Functional exploration of bacteriophage T5 pre-early genes in the host takeover : a route towards new antibacterial strategies ?

Luis Maria Ramirez Chamorro

université PARIS-SACLAY

Functional exploration of bacteriophage T5 pre-early genes in the host takeover: a route towards new antibacterial strategies?

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°569 : Innovation Thérapeutique : Du fondamental à l'appliqué (ITFA)
Spécialité de doctorat : Biochimie et biologie structurale
Unité de recherche : Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.
Référent : Faculté de Pharmacie

Thèse présentée et soutenue en visioconférence totale, le 11 décembre 2020, par

Luis Maria RAMIREZ CHAMORRO

Composition du Jury

Olga SOUTOURINA Professeure, Université Paris Saclay	Présidente
Mart KRUPOVIC Directeur de recherche, Institut Pasteur	Rapporteur & Examinateur
Rob LAVIGNE Professeur, KU Leuven	Rapporteur & Examinateur
Marie-Agnès PETIT Directrice de recherche, INRAE	Examinatrice
David BIKARD Chargé de recherche, Institut Pasteur	Examinateur
Frédérique LE ROUX Directrice de recherche, Ifremer	Examinatrice
Pascale BOULANGER Directrice de recherche, CNRS, Université Paris Saclay	Directrice de thèse
Ombeline ROSSIER Maîtresse de conférences, Université Paris Saclay	Co-Encadrante de thèse

Thèse de doctorat

NNT : 2020UPASQ020

MEMBERS OF THE JURY

President of the Thesis Committee

Olga SOUTOURINA, PhD
ARNs régulateurs chez les Clostridies
I2BC - Institut de Biologie Intégrative de la Cellule
Bât 400, Université Paris Saclay
91405, Orsay, France
olga.soutourina@i2bc.paris-saclay.fr

Thesis reviewed by

Mart KRUPOVIC, PhD
Unité Virologie des Archées
Institut Pasteur
25, rue du Dr. Roux
75015, Paris, France
mart.krupovic@pasteur.fr

Rob LAVIGNE, PhD
Laboratory of Gene Technology
KU Leuven
3001, Leuven (Arenberg), Belgium
rob.lavigne@kuleuven.be

Members of the thesis committee

Marie-Agnès PETIT, PhD
Dynamics of Bacteriophage Genomes
INRAE - Institut national de recherche pour l'agriculture,
l'alimentation et l'environnement
Domaine de Vilvert
78350 Jouy en Josas, France
marie-agnes.petit@inra.fr

David BIKARD, PhD
Synthetic Biology
Institut Pasteur
25, rue du Dr. Roux
75015, Paris, France
david.bikard@pasteur.fr

Frédérique LE ROUX, PhD
Equipe Génomique des vibrios
Laboratoire de Biologie Intégrée des Modèle Marins (LBI2M)
UPMC-CNRS UMR 8227, Station Biologique de Roscoff
Place Georges Tessier, 29680 Roscoff, France
frederique.le-roux@sb-roscoff.fr

PhD supervised by

Pascale Boulanger, PhD
Team Bacteriophage T5
I2BC - Institut de Biologie Intégrative de la Cellule
Bât 430, Université Paris Saclay
91405, Orsay, France
pascale.boulanger@i2bc.paris-saclay.fr

Ombeline ROSSIER, PhD
Bacteriophage T5
I2BC - Institut de Biologie Intégrative de la Cellule
Bât 430, Université Paris Saclay
91405, Orsay, France
ombeline.rossier@universite-paris-saclay.fr

ACKNOWLEDGEMENTS

I want to thank my thesis supervisors Pascale Boulanger and Ombeline Rossier, for having trusted in me to start this project, also for their support and advise.

I am grateful to Mart Krupovic and Rob Lavigne, who agreed to be evaluators of this manuscript. Thanks to Marie-Agnès Petit, David Bikard, Frédérique Le Roux, and Olga Soutourina for accepting to make part of the thesis jury.

To the Paraguayan Ministry of Education, the scholarship agencies "Becas Don Carlos Antonio Lopez", and SFERE. I am heavily indebted to you. Thanks for made this possible.

To Müge Senarisoy, Nicolas Ducrot and Dominique Durand, Leo Zangelmi, Emeline Vernhes, and Madalena Renouard, for your help and for creating such a nice environment in the team.

My thanks to Lory-Anne Baker and Baptiste Errout, with whom I had the opportunity to work during their internship.

I am grateful to the Tavares lab team, Paulo Tavares, Audrey Labarde, Lia Godinho, Jack Dorling, Mehdi, Laura, Stephane. To the colleagues in the I2BC Virology department for the discussions and brilliant ideas provided during the seminars.

Thanks to my wife Heidy, my family, and friends. Thanks for being there.

Contents

1. Introduction 1
1.1. Bacteriophages 1
1.1.1. Phage host interactions 3
1.1.2. Bacteriophages alter bacterial housekeeping processes 4
1.1.3. Targeting the bacterial defense systems 11
1.2. Bacteriophage T5 15
1.2.1. Infectious process 15
1.2.2. Genetic map of the First-Step Transfer DNA 17
1.2.3. Transcription of the pre-early genes. 19
1.2.4. Expression of pre-early genes 21
1.2.5. Other activities attributed to pre-early genes 25
1.3. Aims of this study 29
2. Results 31
2.1. Developing new tools for T5 genome engineering 31
2.1.1. Testing restriction of T5 Infection by CRISPR-Cas9 31
2.1.2. Bacteriophage T5 genome editing through HR plus CRISPR/Cas9 32
2.1.3. Retron mediated gene editing in phages 33
2.1.4. Dilution-Amplification-Screening (DAS) to isolate mutants from phage T5 34
2.1.5. Reversion of single and double amber mutants in *A1* and *A2* 36
2.1.6. Blue/White screening in the T5 genome editing 38
2.1.7. Generating red-fluorescence virion particles by gene fusion with mCherry 40
2.2. Deletion of phage T5 pre-early genes 42
2.2.1. T5 pre-early genes conserved among *Tequintavirus* members 42
2.2.2. At least 14 pre-early genes are dispensable for the infection of Phage T5 44
2.2.3. Whole genome sequencing 45
2.3. Characterization of T5 mutants 48
2.3.1. Deletion of pre-early genes alter the infection kinetics 48
2.3.2. Deletions of pre-early genes impact the phage virulence 50
2.3.3. Cytological analysis of infected bacteria 53
2.4. Functional exploration of T5 pre-early genes 58
2.4.1. Bioinformatic analysis of ORFan pre-early genes 58
2.4.2. Ectopic expression uncovers toxic pre-early proteins 62
2.4.3. Alterations in cell morphology linked to pre-early gene ectopic expression 64
2.4.4. Gp13 is a membrane protein 69
2.4.5. Revisiting the T5 resistance to EcoRI 71

2.5. Defining the Injection stop signal essential for infection 75
3. Discussion 77
3.1. Reverse genetics 77
3.2. How the ectopic expression of T5 pre-early genes impacts the host cell 81
3.3. Towards the understanding of Bacteriophage T5 pre-early genes 82
4. Perspectives 87
4.1. Role of pre-early genes in the host takeover 87
4.2. Identity and function of the ISS 89
5. Conclusion 91
6. Materials and Methods 93
6.1. Strains and culture conditions 93
6.1.1. Bacterial strains 93
6.1.2. Plasmids 94
6.1.3. Phage strains 98
6.1.4. Scripts used in this study 99
6.2. Deletion of phage T5 pre-early genes 100
6.2.1. T5 pre-early genes conserved among tequintaviruses 100
6.2.2. Site Directed mutagenesis 100
6.2.3. Mutants screening with CRISPR-Cas9 100
6.2.4. Mutants screening by dilution 101
6.2.5. Whole genome sequencing of T5 mutants 101
6.3. Characterization of T5 mutants 102
6.3.1. One step growth curve 102
6.3.2. Virulence index 103
6.3.3. Fluorescence microscopy 104
6.4. Ectopic expression of phage genes 106
6.4.1. Cell growth parameters 106
6.5. Functional exploration of pre-early genes 108
6.5.1. Production of Gp13 108
7. Annexes 111
7.1. T5 pre-early genes conserved among *Tequintavirus* members 111
7.2. Uncloned regions 112
7.3. Distribution of DAPI-dye foci in bacterial cells 113
7.4. Sequences of interest 115
7.5. Whole genome sequencing 121
7.5.1. Differences between the T5wt and the NCBI deposited sequence. 121

7.5.2. Phage termini analysis with PhageTerm 122
7.6. Oligonucleotides used in this study 124
7.7. Publication of results 127
8. References 147

Table of Figures

Figure 1: Life cycle of phages. 2

Figure 2. Phage bacteria interaction during the host takeover. 4

Figure 3: Steps of the infection of *E. coli* by bacteriophage T5. 16

Figure 4: Major events during the T5 infection. 17

Figure 5: Genetic map of the Phage T5 LTR (FST-DNA). 18

Figure 6: Annotations at the LTR of some T5 viruses. 18

Figure 7. Putative operons at the First-Step Transfer locus 19

Figure 8. Production of T5 transcripts during infection. 20

Figure 9. Detection of the A2 protein by 2D gel electrophoresis. 24

Figure 10. DNA synthesis after infection of *E. coli* F with T5 mutants. 26

Figure 11. T5 evades the *Escherichia coli* Ssp system. 27

Figure 12: The effect of CRISPR–Cas type I-E targeting on T5 infection. 28

Figure 13: Restriction of phage T5 infection by CRISPR-Cas9. 32

Figure 14. Construction of mutants by homologous recombination and mutant enrichment using CRISPR/Cas9. 33

Figure 15. Sequence of a mutant T5 *amA1* SS84 constructed by retron-mediated recombination 34

Figure 16. Dilution-Amplification Screening of mutants of T5. 35

Figure 17. Screening of T5 mutants by Dilution-Amplification Screening (DAS). 36

Figure 18. Sequencing of amber mutants in genes *A1* and *A2* 36

Figure 19. Amber mutation reversion in T5 amber mutants. 37

Figure 20. Efficiency of plaquing of T5 mutants in essential genes *A1* and *A2*. 38

Figure 21. Construction of a Blue T5 (T5 *lacZ*α) strain. 39

Figure 22. Construction of the mutant T5 PNmC. 41

Figure 23: Presence of T5 genes among *Tequintavirus* members. 42

Figure 24. Presence of T4 genes in virus from the same taxa. 43

Figure 25. Deletion of pre-early genes in T5. 44

Figure 26. Sanger sequencing of mutants obtained in this study. 45

Figure 27. T5 genome sequence was modified for read mapping. 46

Figure 28. Whole genome sequencing of T5 mutants. 47

Figure 29. Infection kinetics of single-gene mutants of T5. 49

Figure 30. Infection kinetics of multiple-deletion mutants of T5. 49
Figure 31. Optical density reduction curves from infections in liquid medium. 51
Figure 32. Virulence of mutants by deletion in T5. 52
Figure 33. Microscopy images of bacteria infected by T5 mutants. 55
Figure 34. Fluorescence intensity of cells infected by T5 mutants. 56
Figure 35. Comparison of shape parameters of T5-infected cells at T0 and T10. 57
Figure 36: Alignment of Dmp with other phage-encoded similar proteins. 58
Figure 37. Amino acid sequence alignment of Gp2 and PARP enzymes. 58
Figure 38. Protein sequence alignment of A1 and selected enzymes with known activity. 59
Figure 39. A2 alignment with DNA binding proteins. 59
Figure 40. Transmembrane domains prediction for Gp5. 59
Figure 41. Transmembrane domains and signal peptide prediction for Gp13. 60
Figure 42.Alignment of Gp13 sequence to proteins with known activity. 60
Figure 43. Workflow scheme for the ectopic expression of pre-early genes. 62
Figure 44. Bacterial growth curves for strains expressing T5 pre-early genes. 63
Figure 45. Survival of bacterial strains expressing T5 pre-early genes. 64
Figure 46. Cells expressing T5 pre-early gene products display different morphologies. 65
Figure 47. Pairwise comparison of parameters from cells expressing phage pre-early gene products. 66
Figure 48. Ectopic expression of the gene 13 and its tagged version. 69
Figure 49. Protein electrophoresis of cell extracts producing Gp13. 70
Figure 50. Transformation of cells encoding the EcoRI Restriction Methylation system. 72
Figure 51. Effects of EcoRI restriction of T5 mutants carrying EcoRI sites in the FST 73
Figure 52. Effects of EcoRI restriction on T5 mutants. 73
Figure 53. Reduction of the Injection Stop signal. 75
Figure 54. Sequencing of the T5 ΔISS::3 LTR. 76
Figure 55. Growth parameters of mutant T5 ΔISS::3. 76
Figure 56. Characteristics of T5 pre-early genes. 81
Figure 57. Proposed insertion of the putative Injection Stop Signal at the left terminal repeat. 89
Figure 58: Pedigree of plasmids constructed. 97
Figure 59. Pedigree of phages obtained during this study. 99

Figure 60. One Step growth protocol used in this study..... 102
Figure 61. Fluorescence microscopy of infected bacteria. 104
Figure 62. Protocol for ectopic expression. 106
Figure 63. Production of Gp13. 108
Figure 64. Summary of the cloned regions of the FST. 112
Figure 65. Distribution of foci in cells infected by mutants of T5. 113
Figure 66. Distribution of DAPI dye foci in cells that express early T5 genes..... 114
Figure 67. Phage termini in deletion mutants of T5..... 123

1. Introduction

1.1. Bacteriophages

Bacteriophages, or phages, are viruses that infect bacteria. Phages, altogether with other viruses (archaeal and eukaryotic) constitute the virosphere. The estimated size of the virosphere is of ~10^7 viruses per mL of water in the oceans (versus 10^6/mL for bacteria), which make them the most abundant microorganisms on earth (Bergh et al., 1989; Wommack and Colwell, 2000). Viruses kill one fifth of the oceanic microbial biomass per day, therefore playing an important role in the organic matter recycling and on biogeochemistry cycles (Suttle, 2007). Such abundance comes along with a diversity in lifestyles.

Phages show one of three life cycles: lytic, lysogenic, or pseudolysogenic. Those that only display lytic cycles are virulent, while phages that can follow lytic or lysogenic cycles are considered temperate. During the lytic cycle, the viral reproduction leads to host cell destruction. Upon genome delivery into the host, virulent phages reorient the host metabolism towards the production of new viral particles, which are released upon cell lysis at the end of the infectious process (Figure 1) (Weinbauer, 2004).

Lysogeny implies viral genome replication without virion production. It unfolds in three steps: establishment, maintenance, and induction of productive cycles. After countering bacterial defenses, temperate phages establish a lysogenic cycle as prophages after the evaluation of host metabolic state or phage density among other parameters. Then, prophages maintain their replication as part of the bacterial chromosome, are vertically transmitted, and undergo evolutionary changes. Induction of productive cycles can be triggered spontaneously or by stress conditions. Prophages in lysogenic cycles affect bacterial metabolism and can affect the equilibrium of microbial communities (Howard-Varona et al., 2017).

Pseudolysogeny encompasses different viral lifestyles hybrid to the lytic or lysogenic ones (Weinbauer, 2004), it is the stalled development of a bacteriophage in the host cell, with either no phage genome multiplication or asynchronous with the cell cycle, and stable maintenance in the cell line, allowing the subsequent resumption of the virus development (Łoś and Węgrzyn, 2012). Pseudolysogeny includes chronic infection. Chronic infections involve the infection and release without lysis, but by budding or extrusion (Weinbauer, 2004). In all cases, irrespective of the type of viral cycle, the early interactions are decisive for the establishment of the infection and the completion of the viral cycle.

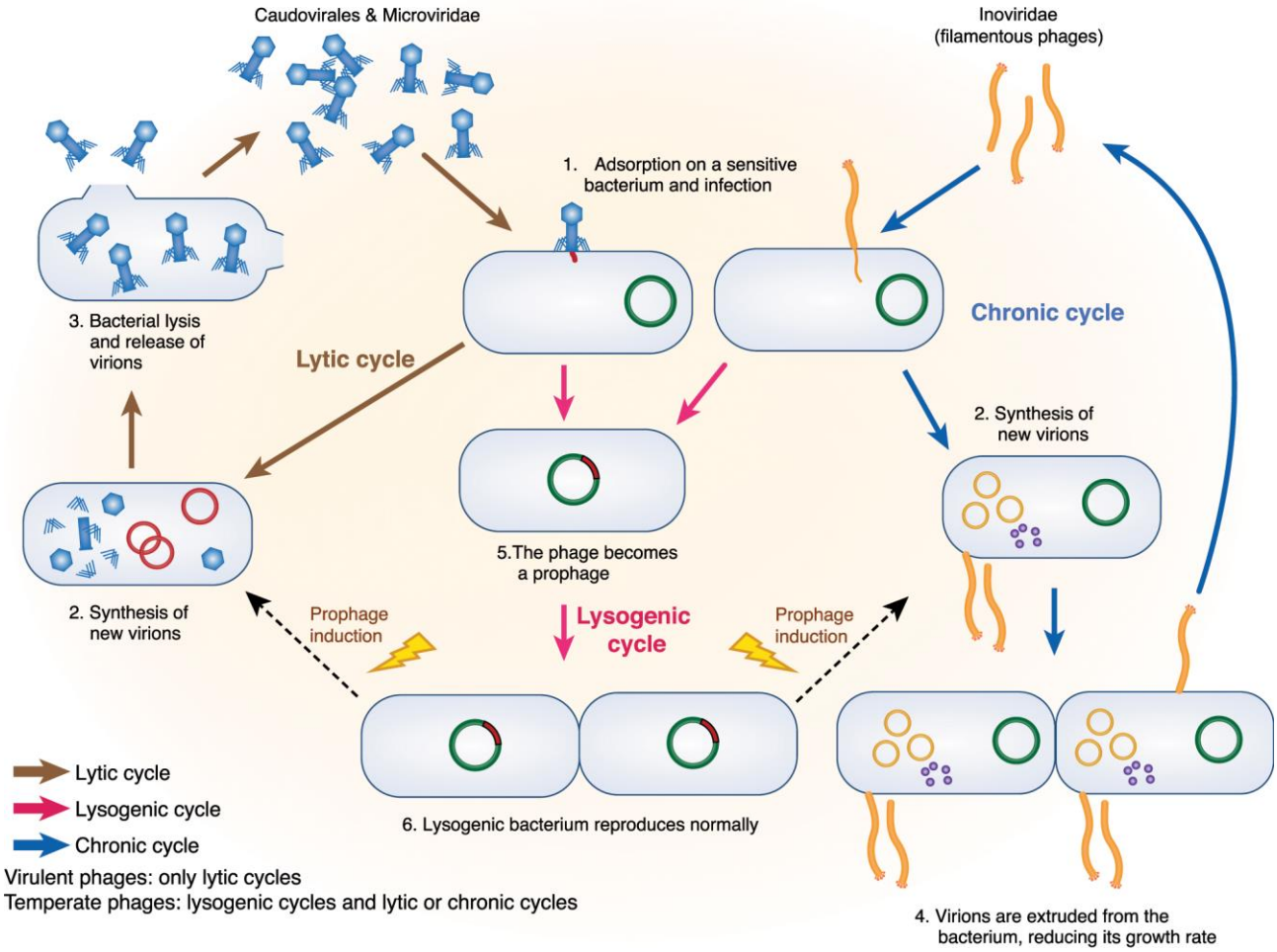


Figure 1: Life cycle of phages.
There are at least three types of phage life cycles: lytic, lysogenic, and chronic infections (one example of pseudolysogenic cycles). In all cases, the cycle starts with host recognition (1) and phage genome delivery. During the lytic cycle, phage multiplication (2) leads to host lysis (3). In chronic infections virions are released while conserving host viability (4). In lysogenic cycles, temperate phage genomes make part of the host genome and replicate along as prophages (6). Prophages can be reactivated and release virions either by a lytic or a pseudolysogenic pathway. This figure is reproduced with permission from Sausset, R., Petit, M.A., Gaboriau-Routhiau, V., and De Paepe, M. (2020). New insights into intestinal phages. Mucosal Immunol. 13, 205–215.

During the lytic infection, three stages of temporal regulation can be identified. After the recognition of their receptor, phages deliver their genetic material into the host, the first stage of the infection starts with the expression of early genes, by the recognition of strong promoters by the host RNA polymerase. The early proteins counteract bacterial defenses and set the conditions for the expression of middle genes, whose products are mainly focused on genome replication. Finally, the late genes are expressed to produce the structural proteins, package the viral DNA into the capsids, assemble new infectious particles and finally lyse the host to release the progeny (Yang et al., 2014).

Since phages are natural predators of bacteria, the notion of using bacteriophages for therapeutic purposes against bacterial infection was born immediately following the discovery of the bacteriophages themselves. These viruses were discovered independently twice by Frederick Twort

in 1915 (Twort, 1915) and Félix d'Hérelle in 1917 (d'Herelle, 1917). d'Hérelle conducted some experiments in humans to control the spreading of cholera with some success. Nevertheless, the main reason why phage therapy was not accepted into the therapeutics was the lack of rigorous evidence to support proof of their efficacy along with the poor knowledge about the phage biology at that time (Sulakvelidze et al., 2001). A lack of acceptation of viruses as a cure, on top of the availability of antibiotics in the 1930s foreshadowed the sunset for the phage therapy for a long time.

