Article

Multi-interface licensing of protein import into a phage nucleus

https://doi.org/10.1038/s41586-024-08547-x

Received: 19 March 2024

Accepted: 17 December 2024

Published online: 5 February 2025

Check for updates

Claire Kokontis[1], Timothy A. Klein[1], Sukrit Silas[1] & Joseph Bondy-Denomy[1,2]✉

Bacteriophages use diverse mechanisms to evade antiphage defence systems. ΦKZ-like jumbo phages assemble a proteinaceous, nucleus-like compartment that excludes antagonistic host nucleases and also internalizes DNA replication and transcription machinery[1–4]. The phage factors required for protein import and the mechanisms of selectivity remain unknown, however. Here we uncover an import system comprising proteins highly conserved across nucleus-forming phages, together with additional cargo-specific contributors. Using a genetic selection that forces the phage to decrease or abolish the import of specific proteins, we determine that the importation of five different phage nuclear-localized proteins requires distinct interfaces of the same factor, Imp1 (gp69). Imp1 localizes early to the nascent phage nucleus and forms discrete puncta in the mature phage nuclear periphery, probably in complex with direct interactor Imp6 (gp67), a conserved protein encoded in the same locus. The import of certain proteins, including a host topoisomerase, additionally requires Imp3 (gp59), a conserved factor necessary for proper Imp1 function. Three additional non-conserved phage proteins (Imp2 and Imp4/Imp5) are required for the import of two queried nuclear cargos (nuclear-localized protein 1 and host topoisomerase, respectively), perhaps acting as specific adaptors. We therefore propose a core import system that includes Imp1, Imp3 and Imp6, with multiple interfaces of Imp1 licensing transport through a protein lattice.

Segregation of cytoplasmic and organelle activity, with regulated movement of biomolecules between them, is a fundamental aspect of eukaryotic life. DNA-containing organelles, long thought to be exclusive to eukaryotes, have recently been discovered during bacteriophage infection in multiple genera of Gram-negative bacteria, including *Pseudomonas*, *Escherichia*, *Serratia*, *Erwinia, Vibrio* and *Salmonella*[1,2,4–8]. A phage-produced, proteinaceous, nucleus-like structure compartmentalizes DNA replication and transcription, with translation occurring in the cytoplasm[1]. This compartment primarily comprises a single protein—chimallin (ChmA) or PhuN—and excludes host nucleases, such as CRISPR–Cas and restriction enzymes[3], but imports various phage proteins and at least one host protein[1,9]. How this selectivity is achieved, and the mechanism of movement through the protein-based lattice that forms the phage nucleus[5,10,11], remain unaddressed. Understanding how proteins move from the cytoplasm into the phage nucleus, analogous to that achieved by secretion systems or eukaryotic nuclear import, may demonstrate new fundamental biological mechanisms.

Transport of cargo proteins through the eukaryotic lipid nuclear membrane is mediated by binding between linear sequence motifs and adaptor proteins, called importins, that shuttle them through a massive nuclear pore of roughly 100 MDa (ref. 12). In ΦKZ-like jumbo phages, there are no known import adaptors/importins that bind to cargo and no known nuclear shell constituents that function as specificity determinants for phage protein entry. Given the abundance of uncharacterized genes in large ΦKZ-like genomes (above 200 kb), and the general insolubility of the protein-based nucleus limiting basic interaction approaches[13], we use an unbiased genetic selection to identify the requirements for protein import. The screen identified five proteins with a role in protein import: two proteins (now named Imp1 and Imp3) that are broadly conserved in nucleus-forming phages, and three (Imp2, Imp4 and Imp5) that appear to be conserved only among related *Pseudomonas*-infecting, nucleus-forming phages. Imp1–5 have no obvious predicted molecular function. Imp6, an RNA-binding protein also called ChmC[14], is additionally identified here as a direct physical interactor of Imp1. Taken together, we propose that these proteins constitute a unique phage protein trafficking system that enables transit into the phage nucleus.

Imp1 is required for protein import

The ΦKZ phage nucleus excludes the restriction enzyme EcoRI (278 amino acids, 31.5 kDa), but fusion of EcoRI to nuclear-localized protein (Nlp) gp152 (UvsX/RecA homologue, hereafter referred to as Nlp1) facilitates the entry of EcoRI into the nucleus and cleavage of phage DNA[3]. This suggests that import is licensed by recognition of nuclear cargo, rather than actively excluding antagonistic host proteins. To identify phage factors that are required for movement of proteins into the nucleus following synthesis in the cytoplasm, we used a genetic selection experiment. We individually fused nine nuclear proteins to

[1]Department of Microbiology and Immunology, University of California San Francisco, San Francisco, CA, USA. [2]Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA, USA. ✉e-mail: Joseph.Bondy-Denomy@ucsf.edu

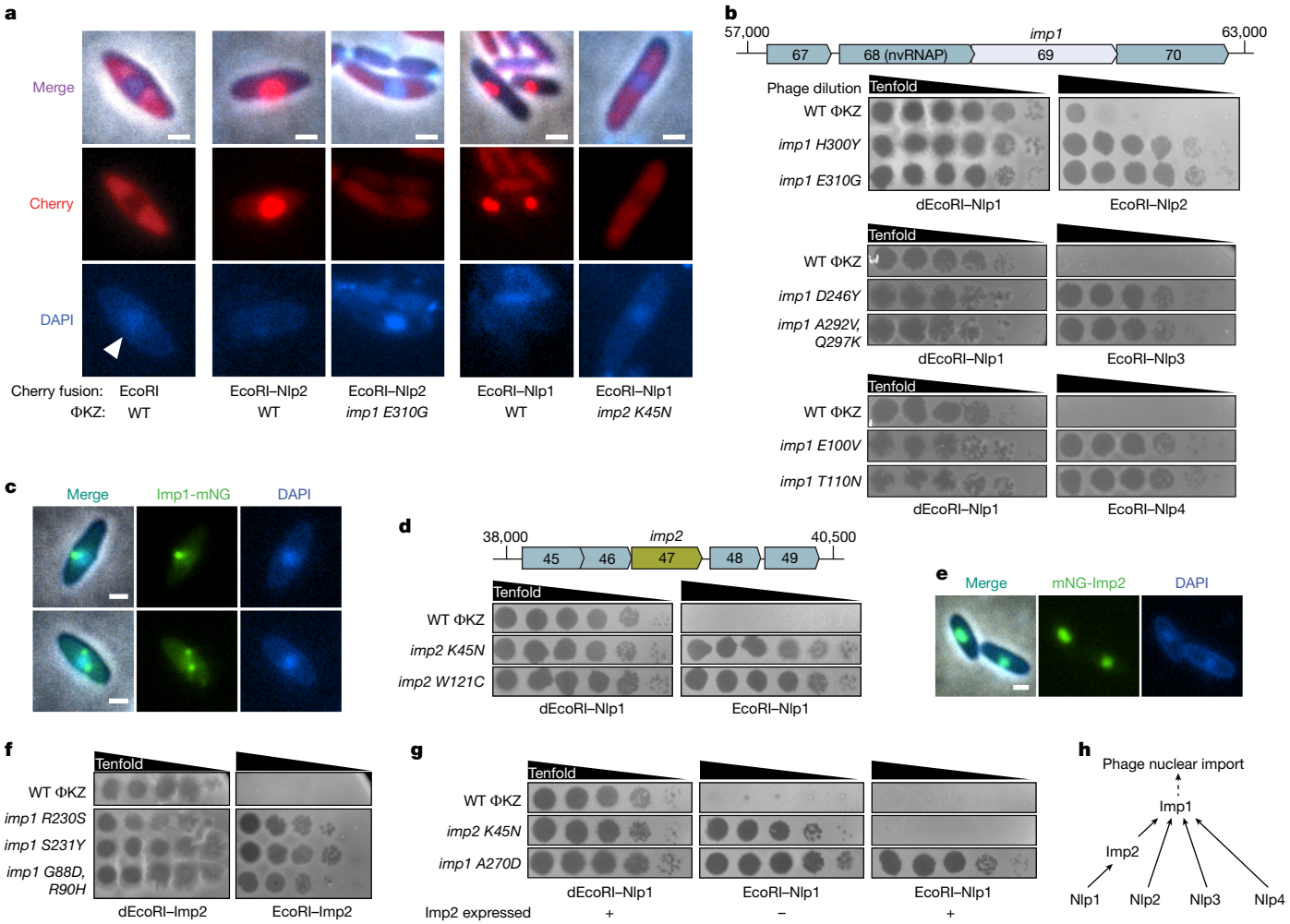


Fig. 1 | Mutations in previously uncharacterized phage genes *imp1* and *imp2* reduce protein import into the phage nucleus. a, Live-cell fluorescence microscopy of *Pseudomonas aeruginosa* strain PAO1 expressing the indicated sfCherry2 fusion proteins infected with WT ΦKZ or the indicated phage mutant (EcoRI + WT ΦKZ, $n = 118$ cells; EcoRI–Nlp2 + WT ΦKZ, $n = 102$ cells; EcoRI–Nlp2 + *imp1 E310G* ΦKZ, $n = 230$ cells; EcoRI–Nlp1 + WT ΦKZ, $n = 69$ cells; EcoRI–Nlp1 + *imp2 K45N*, $n = 58$ cells). DAPI staining indicates phage DNA within the phage nucleus (white arrow). **b**,**d**,**f**, Plaque assays with the indicated WT or mutant phage spotted in tenfold serial dilutions on a lawn of strain PAO1 expressing indicated Nlp2/3/4 (**b**), Nlp1 (**d**) or Imp2 (**f**) fusions to sfCherry2. **c**,**e**, Live-cell fluorescence microscopy of Imp1 fused to mNeonGreen (mNG) infected with WT ΦKZ ($n = 154$), imaged at 50–60 min post infection (**c**), and of Imp2 fused to mNeonGreen and infected with WT ΦKZ ($n = 129$) (**e**). **g**, Plaque assays with the indicated WT or mutant phage spotted in tenfold serial dilutions on a lawn of PAO1 expressing the indicated sfCherry2 fusion, with (+) or without (−) expression of Imp2 in trans. **h**, Schematic representing the factors required for nuclear import of proteins queried in EcoRI selections, as determined by escape mutations isolated in *imp1* and *imp2* from each selection (Nlp2 → Imp1 indicates that Nlp2 requires WT Imp1 for its nuclear import). All plaque assays were repeated independently three times in biological replicates, with similar results. For plate source images, see Supplementary Fig. 2. All microscopy experiments are representative of either two (**a**,**e**) or three (**a**,**c**) biologically independent experiments, with similar results. Scale bars, 1 µm. dEcoRI, catalytically inactivated EcoRI.

the C terminus of EcoRI–sfCherry2 constructs (hereafter referred to as EcoRI–NlpX fusions) to import the EcoRI enzyme (Extended Data Fig. 1a), which has 92 recognition sites in the phage genome. Import of this enzyme is expected to impart selective pressure on the phage to reduce or abolish import of these cargo, by mutations in putative import factors. EcoRI fusion to Nlp1 (gp152, RecA-like), Nlp2 (gp155, RNase H-like), Nlp3 (gp104) and Nlp4 (gp171) reduced phage titre by 10^6–10^7-fold, whereas dead EcoRI (E111G) fused to Nlp1 localized within the phage nucleus but provided no reduction in titre[3] (Extended Data Fig. 1a,b). EcoRI fusion to other phage nuclear proteins (gp50, gp70, gp118, gp123 or gp179)[1,9] restricted EcoRI-sensitive control phages but showed no reduction in ΦKZ phage titre, perhaps because this fusion affects folding or nuclear localization (Extended Data Fig. 1b). As a result, these EcoRI fusions were not further explored.

When an unmutagenized phage population with an approximate titre of 10^{11} plaque-forming units (pfu) per millilitre was subjected to targeting by EcoRI fusions to Nlp1, Nlp2, Nlp3 or Nlp4, spontaneous mutant escape phages emerged at a frequency of roughly 10^{-6}–10^{-7} (Extended Data Fig. 1a). EcoRI–Nlp2, EcoRI–Nlp3 and EcoRI–Nlp4 fusions all selected for distinct mutations in a single uncharacterized phage gene, *orf69*—hereafter referred to as *imp1*—for import gene 1 (Fig. 1b, Table 1 and Supplementary Table 1). A total of eight distinct missense mutant alleles in *imp1* were identified in this way, with no overlap between the different Nlp proteins (Table 1 and Supplementary Table 1). Complementation with wild-type (WT) *imp1* expressed in trans again allowed EcoRI–Nlp2/3/4 fusions to restrict *imp1* mutant escape phages (Extended Data Fig. 1c), demonstrating that mutations in *imp1* are causal for escape from each EcoRI–Nlp fusion. To confirm that mutations in *imp1* had decreased protein import, we imaged EcoRI–Nlp2 during infection with WT phage or an *imp1 E310G* mutant. Indeed, EcoRI–Nlp2 import was decreased by this mutation (Fig. 1a and Extended Data Fig. 2a–c).

