**: p. 67.

89. Bonilla, E., Costa, A.R., van den Berg, D.F., van Rossum, T., Hagedoorn, S., Walinga, H., Xiao, M., Song, W., Haas, P.J., Nobrega, F.L., and Brouns, S.J.J., *Genomic characterization of four novel bacteriophages infecting the clinical pathogen Klebsiella pneumoniae.* DNA Res, 2021. **28**(4).

90. Devoto, A.E., Santini, J.M., Olm, M.R., Anantharaman, K., Munk, P., Tung, J., Archie, E.A., Turnbaugh, P.J., Seed, K.D., Blekhman, R., Aarestrup, F.M., Thomas, B.C., and

Banfield, J.F., *Megaphages infect Prevotella and variants are widespread in gut microbiomes.* Nat Microbiol, 2019. **4**(4): p. 693-700.

91. Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gumy, L.F., and Fineran, P.C., *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, 2020. **5**(1): p. 48-55.

92. Mendoza, S.D., Nieweglowska, E.S., Govindarajan, S., Leon, L.M., Berry, J.D., Tiwari, A., Chaikeeratisak, V., Pogliano, J., Agard, D.A., and Bondy-Denomy, J., *A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases.* Nature, 2020. **577**(7789): p. 244-248.

93. Nguyen, K.T., Sugie, J., Khanna, K., Egan, M.E., Birkholz, E.A., Lee, J., Beierschmitt, C., Villa, E., and Pogliano, J., *Selective transport of fluorescent proteins into the phage nucleus.* PLoS One, 2021. **16**(6): p. e0251429.

94. Ujmajuridze, A., Chanishvili, N., Goderdzishvili, M., Leitner, L., Mehnert, U., Chkhotua, A., Kessler, T.M., and Sybesma, W., *Adapted Bacteriophages for Treating Urinary Tract Infections.* Front Microbiol, 2018. **9**: p. 1832.

95. O'Malley, M.A., Leger, M.M., Wideman, J.G., and Ruiz-Trillo, I., *Concepts of the last eukaryotic common ancestor.* Nature Ecology & Evolution, 2019. **3**(3): p. 338-344.

96. Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., Matsui, Y., Miyazaki, M., Murata, K., Saito, Y., Sakai, S., Song, C., Tasumi, E., Yamanaka, Y., Yamaguchi, T., Kamagata, Y., Tamaki, H., and Takai, K., *Isolation of an archaeon at the prokaryote–eukaryote interface.* Nature, 2020. **577**(7791): p. 519-525.

97. Nobs, S.-J., Macleod, F.I., Wong, H.L., and Burns, B.P., *Eukarya the chimera: eukaryotes, a secondary innovation of the two domains of life?* Trends in Microbiology, 2021.

98. Koonin, E.V., *The origin and early evolution of eukaryotes in the light of phylogenomics.* Genome Biol, 2010. **11**(5): p. 209.

99. Bell, P.J., *Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus?* J Mol Evol, 2001. **53**(3): p. 251-6.

100. Bell, P.J.L., *Evidence supporting a viral origin of the eukaryotic nucleus.* Virus Res, 2020. **289**: p. 198168.

101. Gray, M.W., Burger, G., and Lang, B.F., *Mitochondrial evolution.* Science, 1999. **283**(5407): p. 1476-81.

102. Koonin, E.V., *The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?* Biol Direct, 2006. **1**: p. 22.

103. Martin, W. and Koonin, E.V., *Introns and the origin of nucleus-cytosol compartmentalization.* Nature, 2006. **440**(7080): p. 41-5.

104. Vosseberg, J. and Snel, B., *Domestication of self-splicing introns during eukaryogenesis: the rise of the complex spliceosomal machinery.* Biol Direct, 2017. **12**(1): p. 30.

105. Irimia, M. and Roy, S.W., *Origin of spliceosomal introns and alternative splicing.* Cold Spring Harb Perspect Biol, 2014. **6**(6).

106. Lambowitz, A.M. and Belfort, M., *Mobile Bacterial Group II Introns at the Crux of Eukaryotic Evolution.* Microbiol Spectr, 2015. **3**(1): p. MDNA3-0050-2014.

107. Forterre, P. and Raoult, D., *The transformation of a bacterium into a nucleated virocell reminds the viral eukaryogenesis hypothesis.* Virologie (Montrouge), 2017. **21**(4): p. 28-30.

108. Gruenbaum, Y., Goldman, R.D., Meyuhas, R., Mills, E., Margalit, A., Fridkin, A., Dayani, Y., Prokocimer, M., and Enosh, A., *The nuclear lamina and its functions in the nucleus.* Int Rev Cytol, 2003. **226**: p. 1-62.

109. Taft, R.J. and Mattick, J.S., Genome Biology, 2003. **5**(1): p. P1.

110. Wolf, Y.I., Katsnelson, M.I., and Koonin, E.V., *Physical foundations of biological complexity.* Proc Natl Acad Sci U S A, 2018. **115**(37): p. E8678-E8687.

111. Grabski, D.F., Broseus, L., Kumari, B., Rekosh, D., Hammarskjold, M.L., and Ritchie, W., *Intron retention and its impact on gene expression and protein diversity: A review and a practical guide.* WIREs RNA, 2021. **12**(1).