Nowadays, there is a renewed interest in phage therapy, mainly because of the gloomy predictions about multidrug resistance in bacteria, which coincides with the drying antibiotic pipelines. Undoubtedly, a better understanding of the phage biology, diversity and mechanisms of infection at a molecular level is fundamental for the use of phages in the clinic or in the industry.

1.1.1. Phage host interactions

Over the course of their evolution, bacteriophages have gained mechanisms to co-opt the bacterial metabolism, redirecting metabolic pathways, biosynthetic machineries, and energy sources towards the production or maintenance of the viral progeny through effectors, mainly small proteins (Roucourt and Lavigne, 2009).

Exploring the bacteriophage effectors that subvert host cellular functions uncovers fundamental mechanisms of viral infection and can help identify critical targets of bacteriophages and thus guide the design of new antimicrobial molecules (Liu et al., 2004). Few host molecules are recognized by antibiotics, therefore, identifying new mechanisms capable of steering or altering sensitive cellular targets become important to fight infectious diseases (Molshanski-Mor et al., 2014).

This chapter provides a brief overview of the bacterial process targeted by bacteriophage effectors during the takeover process, with an emphasis on *Escherichia coli* bacteriophages, but paying attention to other models *en passant*. We classify the phage host interactions in those interfering with bacterial housekeeping processes or bacterial defense systems (Figure 2). Phage effectors disrupting bacterial housekeeping metabolism are poisonous for the host by themselves, in contrast, phage effectors interfering bacterial defenses do not tend to be toxic.

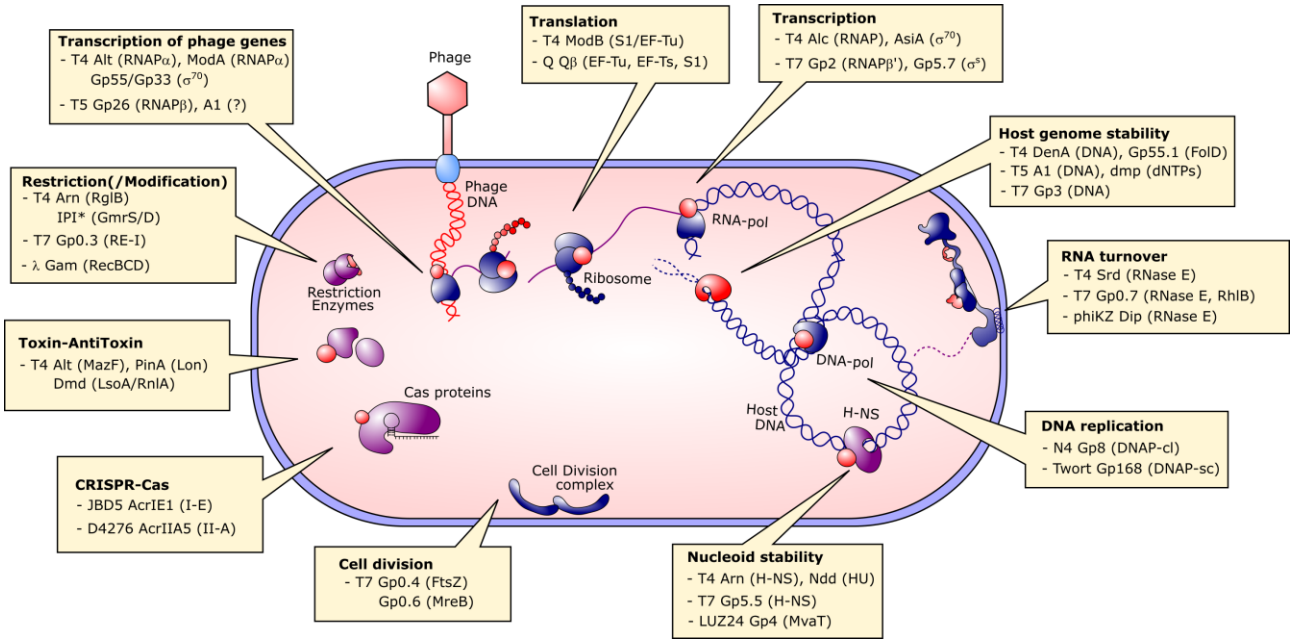


Figure 2. Phage bacteria interaction during the host takeover.
Bacteriophage effectors target bacterial molecules during the host takeover process. Phage proteins (red) target host defense systems (violet) or housekeeping (blue) bacterial proteins. Boxes include the bacteriophage, protein (and target). Phages infecting *Escherichia coli* (λ, T4, T5, T7, N4, Q), *Pseudomonas aeruginosa* (LUZ24, JBD5, phiKZ), *Streptococcus* spp. (D4276), *Staphylococcus* (Twort) were included. The boxes are explained further in this chapter, following a clockwise order.

1.1.2. Bacteriophages alter bacterial housekeeping processes

1.1.2.1. Targeting the host translation

As bacteriophages cannot synthesize proteins outside of their host cell, the control of the bacterial translation machinery is crucial for the success of bacteriophage multiplication. Translation takes place in three steps: i) initiation, promoted by proteins IF-1, IF-2, and IF-3, involving the interaction between the mRNA Shine-Dalgarno (SD) and the 16s rRNA anti-SD sequences within the small ribosomal 30s subunit. ii) elongation, promoted by EF-Tu, EF-Ts, EF-G, and EF-S after the assembly of the mRNA plus ribosomal subunits 30s and 50s composed by proteins S1-S21 and L1-L36, respectively. iii) termination, in which release factors 1 (RF-1) or 2 (RF-2) recognize the stop codon and terminate the elongation, and then, RF-3 frees the ribosome (Kaczanowska and Rydén-Aulin, 2007; Shajani et al., 2011). Bacteriophages target these factors to redirect their specificity (e.g., phages T4 and T7) or changing their activity (e.g., phage Q).

T4 ModB is an ADP ribosyl transferase targeting S1 (Tiemann et al., 1999) and EF-Tu (Depping et al., 2005; Roucourt and Lavigne, 2009). It was suggested that ModB is lethal when recombinantly produced in *E. coli* (Tiemann et al., 1999). ADP-Ribosylation of S1 and EF-Tu might improve the translation of phage mRNA in detriment of those from the host, but the actual function remains unknown (Miller et al., 2003).

T7 Gp0.7 shuts down the translation of host mRNAs. This phage dispensable kinase phosphorylates EF-G and S6 albeit it is not clear whether it is directly responsible for such inhibition (Häuser et al., 2012).

The Q β replicase interacts with the ribosomal protein S1 and elongation factor EF-Tu and EF-Ts to form an RNA-dependent RNA polymerase. S1 acts as a termination factor that permits the rapid release of ssRNA products of the replication and is dispensable for the function of the replicase. On the contrary, EF-Tu and EF-Ts are essential for the replicase function although their role remains unknown (Chetverin et al., 2019).

Translation can also be switched off by host toxin-antitoxin systems and this aspect will be discussed in the section 1.1.3.2 along with bacterial defense mechanisms.

1.1.2.2. Subversion of the host transcription

Many bacteriophages identified so far do not encode their own RNA polymerase and target the bacterial DNA dependent RNA polymerase (RNAP), to control and regulate the expression of viral genes. RNAP is a multimeric protein, composed in *E. coli* of five polypeptides: two subunits α, and one of each β, β', and ω. They constitute the core enzyme with the catalytic activity of RNA polymerization but are incapable of recognizing promoters. Promoter recognition is mediated by sigma factors. One sigma factor plus the core enzyme constitutes the holoenzyme. (Ishihama, 2000; Minakhin and Severinov, 2005). In phages that rely upon the host RNAP, specific phage effectors redirect transcription using different strategies:

- By chemical modifications. ADP-ribosylation is the only covalent modification described thus far at this level. Chemical modifications are intended to redirect the preference of transcription by the core RNAP towards the phage DNA. Some examples are: Alt and ModA (Nechaev et al., 2002), both dispensable proteins in T4 (Goff and Setzer, 1980). It was suggested that ModA is lethal when recombinantly produced in *E. coli* (Tiemann et al., 1999).
- By the emulation of sigma factors: Some phage proteins bind to the RNAP to either redirect the specificity of the holoenzyme or to completely inhibit the polymerase activity. For example T4 gp55 that initiates the transcription from late promoters (Nechaev and Severinov, 2003; Roucourt and Lavigne, 2009). Instances of the RNAP inhibitors are the dispensable T7 protein gp5.7 (Tabib-Salazar et al., 2017) and the essential protein gp2 (Mekler et al., 2011), which is also found in T7-like viruses, e.g., ϕKMV-like phages of *Pseudomonas* (Klimuk et al., 2013) or P7 in *Xanthomonas oryzae* phage Xp10 (Nechaev et

al., 2002). The production of T7 Gp2 is deleterious for *E. coli* but confers escapers an enhanced ability to tolerate diverse types of antibacterial stress, for instance, by raising the minimum inhibitory concentration of gentamicin. These effects result from the binding of Gp2 to the RNAP and impairing the σ^{70} activity but not that of the other sigma factors, such as σ^{38}, which is required to establish transcriptional adaptative changes that lead to adaptation to stress (Sarkar et al., 2017).

- By emulating Anti-sigma factors: In *E. coli*, there are seven species of sigma factors that compete for the RNAP to modulate the transcription of specific bacterial genes. Many eubacteria also produce anti-sigma factors that regulate the core enzyme by binding to specific sigma factors. Phage effectors imitate anti-sigma factors to manipulate the RNAP, especially those relying upon the host RNA polymerase. One example is T4 AsiA, which binds to σ^{70}, the housekeeping sigma factor and the one with the highest affinity to the core enzyme. AsiA binds to sigma 70 to lower its affinity for the core enzyme so other phage sigma factors can outcompete σ^{70}, such as gp55 and gp33 (Minakhin and Severinov, 2005; Sharma and Chatterji, 2008). So far, σ^{70} is the only host sigma factor known to be targeted by T4 during infection (Nechaev and Severinov, 2003).
- By emulating terminators: The RNAP core can terminate the transcription process at defined sequences, resulting in an intrinsic termination. The RNAP core can also be forced by certain proteins, namely Rho or MFD, to terminate or interrupt the transcription if needed (Nudler and Gottesman, 2002). T4 Alc holds a similar activity to Rho and MFD: it selectively interrupts transcription of host DNA (unmodified) rather than that of phage DNA (with methylated cytosines) (Kashlev et al., 1993).

The inhibition of the RNA polymerase depends on the presence of a phage RNA polymerase, in phages producing their own enzyme, the host RNAP is generally inhibited, as it occurs in T7. RNAP-recognizing phage effectors can be identified by copurification during infection (Klimuk et al., 2020), pull-down, or yeast two-hybrid systems (Wan et al., 2021).

1.1.2.3. RNA Turnover

Post-transcriptional regulation is one of the most important ways in which gene expression is controlled. It involves the regulation of structure and stability of cellular transcripts. (Dendooven and Lavigne, 2019; Tejada-Arranz et al., 2020). RNases are main enzymes in the post-transcriptional regulation, some of them act in multiproteic complexes called RNA degradosomes, which core components are at least one RNase and an RNA helicase of the DEAD-box family (Tejada-Arranz et al., 2020).

In *E. coli* the RNA degradosome is a complex composed by the RNase E, the helicase RhlB, the phosphorolytic 3'-5' exoribonuclease PNPase, and the glycolytic enzyme enolase. The RNse E contains two domains: an N-terminal catalytic domain and a C-terminal unstructured one. The C-terminal domain carry short linear motifs and serves as the scaffold for the other three enzymes of the complex (Tejada-Arranz et al., 2020).

Bacteriophage effectors target the degradosome to either inhibit the turnover of transcripts or to promote the cleavage of host RNAs. To this date, three phage effectors were described to target the bacterial degradosome: Dip of *Pseudomonas* phage phiKZ (Van den Bossche et al., 2016) and Gp0.7 of phage T7 (Marchand et al., 2008), both inhibit turnover of nascent transcripts to ensure the infection (Dendooven and Lavigne, 2019), and Srd of phage T4 promotes the cleavage of host transcripts by RNase E early on in the infection (Qi et al., 2015).

1.1.2.4. Inhibition of the host DNA replication

DNA replication is carried out by DNA dependent DNA polymerases, among which those from *Escherichia coli* were the most studied. *E. coli* codes for five different DNA polymerases, the chromosomal replicase is the DNA Polymerase III (DNAP-III). DNAP-III functions are integrated in a multiprotein complex called the DNAP-III Holoenzyme that contains ten proteins grouped functionally in the core, the β sliding clamp (DNAP-sc), and the γ/τ complex clamp loader (DNAP-cl) (O'Donnell, 2006). Bacteriophages target DNA replication through diverse strategies:

- By hindering the assembly of DNAP-III Holoenzyme: Twort Gp168 and G1 Gp240 bind to the DNAP-sc impairing their integration into the holoenzyme, which reduces the processivity of the *Staphylococcus aureus* DNAP (Belley et al., 2006).
- By impairing the Holoenzyme processivity: Coliphage N4 gene product Gp8 binds to DNAP-cl thus inhibiting the elongation process during the host DNA replication. Gp8 is not essential for the development of N4 in *E. coli* but its absence lowers the phage burst size (Yano and Rothman-Denes, 2011).

- By targeting DNA topology or the enzymes responsible for it. Thus, phage phiX174 protein A* shuts off DNA replication in *E. coli* probably by impairing DNA unwinding (Eisenberg and Ascarelli, 1981; Martin and Godson, 1975). While the phage 77 gene *104* protein inhibits the *Staphylococcus aureus* helicase loader DnaI (Liu et al., 2004).

In all these cases, the ectopic production of the phage proteins leads to the inhibition of the bacterial growth attesting deleterious effects of these bacteriophage factors.

1.1.2.5. Targeting the host DNA stability

Apart from DNA replication, the host DNA molecule itself is one privileged target that can be attacked by bacteriophage effectors, either directly or indirectly. Some phages degrade the host DNA to reuse the breakdown products, which is essential for the infection, while in other phages capable of *de novo* nucleotide synthesis, the host DNA degradation is a dispensable process for infection.

Direct targeting implies host DNA degradation. Phages T7 and T3 rely entirely on the host DNA degradation products for their own DNA replication, and this process is essential. Host DNA breakdown by T7 takes place in two steps, the Endonuclease I (Gp3) produces fragments of host DNA that are fully digested by the Exonuclease (Gp6) (Krüger and Schroeder, 1981). At the early stage of phage T5 infection, host DNA is massively and rapidly degraded, under the control of at least the A1 protein (Lanni, 1969). However, T5 does not fully need the pool of host DNA breakdown products for its DNA replication. The dispensable phosphatase T5 Dmp dephosphorylates dNMPs, which are then exported outside the host cell by an unknown mechanism (Mozer et al., 1977). In T4, DenA (endonuclease II) is a dispensable exonuclease that digests only dC-containing DNA (Miller et al., 2003).

Indirect targeting implies blocking some steps in DNA metabolism. The ectopic production of the T4 dispensable protein Gp55.1 in *E. coli* leads to an increased sensitivity to UV radiation. The mechanisms was not studied in detail but Gp55.1 interacts with UvrA and UvrB, two enzymes involved in the DNA nucleotide excision repair (NER) pathway; moreover, Gp55.1 binds to FolD, a bifunctional enzyme involved in the folate metabolism, leading to lower trimethoprim minimal inhibitory concentrations (Mattenberger et al., 2011).

Taken together, bacteriophage proteins may target the host DNA directly by degrading it, inhibiting its replication, or even its repair, but they can also alter the bacterial nucleoid.

1.1.2.6. Alteration of the nucleoid organization

The chromosome of *E. coli* occupies a quarter of the rod's volume, it is highly compacted and at the same time dynamically accessible. In fact, this equilibrium is important for the regulation of replication, chromosome segregation, and cell division, while keeping the entire chromosome available for transcription (Dame, 2005). Nucleoid-associated proteins are numerous and their abundance depends on the growth phase: HU, Lrp, and MukB are found in both Gram-negative (GN) and Gram-positive (GP) bacteria; while H-NS, IHF, and Fis are present only in GN bacteria (Dillon and Dorman, 2010). These proteins represent potential targets for phage effectors.

HU in *E. coli* is targeted by Ndd (D2b) during infection by T4. Ndd has a higher affinity to the *E. coli* DNA than to the T4 DNA and induces a disruption of the nucleoid (Bouet et al., 1998; Roucourt and Lavigne, 2009).

P. aeruginosa MvaT is an H-NS-like protein by Gp4 during phage LUZ24 infection. MvaT is a transcriptional repressor that binds selectively to AT-rich DNA regions and controls motility, virulence and pathogenicity of *P. aeruginosa*. MvaT is also a transcriptional repressor that harnesses the GC-content difference between *P. aeruginosa* genome and foreign DNA to repress the expression of the latter (Wagemans et al., 2015). Other examples include T7 gp5.5 (Liu and Richardson, 1993) and T4 Arn (Ho et al., 2014), both inhibitors of the DNA binding activity of H-NS in *E. coli*.

Phage T4 *55.2* is a dispensable gene whose product alters the supercoiling regulation of *E. coli* DNA by inhibiting the Topoisomerase I. Gp55.2 is toxic for *E. coli* only at high concentrations. T4 gene *55.2* inactivation leads to a delay in the latency of 5 min and a concomitantly important fitness cost (Mattenberger et al., 2015).

The nucleoid stability is, among the other aforementioned bacterial processes, needed for cellular homeostasis as well as for the cell division. Unsurprisingly, cell division is also targeted during the host takeover by some bacteriophages.

1.1.2.7. Impairing host cell division

Bacteria follow three types of size-control strategies: sizers, timers, or adders (Willis and Huang, 2017). In sizer regulation, cells progress through the cell cycle when a critical size is reached. In timer regulation, a fixed time separate two events of division. In adder size regulation, cells divide after the addition of a critical size to the birth size. Examples of adders are *E. coli*, *P. aeruginosa*, and *Bacillus subtilis*. In *E. coli*, the size is regulated according to the volume, i.e., division takes place when a critical volume is added to the parental rod (Willis and Huang, 2017).

Bacterial size and shape are determined mainly by the actin-like MreB that forms helicoids across the cells. MreB helicoids coordinate bacterial elongation (mainly in rod-shaped bacilli) and serves as the scaffold for the large multi-protein peptidoglycan synthesizing holoenzyme (White and Gober, 2012), for the tubulin–like FtsZ responsible for the formation of the division septum (Margolin, 2009), and for the initiation regulator DnaA (Hansen and Atlung, 2018).

E. coli cells depleted in MreB fail to elongate and acquire irregularly spherical shapes, while FstZ-depleted cells form large filaments and cannot divide (Margolin, 2009). Viral factors altering the size and shape of *E. coli* have been identified in T7. T7 Gp0.6 produced in *E. coli* binds to MreB and reshape the cells from rod to lemon–shaped bacteria (Molshanski-Mor et al., 2014), while T7 Gp0.4 binds to FtsZ inhibiting its polymerization and leading to a filamentous and elongated cellular morphology (Kiro et al., 2013). In both cases, the mechanism remains unknown.

Bacteriophages target cell homeostasis to redirect the energy and infrastructure of the host towards the synthesis and assembly of new virions. In most cases these effectors excerpt some effect on the host when recombinantly produced in it. In the next section we focus on the different strategies that phages rely on to ensure the host takeover.

1.1.3. Targeting the bacterial defense systems

In the evolutionary arms race, bacteria have evolved many defense systems against phages, and the viruses have in turn acquired effectors to inactivate these systems. In this section we will present some of the bacterial defense systems that might be relevant to the study of bacteriophage effectors. This section does not pretend to be exhaustive since many new bacterial defense systems have been recently discovered but their precise mechanisms of action have not been uncovered yet. A summary of anti-phage bacterial systems has been published in recent reviews (Millman et al., 2020a, 2020b; Rostøl and Marraffini, 2019).

1.1.3.1. Inactivation of the CRISPR-Cas systems

CRISPR-Cas systems function as the adaptive immune system of bacteria by protecting them from potentially threatening foreign mobile genetic elements, such as phages or plasmids. The protection by CRISPR-Cas comprises three steps: Adaptation, expression, and interference. During the adaptation step, a set of CRISPR associated (Cas) proteins recognizes the foreign DNA/RNA, in some cases, by recognizing a short sequence called Protospacer Adjacent Motif (PAM), cleaving the contiguous sequence, or protospacer, and integrating a DNA copy in the genome, now called spacer. Spacers make part of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) in the genome and build the memory of the cell (McGinn and Marraffini, 2019).