Table 1 | Summary of mutant alleles selected for by EcoRI fusions

Selection	Mutated gene	Accession no.	No. of unique alleles
EcoRI–Nlp1	*imp2* (*orf47*)	NP_803613.1	8
Tn7::Imp2 (EcoRI–Nlp1)	*imp1* (*orf69*)	NP_803635.1	1
EcoRI–Nlp2	*imp1*		3
EcoRI–Nlp3	*imp1*		2
EcoRI–Nlp4	*imp1*		3
EcoRI–Imp2	*imp1*		14
	imp3 (*orf59*)	NP_803625.1	1
EcoRI–Nlp2 + EcoRI–Imp2[a]	*imp1*		10
Tn7::Imp1 (EcoRI–Imp2)	*imp1*		2
	imp3		9
EcoRI–Imp1$_{\Phi PA3}$	*imp1*		2
	imp3		4
EcoRI–TopA	*imp1*		2
	imp3		2
	imp4 (*orf48*)	NP_803614.1	4
	imp5 (*orf287*)	NP_803853.1	1

[a]Indicates that selections were performed sequentially.
Summary of ΦKZ genes mutated under selection by the indicated EcoRI fusion, including gene accession numbers and how many unique alleles were isolated. Tn7::ImpX (EcoRI–Y) indicates that the indicated EcoRI fusion was used for selection, in the presence of ImpX expressed from the bacterial chromosomal attTn7 site.

We next determined the subcellular localization of Imp1 with an mNeonGreen (mNG) fusion expressed from the host chromosome during infection with WT ΦKZ. Following infection, Imp1 transited from being diffuse in the cell to forming one or two puncta at the periphery of the phage nucleus (Fig. 1c and Extended Data Fig. 3a) in 100% of cells examined (n = 154). Time-lapse microscopy showed that Imp1–mNG localizes early in infection to the pole, where the phage assembles the nascent proteinaceous nucleus and probably first initiates protein import (Extended Data Fig. 3b and Supplementary Video 1). Imp1–mNG then moves with the growing phage nucleus, ultimately stabilizing as distinct puncta. The Imp1-mNG fusion protein, however, did not complement *imp1* mutant phages (Extended Data Fig. 3c), perhaps because the fluorescent protein fusion blocks interactions with cargo or other binding partners. Interestingly, the phage ΦPA3 homologue of Imp1 (ΦPA3 gp63) was recently identified as an interactor with ChmA (the major nuclear shell protein) through proximity labelling, and by immunoprecipitation and mass spectrometry in ΦPA3-infected cells[9]. Microscopy experiments with the ΦPA3 Imp1 homologue, and with the phage 201Φ2-1 Imp1 homologue (201-Φ2-1 gp125), similarly showed puncta in the nuclear periphery[1,9]. Taken together, these data suggest a protein import role for Imp1 in association with the nuclear lattice.

Despite Imp1 being required for the import of Nlp2–Nlp4, phages that escaped EcoRI–Nlp1 restriction had eight distinct missense or nonsense mutations in a second uncharacterized gene, *orf47* (*imp2*) (Fig. 1d, Table 1 and Supplementary Table 1). Phages with *imp2* nonsense mutations produced very low titre from picked plaques, and following attempted amplification (roughly 10^5–10^6-fold lower than *imp2* missense mutants), suggesting that this gene is required for optimal fitness. EcoRI–Nlp1 import into the phage nucleus, as determined by fluorescence microscopy, was indeed disrupted during infection with an *imp2 K45N* mutant phage (Fig. 1a and Extended Data Fig. 2d,e). *imp2* expression in trans restored EcoRI–Nlp1 targeting of the phage, confirming the causality of this mutation (Extended Data Fig. 1d). *imp2* mutations were uniquely selected for by EcoRI–Nlp1 (but not by EcoRI–Nlp2/3/4), and *imp2* mutant phages still imported and succumbed to EcoRI fusions to Nlp2, Nlp3 or Nlp4, demonstrating that these mutations specifically perturb Nlp1 import (Extended Data Fig. 1e). Based on sequence searches, Imp2 appears to be conserved only among related *Pseudomonas*-infecting, nucleus-forming phages, and none of its homologues has any known function (Extended Data Fig. 1f). The importance of Imp2 in Nlp1 import for this phage family was confirmed by targeting the related phage ΦPA3 with EcoRI–Nlp1, which similarly showed phage escape mutations in *orf44* encoding an Imp2 homologue (Extended Data Fig. 1g).

To determine the role of Imp2 during phage infection, we next assessed its localization (mNG–Imp2) and observed that it does not form Imp1-like puncta but appears fully imported into the phage nucleus (Fig. 1e). To determine whether Imp2 import is dependent on Imp1 or a new factor, phages were targeted with EcoRI–Imp2, which decreased phage titre by about 10^6, whereas dead EcoRI–Imp2 had no effect (Extended Data Fig. 1a,c). Phages that escaped EcoRI–Imp2 targeting contained missense mutations in *imp1* (Fig. 1f, Table 1 and Supplementary Table 1), none of which overlapped with those previously isolated. To understand why mutations in *imp1* were not initially seen under EcoRI–Nlp1 selection, we repeated this selection but with Imp2 expressed in trans, preventing successful phage escape by mutations in this gene. Indeed, this approach selected directly for a mutation in *imp1* (*A270D*) (Fig. 1g). This new class of Nlp1 import-deficient mutant phage emerged at a lower frequency of under 10^{-8} (compared with around 10^{-7} previously; Extended Data Fig. 1a), explaining why it was not seen initially. Together, these data demonstrate that Nlp1 requires Imp2 for its import, whereas import of Nlp1–4 and Imp2 all require Imp1 (Fig. 1h).

Imp1 possesses cargo-specific interfaces

The selective pressure to reduce import of Nlp1 and Imp2 selected for 15 unique mutant alleles in *imp1*, none overlapping with the eight *imp1* mutant alleles previously isolated by EcoRI fusions to Nlp2, Nlp3 or Nlp4 (Table 1 and Supplementary Table 1). This mutational spectrum, with no nonsense mutations or frameshifts identified, suggests that *imp1* is essential. Given the convergence on Imp1, we next queried whether the unique mutations decreased import of all cargo or only some. Cells expressing each of the EcoRI–Nlp fusions were infected with WT ΦKZ or various *imp1* mutant phages. Interestingly, individual *imp1* mutations only decreased the phage-targeting activity of the EcoRI fusion construct they were initially selected under, demonstrating specific perturbations to import (Fig. 2a). This suggests that the *imp1* mutation acquired by each escape phage maintains sufficient expression, localization and import function for other cargo but specifically perturbs the import of only one queried protein. To visualize the location of functional residues, an AlphaFold2 (ref. 15)-predicted structural model of Imp1 was generated and each set of cargo-specific mutations labelled (Fig. 2b). The four imported cargo selected for mutations that generally clustered into five regions on the Imp1 model. Consistent with discrete functional interfaces on Imp1, double mutants could be isolated by iterative exposure to different EcoRI–Nlp fusions. For example, an Imp2 import-deficient mutant phage (*imp1 G88D/R90H*) acquired a new mutation in *imp1* (*R306L*) when selected on EcoRI–Nlp2 (Fig. 2c). The inverse was also true—a phage that resisted import of EcoRI–Nlp2 (*imp1 H300Y*) acquired a second mutation in *imp1* (*I226T*) when selected on EcoRI–Imp2. We therefore propose that Imp1 is equipped with distinct interfaces to achieve import specificity and licensing.

Sequence- and structure-based searches using an AlphaFold2-generated predicted structural model of Imp1 did not show homologues with any known molecular function. Imp1 is widely conserved in nucleus-forming phages infecting diverse genera, including *Pseudomonas*, *Serratia*, *Xanthomonas*, *Aeromonas*, *Vibrio* and *Erwinia*, and members of the Enterobacteriaceae family (Fig. 2d). Nearly every phage genome encoding an Imp1 protein also encodes a homologue

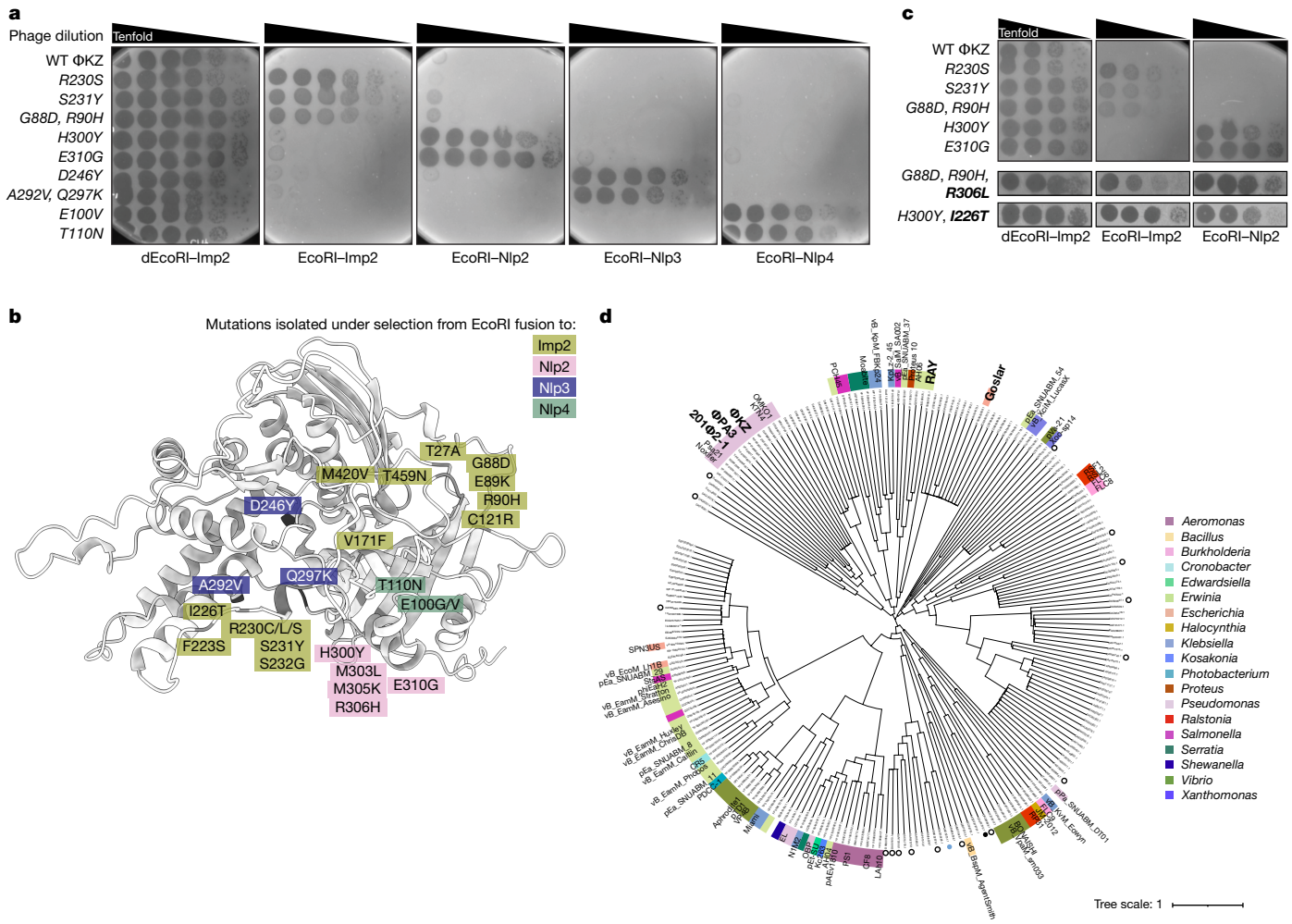


Fig. 2 | Imp1 possesses distinct functional interfaces to enable protein import specificity. a,**c**, Plaque assays with indicated WT or *imp1* mutant phages, isolated after a single selection (**a**) or after two selections, with the secondary mutation in bold (**c**), spotted in tenfold serial dilutions on a lawn of PAO1 expressing the indicated sfCherry2 fusion. Plaque assays were conducted as in Fig. 1b. **b**, AlphaFold2-predicted structural model of Imp1, with mutated residues from isolated *imp1* mutant phages colour coded by the EcoRI selection from which they were isolated. The top output model is shown. **d**, Phylogenetic tree of Imp1 homologues. Cultured phages are coloured by host species, and several model phages are also labelled by name. Uncoloured phages are from metagenomic sequencing. Filled black circle indicates that no phage nuclear shell homologue was detected in the corresponding phage genome, by either three iterations of PSI-BLAST or local blastp. Open black circles indicate that the associated genome record is incomplete, and no phage nuclear shell homologue was detected by the same methods. Filled blue circle indicates that no nuclear shell homologue was detected by the same methods, but the tubulin homologue responsible for centring the phage nucleus in the cell, PhuZ, was detected by blastp. All plaque assays were repeated three times independently in biological replicates, with similar results.

of ΦKZ gp54/ChmA/PhuN, the major phage nucleus protein. Only a few exceptions are noted and are probably due to incomplete genomic records. Together with genetic data demonstrating the importance of multiple Imp1 interfaces in the import of distinct cargo, its localization during infection and proximity to ChmA[9], and its broad conservation in nucleus-forming phages, we propose that Imp1 is a key determinant of protein transport into the phage nucleus for this phage family.