During the expression step, the CRISPR array is transcribed as a single transcript and processed to yield multiple mature CRISPR RNAs (crRNA), which sequences include those of the spacers. These crRNAs remain anchored to the complex of Cas proteins that processed the original transcript. Sometimes, an accompanying RNA helps the crRNA to remain bound to the Cas complex, a transactivating CRISPR RNA (tracrRNA) (Makarova et al., 2020).

During the interference step, the crRNA guides the Cas proteins to cleave the protospacer or a similar sequence in new foreign DNA or RNA (Makarova et al., 2020).

To this date, the classification of CRISPR-Cas systems includes 2 classes (Class 1 and Class 2), 6 types (Types I, III, IV inside the Class 1, and Type II, V, VI inside the Class 2), and 33 subtypes. Class 1 and 2 differs in the expression and interference steps. Class 1 CRISPR-Cas systems are made up by complexes of at least three different Cas proteins responsible for transcript processing and crRNA binding to target the protospacer. Class 2 systems include a multiple-domain single Cas protein that processes the transcript and binds the crRNA to target the protospacer (Makarova et al., 2020).

Bacteriophages target CRISPR-Cas systems through Anti-CRISPR (Acr) proteins. Inhibitors of CRISPR-Cas type I, II, and V were the most studied and the structure of 16 Acrs was solved, however,

none of them is similar to the others. Acr proteins are classified based on the CRISPR-Cas system they inhibit, for instance, AcrIF1$_{Pae}$ inhibits the system type I-F and was found in a *Pseudomonas aeruginosa* genetic element (including phages). Despite this structural diversity, two common mechanisms stand out: DNA mimicry and blockage of DNA cleavage (Davidson et al., 2020). For instance, AcrIIA5 keeps *Streptococcus* spp. Cas9 (class 2, type II-A5) from recognizing the DNA, while AcrIE1 directly inhibits *P. aeruginosa* Cas3 (class 1, type I-E1) thus preventing a target cleavage. A third and unique case described yet, involves the acetylation of Cas12a (type V) from *Moraxella bovoculi* by AcrVA5 (Dong et al., 2019).

1.1.3.2. Toxin-AntiToxin systems

Protein pairs acting as toxin-antitoxin (TA) systems are composed of a toxin that causes the growth arrest by hindering a vital bacterial process, and a cognate antitoxin that neutralizes the effects of the toxin under normal growth conditions. These systems are involved in bacterial persistence under stressful conditions. There are six types of TA systems: In the type I, the antitoxin is a mRNA that anneal to the toxin mRNA and blocks its translation, which otherwise results in hydrophobic peptides that cause loss of membrane potential. In the type II, the antitoxin is a protein that binds to the DNA and the toxin at the same time; stressing conditions lead to the selective proteolysis of the antitoxin and disinhibition of the toxin, which halts replication or translation. In type III, the antitoxin is a small RNA that binds to and blocks the toxin activity. In type IV, the antitoxin indirectly antagonizes the toxin by stabilizing a common target, for instance MreB or FstZ. In type V, the antitoxin degrades the mRNA of the toxin, a protein able to damage the cell membrane. In type VI, the antitoxin promotes the degradation of the toxin, which otherwise stabilized blocks DNA replication (Page and Peti, 2016). TA systems helps to counteract phage infection by favouring suicide and abortive infection.

Phages effectors inhibit the TA system mainly by targeting and inactivating the activity of the toxin. For instance, MazF/MazE pair is a type II ribosome-independent TA in *E. coli* in which MazF is a ribonuclease (Zhang et al., 2003). During the infection by T4, Alt is delivered into the host along with the phage DNA and inhibits MazF through ADP-ribosylation (Alawneh et al., 2016). Another example is Lon, a protease that preferentially cleaves antitoxins and incorrectly folded proteins (Page and Peti, 2016), whose ATP-dependent activity is blocked by T4 PinA protein (Hilliard et al., 1998). Also the type II toxins LsoA and RnlA are inhibited by T4 Dmd (Otsuka and Yonesaki, 2012). In all these cases, the inhibition of the host toxin leads to a better phage fitness.

1.1.3.3. Counteracting the Restriction Modification system

Restriction-Modification systems are a common mechanism of bacterial defense against phage infection by cleaving newly acquired DNA while the host DNA is modified to protect it from cleavage (Samson et al., 2013). It was proposed that RM systems are akin to TA systems: the restriction enzyme acts as the toxin while the modification enzyme plays the role of an antitoxin (Mruk and Kobayashi, 2014). Here we consider RM system separately regarding their importance as anti-phage bacterial defenses.

There are four types of RM systems. Type I RM systems are protein complexes that contain two R sub-units, two M sub-units, and one S sub-unit. The S sub-unit determines the specificity for the targeted sequence, while the R (restriction) cleaves the DNA that was not methylated previously by the M sub-unit. EcoKI is a type I enzyme (Roberts, 2003).

Type II RM systems cleave at their cognate DNA sequence or outside of it at a defined distance. The restriction and modification enzymes are two independent proteins with the same target. EcoRI is a type IIP enzyme. Type III systems are composed by enzymes that recognize and modify (Mod) or cleaves (Res) the DNA. Mod and Res are both required to cleave the DNA (Roberts, 2003). Type IV systems are composed by proteins that cleave only modified DNA (Roberts, 2003).

Phage effectors block RM systems activity actively or passively. In active strategies, phage proteins inhibit restriction enzymes directly. One example is the T7 protein Ocr (needed to overcome classical restriction), encoded by the early gene *0.3*, that mimics the B form of DNA (Walkinshaw et al., 2002) to competitively inhibit the Type-I restriction enzymes EcoB and EcoK (Bandyopadhyay et al., 1985). In T4, IPI* is a nuclease inhibitor injected along with the DNA that inhibits GmrSD (glucose-modified hydroxymethylcytosine restriction endonuclease) (Rifat et al., 2008)

In passive mechanisms, phages protect their genetic material against bacterial nucleases by chemically modifying the genome or blocking the access to it. T4 and other T-even phages bear hydroxy-methylated or glycosylated bases in their DNA, which evades the cleavage of REases (Samson et al., 2013), and by the way render CRISPR-Cas9 less effective on T4 DNA as well (Tao et al., 2017). *Pseudomonas* phage phiKZ and *Serratia* phage PCH45 (a phiKZ-like phage) DNA are not reached by REases and CRISPR-Cas systems because they form nucleus-like structures around the phage DNA (Malone et al., 2020; Mendoza et al., 2020).

A new system was recently found in *Escherichia coli*. It is a ssDNA phosphorothioate (PT) modification system (Ssp) that provides protection against phages through sequence-specific PT sensing and by nicking the non-modified DNA to restrict phage DNA replication. The Ssp system is

composed by SspABCD-SspE where SspA might be a cysteine desulfurase, SspB is a nickase, SspC an ATPase, SspD an ATP pyrophosphatase, and SspE a PT-sensing NTPase and DNA-nicking endonuclease. When *E. coli* expresses a heterologous *sspABCD-sspE* system it becomes resistant to bacteriophages T4, T1, JMPW1, JMPW2, EEP and T7, but not bacteriophage T5 (Xiong et al., 2020). It remains unknown the mechanism by which T5 escape the Ssp restriction.

Other systems we have not mentioned include BREX (Goldfarb et al., 2015), DISARM (Ofir et al., 2018), and an entirely chemical protection of *Streptomyces* (Kronheim et al., 2018). Future work is likely to uncover new phage effectors capable to inhibit these defense systems as well.

In the next section we will present the model we have used in this thesis: bacteriophage T5.

1.2. Bacteriophage T5

Bacteriophage T5 (NCBI:txid2695836) is the representative microorganism of the genus *Tequintavirus* (NCBI:txid187218), belonging to the subfamily *Markadamsvirinae* (NCBI:txid2732013), family *Demerecviridae* (NCBI:txid2731690) and order *Caudovirales*. T5 is a virulent phage that infects *Escherichia coli*. It should be noted that, most studies done on T5 after 1958 were carried out in *Escherichia coli* F, isolated by M. H. Adams (Lanni, 1958) since this virus shows the highest adsorption rates onto this bacterial strain (Heller and Braun, 1979) which facilitates the study of their interaction. Accordingly, we have also used *E. coli* F in this work whenever possible.

The T5 capsid has an average diameter of 930 Å and is connected to a long, non-contractile tail of 120 Å in diameter and 2,500 Å in length (Effantin et al., 2006). Its genome is constituted by a linear 121,752-bp unmodified dsDNA, encompassing single-strand interruptions or "nicks" at genetically defined positions and present only on the minus strand (Abelson and Thomas, 1966; Bujard, 1969; Wang et al., 2005) and a total GC proportion of 39.27%. There is however one report in which ssRNA was detected from extraction of total nucleic acids from the virus (Rosenkranz, 1973) but these results have never been confirmed in another study.

Sequences of three genomes of T5 are found in the GenBank database: T5 (NC_005859 / AY543070), T5 ATCC 11303-B5 (AY587007) (Wang et al., 2005) and T5 strain st0 (AY692264), all submitted in 2004. The first two are wild-type strains, while the third one is a deletion mutant lacking an 8.5-kb locus (Scheible and Rhoades, 1975) that codes for tRNAs. We have used T5 (NC_005859 / AY543070) genome as a reference throughout this report.

1.2.1. Infectious process

The bacteriophage T5 adsorption to the host surface of *E. coli* is mediated by two types of interactions: a reversible interaction succeeded by an irreversible one. First, the distal tip of the long L-shaped tail fiber (formed by the protein pb1) reversibly binds to the O8- or O9-type O-antigen of the bacterial lipopolysaccharide (LPS) of certain *E. coli* strains (Heller and Braun, 1982). This primary interaction is auxiliary but increases 15 times the adsorption rate onto the bacterial surface and allows the phage to "walk" on the outer membrane, until it recognizes and binds irreversibly to its receptor, the iron transporter FhuA (Heller and Braun, 1979). This irreversible interaction is mediated by the protein pb5 (Mondigler et al., 1996), located at the end of the central fiber (Zivanovic et al., 2014), and leads to the delivery of the phage DNA into the host cell. Although the molecular mechanisms by which T5 perforates the bacterial envelope to channel its DNA into the host remains unknown, biochemical and structural studies have shown that the irreversible binding of T5 to FhuA triggers a

structural reorganization of the tail tip resulting in the ejection of tail the tape-measure protein (Arnaud et al., 2017; Zivanovic et al., 2014), the opening of the capsid connector and subsequent DNA release.

The DNA of T5 is delivered into the host in two steps: First, part of the left terminal repeat (LTR), ~8% of the genome, is injected (Shaw and Davison, 1979) and the injection stops (Figure 3, left). During this pause, the so-called pre-early genes are rapidly expressed and their products inhibit bacterial nucleases (Davison and Brunel, 1979a; Sakaki, 1974), degrade the bacterial DNA (Lanni, 1969), dephosphorylate monophosphate nucleotides for them to be excreted as nucleosides (Mozer et al., 1977) and resume the DNA delivery after the first 4-5 min post-infection (Lanni, 1969). The second step transfer (SST) permits the entire phage DNA to enters the cell (Figure 3, right). Early genes (corresponding to middle genes in other phages) and late genes are then expressed, and their products are detected 5 and 10 min post-infection, respectively, leading to viral DNA replication and synthesis of the structural proteins. The assembly of procapsids and packaging of the DNA occurs in parallel with the assembly of the tails and these two sub-particles are connected to form infectious particles that are released upon lysis the cell.

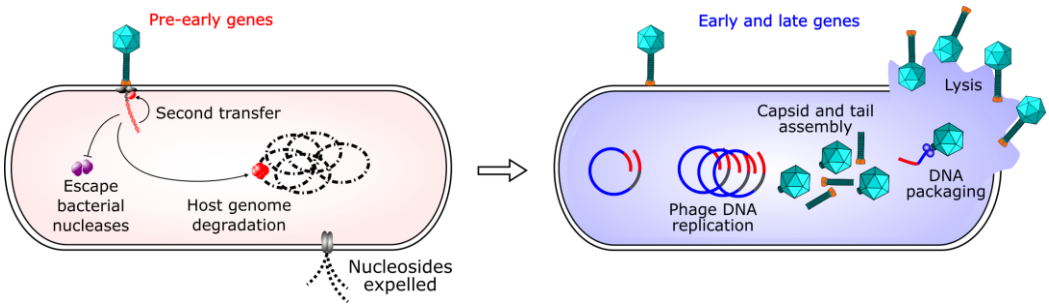


Figure 3: Steps of the infection of *E. coli* by bacteriophage T5.
T5 delivers 8% of its DNA during the first-step transfer (FST) and stops. The encoded pre-early proteins degrade the host DNA, inhibit nucleases, and resume the transfer. The second-step transfer permits the entrance of the entire DNA in the cell. Pre-early proteins are encoded in the left and right-terminal repeat (LTR and RTR, respectively, in red), early and late proteins are encoded in the rest of the genome (blue).

An optimal infection depends on the availability of calcium ions and good aeration (Adams, 1949). The DNA delivery is independent of calcium (Lanni, 1960), however, calcium concentrations below 0.1 mM impair the closing of the transiently opened channel during DNA delivery, which affects the host metabolic state and results in halting phage replication (Bonhivers and Letellier, 1995). On the other hand, accessibility to this ion might help synchronize the cell lysis by the endolysin (gene *lys*) (Kovalenko et al., 2019; Mikoulinskaia et al., 2009).

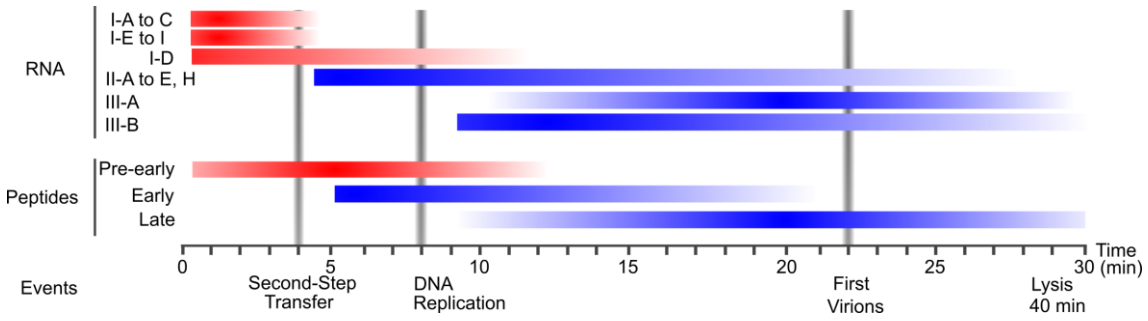


Figure 4: Major events during the T5 infection.
Pre-early proteins, transcripts and peptides/proteins were highlighted in red. Early and late transcripts as well as peptides were highlighted in blue. Adapted from (Beckman et al., 1971; McCorquodale and Buchanan, 1968; Sirbasku and Buchanan, 1970a, 1970b).

At 37 °C, in presence of calcium ions and a good aeration, pre-early proteins are synthesized during the 4 to 5-min pause following the FST (Figure 4). After resuming of DNA delivery or SST, the early proteins are produced for up to 20 minutes, while the viral DNA starts being replicated at 8 minutes post-infection. Finally, late protein synthesis begins at 12 minutes and continues until lysis, at 35 to 50 minutes. The first virions are assembled as early as 22 minutes. The infection yields burst sizes between 200 to 500 progeny per infected cell (McCorquodale and Warner, 1988). Ribonucleic acid production precedes that of proteins (Sirbasku and Buchanan, 1970a). Pre-early genes code for class-I transcripts/peptides, early genes for class II products, while late genes code for class III products (Figure 4).

1.2.2. Genetic map of the First-Step Transfer DNA

The FST-DNA corresponds to the left terminal repeat (LTR). In T5wt (NC_005859) this locus contains 17 annotated open reading frames, two of them (*A1* and *A2*) were identified as essential (Lanni, 1969), one as dispensable (dmp, (Mozer et al., 1977)) and 14 remain putative (Figure 5). Four promoters are included in the FST locus: two of them were experimentally characterized as strong promoters (P-D/E 20, P-H 22) (Gentz and Bujard, 1985) whereas the other two are putative (P_{hegG}, P_{11}). There are two putative intrinsic terminators: one between genes *11* and *12*, the other downstream *17*. The pre-early genes are all in the FST-DNA (Hayward and Smith, 1973) (Figure 5). The FST-DNA is separated from the rest of the genome by a 1.5-Kb non-coding region downstream gene *17*.

Genes *dmp* to *hegG* are oriented leftward, probably under the control of one of three promoters, while genes *11* to *17* might be transcribed rightward, probably under the control of the putative promoter P_{11}. However, there is so far no experimental evidence to support any specific operon layout.

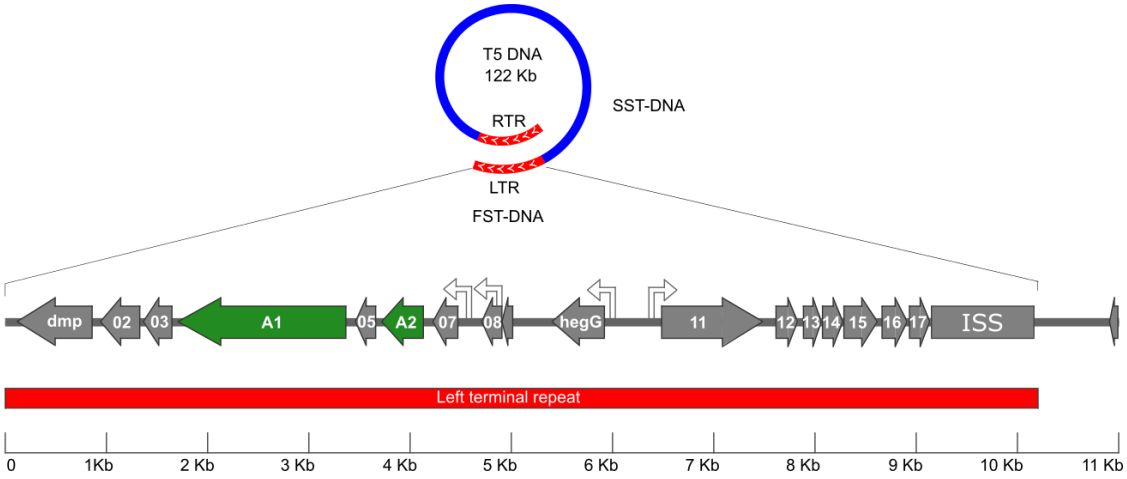


Figure 5: Genetic map of the Phage T5 LTR (FST-DNA).
The LTR (red) enters the first during DNA delivery. The injection stops between 9 and 11 kb (ISS). Essential genes (green) are needed for host DNA degradation and the second step transfer. Adapted from T5 annotation (NC_005859).

Most *Tequintavirus* phage genomes bear the same gene arrangements in the LTR, all of them bear orthologous to T5 genes *A1*, *A2*, *dmp* and *07* in the same orientation. In almost all cases there are two sets of genes headed to opposite directions neighbored by an intergenic region of 1.5-2.0 kb, part of which is included in the left terminal repeat. In some phages, mainly unclassified *Tequintavirus*, the group of genes orthologous to T5 *11* to T5 *17* is poorly conserved (Figure 6), as in chee130_1 (MF431736.1), or even lost as in phagemcphageface (MT074447.1). Orthologous to the T5 gene *A1* is present but was not annotated in SE3 (MK770410.1) because it is truncated: part of *A1* is at the beginning of the genome and the other part at the end. The 1.5-Kb non-coding region between genes *17* and *18* in T5 is absent in the phage genomes of vB_SenS_SB10 (MK947458.1) and vB_SenS_SB13 (MK947459.1).

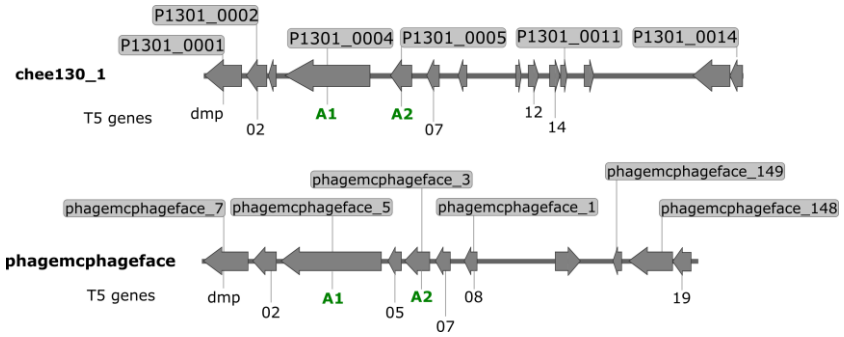


Figure 6: Annotations at the LTR of some T5 viruses.
Phage chee130_1 (MF431736.1) and phagemcphageface (MT074447.1) genes and their homologues in T5 (NC_005859). Labels above the genes refer to the phage gene nomenclature while the labels underneath correspond to the names of their homologues in T5.