Imp3 is required for Imp1 function

To identify factors required for Imp1 function or localization, we fused EcoRI to Imp1 to assess its proximity to phage DNA and select for mutants that disrupt its localization. EcoRI–Imp1 abolished phage replication such that no escape mutants could be isolated (limit of detection, 8.3×10^{-10}; Extended Data Fig. 1a), whereas dead EcoRI–Imp1 had no impact (Extended Data Fig. 4a). This demonstrates that the Imp1 N terminus is localized within the phage nucleus and suggests that mutations that alter this localization may be lethal. Fusion of EcoRI to 11 different Imp1 mutant proteins, with representatives chosen from selection experiments with the various EcoRI–Nlp fusions, also maintained full phage targeting, with no escape mutants isolated (Extended Data Fig. 4a). This suggests that mutated Imp1 proteins express and localize correctly, as proposed above.

To identify additional *imp1* genetic interactions not seen in previous selections, we used two alternative selection approaches, both of which led to the same new, uncharacterized gene, *orf59* (*imp3*) (Fig. 3a–c). Using the EcoRI–Imp2 fusion that previously yielded *imp1* mutant phages, we repeated the selection but expressed WT Imp1 in trans to enable isolation of phages with mutations in other import factors (frequency approximately 10^{-8}–10^{-9}; Extended Data Fig. 1a). One class of mutant phages emerged that acquired spontaneous missense mutations in *imp1* (Extended Data Fig. 4b, Table 1 and Supplementary Table 1), which are genetically dominant. However, 9 out of 11 phage mutants that escaped this targeting construct acquired mutations in *imp3* (Fig. 3a). Mutations were either coding changes or ten base pairs upstream of the ATG start in the probable Shine–Dalgarno sequence

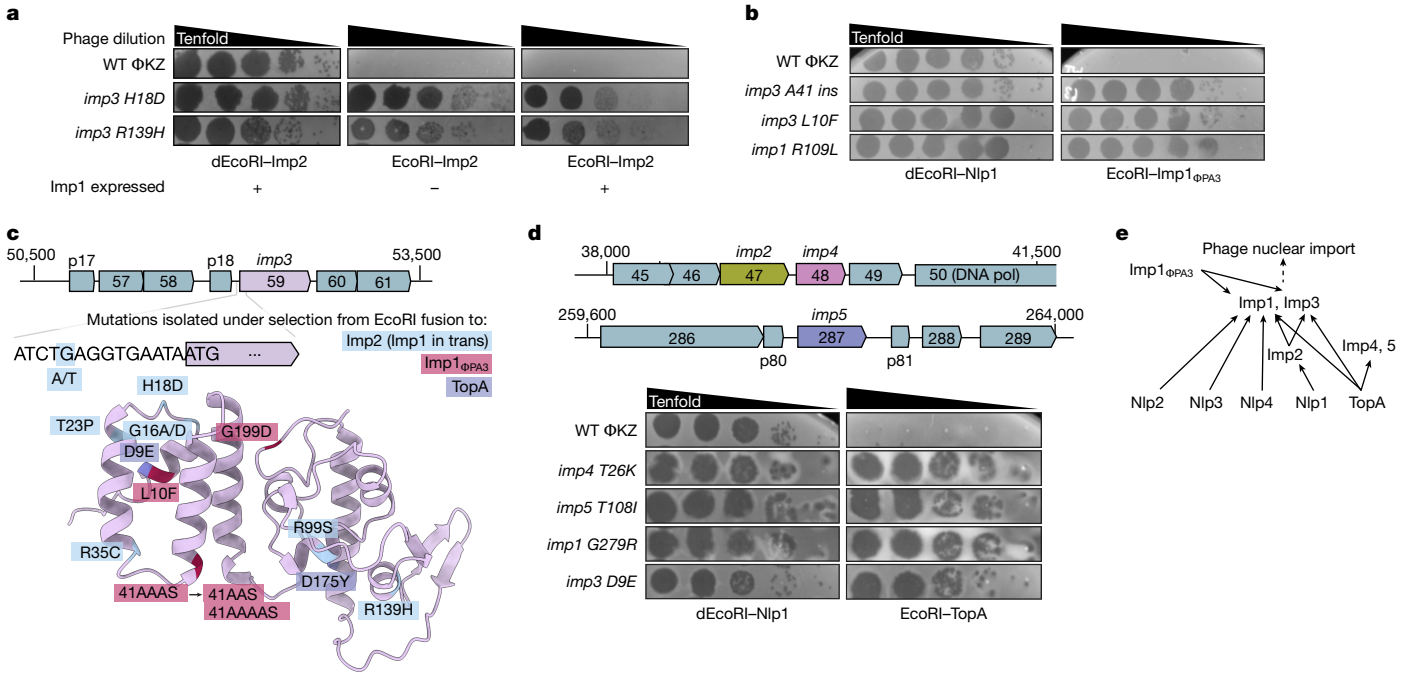


Fig. 3 | Imp3 is required for proper Imp1 function. a,**b**, Plaque assays with the indicated WT or mutant phage spotted on PAO1 expressing the indicated sfCherry2 fusion, with (+) or without (–) Imp1 expressed in trans (**a**); $Imp1_{ΦPA3}$ indicates the phage ΦPA3 Imp1 homologue (**b**); plaque assays were performed as in Fig. 1b. **c**, AlphaFold2-predicted structural model of Imp3, with mutated DNA bases or residues from isolated *imp3* mutant phages colour coded by the EcoRI selection from which they were isolated. The top output model is shown. **d**, Plaque assays with the indicated WT or mutant phage spotted on PAO1 expressing the indicated sfCherry2 fusion. **e**, Schematic representing the factors required for nuclear import of all proteins queried in EcoRI selections, as determined by the escape mutations isolated in *imp1–5* from each selection. All plaque assays were repeated three times independently in biological replicates, with similar results.

(Fig. 3c, Table 1 and Supplementary Table 1). Coding changes are mapped on a predicted Imp3 AlphaFold2 model, where the C-terminal domain is predicted with high confidence and the N-terminal domain with lower confidence (Extended Data Fig. 5a). In a complementary approach, we fused EcoRI to gp63 from phage ΦPA3, an Imp1 homologue with 52% amino acid identity ($Imp1_{ΦPA3}$). EcoRI–$Imp1_{ΦPA3}$ provided weaker selection pressure against ΦKZ than EcoRI–$Imp1_{ΦKZ}$, and enabled the emergence of mutant ΦKZ phages (frequency around 10^{-5}; Fig. 3b and Extended Data Fig. 1a). This approach also selected for two dominant *imp1* mutations and four mutations in *imp3*, which possibly perturb $Imp1_{ΦPA3}$ localization or assembly (Fig. 3b,c, Table 1 and Supplementary Table 1). These data suggest that Imp3 is required for the Imp1-dependent import of Imp2 and potentially for proper localization/assembly of Imp1 itself.

Import requirements for a host protein

To determine whether host proteins require Imp1 for import or operate through a distinct pathway, EcoRI was fused to the host topoisomerase (TopA), previously shown to be imported[1]. The EcoRI–TopA fusion also selected for escape phages (roughly 10^{-6} frequency; Extended Data Fig. 1a), with mutations in *imp1* and *imp3* emerging (Fig. 3c,d, Table 1 and Supplementary Table 1). For a structural model with 33 unique Imp1 mutations shown, in addition to the AlphaFold2 confidence scores and an electrostatic representation, please refer to Extended Data Fig. 4b–d. In addition, EcoRI–TopA selected for mutations in two new uncharacterized genes, *orf48* (*imp4*) and *orf287* (*imp5*) (Fig. 3d). Like Imp2 (encoded by *orf47*), identified above as a protein specifically required for Nlp1 import, Imp4 and Imp5 are conserved only in ΦKZ-like phages infecting *Pseudomonas* (Extended Data Fig. 5b), and no other EcoRI fusion selected for mutations in these genes. We thus interpret this to mean that Imp4 and Imp5 are required for TopA import, in addition to core import proteins Imp1 and Imp3.

To understand the contribution of Imp3 to protein import and Imp1 function, we assessed its phylogenetic distribution and localization. Like Imp1, Imp3 is well conserved among nucleus-forming phages beyond those that infect *Pseudomonas*[6] (Extended Data Fig. 5c). However, fluorescently tagged Imp3 did not express well and thus its localization could not be assessed. In addition, EcoRI–Imp3 fusions did not restrict ΦKZ but did restrict EcoRI-sensitive phage 14-1 (Extended Data Fig. 5d), suggesting that either Imp3 is not localized in the nucleus or the fusion to EcoRI disrupts its localization. Coencoded genes in a putative operon[16] with *imp3* are probably required for proper expression or function, because these genes (*p18*, *imp3*, *orf60* and *orf61*) were required to achieve partial complementation of EcoRI–$Imp1_{ΦPA3}$ escape mutants (Extended Data Fig. 4e). Moreover, despite EcoRI–TopA selection leading to different phages acquiring mutations in any one of four different genes (*imp1*, *imp3*, *imp4* and *imp5*), none of these individual genes when provided in trans resensitized the mutants to EcoRI–TopA (that is, no complementation). Notably, a large synthetic construct expressing the *imp3* operon, together with *imp1*, *imp4* and *imp5*, provided full complementation of EcoRI–TopA import for every mutant phage from this selection (Extended Data Fig. 5e). These data suggest that *imp1*, *3*, *4* and *5*, potentially together with *imp3* gene neighbours, form a core import unit for TopA.

Imp1 binds to Imp6 and Nlp2

We next reasoned that the process of protein import may require additional proteins that associate with Imp1 that were not identified by genetic selection experiments. To determine candidate interactors, we examined the operon structure and conservation of the locus around *imp1* (*orf69*), which encodes gp67–70 (Fig. 1b). Previous work has established that gp68 (together with gp71–73 and gp74) is part of the non-virion RNA polymerase (nvRNAP) complex[17] and therefore we did

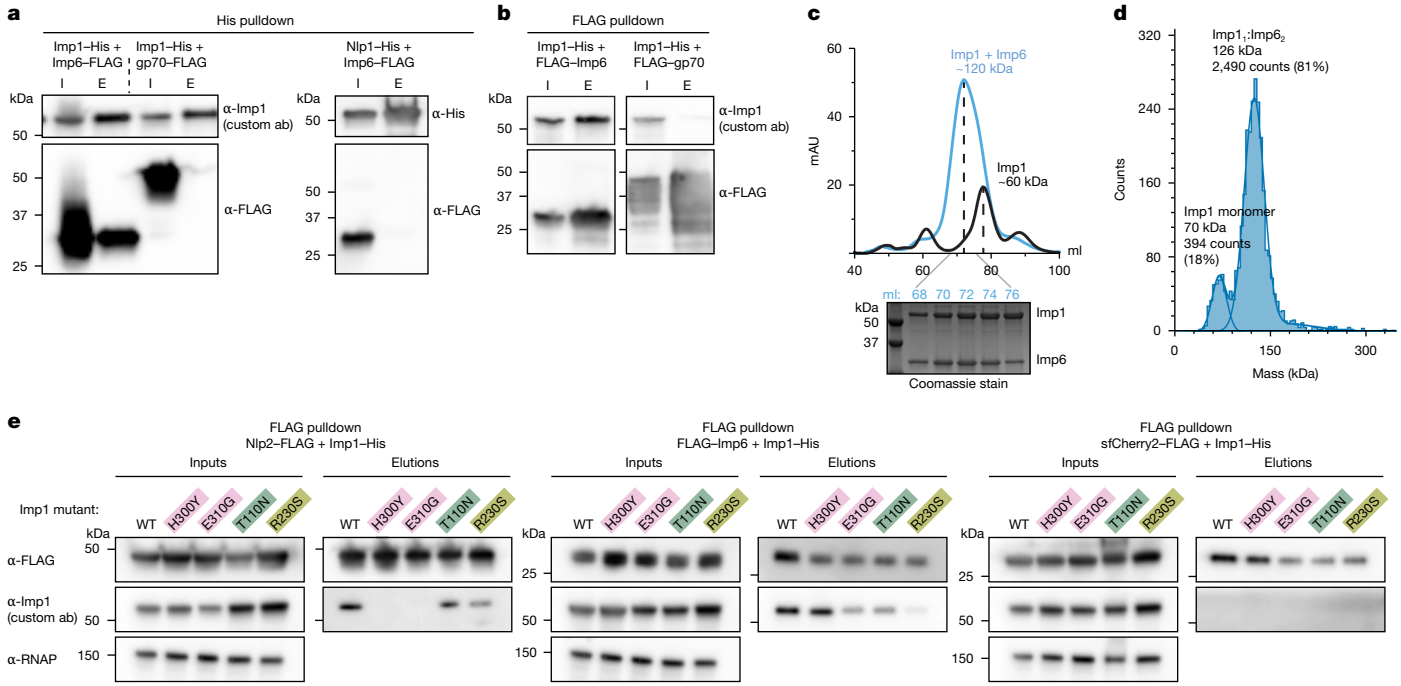


Fig. 4 | Imp1 binds to Imp6 and cargo protein Nlp2. a,**b**, Immunoblot analysis (anti-Imp1) of His pulldown of His-tagged Imp1 with FLAG-tagged Imp6 or gp70, and Nlp1-His with Imp6–FLAG as negative control (**a**); FLAG pulldown of Imp1–His with FLAG–Imp6 or FLAG–gp70 (**b**). **c**, Size-exclusion chromatography (SEC) of Imp1–His alone or in complex with Imp6–FLAG. Peak Imp1–His/Imp6–FLAG SEC fractions (indicated by volume (ml) collected from the column) were run on an SDS–PAGE gel and stained with Coomassie. **d**, Estimation of complex Imp1–His–Imp6–FLAG molecular weight by mass photometry. **e**, Immunoblot analysis of FLAG-tag pulldown of Nlp2–FLAG, FLAG–Imp6 and sfCherry2–FLAG with Imp–His WT or indicated mutants. RNAP serves as input loading control. Pulldowns were performed either two times (**a**,**b**) or three times (**e**) independently, with similar results. See Supplementary Fig. 1 for uncropped and unprocessed gel source data. I, input; E, elution.