The FST locus is neighbored by a 1.5-Kb intergenic region called the injection stop signal (ISS). The ISS region was proposed to be the locus laying in the interphase between the bacterial cell and the bacteriophage (Heusterspreute et al., 1987) during the first step transfer, i.e., the locus at which

the DNA injection stops. This region locates between genes *17* and *18* (Figure 5). It is made up by 1,777 bp and probably lengths 569 nm (linear DNA-B). The *E. coli* membrane and cell envelope thickness rounds 25 nm. This intergenic region presents some heterogeneities: two thirds is part of the LTR and are rich in direct and inverted repeats, while the remaining one third is not full of such sequences (Davison, 2015; Heusterspreute et al., 1987). A more precise investigation of this region is lacking to characterize this region.

Likewise, no experimental investigation on the transcriptional organization and regulation of FST genes was performed since the mid-1980s. In the next section, we will review the evidence available the evidence about the transcription of the FST locus.

1.2.3. Transcription of the pre-early genes.

Of the four annotated promoters in the FST locus, two of them, P-D/E 20 and P-H 22, were studied in the past and showed a strong activity *in vitro* and *in vivo* (Gentz and Bujard, 1985). The other two, P_{hegG} and P_{11}, were never characterized. Two transcriptional terminators were predicted at position 7534 and 9275, the first is between genes *11* and *12* and the second is downstream of *17*.

Three classes of transcripts are produced during the T5 infection (Sirbasku and Buchanan, 1970a, 1970b) (Figure 4): class-I transcripts are synthetized immediately at the onset of the infection and most are detectable until 4 min post-infection, before the Second-step transfer (SST). Class-II transcripts appear between 4 and 5 min and remain detectable for up to 36 min. Class-III transcripts are produced from 9 min until the end of the viral cycle (45-60 min) (Sirbasku and Buchanan, 1970a, 1970b). These reports indicate that class-I transcripts come from the FST-DNA and lead to the production of pre-early proteins.

Table 1: Pre-early transcripts.

Weight (Da)	Length (Kb)
1.0×10^6	3.1
6.2×10^5	1.9
3.5×10^5	1.1
2.4×10^5	0.75
2.3×10^5	0.72
1.5×10^5	0.47
8.5×10^4	0.26
7.6×10^4	0.24
7.1×10^4	0.22
6.0×10^4	0.19

Adapted from (Hayward and Smith, 1973; Sirbasku and Buchanan, 1970a, 1970b). Molecular weight (Da) was converted to size (Kb) with the formula:
$Size_{(Kb)} = (MW_{(Da)} - 159)/320.5$.

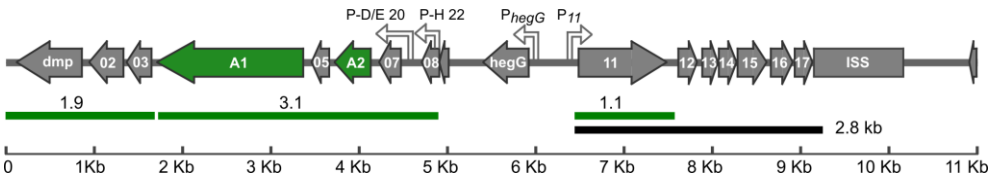


Figure 7. Putative operons at the First-Step Transfer locus.
The distance between annotated promoters and terminators is depicted in black. Some pre-early transcripts depicted in green are mapped arbitrarily to suggest correspondence with the annotated genes. Essential genes are highlighted in green.

Studies in the 1970s detected at least ten pre-early transcripts, twelve early and four late ones. Their molecular weight was determined under denaturing conditions (Hayward and Smith, 1973; Sirbasku and Buchanan, 1970a, 1970b). A comparison between a calculated size of pre-early transcripts (Table 1) and putative operon lengths (Figure 7) suggests that the transcripts might be processed during infection, that not all the annotated pre-early genes are expressed.

The RNA processing during T5 infection is plausible since other phages have been described which also encode processed transcripts. Hence, *E. coli* RNase P or RNase III cleaves *in vivo* transcripts from phages P1 or T7, respectively (Dunn and Studier, 1973; Hartmann et al., 1995). However, it is intriguing that no trace of the unprocessed T5 transcripts could be detected at the time (Sirbasku and Buchanan, 1970a, 1970b)

Predicting which genes are encoded in the pre-early transcripts detected in 1970 is highly speculative (Figure 7). At the minimum, the essential pre-early genes *A1* and *A2* should be expressed and they were suggested to be in the same operon (Davison and Lafontaine, 1984). Moreover, detection of deoxynucleotide monophosphatase activity before the SST also indicated that the *dmp* gene is expressed (Mozer et al., 1977). Finally, an analysis of the proteins synthetized in a cell-free system from the FST locus suggests that genes *11* and *hegG* might also be expressed (Blaisdell and Warner, 1986).

Some studies carried out in the past give some hints about the interaction between pre-early genes and pre-early proteins (Sirbasku and Buchanan, 1970a). We have created a figure that summarizes these finding (Figure 8). When T5 infects chloramphenicol-treated cells, pre-early transcripts (class I) appear immediately at the beginning of the infection, but remain stable for up to 28 minutes, while early and late transcripts (classes II and III) cannot be detected. When bacterial cells are infected by T5 and the infection is interrupted by mechanical shearing, only the FST fragment is

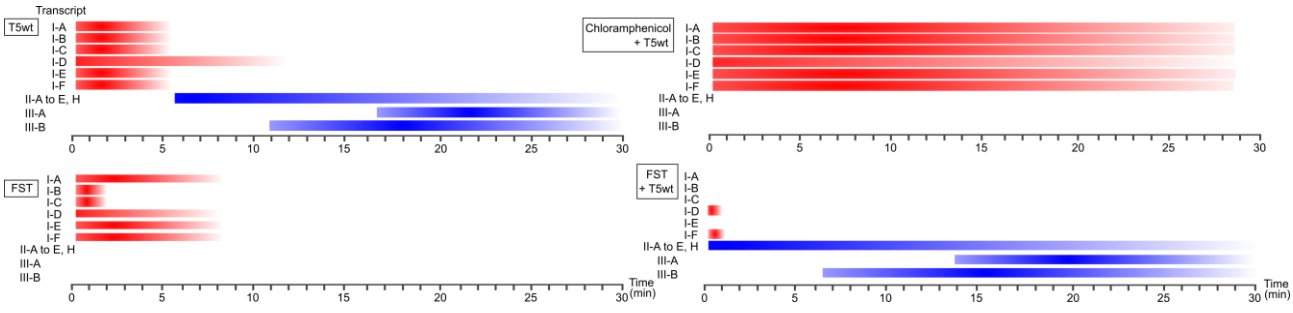


Figure 8. Production of T5 transcripts during infection.
Pre-early (in red), early and late transcripts (in blue) were studied by electrophoresis under different conditions of infection. T5wt, uninterrupted infection by the wild type phage. chloramphenicol+T5wt, cells exposed to chloramphenicol were infected with T5. FST, T5 infection was interrupted by mechanical shearing. FST+T5wt, cells from interrupted infection were infected again with T5. Adapted from (Sirbasku and Buchanan, 1970a).

injected inside the cells. Then, pre-early transcripts are detected for slightly longer intervals than in a normal infection, but early and late transcripts are not detected. If cells first infected only with the FST fragment (treated by mechanical shearing), are re-infected by T5, pre-early transcripts are not detectable, while early and late transcripts are immediately detectable (Figure 8).

In conclusion, the transcription of pre-early genes is independent from the production of pre-early proteins. On the contrary, pre-early proteins are necessary to shut off pre-early gene transcription. Early and late gene transcription depends on the pre-early proteins. Moreover, the production of pre-early proteins before the infection accelerates the expression of the early and late genes (Sirbasku and Buchanan, 1970a). In the next section we present the pre-early genes annotated at the FST locus as well as activities attributed to the pre-early proteins.

1.2.4. Expression of pre-early genes

Information on the putative products of the 17 pre-early genes annotated in the FST locus is presented in Table 2. The first genetic studies on the FST started in the 1960s with the isolation of mutants of T5. These mutants were generated by random chemical mutagenesis using 5-bromodeoxyuridine to primarily obtain amber mutants on pre-early genes (Hendrickson and McCorquodale, 1971a; Lanni, 1969). Genes with amber mutations bear specific termination codons that can be misread as another residue by some bacterial strains known as amber suppressors to avoid premature translation termination. Phages with amber mutations can be amplified in amber suppressors that allow the synthesis of the full-length protein from the mutated gene or they can be confronted with a non-suppressor strain in which a truncated version of the protein will be produced. Among all the strains isolated in the 1960s and 1970s, 20 amber mutations in pre-early genes impaired viral replication and they all mapped to only two genes, *A1* and *A2* (Hendrickson and McCorquodale, 1971a; Lanni, 1969). Biochemical screening of other mutants allowed to isolate a phage strain with a non-sense mutation in the dispensable gene *dmp* (Mozer et al., 1977). Taken together, the genetic analysis performed in the last century defined the putative functions of the pre-early genes *dmp*, *A1*, and *A2*.

Table 2. Putative proteins encoded in the FST-DNA

Protein	MW (KDa)	pI	Putative function
Dmp	28.2	5.27	Phosphatase [a]
Gp2	14.2	5.47	
Gp3	10.7	4.71	
A1	61.5	6.12	Second Transfer [b] Host DNA degradation [b]
Gp5	7.2	9.10	
A2	14.3	8.78	Second transfer [b]
Gp7	9.1	9.05	
Gp8	8.0	10.58	
Gp9	3.9	9.78	
HegG	20.0	9.43	
Gp11	37.0	5.32	
Gp12	8.6	5.61	
Gp13	6.2	7.78	
Gp14	8.1	10.00	
Gp15	13.0	5.82	
Gp16	9.6	9.41	
Gp17	7.7	4.39	

[a] (Mozer et al., 1977), [b] (Lanni, 1969).

T5 amber mutants on *A1* or *A2* fail to transfer their DNA into the host, and mutants on *A1* do not degrade the host genome (Lanni, 1969). Another mutant was obtained, T5 *dmp*$^{-}$, with a non-sense mutation in the gene *dmp* (Mozer et al., 1977). A comparative study of these mutants with the wild type has permitted to assert the function of some pre-early proteins. Pre-early proteins are detectable until 12 minutes post-infection (Figure 4).

Synthesis of pre-early proteins was studied by radioactive labelling of bacteria infected by T5 (Beckman et al., 1971; Duckworth and Dunn, 1976; McCorquodale et al., 1977) or *in vitro* transcription-translation of the FST-DNA (Blaisdell and Warner, 1986). The peptides were then detected by electrophoresis under denaturing or non-denaturing condition. In denaturing experimental conditions, detergents are added to unfold proteins allowing the estimation of the molecular weight, while in non-denaturing conditions proteins keep their native structure and can be purified complexed with their protein or DNA partners. A maximum of nine pre-early peptides were detected (McCorquodale et al., 1977). In Table 3, we sorted the detected pre-early peptides by molecular weight (MW), identity, cite the putative products of annotated pre-early genes with similar MW (MW from Table 2)

Table 3: Pre-early peptides detected during T5 infection.

Native Total [a]	Denatured Total [a]	Denatured Total [b]	Denatured Membranes [c]	Denatured Total (*in vitro*) [d]	Denatured Total 2D [g]	Protein	Similar Molecular Weight
		65		65		A1 [b, d]	A1 (Start at 3376)
244	57	64	59			A1 [b, c]	
364	57 + 15					A1+A2 [a]	A1+A2
				40			
		30			30		Gp11
				27		Dmp [e]	Dmp
18.5	19	23					hegG
	15	15		16	15	A2 [b, d]	A2 / A2+Gp2
18.2	11.4	13		14	14	h, A3 [f]	Gp15
		8.3			10		Gp3 / Gp7 / Gp8 / Gp12
		7.1					Gp14 / Gp16 / Gp17
		5.5					Gp5 / Gp13
		3.3					Gp9

[a] (Beckman et al., 1971), [b] (McCorquodale et al., 1977), [c] (Duckworth and Dunn, 1976).
[d] (Blaisdell and Warner, 1986), [e] (Mozer et al., 1977), [f] (Szabo et al., 1975), [g] (Snyder and Benzinger, 1981).
Fraction used in each study indicates native or denaturing conditions from total cell extracts or membranes.
The identified peptide is indicated with the respective reference. Estimated weights are given in kDa.
Putative pre-early proteins with approximate molecular weight are indicated on the rightmost column.

1.2.4.1. A1: a multi-tasking protein?

Two pre-early peptides corresponding to A1 were identified, with MW of 65-kDa and 64-kDa (Blaisdell and Warner, 1986; McCorquodale et al., 1977). These two forms could result from two start codons separated with 39 bp, or from a proteolytic cleavage. Under non-denaturing conditions, a 244-kDa multimeric protein was recovered, this protein yielded only one peptide after denaturation, A1 (Blaisdell and Warner, 1986). Blaisdell and Warner suggested that A1 might oligomerize *in vivo*. In our laboratory, it was demonstrated that A1 is an octameric DNase (thesis Zangelmi, 2018), while in other study A1 was copurified with the RNAP from T5-infected bacteria (Klimuk et al., 2020).

When *E. coli* F is infected by a mutant T5*amA1*, only the FST-DNA is delivered (Lanni, 1969), the host DNA is not degraded (Lanni, 1969), the host DNA replication is interrupted (Hendrickson and McCorquodale, 1971b; Mozer et al., 1977), but host protein synthesis is not interrupted (McCorquodale and Lanni, 1970).

Altogether, these data strongly suggest that A1 might direct or indirectly have more than one function: host protein shut-off, host DNA degradation, second step transfer.

1.2.4.1. The dispensable phosphatase Dmp

Bacteriophage T5 promote the excretion of DNA breakdown products and re-synthetize most of the nucleotides needed for its replication. Dmp is a deoxynucleoside 5'-monophosphatase (dNMPase) (Warner et al., 1975) that dephosphorylates nucleotides, which are then exported by an unknown mechanism. Indeed, *E. coli* cells infected by T5 *dmp*$^-$ exported around 60 % of the nucleosides from the host DNA and reincorporated the remaining 40% into the newly-synthesized

DNA. Such mutation does not affect neither the host DNA degradation nor the second-step transfer of DNA (Mozer et al., 1977).

1.2.4.2. The essential DNA binding protein A2

The amino-acid sequence of the protein A2 was determined before the complete sequencing of the T5 genome (Fox et al., 1982; Snyder, 1991). A2 was suggested to interact with the membrane (Snyder, 1984) and with A1 (Beckman et al., 1971). A2 is involved in the second step transfer since T5 *amA2* is not capable to complete the DNA transfer when infecting amber non-suppressive cells (Lanni, 1969). Moreover, this protein was also proposed as a dimeric protein with non-specific DNA binding activity (Snyder and Benzinger, 1981). *In vitro*, purified A2 is a dimeric protein indeed (thesis Zangelmi, 2018).

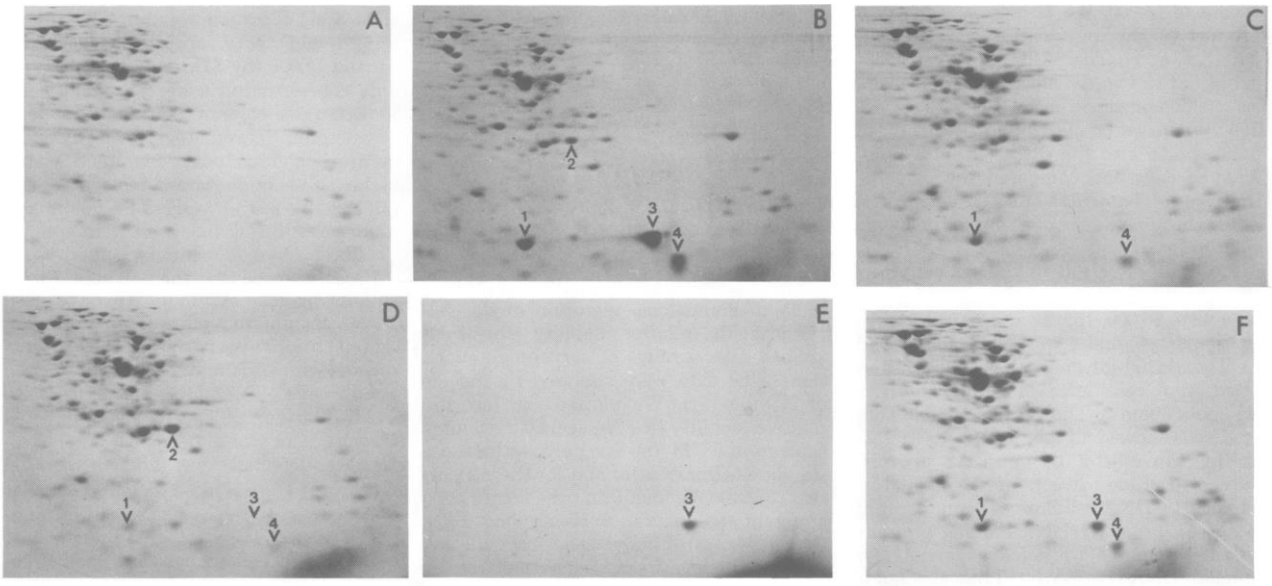


Figure 9. Detection of the A2 protein by 2D gel electrophoresis.
Two-dimensional gel electrophoresis obtained from extracts of (A) uninfected *E. coli* cells, (B) cells infected with T5 *amA1*, (C) cells infected with T5 *amA2*, (D) cells infected with T5wt, (E), 8 μg of purified A2 protein, and (F) 8 μg of purified A2 protein plus cells infected with T5 *amA2*. The direction of migration along the first dimension was from left to right (the pH of gel slices ranged from 4 on the left to 8.5 on the right). The spots numbered 1 to 4 were present on gel B but absent on gel A. Gels were stained with Coomassie brilliant blue. This figure is reproduced with permission from Snyder, C.E., and Benzinger, R.H. (1981). Second-step transfer of bacteriophage T5 DNA: purification and characterization of the T5 gene A2 protein. J. Virol. 40, 248–257.

It is worth noting that a peptide of ca. 30 kDa (McCorquodale et al., 1977; Snyder and Benzinger, 1981) was identified among the pre-early proteins (Table 3). It might be the 37 kDa protein Gp11. When bacterial cells are infected by a T5 *amA2* strain, 14-kDa and 30-kDa pre-early peptides cannot be detected (spots 3 and 2 in Figure 9, B, D) (McCorquodale et al., 1977; Snyder and Benzinger, 1981). The 14-kDa peptide in spot 3 corresponds to A2 while the 30-kDa peptide in spot 2 is probably Gp11 (37 kDa), based on its size and isoelectric point. Another possibility for spot 2 would be Dmp (28 kDa), but with reduced likelihood since its phosphatase activity can be detected in T5 *amA2*

infected cells (Mozer et al., 1977). Thus, in addition to its requirement for the Second-step transfer, A2 might be needed for the synthesis of a pre-early protein of ca. 30 kDa, presumably Gp11.

1.2.5. Other activities attributed to pre-early genes

In addition to the phenotypes associated to Dmp, A1 and A2, the FST DNA was suggested to code for effectors that control transcription shut-off, inhibit host replication and promote the resistance of T5 DNA to several nucleases.

1.2.5.1. Transcription shut-off

While the transcription of pre-early genes is independent from protein synthesis, on the contrary, the shut-off of FST-DNA transcription requires the synthesis of pre-early proteins (Section 1.2.3) (Sirbasku and Buchanan, 1970a).

Hendrickson and Bujard studied the transcription asymmetry using hybridization assays between RNA from T5 and its genome. Their results suggest that there are as much as twice pre-early transcripts from the uninterrupted strand (H-strand) than from the nicked strand (L-strand), and 2.6 times more early-RNAs from the H-strand than from the L-strand. Remarkably, they only detected late transcripts annealing to the uninterrupted strand, meaning that the late genes might have a unidirectional transcription (Hendrickson and Bujard, 1973). Nicked DNA can be regarded as transcriptional enhancers and are present in some phages, like T4, in which the nicked DNA close to late genes imposes a preferential orientation for the tracking clamp protein gp45 to enhance late gene transcription (Sanders et al., 1995).

Several hypotheses were proposed that could explain the shut-off of FST-DNA transcription, namely, FST-DNA partial digestion by A1, a natural decay of pre-early transcripts or an active DNA binding system capable to recognize and shut down gene expression (McCorquodale and Lanni, 1970). The results of superinfection experiments performed by Sibarsku and Buchanan (Figure 8) are more compatible with the third hypothesis. Yet to date, no experimental data is available to support any of these hypotheses.

1.2.5.2. Inhibition of Host DNA replication

One report showed that *E. coli* infected by a triple mutant *A1-A2-dmp* of T5, cannot replicate. In absence of these three genes, the host DNA was neither degraded nor replicated (Figure 10), which implies that there is some phage effector capable to halt the DNA replication besides A1, A2, and Dmp, and this effector is encoded in the FST-DNA (Mozer et al., 1977). The gene and mechanism responsible for this inhibition remain to be identified.