not pursue it further. Homologues of gp67, gp69 and gp70 from phage ΦPA3 were previously found to be associated with the nuclear lattice[9], with the gp67 homologue also displaying non-specific RNA-binding activity and being named ChmC[14]. Therefore, we assessed whether Imp1 interacts with gp67 or gp70. 6×His-tagged Imp1 expressed in *Escherichia coli* immunoprecipitated with gp67–FLAG, but not with gp70–FLAG (Fig. 4a), and a reciprocal FLAG immunoprecipitation yielded the same result (Fig. 4b). To confirm that gp67–FLAG does not adhere to Ni-NTA beads non-specifically, 6×His-tagged Nlp1 was used, which did not pull down gp67–FLAG (Fig. 4a). The Imp1–gp67 interaction was stable in size-exclusion chromatography (Fig. 4c); its elution position and mass photometry suggested a complex of around 120 kDa, consistent with a stoichiometry of $gp67_2$:$Imp1_1$ (Fig. 4d). Because of the association of gp67 with Imp1, we now refer to it as Imp6, as a putative member of a phage nuclear import complex.

In addition to the interaction with Imp6 shown above, Imp1 may bind directly to imported cargo to facilitate transport. We therefore conducted immunoprecipitations of cargo proteins Imp2, Nlp2, Nlp3 and Nlp4 expressed in *E. coli* to query direct binding to Imp1. Control experiments showed that Nlp3 did not express well, whereas Imp2 and Nlp4 interacted non-specifically with negative control baits. Interestingly, Nlp2–FLAG immunoprecipitated 6×His-tagged Imp1 in a FLAG pulldown, but not a 6×His-tagged negative control bait (Fig. 4e), suggesting a direct interaction between Imp1 and the Nlp2 cargo protein. An AlphaFold3 (ref. 18) structural model predicted a high-confidence Imp1–Nlp2 dimer, with a binding interface containing the Imp1 residues vital for Nlp2 import in vivo (H300 and E310) (Extended Data Fig. 6a). Interestingly, the C-terminal domain of Nlp2, which is of unknown function, is the region predicted to bind to Imp1 whereas the N-terminal domain has recently validated RNase H activity, which is essential for phage replication[19]. These data suggest that the Nlp2 C-terminal domain contains an import signal at the interface between Nlp2 and Imp1, which mediates direct Imp1 binding. Using our EcoRI fusion import assay, the Nlp2 C-terminal domain (residues 201–482) was indeed necessary and sufficient for import when fused to EcoRI (Extended Data Fig. 6b).

Focusing on the Nlp2–Imp1 interaction, we next reasoned that mutations in Imp1 that specifically decrease import of Nlp2 might weaken direct binding between them whereas other Imp1 mutations would not. We queried binding of Nlp2 to Imp1 mutants H300Y and E310G, which decreased Nlp2 import in vivo. These Nlp2-specific mutations in Imp1 abolished the binding between Nlp2 and Imp1, whereas control mutants T110N (decreased import of Nlp4) and R230S (decreased import of Imp2) maintained interaction with Nlp2 (Fig. 4e). Notably, as a positive control, we presumed that Imp6 would retain binding to Imp1 mutants that abolish a specific Nlp2 cargo interaction. Imp1 H300Y and E310G mutants indeed still coimmunoprecipitated with Imp6, demonstrating that lost Nlp2 binding is due to perturbation of the binding site rather than a result of misfolding or poor expression (Fig. 4e). We therefore conclude that the genetic selection approach has shown direct interactions between a key protein import specificity determinant (Imp1) and imported cargo proteins (for example, Nlp2) and/or their adaptors. Moreover, we propose that a roadmap for future biochemical and structural work, to understand this import pathway at the atomic level, has now been established.

Discussion

The ΦKZ-like family of jumbo phages assembles a notable protein-based nucleus during infections, which selectively internalizes or excludes proteins, demonstrating a eukaryotic nucleus-like segregation ability. These phages often have very large genomes (above 200 kb), with many genes of unknown function that probably contribute to this and other processes executed by this phage family, necessitating unbiased approaches to identification of gene functions. Through the

deployment of a genetic selection approach, we have identified five factors (Imp1–5) required, in different combinations, for the import of six proteins (Nlp1–4, Imp2 and TopA; Table 1).

A key finding from this work is the functional identification of a broadly conserved import and specificity factor, Imp1, a protein that has distinct interfaces for multiple imported cargo and directly interacts with at least one cargo protein (see model in Extended Data Fig. 6c). In addition, microscopy experiments and previous proximity labelling with mass spectrometry[9] are consistent with an interaction between Imp1 and the nuclear lattice. While this study was in revision, an independent paper also demonstrated that Imp1 (referred to as PicA) is required for import of a phage cargo (gp104/Nlp3) by a proposed import signal[20]. Together with the many Imp1 interfaces identified here, these data suggest that import is licensed by distinct import signals possessed by different cargo (sequences for Nlp2 residues at the predicted Imp1–Nlp2 interface and the proposed Nlp3 import signal are presented in Extended Data Fig. 6d).

In addition to identification of multiple Imp1 interfaces that license import, we report a functional role for the broadly conserved protein of unknown function, Imp3, and a direct physical interaction between Imp1 and Imp6. An Imp6 homologue in phage ΦPA3 was recently described as an essential non-specific RNA-binding protein localized to the periphery of the phage nucleus[14]. *imp1* and *imp6* have a nvRNAP subunit embedded between them in the genome, with the remaining subunits encoded close by. Future work is needed to determine the potential connections between transcription within the phage nucleus (which is executed by the same nvRNAP), Imp6 non-specific RNA binding, RNA export/localization and protein import. Likewise, it currently remains unclear how proteins physically transit into the phage nucleus, which will be the subject of future work. Possible options include a bona fide pore, a flippase-like mechanism or an unfoldase that participates in threading the cargo through the protein lattice.

The fundamental biological challenge of protein localization in segregated organellar compartments has largely been studied in eukaryotic systems. This phage family and its protein-based nucleus provide a new challenge and opportunity for understanding the basic mechanisms of protein movement. The identification of phage genes required for import, along with a method for their discovery in other phage systems, are key first steps to unravelling this fascinating mystery.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-024-08547-x.

1. Chaikeeratisak, V. et al. Assembly of a nucleus-like structure during viral replication in bacteria. *Science* **355**, 194–197 (2017).
2. Chaikeeratisak, V. et al. The phage nucleus and tubulin spindle are conserved among large *Pseudomonas* phages. *Cell Rep.* **20**, 1563–1571 (2017).
3. Mendoza, S. D. et al. A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature* **577**, 244–248 (2020).
4. Malone, L. M. et al. A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nat. Microbiol.* **5**, 48–55 (2020).
5. Laughlin, T. G. et al. Architecture and self-assembly of the jumbo bacteriophage nuclear shell. *Nature* **608**, 429–435 (2022).
6. Prichard, A. et al. Identifying the core genome of the nucleus-forming bacteriophage family and characterization of *Erwinia* phage RAY. *Cell Rep.* **42**, 112432 (2023).
7. Weintraub, S. T. et al. Global proteomic profiling of *Salmonella* infection by a giant phage. *J. Virol.* **93**, e01833-18 (2019).
8. Jacquemot, L. et al. Therapeutic potential of a new jumbo phage that infects *Vibrio coralliilyticus*, a widespread coral pathogen. *Front. Microbiol.* **9**, 2501 (2018).
9. Enustun, E. et al. Identification of the bacteriophage nucleus protein interaction network. *Nat. Struct. Mol. Biol.* **30**, 1653–1662 (2023).
10. Nieweglowska, E. S. et al. The ϕPA3 phage nucleus is enclosed by a self-assembling 2D crystalline lattice. *Nat. Commun.* **14**, 927 (2023).
11. Heymann, J. B. et al. The mottled capsid of the *Salmonella* giant phage SPN3US, a likely maturation intermediate with a novel internal shell. *Viruses* **12**, 910 (2020).
12. Lin, D. H. & Hoelz, A. The structure of the nuclear pore complex (an update). *Annu. Rev. Biochem.* **88**, 725–783 (2019).
13. Fossati, A. et al. Next-generation proteomics for quantitative jumbophage-bacteria interaction mapping. *Nat. Commun.* **14**, 5156 (2023).
14. Enustun, E. et al. A phage nucleus-associated RNA-binding protein is required for jumbo phage infection. *Nucleic Acids Res.* **52**, 4440–4455 (2024).
15. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
16. Putzeys, L. et al. Refining the transcriptional landscapes for distinct clades of virulent phages infecting *Pseudomonas aeruginosa*. *Microlife* **5**, uqae002 (2024).
17. Yakunina, M. et al. A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage. *Nucleic Acids Res.* **43**, 10411–10420 (2015).
18. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **63**, 493–500 (2024).
19. Gerovac, M. et al. Non-genetic messenger RNA silencing reveals essential genes in phage-host interplay. Preprint at *bioRxiv* https://doi.org/10.1101/2024.07.31.605949 (2024).
20. Morgan, C. et al. An essential and highly selective protein import pathway encoded by nucleus-forming phage. *Proc. Natl Acad. Sci. USA* **121**, e2321190121 (2024).

Methods

Bacterial growth and genetic manipulation

Escherichia coli strains XL1-Blue and BL21 (DE3) and *Pseudomonas aeruginosa* strain PAO1 were grown in lysogeny broth medium at 37 °C, with shaking at 175 rpm. Bacteria were plated on lysogeny broth solid agar, and any necessary antibiotics to maintain plasmids, with 10 mM $MgSO_4$ when plating for phage infection. When growing for phage infection, PAO1 overnight cultures were inoculated from a single colony from a struck out glycerol stock or fresh transformation, and the culture was grown for approximately 15–16 hours. Expression of genes inserted in the chromosomal attTn7 site was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) in lysogeny broth solid agar. Basal expression of EcoRI–sfCherry2–NlpX or –ImpX fusions cloned into the pHERD30T[21] plasmid under the leaky arabinose-inducible promoter was sufficient to induce phage restriction and enable visualization by microscopy, and thus no L-arabinose was added to plates or liquid cultures.

Phage growth

Phages were grown at 30 °C on lysogeny broth solid agar with 10 mM $MgSO_4$, plus any necessary antibiotics and inducer. Bacteria (150 µl) and phage (10 µl) were mixed in 3.5 ml of 0.35% top agar with 10 mM $MgSO_4$ and plated on lysogeny broth solid agar. Plates were incubated at 30 °C overnight. The following day, individual plaques were picked and stored in 200 µl of SM phage buffer. Escaper phages were plaque purified three times by repeating this method. High-titre lysates were generated by infection of PAO1 expressing the fusion construct that was used in selection overnight in liquid lysogeny broth plus 10 mM $MgSO_4$, with appropriate antibiotics and inducers, at 37 °C. The supernatant was collected and treated with 5% vol. chloroform, shaken gently for 2 min and spun down for 5 min at maximum speed to remove cell debris. This was repeated, and the final phage lysate was stored with 1–5% vol. chloroform.

Phage spot titration plaque assay

P. aeruginosa strain PAO1 overnight culture (150 µl) was mixed with 3.5 ml of 0.35% top agar and poured on solid lysogeny broth agar plates. Once solidified, tenfold dilutions of phage in SM phage buffer were spotted on the surface in 3-µl spots.

Efficiency of plating assay

P. aeruginosa strain PAO1 overnight culture (150 µl) was mixed with 10 µl of phage, with the subsequent addition of 3.5 ml of 0.35% top agar, and the mixture was poured on solid lysogeny broth agar plates. Plates were incubated at 30 °C overnight and plaques were counted the following day. Efficiency of plating was calculated as PFU ml^{-1} on EcoRI targeting fusion relative to PFU ml^{-1} on a non-targeting strain. The efficiency of plating graph was generated in Prism (v.10.3.1).

Construction of fusion proteins

The shuttle vector pHERD30T[21] was used for cloning and expression of EcoRI–sfCherry2–Nlp/Imp and mNeonGreen fusions in *P. aeruginosa* strain PAO1. This vector has a gentamicin-selectable marker and an arabinose-inducible promoter. For mNeonGreen fusions, the vector was digested with NcoI and HindIII restriction enzymes. For EcoRI–sfCherry2 fusions, the vector backbone[3] was either amplified by PCR or digested with SacI and SpeI restriction enzymes and purified. Inserts were amplified by PCR from diluted ΦKZ lysates as the DNA template, and joined into the linearized vector by Hi-Fi Gibson DNA Assembly (NEB) according to the manufacturer's protocol. The resulting reactions were used to transform *E. coli* XL1-Blue competent cells. EcoRI–sfCherry2 fusions to gp59, 70, 104, 171, 179, and TopA were generated by Genscript's custom cloning service. All plasmid constructs were sequenced using primers that either annealed to the vector outside of the multiple cloning site (Quintara) or were whole-plasmid sequenced (Plasmidsaurus). Plasmids were then electroporated into competent PAO1 cells and selected on gentamicin.

Construction of chromosomal insertion strains

For chromosomal insertion of phage import factor genes into PAO1, genes of interest were cloned into pUC18-miniTn7T-LAC[22], linearized once with SpeI and SacI and purified. Inserts were amplified by PCR from diluted ΦKZ lysates and joined into the linearized vector by Hi-Fi DNA Assembly (NEB). The resulting vectors were used to transform *E. coli* XL1-Blue competent cells and verified by sequencing using primers that anneal to the vector outside of the multiple cloning site. The transposase helper vector pTNS3 (ref. 23) was used, with miniTn7 constructs, to transform PAO1 and insert genes of interest into the PAO1 chromosome at the Tn7 locus, with a –pTNS3 control in parallel, and the transformation was selected on gentamicin. Candidate integrants were screened by either whole-genome sequencing (described below) or colony PCR using PTn7R and PglmS-down, as well as an internal gene-specific primer paired with PglmS-up[22]. Colony PCR-screened integrants were then amplified and sequenced with primers that annealed outside the attTn7 site, to verify integration. Electrocompetent cell preparation, electroporation, integration, Flp recombinase-mediated gentamicin marker excision using the pFLP2 plasmid, and plasmid curing by sucrose counterselection were performed as described previously[22].

Whole-genome sequencing

Genomic DNA extraction from phage lysates was performed by the addition of 200 µl of lysis buffer (final concentration 10 mM Tris pH 7.5, 1 mM EDTA, 100 µg ml^{-1} Proteinase K, 100 µg ml^{-1} RNaseA and 0.5% SDS) to 200 µl of high-titre phage lysate (over 10^9 PFU ml^{-1}), with incubation at 37 °C for 30 min and then 55 °C for 30 min. Preparations were then purified by phenol chloroform extraction followed by chloroform extraction and ethanol precipitation, or using the DNA Clean & Concentrator Kit (Zymo Research). Genomic DNA extraction from bacteria was performed by lysis of 200 µl of an overnight culture, with the lysis buffer conditions as above, and DNA was purified with the Genomic DNA Clean & Concentrator Kit (Zymo Research). DNA was then quantified using the Qubit 4.0 Fluorometer (Life Technologies). Genomic DNA (50–100 ng) was used to prepare whole-genome-sequencing libraries using the Illumina DNA Prep Kit (previously, Nextera Flex Library Prep Kit). A modified protocol was used, with fivefold reduced quantities of tagmentation reagents per preparation, except for the bead-washing step, in which the recommended 100 µl of tagment wash buffer was used. On-bead PCR indexing amplification was performed using custom-ordered indexing primers (IDT) matching the Illumina Nextera Index Kit sequences and 2× Phusion Master Mix (NEB). PCR reactions were amplified for 9–11 cycles, and subsequently resolved by agarose gel electrophoresis. DNA products were excised around the 400-base pair size range and purified using the Zymo Gel DNA Recovery kit (Zymo Research). Libraries were quantified by Qubit. Libraries were pooled in equimolar ratios and sequenced on Illumina MiSeq using 150 cycle v.3 reagents (single end: read 1, 150 cycles; index 1, eight cycles; index 2, eight cycles. Paired end: read 1, 75 cycles; index 1, eight cycles; index 2, eight cycles; read 2, 75 cycles). Data were demultiplexed on instrument and trimmed using cutadapt (v.1.15) to remove Nextera adaptors. Trimmed reads were mapped using Bowtie 2.0 (ref. 24) (–very-sensitive-local alignments) and alignments visualized using IGV (v.2.11.0). Mutations were called if present in over 90% of sequencing reads at loci with at least 20× coverage.

Live-cell fluorescence microscopy

Lysogeny broth 0.8% agar pads (25% lysogeny broth, 2.5 mM $MgSO_4$) were supplemented with 0.5 µg ml^{-1} DAPI for phage DNA staining. PAO1 strains expressing each of the fluorescent protein constructs were

grown in liquid culture to an approximate optical density at 600 nm (OD_{600}) of 0.5, supplemented with inducer if necessary (0.05% arabinose for pHERD30T–mNG-Imp2, 0.1 mM IPTG for Tn7::Imp1–mNG) to induce construct expression, and were subsequently infected with ΦKZ lysate for 50 min at 30 °C before imaging. Microscopy was performed on an inverted epifluorescence device (Ti2-E, Nikon) with the Perfect Focus System and a Photometrics Prime 95B 25-mm camera. Images were acquired using Nikon Elements AR software (v.5.02.00). Cells were imaged through channels of blue (DAPI, 50-ms exposure, for phage DNA), green (FITC, 200-ms exposure, for mNeonGreen constructs), red (Cherry, 200-ms exposure, for sfCherry2 constructs) and phase contrast (200-ms exposure, for cell recognition) at ×100 objective magnification (numerical aperture 1.45). For time-lapse imaging, cells were infected with ΦKZ lysate for 20 min at 30 °C before imaging. During time-lapse imaging, the specimen was imaged at intervals of 3 min for 2.0–2.5 h, with channels and exposure times as described above. Final figure images were prepared in Fiji (v.2.1.0/1.53c)[25].

Structure prediction and structure-based homology search

The predicted structures of Imp1 and Imp3 were generated with AlphaFold2 using the ColabFold Google colab notebook with default settings (Imp1 using ColabFold v.1.2, subsequently Imp3 using ColabFold v.1.5.2), with MMseqs2 and HHsearch to generate sequence alignments and templates[15]. The highest-ranked confidence model (based on average predicted local distance difference test) was used for structure modelling and structure homology searches using DALI[26] with the heuristic PDB search option. The predicted structure of the Imp1–Nlp2 complex was generated with AlphaFold3 using the AlphaFold Google web server (https://golgi.sandbox.google.com/).

Phylogenetic analysis

Homologues of ΦKZ import factors were identified by three iterations of PSI-BLAST against the non-redundant protein database. Hits with over 70% coverage and e-value below 0.005 were included to generate a multiple sequence alignment (MSA) using MAFFT (v.7.490, fast strategy). Genomes not labelled as bacteriophages with homologues were manually inspected for the presence of phage structural genes (that is, to exclude bacterial contigs), and were excluded from MSA input if no phage genes were annotated. Following manual inspection of MSA, results were input to FastTree (v.2.1.11 SSE3, default settings) to generate a phylogenetic tree and were visualized in the Interactive Tree of Life. To determine whether phage genomes containing Imp1 homologues also encode ChmA/PhuN (gp54 in ΦKZ), gp54 homologues were acquired by three iterations of PSI-BLAST, as described above, and genomes from Imp1 homologues were inspected for the presence of gp54 homologues. Genomes lacking an apparent gp54 homologue according to PSI-BLAST were inspected manually to identify a predicted gp54 locus (flanked by two well-conserved, phage-encoded genes: DNA polymerase upstream and RNA polymerase β-subunit downstream). These candidates were confirmed as gp54 homologues by their similarity (over 50% identity and over 70% coverage) to a homologue present in a genome on the Imp1 tree.

Protein expression and purification

6×His-tagged Imp1 and Nlp1 were cloned into pET29b, and FLAG-tagged Imp6, gp70, Nlp2 and sfCherry2 were cloned into pETduet-1. Binary interactions between His-tagged Imp1/Nlp1 and FLAG-tagged Imp6/gp70 were assessed by coexpression in BL21 (DE3) cells and purification by affinity chromatography using Ni-NTA. For expression, 100-ml cultures of cells were grown in lysogeny broth supplemented with kanamycin (50 μg ml^{-1}) and carbenicillin (100 μg ml^{-1}) at 37 °C, with shaking at 175 rpm. When cultures reached OD_{600} of about 0.6, protein expression was induced with 1 mM IPTG followed by incubation at 18 °C for about 16 h. Cells were centrifuged at 5,000*g* and resuspended in lysis buffer (25 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.5 mM TCEP and protease inhibitor (Roche)). Lysis was performed four times by sonication at 20% amplitude for 10 s. Insoluble material was removed from the lysate by centrifugation at 21,000*g* for 30 min. Next, 300 μl of Ni-NTA resin slurry (Qiagen) was washed with 10 ml of wash buffer (25 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole and 0.5 mM TCEP) in a gravity column. Cleared lysate was then run over Ni-NTA resin and non-specific interactors were removed by 10-ml washes using wash buffer (four washes in total). Protein was eluted with 300 μl of elution buffer (25 mM Tris-HCl pH 8.0, 300 mM NaCl and 400 mM imidazole). Size-exclusion chromatography was performed with a Superdex 200 Increase 10/300 GL column (or HiLoad 16/600 Superdex 200-pg when expression volume was scaled up) using an AKTA Pure Protein Purification System (Cytiva). SEC buffer (25 mM Tris-HCl pH 8.0 and 300 mM NaCl) was used for purification, and protein samples collected underwent immunoblotting and mass photometry.

The pulldown experiments on FLAG-tagged proteins shown in Fig. 4b were performed similarly to those on His-tagged proteins, but with several modifications. Cells were resuspended in lysis/wash buffer (25 mM Tris-HCl pH 8.0, 300 mM NaCl and protease inhibitor (Roche)) and sonicated as described above. Then, 25 μl of anti-FLAG magnetic agarose resin (Pierce) was washed and incubated with cleared lysate for 1 h at 4 °C, with constant agitation. Protein-bound resin was washed five times with 1 ml of wash buffer using a magnetic separation rack. Bound protein was eluted with 100 μl of 1.5 mg ml^{-1} 1× FLAG peptide (Millipore) resuspended in wash buffer. Protein interactions were assessed by immunoblot (see below).

The FLAG pulldowns shown in Fig. 4e were performed as above, but with the following modifications. Cells were centrifuged at 5,000*g* and resuspended in lysis buffer (25 mM Tris-HCl pH 8.0–8.3, 150 mM NaCl, 0.5% NP40, 25 units ml^{-1} universal nuclease (Pierce) and protease inhibitor (Roche)). Lysis was performed by sonication at 20% amplitude, 1 s on/1 s off for a total of 10 s on, three times, and insoluble material was pelleted from the lysate by centrifugation at 21,000*g* for 30 min. Next, 25–30 μl of anti-FLAG magnetic agarose resin (Pierce) was washed and incubated with cleared lysate for 1.5 h at 4 °C, with constant agitation. Protein-bound resin was washed five times with 1 ml of wash buffer (25 mM Tris-HCl pH 8.0–8.3 and 300 mM NaCl) using a magnetic separation rack. Bound protein was eluted with 100 μl of 150 μg ml^{-1} 1× FLAG peptide (Millipore) resuspended in wash buffer. Elution was repeated for a total of two times.

Immunoblotting

Protein samples collected through pulldown and SEC analyses were mixed 3:1 with 4× Laemmli buffer, supplemented with β-mercaptoethanol and boiled for 10 min. These samples were then run on precast SDS–PAGE gels (Bio-Rad) and transferred to polyvinylidene difluoride membranes. Membranes were blocked with TBS-T buffer (1× Tris buffered saline and 0.1% Tween-20) supplemented with 5% skim milk. A commercial primary FLAG antibody (α-FLAG, Millipore, catalogue no. F1804), a commercial primary His antibody (α-His, Cell Signaling, catalogue no. 2365S) or a commercial primary RNA polymerase β-subunit antibody (α-RNAPβ, BioLegend, catalogue no. 663903) was then added to the skim milk buffer at a titre of 1:5,000. A custom primary antibody for Imp1 (α-Imp1, Genscript) was used at a titre of either 1:5,000 (Fig. 4a,b) or 1:2,000 (Fig. 4e). Primary antibody was left to incubate at room temperature for 1 h. Membranes were washed three times with 10 ml of TBS-T, then incubated with either commercial α-Rabbit (Cell Signaling, catalogue no. 7074S) or commercial α-Mouse (Invitrogen, catalogue no. 62-6520) secondary antibody at a titre of 1:5,000 for 45 min at room temperature. Membranes were then washed three times with 10 ml of TBS-T and developed with Clarity Max ECL substrate (Bio-Rad). Immunoblot images were captured with an Azure Biosystems C400 imager. See Supplementary Fig. 1 for uncropped and unprocessed gel source data.