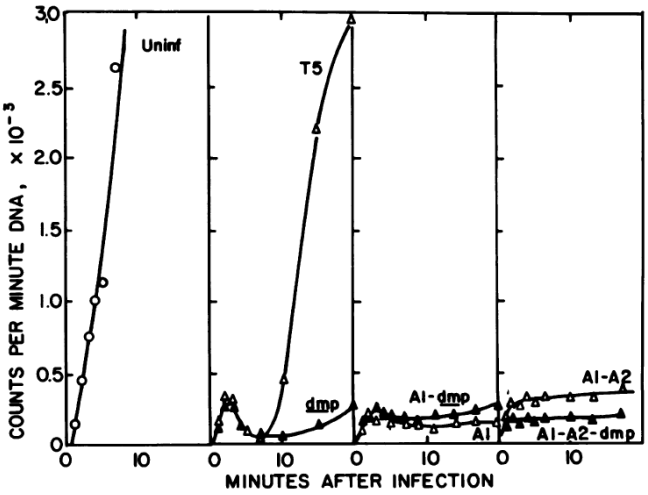


Figure 10. DNA synthesis after infection of *E. coli* F with T5 mutants.
DNA was measured by incorporation of [^{3}H] thymidine into a trichloroacetic acid-insoluble form at various times after infection by the indicated phage. Uninf: Non-infected cells, T5: wild-type phage, dmp: T5 *dmp*$^-$, A1: T5 *amA1*., A1-A2: T5 *amA1 amA2*, A1-A2-dmp: T5 *amA1 amA1. dmp*$^-$. This figure is reproduced with permission from Mozer, T.J., Thompson, R.B., and Berget, S.M. (1977). Isolation and Characterization of a Bacteriophage T5 Mutant Deficient in Deoxynucleoside 5′-Monophosphatase Activity. J. Virol. 24, 9.

1.2.5.3. Resistance to nucleases

- RecBCD

The complex RecBCD of *E. coli* plays an important role in DNA repair, as well as recombination and protection against phage infection (Dillingham and Kowalczykowski, 2008). Infection of *E. coli* with various DNA phages including T4, T5, and T7 led to a reduction in the ATP-dependent DNase activity of *E. coli*. This phage-mediated inhibition did not take place when protein synthesis was prevented by the addition of chloramphenicol before the infection (Sakaki, 1974). Thus, the host ATP-dependent DNase is likely inhibited during infection by T5 but the mechanism for this inhibition was never elucidated.

- Restriction enzymes

The FST-DNA was suggested to encode some mechanism capable of inhibiting the host restriction nucleases hazardous for T5 DNA. T5 wt (NCBI: NC_005859.1) bears five EcoRI (GAATTC), two EcoK (AAC(N_6)GTCG) and nine EcoB (TGA(N)$_8$TGCT) restriction sites in the SST-DNA, but none in the FST-DNA (LTR). The DNA is unmodified and intrinsically sensitive to all these nucleases *in vitro*. Davison and Brunel in 1979 isolated T5 mutants sensitive to EcoRI$^+$ *E. coli* strains and found that the mutations were introduced in the FST DNA, at an undetermined position between *A1* and *A2*. These EcoRI sites lowered the efficiency of plaquing by a factor between 10^2 and 10^5 in an *E. coli* EcoRI$^+$ host (Davison and Brunel, 1979a). Moreover, the modification of the restriction site sequences in the FST-DNA recovered the wild type phenotype (Davison and Brunel, 1979b). This supports the hypothesis that a mechanism inhibiting the restriction enzymes is set up during T5 infection.

On the other hand, the nuclease activity of crude extracts of *E. coli* EcoRII+ and EcoRV+ strains infected by T5 mutants was assessed on the DNA of the bacteriophage lambda. This DNA is sensitive to EcoRII or EcoRV nucleases *in vitro* and was also cleaved by the crude extracts of T5-infected cells (Chernov and Kaliman, 1987, publication in Russian, only the abstract is available in English). This supports the hypothesis that there is no inhibitor of restriction enzymes encoded in the FST DNA but rather suggests that the DNA resists against restriction nucleases by a "passive" evasion.

Bacteriophage T5 is also resistant to a recently described Restriction-Modification system, a single-stranded DNA phosphorothioate sensing system (Ssp) (Figure 11) (Section 1.1.3.3). This system is made up of four proteins SspABCD and SspE. SspE represses phage replication by introducing nicks into phage DNA. Of all the phages tested, T5 is the only resistant to SspABCD-SspE. However, T5 DNA is susceptible to the nicking activity of SspE *in vitro* (Xiong et al., 2020). It is unclear, whether T5 DNA is modified *in vivo* by this system.

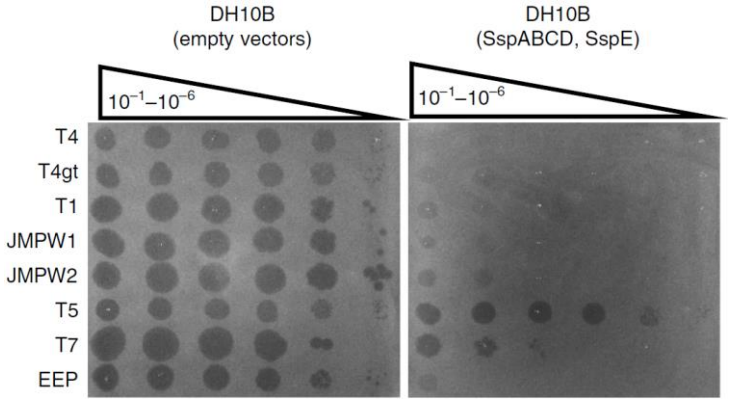


Figure 11. T5 evades the *Escherichia coli* Ssp system.
E. coli harbouring the Ssp system from *E. coli* 3234/A showed a higher level of resistance to most tested phages, except T5. The mechanism of T5 resistant is unknown. This figure is reproduced with permission from Xiong, X., Wu, G., Wei, Y., Liu, L., Zhang, Y., Su, R., Jiang, X., Li, M., Gao, H., Tian, X., et al. (2020). SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities. Nat. Microbiol. 1–12.

- CRISPR/Cas systems

Another recent study suggested that bacteriophage T5 is resistant to CRISPR-Cas immunity in *E. coli*. Thus, following a challenge of *E. coli* cells coding for a CRISPR-Cas type I-E system with T5, survivors of the infection acquired spacers that matched FST-DNA or SST-DNA. However, CRISPR–Cas was able to recognize and interfere with protospacers from pre-early, early, and late T5 genes cloned in plasmids but not effectively within some regions of the phage DNA (Strotskaya et al., 2017) (Figure 12).

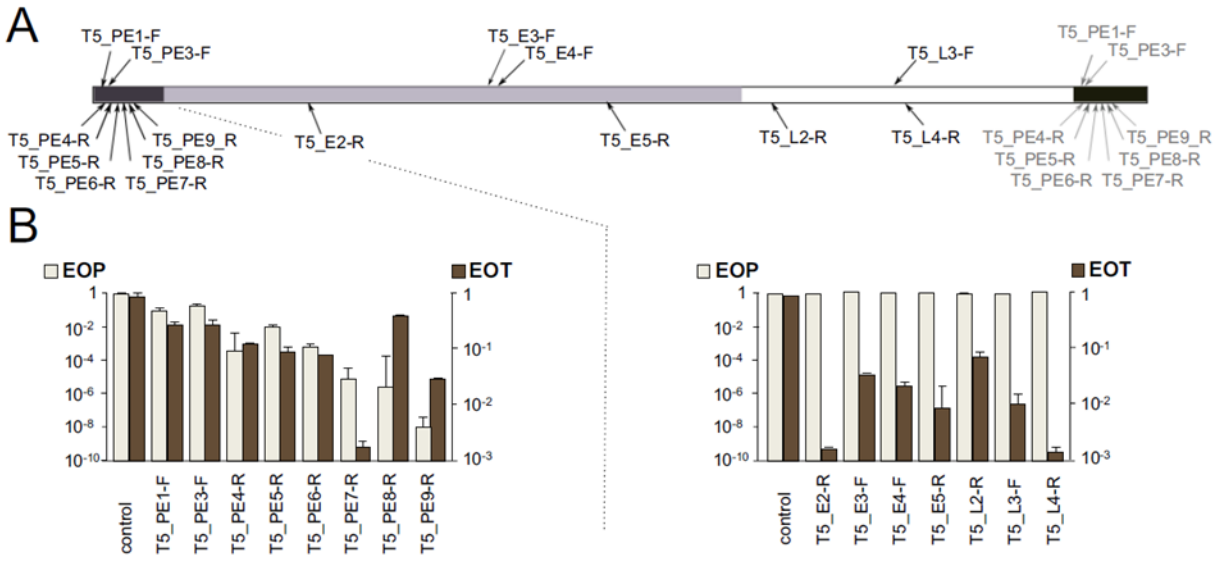


Figure 12: The effect of CRISPR–Cas type I-E targeting on T5 infection.
(A) Protospacer are obtained from the T5 genome come mostly from the terminal repeats, (B) The efficiency of plaquing (EOP) of T5 wt or the efficiency of transformation (EOT) of protospacer containing plasmids were compared on strains coding for CRISPR-cas type I-E. This figure is reproduced with permission from Strotskaya, A., Savitskaya, E., Metlitskaya, A., Morozova, N., Datsenko, K.A., Semenova, E., and Severinov, K. (2017). The action of *Escherichia coli* CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. Nucleic Acids Res. 45, 1946–1957.

Bacteriophage T5 resists a variety of bacterial defense systems. It is improbable that a single protein will inhibit such a large variety of nucleases, and it is also unlikely that the FST-DNA will encode many nucleases inhibitors. For these reasons, we hypothesize that T5 resistance to several bacterial nucleases might stem from a passive mechanism, such as preventing contacts between phage DNA and some of the host nucleases.

1.3. Aims of this study

T5 is a lytic phage that infects *Escherichia coli* in two steps: first, it delivers only 8 % of its DNA and stops, the fragment codes for pre-early proteins that inhibit bacterial defense mechanisms, degrade the host genome and resume the delivery; finally, the rest of the genome enters the cell to replicate the viral DNA and assemble new virions, until cell lysis. With its original strategy of infection and the clustering of its pre-early genes in a short region of the genome, T5 constitutes an attractive model to investigate the mechanisms of host takeover.

Of the 17 annotated pre-early genes of the FST DNA injected first, only three have been described in the past: the dispensable gene *dmp* coding for a phosphatase, and the two *A1* and *A2* essential genes both required for the second step transfer. From which *A1* is also required for the degradation of the host chromosome at the onset of infection. Almost nothing is known about the molecular mechanism by which these 17 pre-early genes take over the host before starting the second step transfer: controlling the host chromosome degradation while protecting the viral genome, circumventing the host nucleases, blocking the host DNA replication, subverting the transcription and translation machineries.

The aims of this thesis were to tackle three main questions concerning the FST-DNA and the pre-early genes of bacteriophages T5: First, which pre-early genes are essential for T5 infection? Two of the 17 genes were identified as essential, but it remains unclear why no other gene was identified in the past by random mutagenesis. Second, what is the contribution of each gene to the phage infection? Dispensable genes might be so under laboratory conditions, buy play a more important role out in the nature and when the phage faces different hosts. Finally, the elucidation of the exact role of each pre-early gene escape to the temporal limitations of this work, reason for which, a proxy will be to test what is the effect of phage pre-early gene expression over in the host.

With few genes holding a known function, many attributable phenotypes, and many more possible functions, we studied the annotated pre-early genes by two approaches: i) By reverse genetics, we constructed and characterized amber mutants on the two essential genes (*A1*, *A2*), single-deletion mutants on four genes (*dmp*, *02*, *05*, *07*), and multiple-deletions mutants on all the pre-early genes but four (*A1*, *A2*, *08*, *09*). ii) On the other hand, we carried out the ectopic expression of the 17 pre-early genes and studied the survival, growth kinetics and morphology of the recipient bacteria.

The FST locus is neighbored by a 1.5-Kb intergenic region called the Injection Stop Signal (ISS) that interacts temporarily with the interphase phage tail-bacterial envelope and it is supposed to

contain a sequence necessary for halting the DNA delivery after the FST. As an attempt to advance potentially convergent subjects, we have constructed and partially characterized some other deletion mutants: A phage with a partially deleted ISS to narrow down the actual stop signal with which some phage/host proteins might interact, a "blue" phage to simplify future assays on recombination efficiency, and a "fluorescent" phage to help on the study of T5 infection timing.

2. Results

2.1. Developing new tools for T5 genome engineering

2.1.1. Testing restriction of T5 Infection by CRISPR-Cas9

Most T5 mutants were isolated in the past century following random chemical mutagenesis (Abelson and Thomas, 1966; Hendrickson and McCorquodale, 1971a; Lanni, 1969). More recently, directed mutagenesis of T5 was performed upon infection of host strains carrying a plasmid with the desired mutation to facilitate homologous recombination (HR) within the genome of T5. Screening for the T5 mutant plaques was performed by hybridization of radioactively labeled oligonucleotides under stringent conditions, a rather cumbersome method (Klimuk et al., 2020; Vernhes et al., 2017). One objective to this thesis was to develop straightforward and simple tools for T5 genome engineering.

In T3, HR yields down to the recovery of ~0.1-1% mutants (Yehl et al., 2019) while in T4, it does so between 0.01-0.5% (Singer et al., 1982; Völker and Showe, 1980). These frequencies in mutant recovery after HR are similar to those estimated for T5 (0.1-0.2%) (Vernhes and Boulanger, unpublished). When CRISPR-Cas9 is added to the protocol to eliminate wild-type T4 phages that have not undergone HR, the frequency of mutant recovery reaches 1-2% (Tao et al., 2017). CRISPR-Cas systems have been successfully used to improve mutant recovery in T7 (Kiro et al., 2014), T4 (Tao et al., 2017), virulent *Lactococcus* phages (Lemay et al., 2017; Martel and Moineau, 2014), and *Listeria* phages (Hupfeld et al., 2018). In most cases, the template for the mutagenesis is provided in a plasmid which recombines with the phage genome during infection, while the CRISPR-Cas system helps at counter selecting the wild-type phages. However, it was observed that CRISPR-Cas system on-target and off-target efficacy is not uniform across the targeted genomes, and is influenced by DNA modifications (Tao et al., 2017), the spacer sequence (Cui et al., 2018; Haeussler et al., 2016) as well as the protein doses (Cui et al., 2018).

To test whether phage T5 infection is restricted by CRISPR-Cas9, we constructed the plasmid pAC9 for constitutive expression of sgRNAs and inducible expression of the nuclease Cas9 to restrict phage T5 infection. We transformed *E. coli* F with a plasmid coding for the nuclease Cas9 and customizable sgRNAs. Single guide RNAs (sgRNA) are 100-bp chimeric molecules that replace the pair crRNA:tracrRNA, in which the 5' end is the spacer. The sgRNAs matched different locations across the first 10 kb of the T5 phage genome. By double-layer agar method, we compared the titer of a T5 stock on two host strains: *E. coli* F expressing both Cas9 and each sgRNA, versus the same host lacking the sgRNA. The ratio of PFUs between the two hosts was considered as the efficiency of

plating (EOP) and reflects the protective capacity of CRISPRs against phages (Strotskaya et al., 2017; Tao et al., 2017) for each sgRNA.

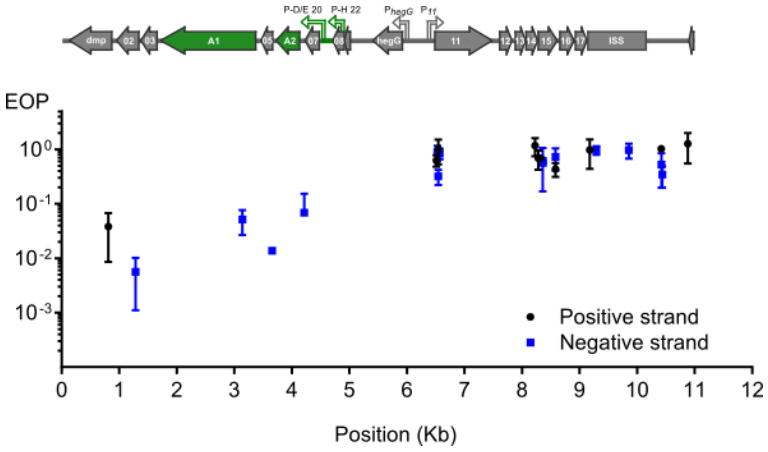


Figure 13: Restriction of phage T5 infection by CRISPR-Cas9.
The restriction of T5 infection was measured as the efficiency of plating (EOP), calculated as the titer of T5 on sgRNA- and Cas9-containing hosts versus the titer on the same host lacking the sgRNA. Titers were determined by plaque assay. EOP using spacers matching the positive or negative strand were highlighted in black or blue, respectively. Each point represents the mean and standard error of the mean (s.e.m.) from three replicas. In the genetic map of the first 11 kb of the 121-kb genome, green arrows represent genes *A1* and *A2* which are predicted to be essential based on the phenotype of amber mutations genetically mapped in the 1960s.

Restriction of T5 infection varied across the tested region regardless of the DNA strand that was targeted. Spacers with the highest interference mapped to the genes *dmp*, *02*, *A1*, *05*, and *07* and lowered the titer of T5wt between 10 to 100 times. The other spacers did not provide protection as the titer of T5wt decreased less than ten times. Therefore, Cas9 can be efficiently used to restrict T5wt infection when it is targeted to a region within the first 4.5 kb of the T5 genome.

2.1.2. Bacteriophage T5 genome editing through HR plus CRISPR/Cas9

To introduce genetic modifications in T5, we focused on the region with effective spacers. In the late sixties, Lanni and coworkers described the isolation of T5 amber mutants that could only be propagated in the CR63 permissive strain (Lanni, 1968; Lanni et al., 1966). Mutations were genetically mapped to genes *A1* or *A2,* but no sequencing data were obtained to confirm this finding. Therefore, we decided to introduce a stop codon early in the open reading frame of the gene *A1*, resulting in the synthesis of truncated A1 (27 amino acids instead of 556). A 1-kb fragment of the T5 genome was amplified and cloned into a pUC19 plasmid. This fragment comprises 500-bp on each side of the *A1* start codon in order to provide ample homologous sequences on either side of the desired mutation, and the codon for T28 on the insert was replaced by a stop codon. Phage mutagenesis was performed by allowing infection of *E. coli* CR63 carrying the resulting plasmid. Counterselection of the T5wt within the crude lysate was carried out by double agar overlay plaque assay onto *E. coli* CR63(pAC9-A1), a strain that expressed Cas9 and a sgRNA directed against the wild-type *A1* gene.

The resulting plaques were then picked and patched onto both *E. coli* strains F (non-permissive) and CR63 (amber suppressive). We recovered six amber mutant plaques in 100 patched ones (Figure 14, A, B). Sequencing confirmed the introduction of the intended mutation in the T5 *amA1* phages (Figure 14, C). Thus, our results demonstrate that it is possible to use HR plus CRISPR/Cas9 to introduce point mutations in the genome of T5.

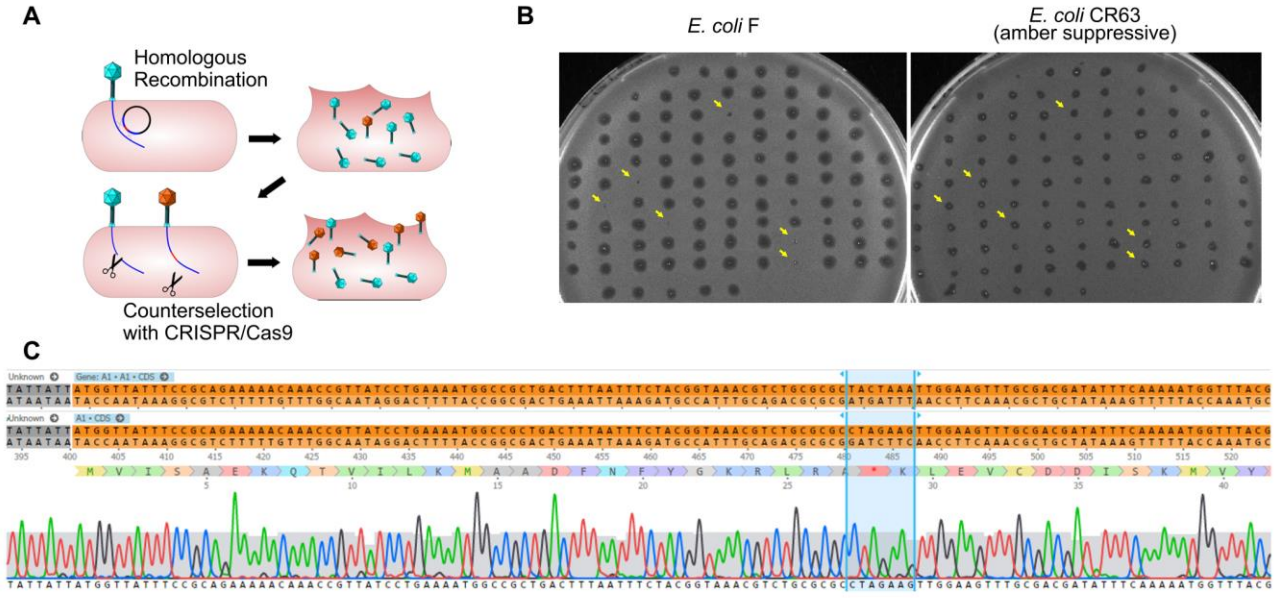


Figure 14. Construction of mutants by homologous recombination and mutant enrichment using CRISPR/Cas9. (A) Upon infection by wild-type phages of a host bearing a template plasmid, HR introduces the genetic modification in a subset of phages. In a second step, mutants among the progeny are enriched by counterselection onto a host producing the Cas9 nuclease targeted against the wild-type sequence. (B) Screening of the phages obtained after mutagenesis and counterselection. Lysis plaques were tested onto a non-permissive *E. coli* strain F and the amber suppressive strain CR63. Plaques with no lysis on *E. coli* F were considered positive and sequenced. (C) Sequencing results for one of the six amber mutants T5 *amA1* T28. Top lane corresponds to the T5 nucleotide sequence with the desired amber mutation within gene *A1* at position (3226..3332) of the T5 genome. Note that *A1* coding sequence is on the complementary strand. Second lane corresponds to amino acid residues 16 to 50 translated from the sequence above (reading from right to left). Third lane shows sequencing results.