Mass photometry

Imp1–6 protein complexes purified by size-exclusion chromatography were analysed using a OneMP mass photometer (Refeyn). Adequate data collection was carried out by mixing 1 µl of 1 µM protein with 15 µl of buffer (25 mM Tris-HCl and 300 mM NaCl). Data collection was performed with AcquireMP software (Refeyn 2024 R1.1). Data were collected for 1 min and yielded 3,055 measurable events. Data processing was done using DiscoverMP software (Refeyn DiscoverMP 2024 R1).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All genotypes from sequenced ΦKZ phages are included in Supplementary Table 1. Source data are available for Figs. 1–4 and Extended Data Figs. 1 and 3–6 in Supplementary Figs. 1 and 2. The Protein Data Bank was queried when running DALI from the DALI web server (http://ekhidna2.biocenter.helsinki.fi/dali/), and can can be accessed here: https://www.rcsb.org/.

21. Qiu, D., Damron, F. H., Mima, T., Schweizer, H. P. & Yu, H. D. PBAD-based shuttle vectors for functional analysis of toxic and highly regulated genes in *Pseudomonas* and *Burkholderia* spp. and other bacteria. *Appl. Environ. Microbiol.* **74**, 7422–7426 (2008).
22. Choi, K.-H. & Schweizer, H. P. mini-Tn7 insertion in bacteria with single attTn7 sites: example *Pseudomonas aeruginosa*. *Nat. Protoc.* **1**, 153–161 (2006).
23. Choi, K.-H. et al. Genetic tools for select-agent-compliant manipulation of *Burkholderia pseudomallei*. *Appl. Environ. Microbiol.* **74**, 1064–1075 (2008).
24. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
25. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
26. Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: one hundred discoveries. *Protein Sci.* **32**, e4519 (2023).

Acknowledgements J.B.-D. is supported by the National Institutes of Health (nos. R01 AI171041 and R01 AI167412). C.K. received support from the UCSF Discovery Fellowship. We thank Bondy–Denomy laboratory members for input into this work, including I. Fedorova for help in running FastTree. We additionally thank D. Agard, E. Nieweglowska, C. Gross and A. Davidson for vital input.

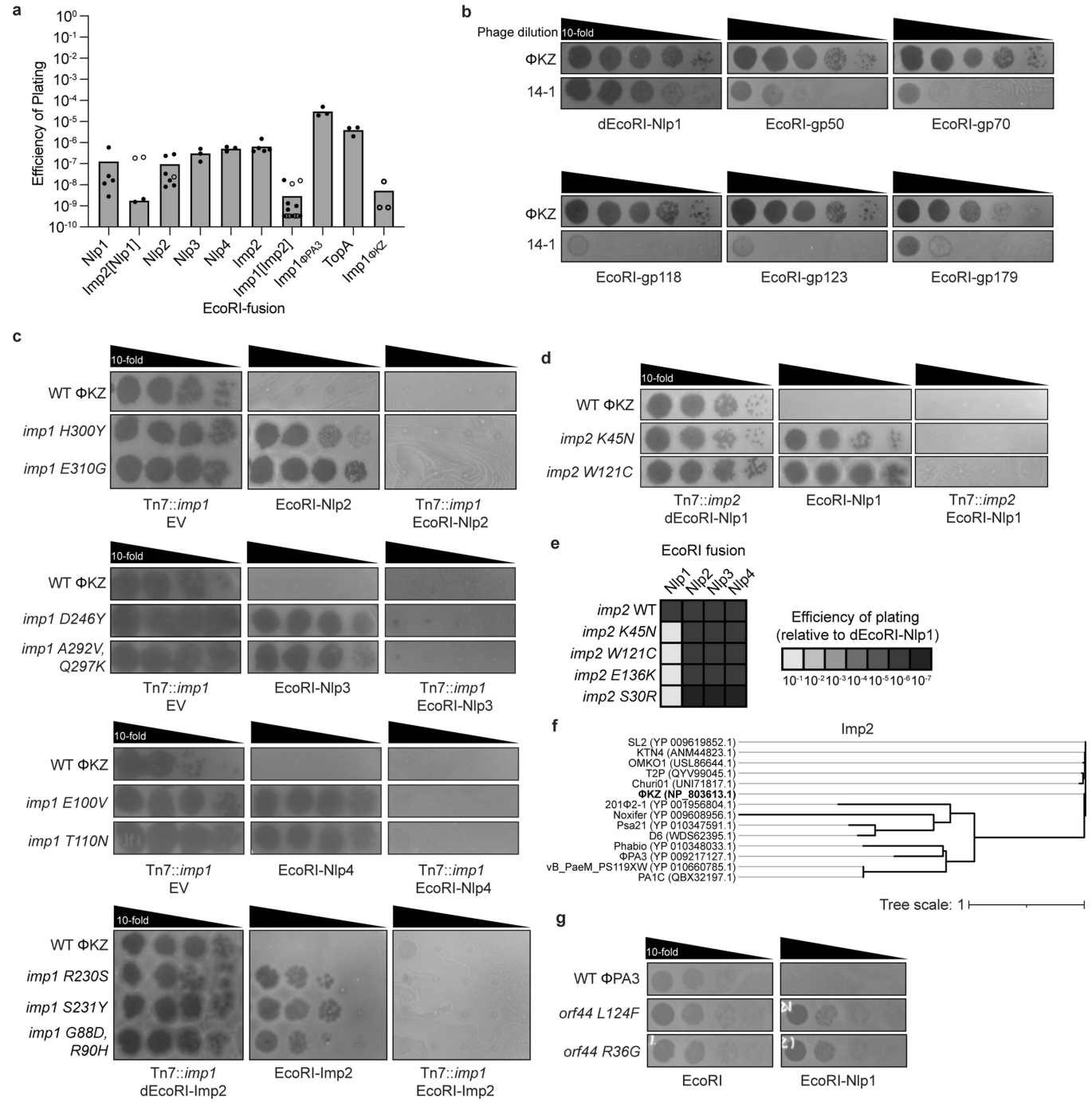
Author contributions C.K. and J.B.-D. conceived the project and designed experiments. C.K. executed phage genetic screens, sequencing, bacterial genetics, fluorescence microscopy, Imp1–Nlp2 biochemical experiments and associated data analysis. T.A.K. executed Imp1–6 biochemical experiments. S.S. provided assistance with sequencing and analysis of phage mutants. J.B.-D. and C.K. wrote the manuscript, and all authors edited it. J.B.-D. supervised experiments and procured funding.

Competing interests J.B.-D. is a scientific advisory board member of SNIPR Biome and Excision Biotherapeutics, a consultant to LeapFrog Bio and a scientific advisory board member and cofounder of Acrigen Biosciences and ePhective Therapeutics. S.S. is cofounder and equity holder in BillionToOne, Inc. and a scientific advisory board member for Junevity, Inc. The remaining authors declare no competing interests. The Bondy–Denomy laboratory received past research support from Felix Biotechnology.

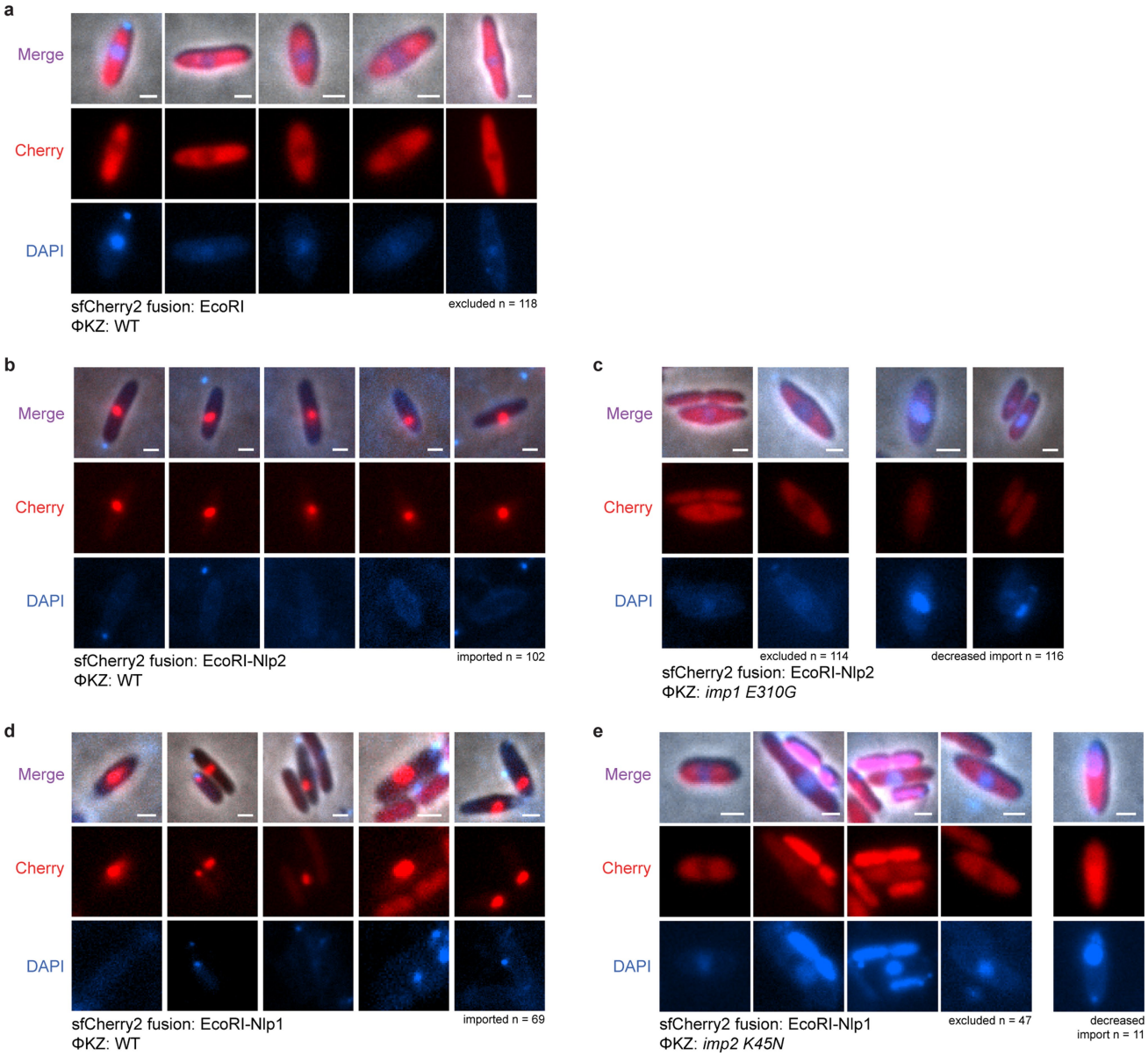
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-024-08547-x.

Correspondence and requests for materials should be addressed to Joseph Bondy-Denomy.

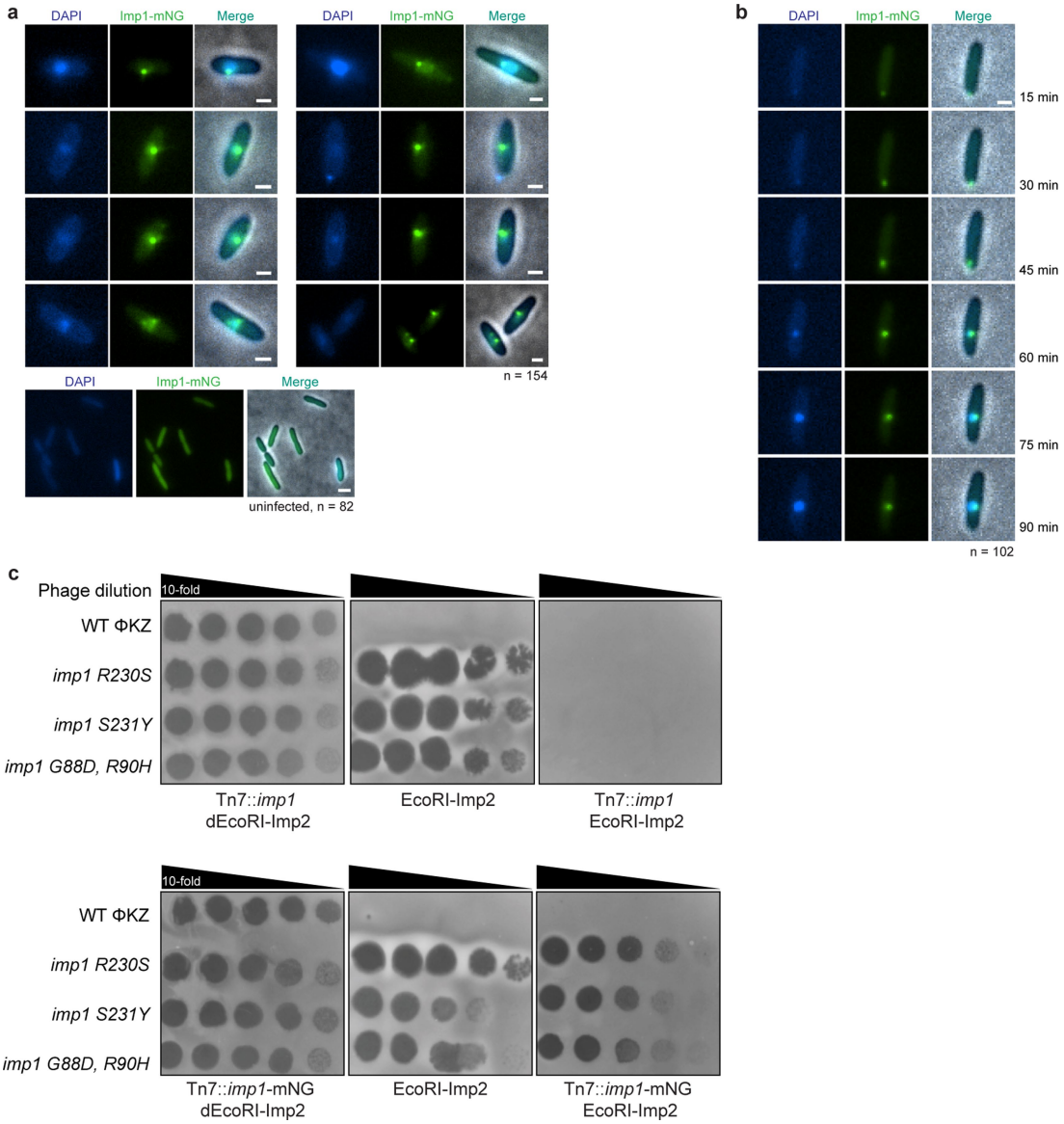
Peer review information *Nature* thanks Alexander Hynes, Ulrich Kubitscheck and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.