2.1.3. Retron mediated gene editing in phages

We explored the use of alternative forms of DNA templates for phage genome engineering. Bacterial retroelements or retrons, are chimeric RNA/DNA molecules composed of covalently linked ssRNA (msr) and ssDNA (msd). Both msr and msd are coded in the same cistron, altogether with a reverse transcriptase responsible for the partial reverse transcription and linking of the msr-msd molecules (Simon et al., 2019). These elements can provide ssDNA *in vivo* for gene modification and were used for bacterial genome engineering in the past. Their mechanism is not well known and probably recombination is carried out by ssDNA annealing during the DNA replication (Farzadfard and Lu, 2014a; Simon et al., 2018).

We have tested if these genetic tools in bacteria were also applicable to phages, in combination with CRISPR-Cas9 counterselection. For the retron template, we modified plasmid pFF745, which codes for the msr, msd, and the reverse transcriptase in a poly-cistron controlled by an IPTG-inducible promoter (Farzadfard and Lu, 2014a). Part of the msd sequence was replaced with a 75-bp segment from the phage *A1* gene centered on the serine codons for S84 and S85. These codons were substituted by stop codons, generating plasmid pFFA1. To introduce the mutation *A1* S84Stop/S85Stop into the phage genome, we cultured *E. coli* CR63(pFFA1) in the presence of IPTG and infected the bacteria with the wild-type phage. The crude lysate was plated onto *E. coli* CR63 bearing plasmid pAC9 and sgRNA matching the unmodified codons S84/S85 of T5 *A1*. For mutant screening, we picked 100 plaques and patched them onto plates seeded with amber-suppressive and non-suppressive *E. coli* strains, CR63 and F, respectively. We recovered two plaques out of 100 plaques that could only lyse the amber-suppressing host strain. Such mutants were designated T5 *amA1* SS84. Sanger sequencing results showed the correct introduction of mutations in *A1* and no other alteration in the targeted locus (Figure 15). Our results suggest that retrons could be effectively used for phage genome engineering.

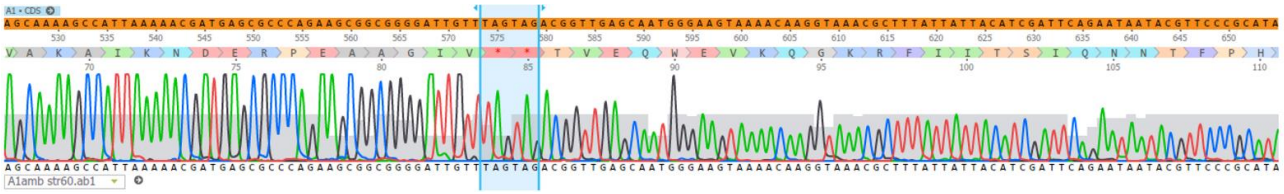


Figure 15. Sequence of a mutant T5 *amA1* SS84 constructed by retron-mediated recombination.
Purified amber mutant phages T5 *amA1* SS84 obtained by retron-mediated mutagenesis were used as a template for PCR amplification. Amplicon sequences showed the presence of two stop codons instead of serine codons (S84 and S85) in the T5 *A1* gene.

2.1.4. Dilution-Amplification-Screening (DAS) to isolate mutants from phage T5

As we have seen above, some regions of the T5 genome are not amenable to the CRISPR/Cas9 counterselection, a feature that enriches the phage progeny in mutants relative to wild type after mutagenesis. We reasoned that a similar enrichment could be achieved with a simple dilution-amplification screening (DAS) (Figure 17, A). As a proof of principle for this procedure, we sought to obtain a phage carrying an amber mutation in the essential gene *A2* at the codon S37.

We allowed the recombination between the phage T5 and the plasmid bearing a mutated copy of the gene *A2* during infection. Following filtration to eliminate uninfected bacteria, the crude lysate was serially diluted from ca. 10^{10} PFU/mL down to 10^{2} PFU/mL. We distributed each dilution into several wells containing the amber-suppressive host and incubated for 3 h to amplify the phages. We screened pools of well aliquots by mismatched amplification mutation assay (MAMA)-PCR to

detect the mutation and the positive wells were plated by a double agar layer method to obtain individual plaques (Figure 16). The plaques were picked and patched successively onto two plates with a non-permissive and a permissive host, respectively. The plaques that lysed only the permissive host were considered positive (Figure 17, B). We recovered six positives out of 50 plaques, from which one was confirmed by Sanger sequencing (Figure 17, C). Since our estimation of T5 recombination rates is around 0.25% (see below), our recovery rate of 12% indicates that we could enrich the mutant in some pools enough to simplify the recovery of T5 *amA2* S37 mutants. Hence, DAS is an efficient method for mutant isolation.

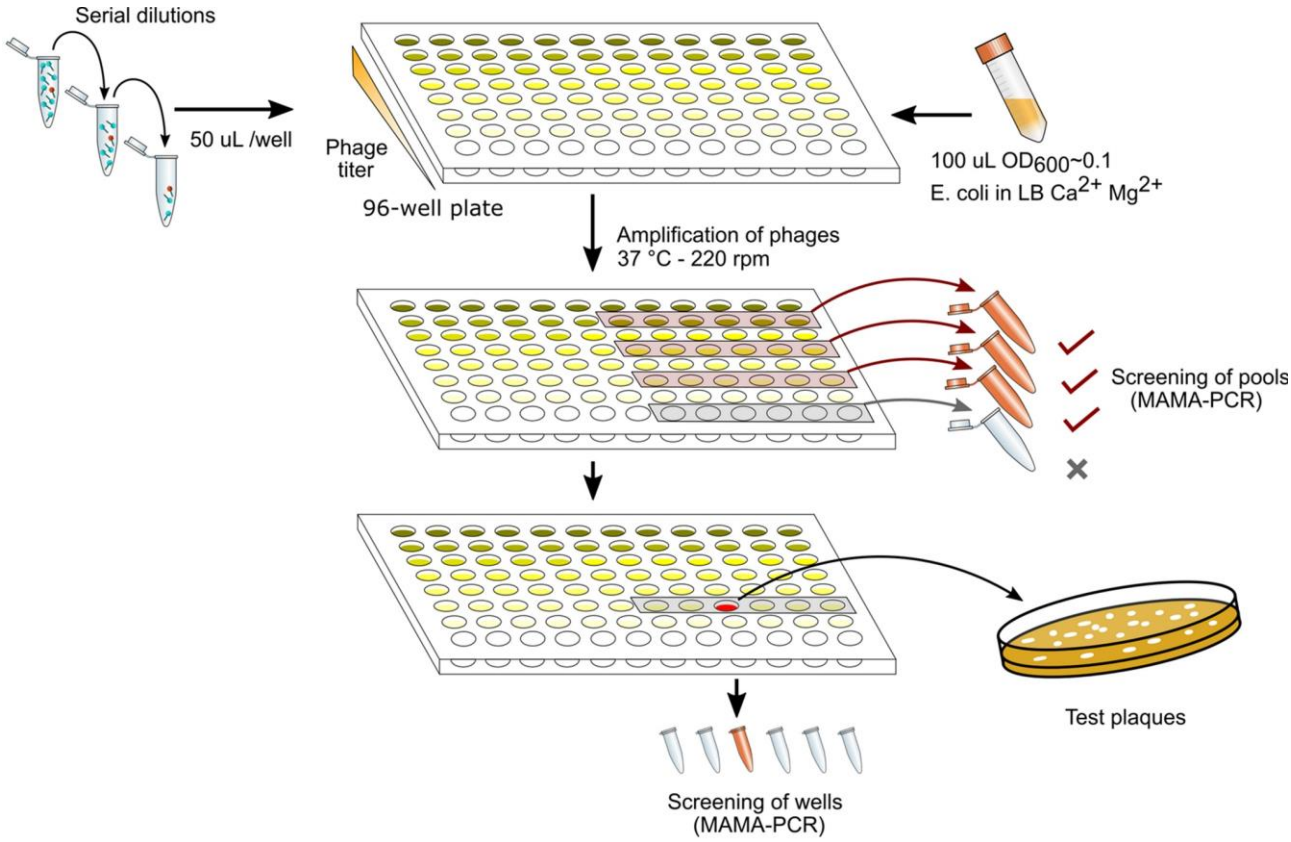


Figure 16. Dilution-Amplification Screening of mutants of T5.
The lysate after homologous recombination (HR) is diluted to get 10-10^3 PFU/mL, then distributed in a 96-well plate (plate 1) and the host is added. After incubation for 3 hs in shaking, 20 uL from six wells are pooled and the mutant T5 is detected by MAMA-PCR. Positive pooled wells are indicated by a red color. Another MAMA-PCR is carried out with individual wells from plate 1 corresponding to the highest positive dilution. Once a single positive well is identified, its content is plated to test individual plaques by MAMA-PCR or pick and patch.

To generate a phage carrying amber mutations in both *A1* and *A2*, we also used DAS to add an amber mutation in *A2* (S37Stop) in phage T5 *amA1* T28. Since, the screening was not possible by differential growing on amber suppressive and non-suppressive strains, all the screening was made by a MAMA-PCR, an allele specific PCR. The construction of T5 *amA1*A2 (*A1* T28Stop, *A2* S37Stop) was confirmed by Sanger sequencing (Figure 18). DAP was also used to add the mutation (S37Stop)

on the gene *A2*, in a phage T5 *amA1.* Since, the screening was not possible by differential growing on amber suppressive and non-suppressive strains, all the screening was made by MAMA-PCR. The construction of T5 *amA1A2* (*A1* T28Stop, *A2* S37Stop) was confirmed by Sanger sequencing (Figure 18).

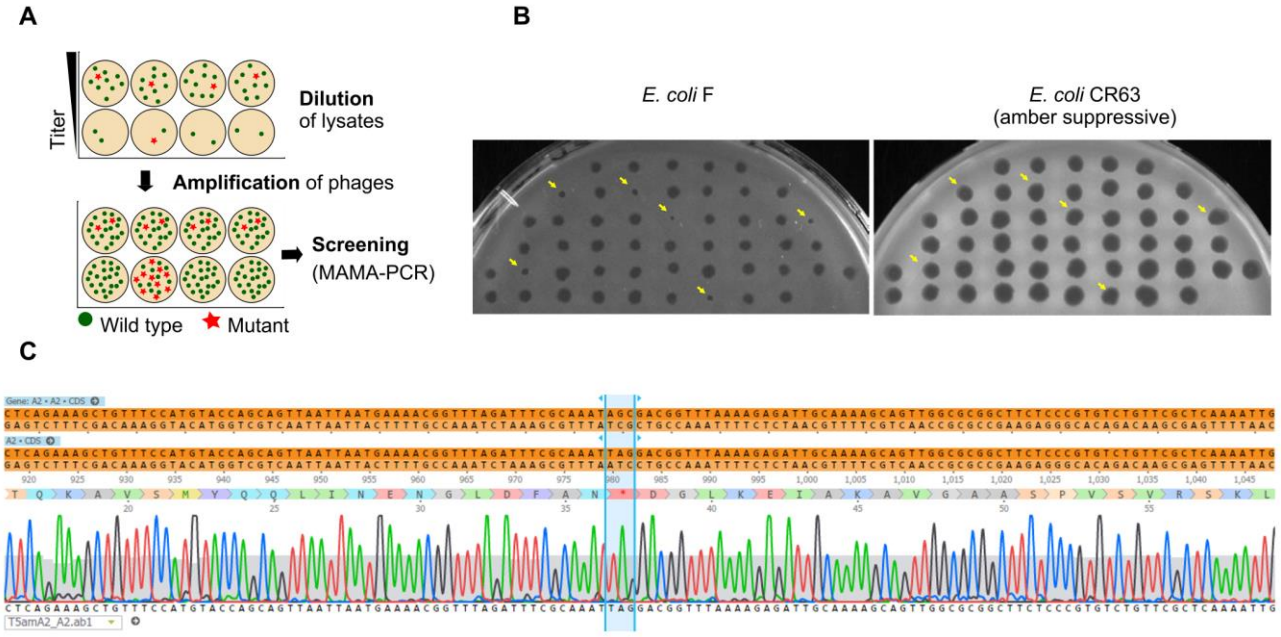


Figure 17. Screening of T5 mutants by Dilution-Amplification Screening (DAS).
In (A) DAS principle: following mutagenesis, mutants are far less abundant than wild-type phages in the crude lysates. In the first step, dilutions of the lysates are performed to lower the phage titers and alter the proportion of mutants present in each pool (above, second row). In the second step, phages in the diluted pools are amplified by adding the host. While the mutants remain scant within amplified pools from initial high viral titers (below, first row), the modified phages become enriched in the progeny obtained with some of the highly diluted pools (below, second row). (B) Plaques obtained from a PCR-positive well were patched onto the non-permissive strain F and the amber-suppressive host CR63. Positive plaques are highlighted with an arrow. (C) Sequencing result shows the correct introduction of point mutations in the T5 genome.

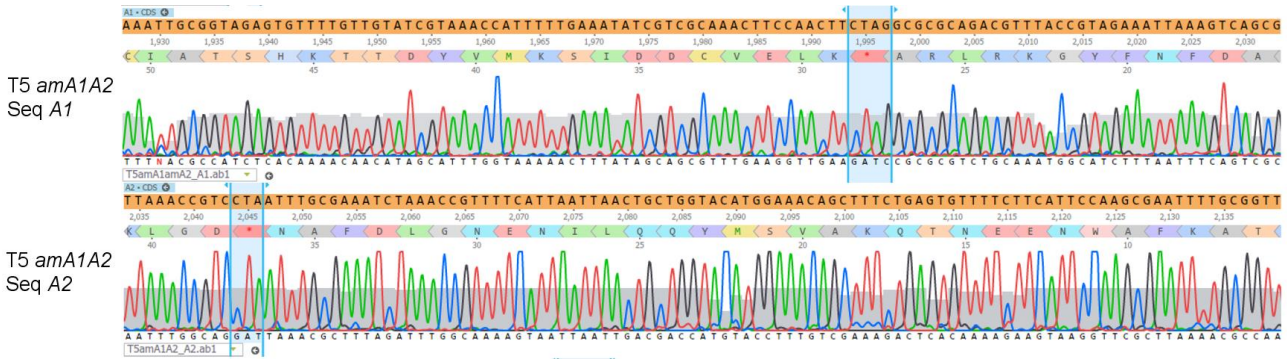


Figure 18. Sequencing of amber mutants in genes *A1* and *A2*.
T5 *amA1A2* was constructed from T5 *amA1* (T28Stop) by adding the modification *amA2* (S37Stop) by HR and positive mutants were screened by DAS. Genes *A1* and *A2* were amplified from T5 *amA1A2* and sequenced. In blue, the introduced mutations.

2.1.5. Reversion of single and double amber mutants in *A1* and *A2*

Each mutant, recovered from one single plague and checked by sequencing of the desired mutation, was amplified by infection of larger cultures of CR63. We titrated the phage stocks (T5wt, T5 *amA1*, T5 *amA2*, and T5 *amA1A2*) on two strains. The amber suppressive strain CR63 measure the phage titer of mutants and revertants, while the non-suppressive strain F only measures the titer of revertants. Revertant phages are usually the result of spontaneous mutations in the amber (stop) codons that restore the wild type phenotype. To test this, for each single gene mutant we isolated, three plaques that formed on the non-permissive strain and sequenced DNA amplicons following PCR using primers specific for each gene (Figure 19). For all six revertants, we found substitutions that eliminated the stop codons in all cases.

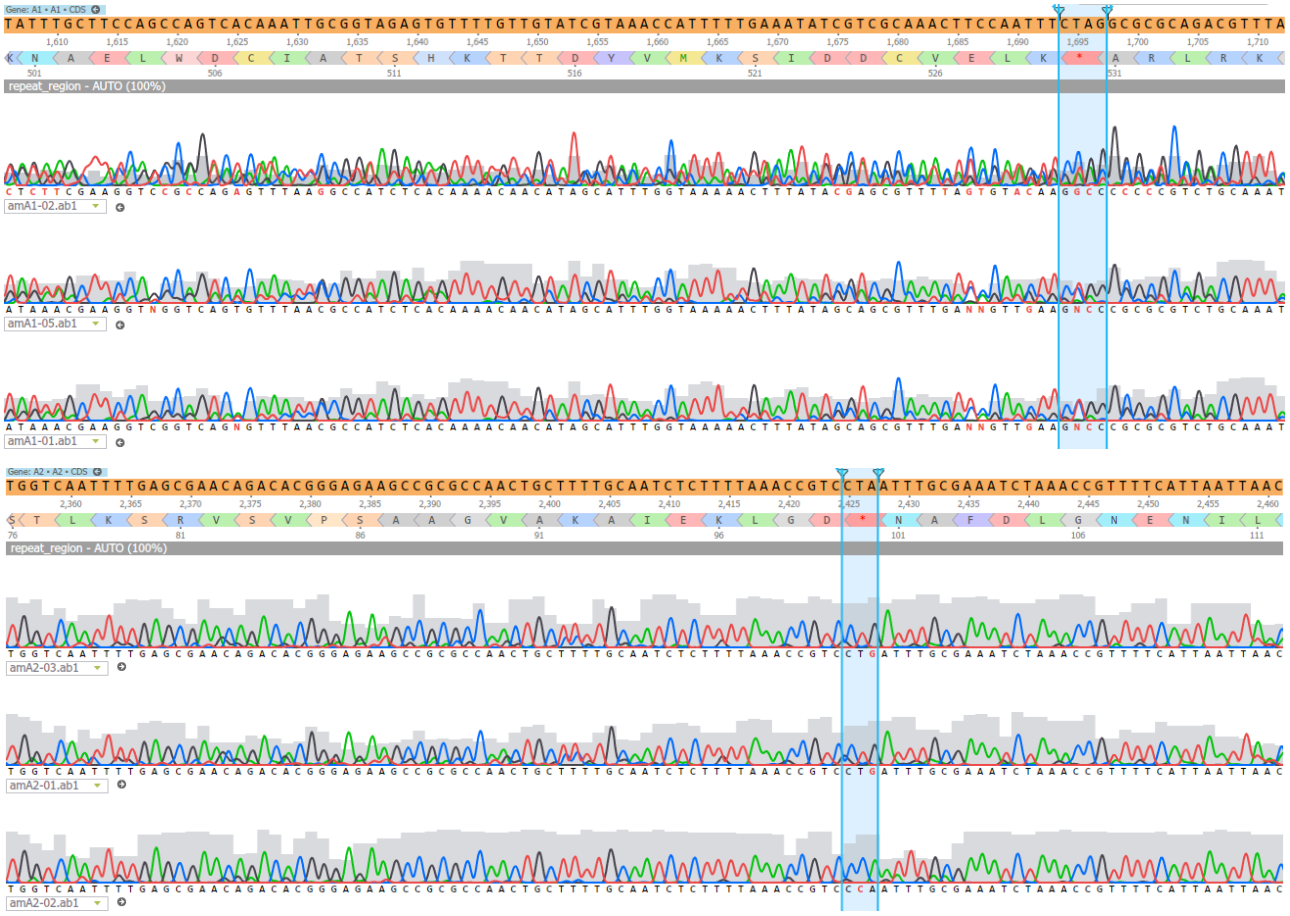


Figure 19. Amber mutation reversion in T5 amber mutants.
Some amber mutants from phages T5 *amA1* and T5 *amA2* able to infect the wild-type *E. coli* F strain were recovered and sequenced at the genes *A1* (above) or *A2* (below). Most of the phages reverting to the wild-type T5 phenotype carry spontaneous mutations at the introduced stop codons.