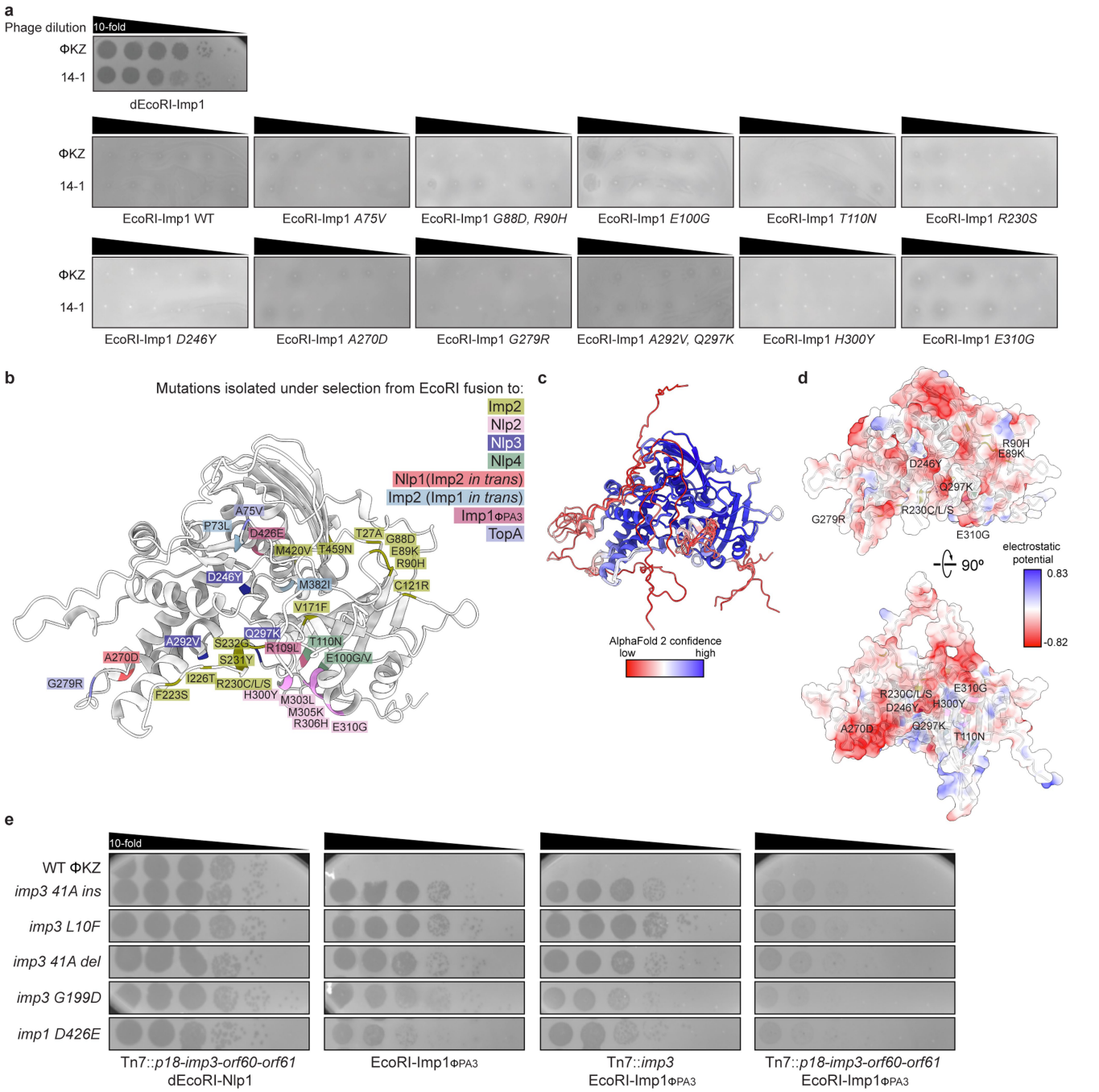
Extended Data Fig. 1 | *imp1* and *imp2* are required for protein import. **a**, Efficiency of plating (EOP) of WT ΦKZ on EcoRI fusions, quantified as the number of plaque forming units (PFU)/mL on PAO1 expressing the indicated EcoRI fusion relative to PFU/mL on a nontargeting EcoRI strain. Closed circles indicate plaques that could be counted. Open circles indicate that no plaques were observed, and an arbitrary value of 1 plaque was recorded to calculate a non-zero limit of detection. Variation in plaquing efficiencies and limits of detection on a given strain result from different titers of the phage stock when the test was done. Bar heights represent the mean EOP between replicates where plaques could be counted. EcoRI-Nlp1 n = 5 independent biological replicates, Imp2[EcoRI-Nlp1] n = 4, EcoRI-Nlp2 n = 5, EcoRI-Nlp3 n = 3, EcoRI-Nlp4 n = 3, EcoRI-Imp2 n = 3, Imp1[EcoRI-Imp2] n = 4, EcoRI-Imp1$_{\Phi PA3}$ n = 3, EcoRI-TopA n = 3, EcoRI-Imp1$_{\Phi KZ}$ n = 2. **b**-**d**, Plaque assays with WT ΦKZ or WT 14-1 (EcoRI-sensitive phage) (**b**) or WT ΦKZ and mutant phage (**c**, **d**) spotted in 10-fold serial dilutions on a lawn of PAO1 expressing indicated sfCherry2 fusions or an empty vector control (EV), with or without expression of a phage gene *in trans* from the bacterial attTn7 site (Tn7::impX) (**c**, **d**). **e**, Representation of EOP of WT ΦKZ and EcoRI-Nlp1 resistant mutant phages on PAO1 expressing EcoRI-Nlp1-4. EOP was calculated as the number of PFU/mL on the EcoRI-Nlp strain relative to PFU/mL on the non-targeting dead EcoRI-Nlp1 (dEcoRI-Nlp1) strain. EOP for each phage/strain pair is colored by mean EOP between three independent biological replicates. **f**, Imp2 phylogenetic tree. **g**, Plaque assays with WT ΦPA3 or mutant phage spotted in 10-fold serial dilutions on a lawn of PAO1 expressing indicated sfCherry2 fusions. Plaque assays were performed two (**b**, **g**) or three (**c**, **d**) independent times in biological replicates with similar results. Please see Supplementary Data 1 for source data underlying graphical representations (**a**, **e**).



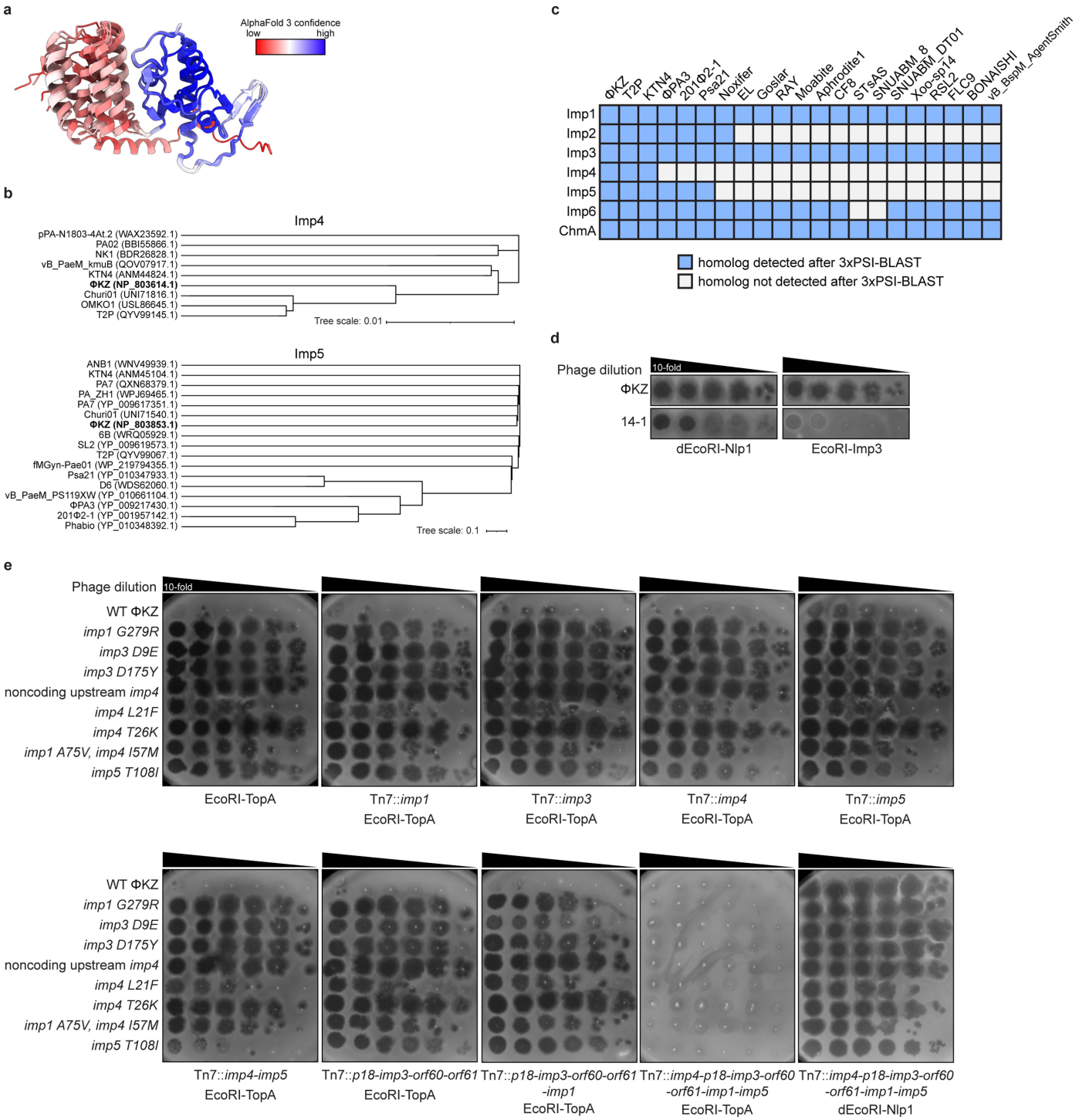
Extended Data Fig. 2 | Mutations in *imp1* and *imp2* decrease nuclear protein localization. a-e, Representative images of live-cell fluorescence microscopy of PAO1 expressing the indicated sfCherry2 fusions, infected with WT or indicated mutant ΦKZ (EcoRI+WT ΦKZ, n = 118 cells. EcoRI-Nlp2+WT ΦKZ, n = 102. EcoRI-Nlp2+*imp1 E310G* ΦKZ, n = 230. EcoRI-Nlp1+WT ΦKZ, n = 69. EcoRI-Nlp1+*imp2 K45N*, n = 58). Scale bars, 1 µm. "Excluded" refers to localization of sfCherry2-fused proteins outside of the phage nucleus. Microscopy was performed as in Fig. 1c and replicated two (EcoRI-Nlp1+*imp2 K45N*) or three (EcoRI+WT ΦKZ, EcoRI-Nlp2+WT ΦKZ, EcoRI-Nlp2+*imp1 E310G* ΦKZ, EcoRI-Nlp1+WT ΦKZ) independent times in biological replicates with similar results.