For both single-gene amber mutants, frequency of reversion to wild-type phenotype was similar, ca. one per million (Figure 20, A), while for the double-gene amber mutant T5 *amA1A2*, no plaque can be found on the non-permissive strain F, from T5 *amA1A2* (Figure 20, B). The presence of plaques on F (revertant, wild-type phages) for single-gene mutants and the absence of them from a multiple-gene mutant suggest it is not possible for the phage to repair both stop codons by

spontaneous mutations in few cycles. Nevertheless, at higher titers, T5 *amA1A2* is capable of impairing the bacterial growth (Figure 20, B), which suggests that there might be other proteins besides A1 and A2 that can hinder bacterial replication.

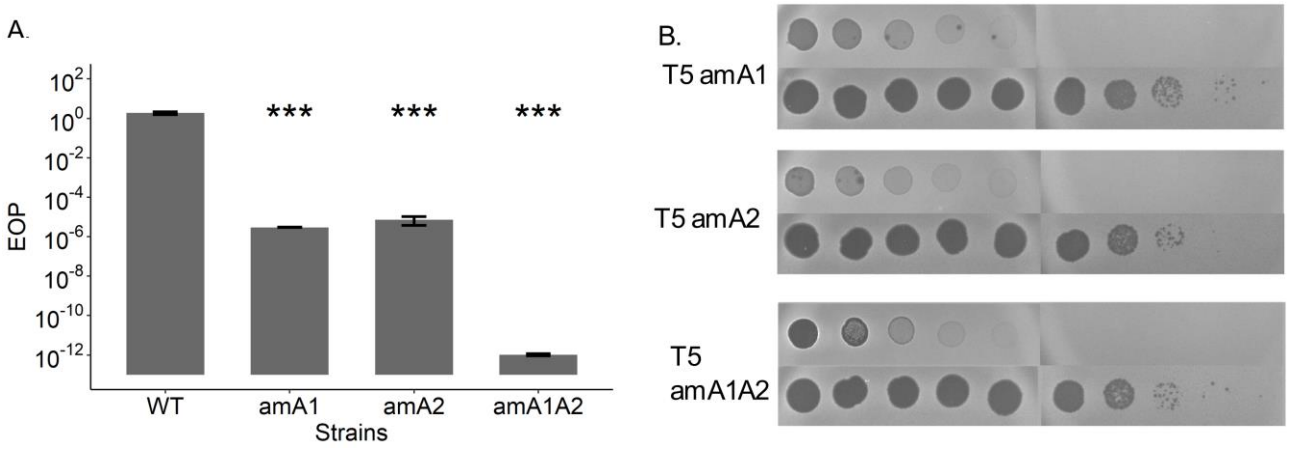


Figure 20. Efficiency of plaquing of T5 mutants in essential genes *A1* and *A2*.
Amber mutants T5 *amA1*, T5 *amA2*, and T5 *amA1A2* stocks were titrated on *E. coli* strains non-permissive F and permissive CR63. (A) The efficiency of plaquing (EOP) was calculated as the phage titer on F over CR63. (B) Spot assay of the mutants on F (top row) over CR63 (bottom row). Tukey post-hoc test for multiple comparisons, *** $p = 1.2 \times 10^{-4}$, $n = 3$. For T5 *amA1A2*, no plaques were found on *E. coli* strain F.

2.1.6. Blue/White screening in the T5 genome editing

Following the successful introduction of point mutations in the T5 genomes, we further tested whether insertion of exogenous genes in the FST was possible. Thus, we tested a color-based marker, the *lacZ*α gene, could be inserted into the T5 genome.

The *lacZ* gene codes for the β-Galactosidase, a tetrameric enzyme made up by 1,023 amino acids. The enzyme can be split in two peptides LacZα (90 aa) and LacZω (933 aa), that can be provided separately to reconstitute the original enzyme. The gene *lacZ*α is used in molecular cloning as a negative marker for the insertion of sequences into certain plasmids: if an insert disrupts *lacZ*α, the otherwise-encoded LacZα cannot split-complement the chromosome-encoded LacZω anymore, leading to the absence of the β-Galactosidase function in *E. coli*. The β-galactosidase activity in *E. coli* can be tested in solid media by adding X-gal to it. The β-Galactosidase breaks down X-gal into 5-bromo-4-chloro-indoxyl and galactose. 5-bromo-4-chloro-indoxyl oxidizes and dimerizes to form 5,5'-dibromo-4,4'-dichloro-indigo, which is insoluble and blue, thereby staining the bacterial colonies. It was shown that a fusion of the *lacZ* gene to the T4 *ipIII* internal protein gene yields phages with β-galactosidase activity (Hong and Black, 1993). We tested whether an insertion of the gene *lacZ*α in the FST region of T5 would yield β-galactosidase producing plaques as well.

We inserted the *E. coli lacZ*α gene between the T5 pre-early genes *05* and *A1* (Figure 21.A) by HR. T5 *lacZ*α plaques displayed β-galactosidase activity (Figure 21.B) when plated with the strain *E.*

coli XL-1 Blue, in presence of X-Gal and IPTG. The insertion was confirmed by sequencing (Figure 21.C).

This system presents some advantages over the system used in T4 presented in (Hong and Black, 1993). Namely, the small size of the gene *lacZ*α (270 bp) versus that of *lacZ* (3072 bp) allows an insertion without overloading the phage genome.

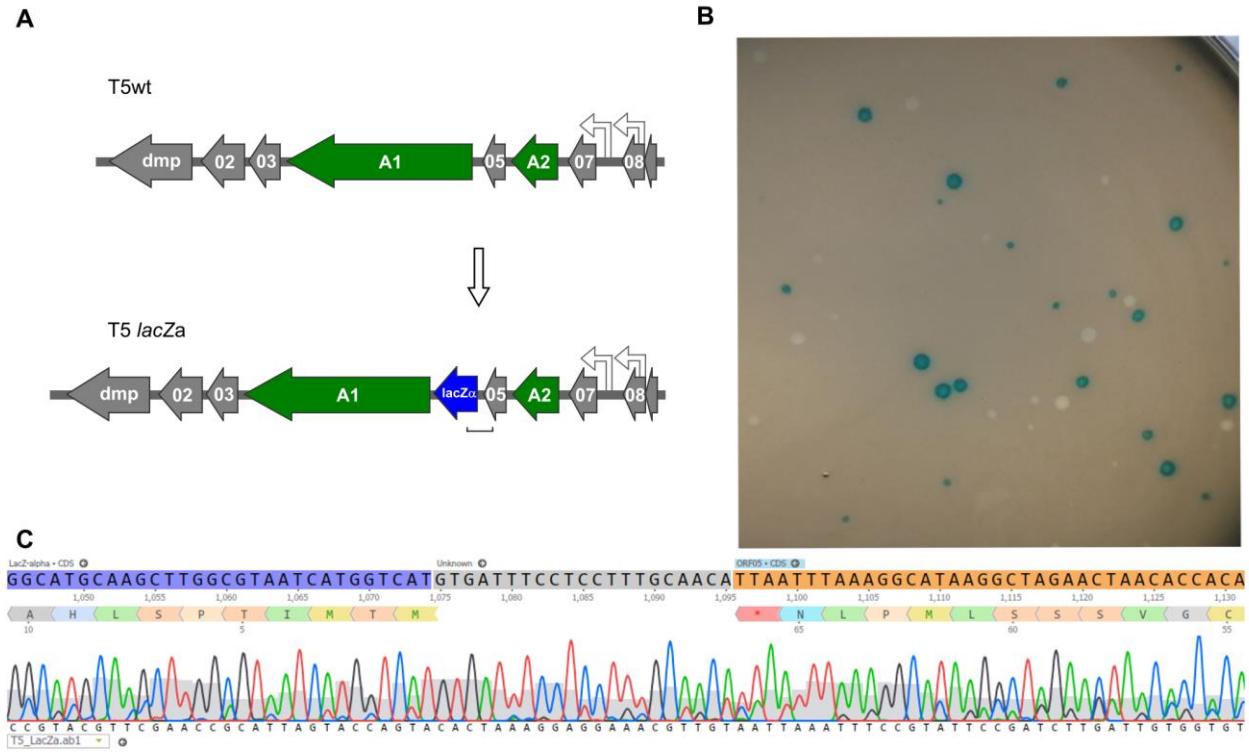


Figure 21. Construction of a Blue T5 (T5 *lacZ*α) strain.
(A) Genetic map showing that the gene *lacZ*α was introduced between genes *A1* and *05* in the genome of T5. (B) Plaques of T5 *lacZ*α and T5wt after plating along *E. coli* XL-1 Blue supplemented with X-Gal and IPTG: only the mutant produced blue plaques. C) Sequencing results of one of the blue plaques, showing the correct insertion of *lacZ*α downstream of gene *05*.

Another advantage of this system is that it allows the colorimetric detection of recombinant phages, and thereby provides an estimation of the yield of T5 recombinants. We tried to determine the recombination rate of T5 after HR using the blue/white marker *lacZ*α with the aim of setting a standard for future comparisons between gene editing methods.

After HR between T5 and the template plasmid (pUC05-lacZalpha2) during the infection of *E. coli* XL1 blue, the lysate was plated with *E. coli* XL1 Blue. We counted the number of blue plaques over the total number of them and found that 0.27 ± 0.01 % (n = 6) of the phages obtained were recombinants. We believe this work constitutes the first systematic evaluation of HR-mediated mutant yield in T5. In T4, the frequency of mutants was reported to be between 0.1-0.5 % (Völker and Showe, 1980), while in T3, it is comprised between 0.1-1% (Yehl et al., 2019). Therefore, the HR in T5 is consistent with that of other phages from the T series.

2.1.7. Generating red-fluorescence virion particles by gene fusion with mCherry

To test whether our mutagenesis protocol was applicable to another region of the T5 genome, we generated a gene fusion between part of gene *151* and a reporter gene encoding a fluorescent protein. We took advantage of the properties of T5 gene *151* product, the decoration protein pb10 of T5 capsid. Although dispensable for capsid assembly, pb10 contributes to its stability under high temperature (Vernhes et al., 2017). Pb10 is a two-domain protein that binds with a high affinity (K_D= 10^{-12} M) to 120 sites at the surface of T5 capsid by its N-terminal domain, while exposing its C-terminal domain to the solvent. The N-ter domain of pb10 could be fused to the mCherry fluorescent protein that replaces the C-ter domain without affecting the binding affinity of N-ter domain ((Vernhes et al., 2017) and Patent PCT/FR2020/051628). Thus, the purified fusion protein Pb10(Nter)-mCherry can decorate the capsid of T5 Δ*pb10* with the same efficiency as wild type pb10. Based on these data, we integrated the fusion *pb10*(Nter)-*mcherry* (Δ[103638..103928]::*mcherry*) into the T5 genome to obtain T5 PNmC (Figure 22, A).. The sequencing results depict the N terminus of pb10 coding sequence in the same ORF with the gene that encodes mCherry (Figure 22, C).

We carried out the infection of *E. coli* F cells grown in exponential phase with T5 PNmC at a MOI of 5. Then, we collected samples and fixed the cells at different time points to study the cells by fluorescence microscopy. In the first 40 minutes, red-spectrum foci could be observed outside and onto the cells with no detectable preference for the poles. These red dots reflect the fluorescent labelling decorating the capsids of T5 PNmC particles (Figure 22, B). From 10 minutes onwards, red fluorescence was also detectable in the cells, consistent with the production of pb10-mCherry at the late stages of the infection. Thus, combining HR and DAS could be successfully implemented for genome editing in the SST-DNA of T5

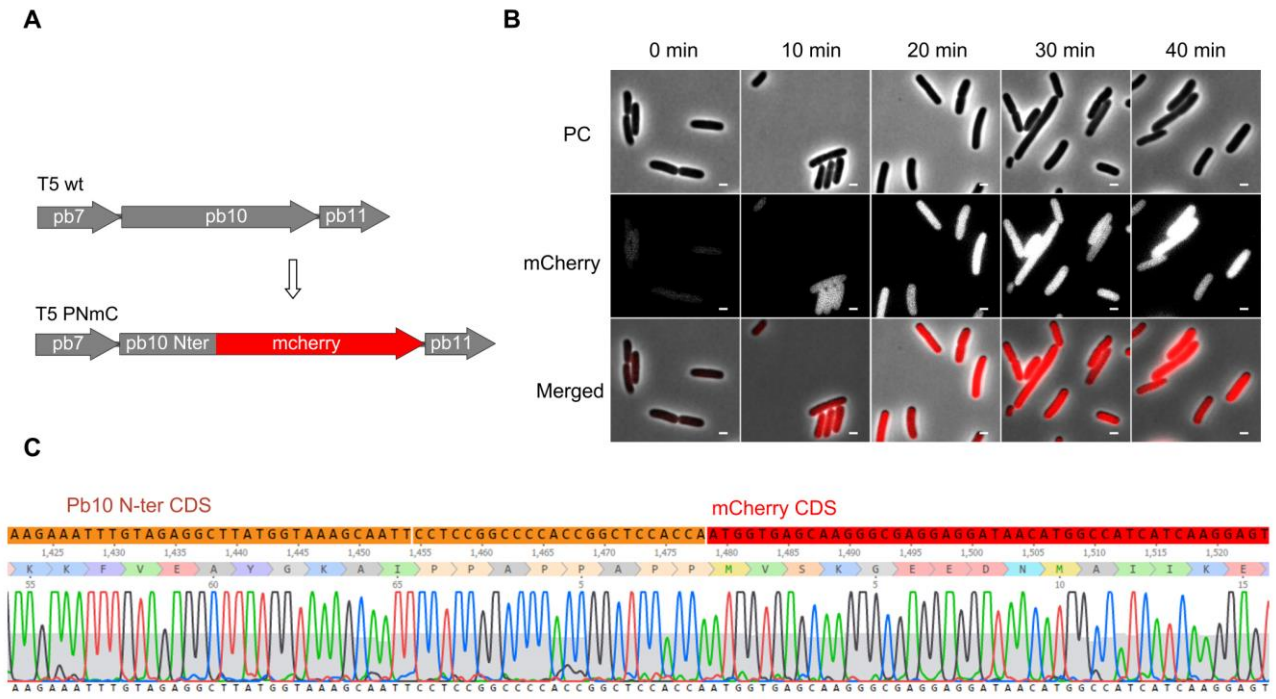


Figure 22. Construction of the mutant T5 PNmC.
(A) Partial map of the capsid gene locus in the genome of T5 PNmC mutant: a copy of the fusion protein *pb10*(Nter)-*mcherry* (Δ[103638..103928]::*mcherry*) replaces the wild-type pb10 gene. (B) Phase contrast (PC) and fluorescence (mCherry) microscopy of *E. coli* F cells infected by T5 PNmC. The time after infection is indicated above. White bars, 2 μm. (C) Sequencing results of the T5 PNmC genome locus highlighted in (A), the orange ORF corresponds to Pb10 (N terminus) while the red to mCherry.

In this chapter, we have successfully investigated several methods for editing the T5 genome. Here, in addition to the classical HR-based method using a plasmid carrying homologous sequences, we also used an alternative template DNA, i.e., bacterial retroelements called retrons. By targeting different locations across the FST locus with sgRNA-guided Cas9, a nuclease normally not present in *E. coli*, I found that interference was not uniform. Therefore, we developed DAS, an alternative method to improved recovery of mutants. DAS might be applicable to other virulent phages that are naturally resistant to CRISPR/Cas nucleases. All these strategies likely constitute welcome additions to the toolbox for the phage genome engineering. Having a hand of methods for T5 genome editing we could use reverse genetics to investigate the function of the pre-early genes. Thus, our next objective was to construct deletion mutants and to compare their phenotype of to that of T5 wt.

2.2. Deletion of phage T5 pre-early genes

The subsequent stages of our work consisted in assessing the significance of other pre-early genes besides *A1* or *A2* for T5 infection. It was previously suggested that some pre-early genes, (such as *02*, *05*, *07*) might be essential according to how well conserved they are among T5-related viruses (Davison, 2015). Since that work, more genome sequences of T5-like viruses were deposited to the databases (Sváb et al., 2018). Therefore, we decided to perform an extended analysis of all genes among *Tequintavirus*, *Markadamsvirinae* and *Demerecviridae*.

2.2.1. T5 pre-early genes conserved among *Tequintavirus* members

When comparing genomes, orthologous genes come from a single ancestral gene in the last common ancestor. Orthology does not necessarily mean conservation of gene function, but the search of orthologs is a proxy to understanding patterns of functional conservation or changes through the course of evolution (Koonin, 2005).

Orthologs can be grouped in orthogroups (Emms and Kelly, 2019). The size of orthogroups can constitute a measure for the gene conservation among different genomes. From fully sequenced and annotated phage genomes classified in the genus *Tequintavirus* (NCBI:txid187218), we determined the number of genes belonging to the same orthogroup as genes from the bacteriophage T5 (NC_005859), more specifically, we counted the number of genomes bearing at least one ortholog to each T5 gene as a proxy for the conservation of T5 genes. We applied the OrthoFinder algorithm (Section 6.2.1) (Emms and Kelly, 2019) on 111 genomes (12,321 genes in total, repeated genes were not included).

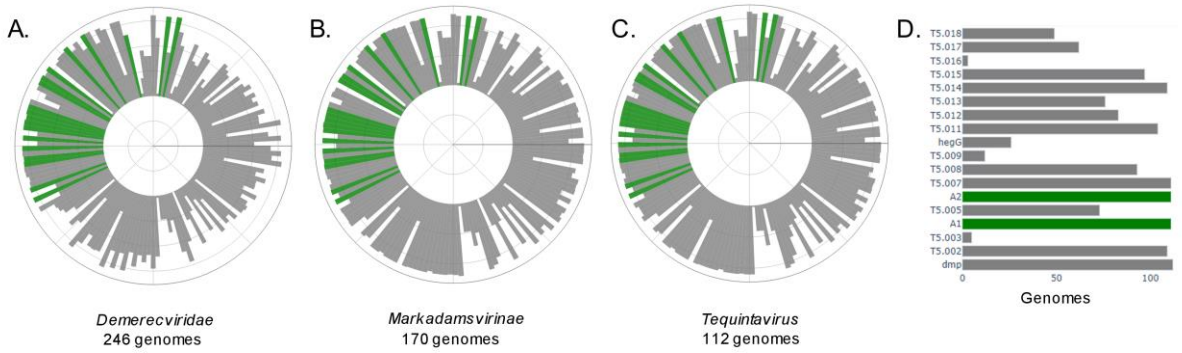


Figure 23: Presence of T5 genes among *Tequintavirus* members.
Genomes in (A) *Demerecvinidae*, (B) *Markadamsvirinae*, or (C) *Tequintavirus* with orthologs to T5 genes. (D) Genomes in *Tequintavirus* with orthologs to pre-early genes. Known essential genes were highlighted in green, unknown in gray. Bar lengths depict the frequency of genomes with at least one gene in the same orthogroup as each T5 gene. The bars in green correspond to essential genes, in gray the dispensable or unknown genes.

From a comparison of all the T5 genes in phages from the same family (Figure 23, A), subfamily (Figure 23, B), and species (Figure 23, C), a closer look at the LTR (Figure 23 C) shows that all the T5

viruses carry at least one ortholog gene to *dmp*, *A1*, *A2*, and 07, as it was suggested previously (Davison, 2015). There are other well-conserved genes: *02* (98%), *08* (84%), *11* (94%), *12* (74%), *14* (98%) and *15* (87%). Less conserved ones are *13* (68%), *05* (66%) and *17* (56%), while *03* (4%), *09* (10%), *hegG* (17%) and *16* (3%) are present in very few *Tequintavirus* members. For details about the rest of the genes, see Section 7.1.

Interestingly, when we performed the same analysis in T4 (AF158101.6) (Figure 24), for which all the essential genes are known (Miller et al., 2003). We found that most T4 genes are well conserved in *Tequatroviridae* (NCBI:txid10663, 200 genomes), disregarding they are essential or dispensable, but not so much among members of the *Tevenvirinae* (NCBI:txid1198136, 349 genomes) subfamily. This elicits that the phage genomes in the same genus might be too similar, while phage genomes from the same family are too diverse. As such, subfamilies are probably more suitable taxonomic level to contrast conserved and non-conserved genes. Although essential genes are, on average, more conserved than dispensable ones, not all the conserved genes are essential (Figure 24).