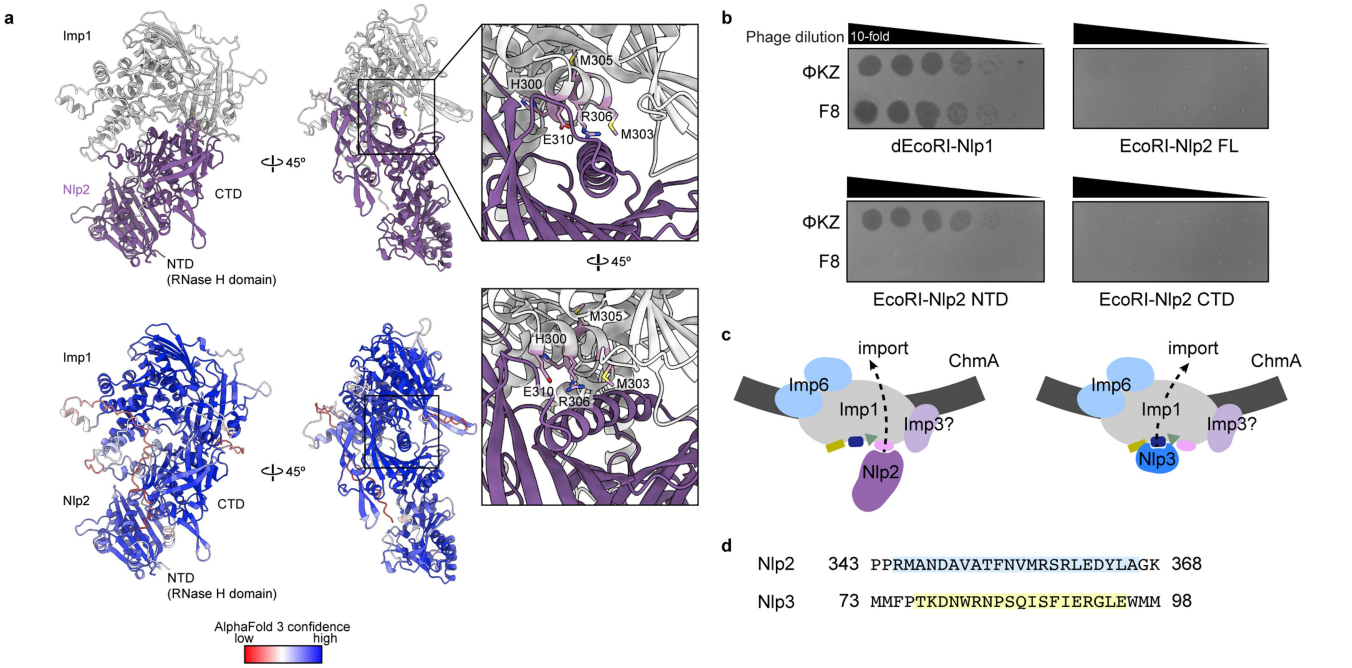
Extended Data Fig. 3 | Imp1-mNG localizes early near the site of genome injection and remains with the phage nucleus. a, Live-cell fluorescence microscopy of PAO1 expressing Imp1-mNeonGreen (Imp1-mNG) from the attTn7 site, infected with WT ΦKZ (Top eight panels, n = 154 cells; scale bar, 1 µm) or uninfected (bottom panel, n = 82 cells; scale bar, 2 µm). **b**, Live-cell time-lapse fluorescence microscopy of PAO1 attTn7::Imp1-mNG infected with WT ΦKZ. n = 102 cells; scale bar, 1 µm. "t = " indicates time in minutes after injected phage DNA is first seen as puncta at the cell pole. Microscopy was performed three independent times in biological replicates with similar results. **c**, Plaque assays with the indicated WT or mutant phage spotted in 10-fold serial dilutions on a lawn of PAO1 expressing the indicated sfCherry2 fusions, with or without expression of the appropriate phage gene *in trans* from the bacterial attTn7 site (Tn7::ImpX). Plaque assays were performed as in Fig. 1b and replicated two independent times in biological replicates with similar results.



Extended Data Fig. 4 | *imp1* and *imp3* mutational analysis. a, Plaque assays with WT ΦKZ or WT 14-1 (EcoRI-sensitive phage) spotted in 10-fold serial dilutions on a lawn of PAO1 expressing the indicated sfCherry2 fusions. **b**, Top output Imp1 model from AlphaFold 2, with mutated residues from isolated *imp1* mutant phages color coded by the EcoRI selection from which they were isolated. **c**, Overlay of all five output Imp1 models from AlphaFold 2, colored by confidence scores. **d**, Surface map of electrostatic potential (semi-transparent overlay) of the top Imp1 predicted structural model from the same view (top), or rotated 90° view (bottom). Several mutated residues are indicated. **e**, Plaque assays with the indicated WT or mutant phage spotted on PAO1 expressing the indicated sfCherry2 fusions, with or without Imp3 or the Imp3 operon (*p18-imp3-orf60-orf61*) expressed *in trans*. Plaque assays were performed as in Fig. 1b and replicated two (**a**) or three (**e**, using 1 or 2 mM IPTG to induce expression from the attTn7 site) independent times in biological replicate with similar results.



Extended Data Fig. 5 | Import of host protein TopA requires *imp1*, *imp3*, *imp4*, and *imp5*. **a**, Overlay of all five output Imp3 models from AlphaFold 2, colored by confidence scores. **b**, Imp4 and Imp5 phylogenetic trees. **c**, Conservation of Imp homologs across nucleus-forming jumbo phages. **d**, Plaque assays with WT ΦKZ or WT 14-1 (EcoRI-sensitive phage) spotted in 10-fold serial dilutions on a lawn of PAO1 expressing the indicated sfCherry2 fusions. **e**, Plaque assays with the indicated WT or mutant phage spotted in 10-fold serial dilutions on a lawn of PAO1 expressing the indicated sfCherry2 fusions, with or without individual or combinations of Imp1-5 or the Imp3 operon (*p18-imp3-orf60-orf61*) expressed *in trans*. Plaque assays were performed as in Fig. 1b and in two (**d**) or three (**e**) independent biological replicates with similar results.



Extended Data Fig. 6 | The Nlp2 C-terminal domain is predicted to bind at the Nlp2-specific Imp1 interface, and is sufficient for import into the phage nucleus. a, Top left, top Imp1-Nlp2 model output from AlphaFold3 is displayed. Imp1 is colored in grey and Nlp2 is colored in purple, and Nlp2 N-terminal domain (NTD) and C-terminal domain (CTD) are indicated. All five models output from AlphaFold3 were aligned and are structurally similar, but only the first model is shown for clarity. Bottom left, top model overlayed with confidence scores, colored from high (blue) to low (red) confidence (ipTM = 0.74, pTM = 0.73). Right, boxed views show Imp1 residue positions mutated under selection with EcoRI-Nlp2 (pink) at the predicted interface between Imp1 and Nlp2. Non-carbon atoms are colored according to identity (oxygen in red, nitrogen in blue, sulfur in yellow). **b**, Plaque assays with WT ΦKZ or WT F8 (EcoRI-sensitive phage) spotted in 10-fold serial dilutions on a lawn of PAO1 expressing the indicated EcoRI fusions. FL, full length. NTD, N-terminal domain (residues 1-200). CTD, C-terminal domain (residues 201-482). Plaque assays were repeated three independent times in biological replicates with similar results. **c**, Model for how multiple cargo-specific Imp1 interfaces may facilitate cargo import into the phage nucleus through the nuclear wall (ChmA). Independent Imp1 interfaces are colored and shaped differently to indicate distinct cargo compatibilities. An Imp6 dimer is shown bound to Imp1, and a hypothetical interaction between Imp1 and Imp3 is also depicted. **d**, Nlp2 residues positioned at the predicted interface with Imp1 (residues 345-366, highlighted in blue) and Nlp3 proposed import signal (residues 77–95[20], highlighted in yellow).

nature portfolio

Corresponding author(s): Joseph Bondy-Denomy

Last updated by author(s): Nov 27, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Live-cell fluorescence microscopy was performed with Nikon Elements AR software (version 5.02.00). Mass photometry was performed using Refeyn AcquireMP 2024 R1.1. DALI was run from the DALI web server (no version information available; http://ekhidna2.biocenter.helsinki.fi/dali/).
Data analysis	Fiji (v2.1.0/1.53c); AlphaFold2 using MMseqs2 (Colabfold v1.2.0 and 1.5.2); AphaFold 3 (no version information available, can be found here: https://golgi.sandbox.google.com/); cutadapt (v1.15); Bowtie 2.0; SAMtools (v1.7); IGV (v2.11.0); ChimeraX (v1.1.1); Refeyn DiscoverMP 2024 R1; Prism (v10.3.1); MAFFT (v7.490); FastTree (v2.1.11 SSE3); iTOL: Interactive Tree of Life

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Supplementary Information is available for this paper. The PDB was queried when running DALI from the DALI web server, and the PDB can be accessed here: https://www.rcsb.org/. Correspondence and requests for materials should be addressed to Joseph Bondy-Denomy. Reprints and permissions information is available at www.nature.com/reprints.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable, no human subjects
Reporting on race, ethnicity, or other socially relevant groupings	not applicable
Population characteristics	not applicable
Recruitment	not applicable
Ethics oversight	not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed–for microscopy experiments, assessing n=50-200 cells is standard in the field, and phenotypes were reproducible within this n of cells and within the number of replicates performed. Live-cell fluorescent microscopy imaging sample sizes were determined by microscope availability as well as observing reproducible phenotypes in n = 50-200 cells per construct imaged. Representative images are presented in Extended Data Figures 2 and 3. For efficiency of plating experiments, results were consistent between n=2 biologically independent experiments, but often more (up to 5, see legend for exact numbers) were performed over the course of the work, where n=2-5 would be standard practice in the field. For pull downs/western blotting, an n=2-3 was performed in accordance with typical standards.
Data exclusions	Data for efficiency of plating (EOP) calculations were excluded when the EOP could not be calculated accurately; for example, when there were too many plaques on a plate to count accurately, if the bacterial lawn grew too poorly for plaques to be distinguished, or if the non-targeting control plate yielded too few (<5) plaques to accurately determine phage titer. During microscopy, cells that were out of focus were not included in the total n= of cells or assessed for relevant phenotypes.
Replication	Live-cell fluorescence microscopy was performed at least two independent times for each construct, with similar results. Imaging was performed over the span of up to two years depending on the fluorescent construct, in each case on separate occasions for each replicate. Western blotting was performed between two to three independent times (detailed in figure legends), with similar results. Plaque assays were performed at least two independent times, but more frequently three or more independent times (detailed in figure legends), with similar results.
Randomization	Experimental groups were not allocated and therefore no randomization was performed.
Blinding	Investigators were not blinded for any parts of the experiments, as this is not a common practice for the experiments performed in this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

custom Imp1 primary polyclonal antibody (α-Imp1, Genscript), primary FLAG antibody (α-FLAG, Millipore cat# F1804), primary His antibody (α-His, Cell Signaling cat# 2365S), primary RNA polymerase β subunit (RNAP antibody (α-RNAPβ, BioLegend cat# 663903) secondary Rabbit IgG HRP-linked antibody (α-Rabbit, Cell Signaling, cat# 7074S), secondary Mouse IgG HRP-linked antibody (α-Mouse, Invitrogen, cat# 62-6520).

Validation

The custom Imp1 primary antibody was validated by the manufacturer (Genscript) for purity >90% by SDS-PAGE, and for specificity for the Imp1 peptide synthesized by Genscript (used to produce the antibody) by ELISA, compared to a negative control IgG. The Imp1 antibody was further validated to confirm specificity to full length Imp1 protein by the authors by western blot of PAO1 cell lysate expressing Imp1 from a plasmid, compared to a lysate expressing the empty vector negative control.

Primary FLAG antibody was validated according to the manufacturer's specifications: Purity: Two major bands with purity >90% when analyzed by microfluidic gel capillary electrophoresis. Specificity: Detects a single band of protein on a Western Blot from mammalian crude cell lysates. Sensitivity: Detects 2 ng of FLAG-BAP fusion protein by Dot Blot using Chemiluminescent Detection.

Primary His antibody was validated according to the manufacturer's specifications: Purification: Polyclonal antibodies are produced by immunizing animals with a 6xHis synthetic peptide. Antibodies are purified by protein A and peptide affinity chromatography. Specificity/Sensitivity: His-Tag Antibody detects recombinant proteins containing the 6xHis epitope tag, compared to a negative control sample.

Primary RNAPβ antibody (clone 8RB13) was validated according to the manufacturer's specifications: Purified antibodies are tested for purity by SDS-PAGE gel electrophoresis. IgG antibodies are required to have purity >95%. Each lot of this antibody is quality control tested by Western blotting.

Secondary Rabbit IgG antibody was validated according to the manufacturer's specifications, which can be found here: https://www.cellsignal.com/products/secondary-antibodies/anti-rabbit-igg-hrp-linked-antibody/7074?utm_term&utm_campaign=SO-Products-SecondaryAntibodies&utm_source=adwords&utm_medium=ppc&hsa_acc=8625036580&hsa_cam=21865491305&hsa_grp=170749758698&hsa_ad=719519792785&hsa_src=s&hsa_tgt=dsa-2370996940456&hsa_kw&hsa_mt&hsa_net=adwords&hsa_ver=3&gclid=EAIaIQobChMI9eS0nPTziQMVXiitBh0ISgSJEAAYASAAEgJ_LPD_BwE

Secondary Mouse IgG antibody was validated according to the manufacturer's specifications, which can be found here: https://www.thermofisher.com/antibody/product/Goat-anti-Mouse-IgG-H-L-Secondary-Antibody-Polyclonal/62-6520

Plants

Seed stocks

not applicable

Novel plant genotypes

not applicable

Authentication

not applicable