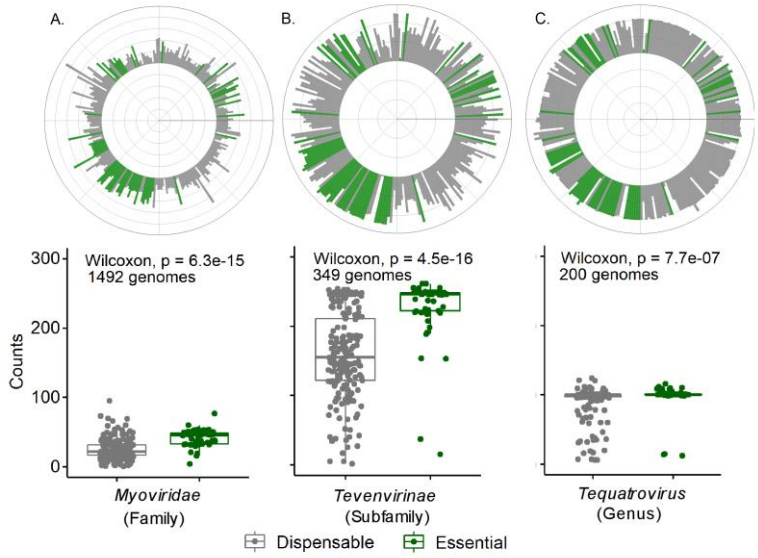


Figure 24. Presence of T4 genes in virus from the same taxa.
Conservation of genes from T4 was measured among phages in (A) *Myoviridae*, (B) *Tevenvirinae*, or (C) *Tequatrovirus*. Bar lengths depict the frequency of genomes with at least one gene in the same orthogroup as each T4 gene. The bars, dots and boxes in green correspond to essential genes, and in gray to dispensable genes. In polar plots, T4 genome starts in the upmost position and continues clockwise.

As a conclusion, in T5, only *dmp*, *A1*, *A2*, and *07* were consistently well conserved, which suggest that the other genes might not be essential. Therefore, we carried out the mutagenesis of all the pre-early phages directly by deletion to test if they are all dispensable.

2.2.2. At least 14 pre-early genes are dispensable for the infection of Phage T5.

We carried out the deletion of the genes *dmp*, *02*, *05*, and *07* (Figure 25, A to D, respectively) by HR and CRISPR-Cas9. In all cases, the ATG and the last 10 codons were kept. Then, we constructed multiple-deletion mutants on the genes *dmp-02-03*, and the genes *hegG-11-12-13-14-15-16-17* by HR and DAS. Finally, a combination of all these modifications yielded a strain lacking all the dispensable genes on the left half of the FST locus, T5 DL or T5 Δ*dmp-02-03* Δ*05* Δ*07* (Figure 25, E), all the genes on the right half of the locus, T5 DR or T5 Δ*hegG-17* (Figure 25, F), and a combination of both, T5 DLDR or Δ*dmp-02-03* Δ*05* Δ*07* Δ*hegG-17* (Figure 25, G). These T5 mutant strains were confirmed by Sanger sequencing (Figure 26) of the region including the deletion (on-target). It is important to notice that Sanger sequencing gives no data about off-target mutations.

We could not mutate the genes *08* and *09*. The reason is most likely the inability to clone the template required for genome editing through HR into a high-copy-number or a mid-copy-number plasmid, even after many attempts. Nevertheless, both genes could be cloned separately in a mid-copy-number plasmid, which expression is under the control of an Arabinose-inducible promoter (Section 2.4.2). Interestingly, some intergenic regions were not possible to clone, probably because of the presence of a strong promoter (Section 7.2). The list and pedigree of mutants obtained in this study is presented in the section 6.1.3.

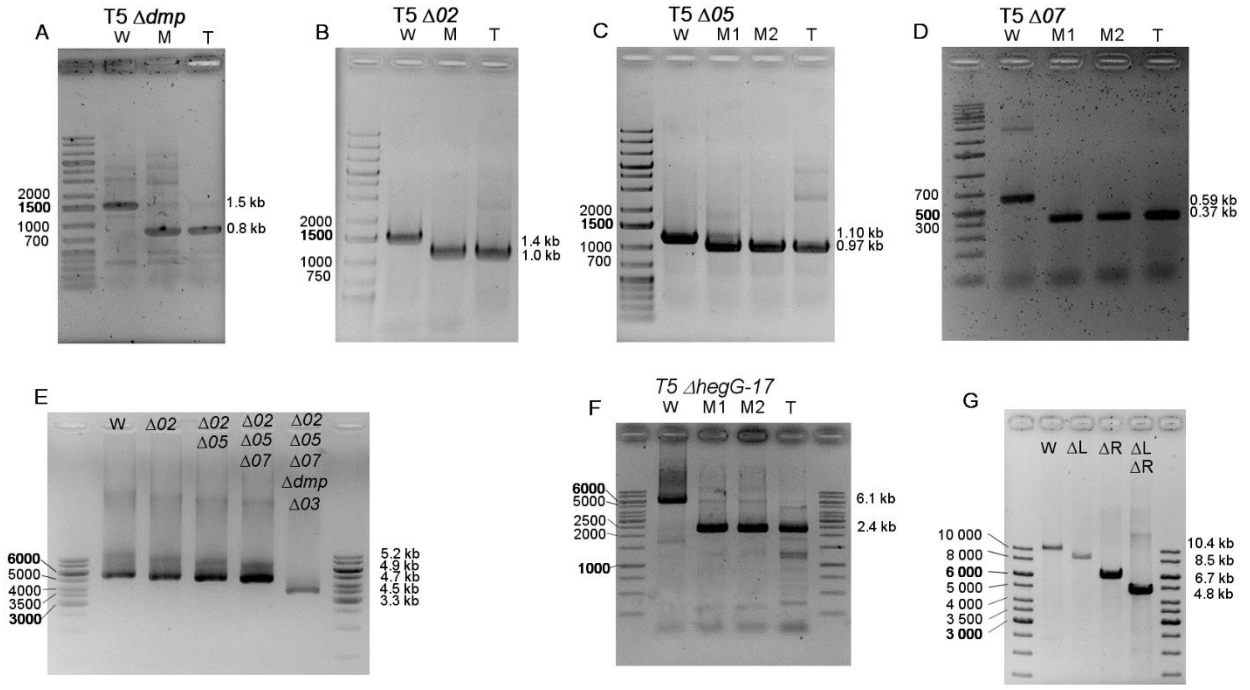


Figure 25. Deletion of pre-early genes in T5.
PCR products amplified by the primers flanking the deleted region in each mutant. Single gene deletions for the genes *dmp* (A), *02* (B), *05* (C), *07* (D) were constructed, along with combinations of them, T5 DL, (E), through HR and CRISPR-Cas9. A T5 lacking the genes from *hegG* to *17*, T5 DR, (F) was constructed by HR and DAS. Finally, all the mutations were combined to delete 13 out of 17 genes, T5 DLDR, (G).

It should be noted that the amplification of the mutant DNAs by PCR showed only one fragment, which suggest that both terminal repeats were modified. To test this, we purified genomic DNA from the mutants and the wild type and sent them to sequence.

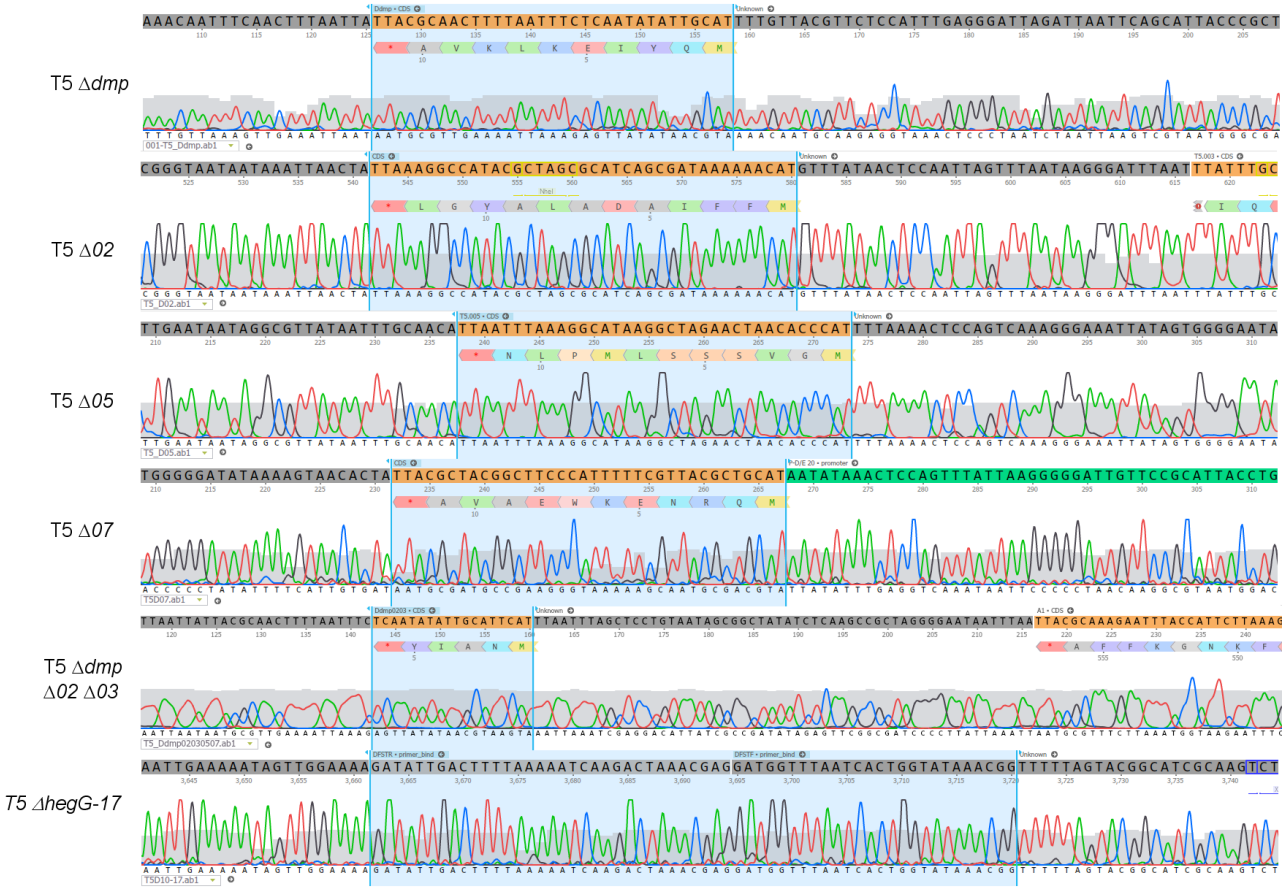


Figure 26. Sanger sequencing of mutants obtained in this study.
Phage DNAs were amplified by PCR, purified, and sent to be sequenced. The references depict the expected modification, the zones highlighted in blue constitute either the deleted gene or the closest regions to the newly formed sequence.

2.2.3. Whole genome sequencing

Four phage DNAs were sequenced in the Sequencing platform at the I2BC. The results showed a good quality of the reads. These reads were then mapped through Bowtie2 and indexed with SAMtools. As T5 bears two direct terminal repeats, so, we constructed each T5 mutant genome map with the LTR bearing the intended mutations, this would permit to visualize newly formed sequences (scars from the gene editing), and kept the RTR unaltered so the wild type genes would be mapped by the reads in case they are still present in the genome, i.e., to detect if at least one of the terminal repeats remained unedited or in case the wild-type phage is still present in the sample (Figure 27).

As a control, the T5wt genome was sequenced and according to the results, it showed no alteration regarding the sequence deposited to the NCBI on both terminal repeats (Figure 28). Short reads mapping was set to the default so every read could be present in all the sites it fully matches.

Short reads from the T5 DL genome mapped correctly to the mutated LTR, while on the RTR, with a wild type version of the FST locus and part of the ISS, no read mapped to the genes *dmp*, *02*, *03*, *05*, and *07*, which indicates that both direct terminal repeats in the strain T5DL were mutated.

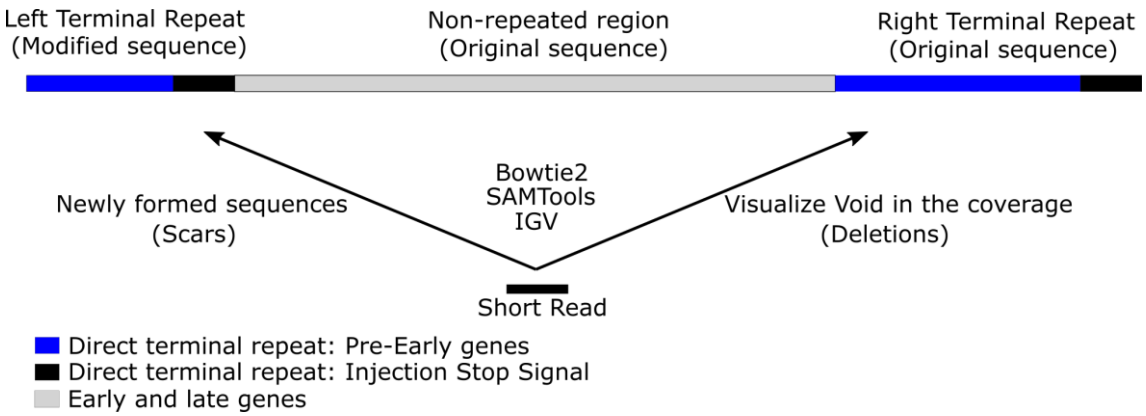


Figure 27. T5 genome sequence was modified for read mapping.
The genomic sequence of the different strains of T5 was constructed according to the results obtained from Sanger sequencing. The mutations were introduced on the left terminal repeat to map the newly formed sequences (deletions) while the right terminal repeat was left as the wild type to visualize lack of coverage. Every short read could map both terminal repeats whenever a high quality is achieved.

Regarding the T5 DR, the newly formed sequences are present in the phage genome. It mapped all the genes on the left half of the FST, namely *dmp*, *02*, *03*, *05*, and *07*, but none from *hegG* to *17*. Nevertheless, we found an unexpected deletion at the ISS (9,875..10,340). The template plasmid pUCDFSTR used for HR had no modifications at the ISS. However, T5 DR comes from the strain T5 ris16IS (Section 6.1.3), which results from the HR between T5wt and pUCEco16IS, and this last descends from the plasmid pUCISSinv (Figure 58). Thus, we traced the unintended modification on the ISS back to pUCISSinv, which was constructed by HR cloning. pUCISSinv spontaneously removed the region 9,570 to 10,216 from T5. It is worth noticing that there are 630 bp of unaltered sequence between the "missing ISS fragment" and the *hegG-17* deletion. The unexpected deletion at the ISS, prove that, this region is in fact, dispensable for the phage infection, a topic we will address later on in the section 2.5.

The strain T5 DLDR was constructed independently from the strain T5 DR (Section 6.1.3), and does not bear the same mutation on the ISS as T5 DR does. On top of the on-target deletions made on the genes *dmp*, *02*, *03*, *05*, and *07*, T5DLDR lacks genes *hegG* to *17*.

Off-target mutations in the constructed strains were identified by the I2BC Sequencing Platform. These are *042* (E199G), *068* (N42S) in the mutants T5 DL and T5DLDR. T5 DR has different mutations as the other two: *080* (D127N) and *132* (Indel). T5 DLDR has *085* (T95N), *D13* (T566K), plus *D15* (H83Y) (Table 4).

T5.042 (YP_006870.1) codes for a putative Clp protease. T5.080 (YP_006908.1) codes for a protein with one putative transmembrane domain, it holds a similar HMM profile to stomatins from *Pyrococcus horikoshii* (P: 99%) and *Mus musculus* (P: 98%), as well as flagelar proteins from *Vibrio* spp. T5.085 product (YP_006873) is similar to an unknown protein from *Corynebacterium diphteriae* (3KDQ_A, Probability 99.9, E-value 1.6e-21). D13 product is a endonuclease, similar to T4 g46 (AAC05392.1) (Blinov et al., 1989). D15 codes for the T5-exonuclease, needed for the elongation of single-strand interruptions in the T5 DNA (Moyer and Rothe, 1977).

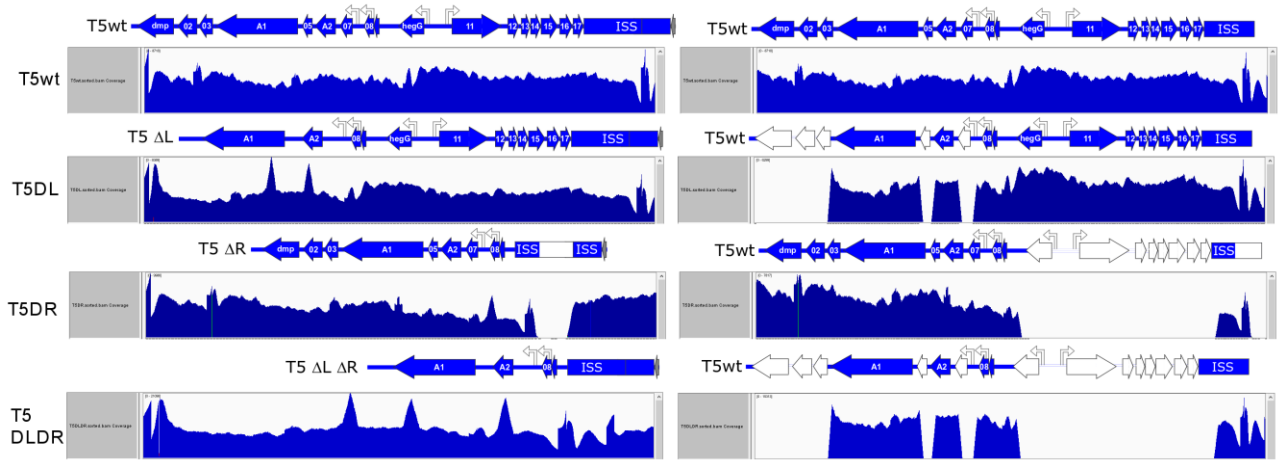


Figure 28. Whole genome sequencing of T5 mutants.
T5 mutant maps were constructed based on the Sanger sequencing results, they include on the (left) LTR the mutated version of the FST (DL, DR, DLDR) and on the (Right) RTR, the wild-type one. The regions in blue at the coverage track were well mapped, while the regions in white did not or were poorly covered. Sequencing was carried out in NextSeq Illumina (80 cycles, paired ends), aligned with bowtie2, and visualized in IGV.

Table 4. Off-target mutations on the T5 genome.

Mutant			T5DLDR	T5DL	T5DR	Gene	Residue	Product
Position	**ref**	**mut**						
10696	G	C			x	NA		
21158	T	C	x	x		*042*	E199G	Protease
33725	T	C	x	x		*068*	N42S	Unknown
40360	C	T			x	*080*	D127N	Transmembrane
43860	G	T	x			*085*	T95K	Unknown
54515	C	A	x			*nrdD*	V185V	Ribonucleoside triphosphate reductase
79465	C	A	x			*D13*	T566K	Endonuclease
80324	C	T	x			*D15*	H83Y	T5-exo
81540	GTT	GTTT			x	*132*		Unknown

More studies in the future will shed light on the fitness cost or compensation effects of those off-targets modifications. We characterized the strains T5 DL, T5 DR, T5 DLDR, along with the single-gene mutants by different methods, the results of such characterization are exposed in the next section.

2.3. Characterization of T5 mutants

2.3.1. Deletion of pre-early genes alter the infection kinetics

The life cycle of a virulent phage can be broken down to the following steps: Adsorption, attachment and genome delivery, latent period, and virus release. The eclipse period starts at the genome delivery and correspond to the duration of virion assembly, while the latent period spans between virus attachment and the progeny release. The burst size corresponds to the number of phage progeny produced per infection. These three parameters can be measured with the One-Step-Growth experiment (OSG) (Hyman and Abedon, 2009; Kropinski, 2018).

We carried out the characterization of the single-gene (Figure 29) and multiple-gene (Figure 30) mutants by OSG experiment according to protocol adapted from (Kropinski, 2018). The OSG for each mutant was drawn over 90 min by measuring plaque forming units every five minutes post-infection (Figure 29, A), from which we calculated the burst size as the ratio of the titers at the last over the first three time points. The eclipse was measured by adding chloroform to the aliquots to lyse the cells and interrupt the infection. Both eclipse and latency were considered as the last time point before a tenfold rise in the phage titer.

The T5wt produced over 350 phages per infection, after an eclipse of 25 min and a latency of 40 minutes. We found that mutant burst sizes were about 40 % larger than that of the wild type for strains lacking genes *02*, *05*, or *07*, while the mutant on gene *dmp* yielded a smaller burst size (Figure 29, B). In parallel to this, T5 Δ*dmp* had a longer eclipse and latency, whereas the eclipse remained practically unaltered and the latency lasted longer whilst for the mutants Δ*02*, Δ*05*, and Δ*07* (Figure 29, C). Thus, it is probable that the progeny release and not virion assembly is perturbed in Δ*02*, Δ*05*, and Δ*07*, while virion assembly and progeny release are delayed in T5 Δ*dmp*. Moreover, T5 Δ*dmp* titer rise is slower than T5wt's as seen before (Mozer et al., 1977).

A comparison of the characteristics of the multiple deletion mutants T5 D257, DL, DR, and DLDR revealed that T5 D257 produces a larger bursts (Figure 30), its eclipse period remains unaltered, and the latency is delayed, similarly as for the strains T5 Δ*02*, Δ*05*, or Δ*07*.

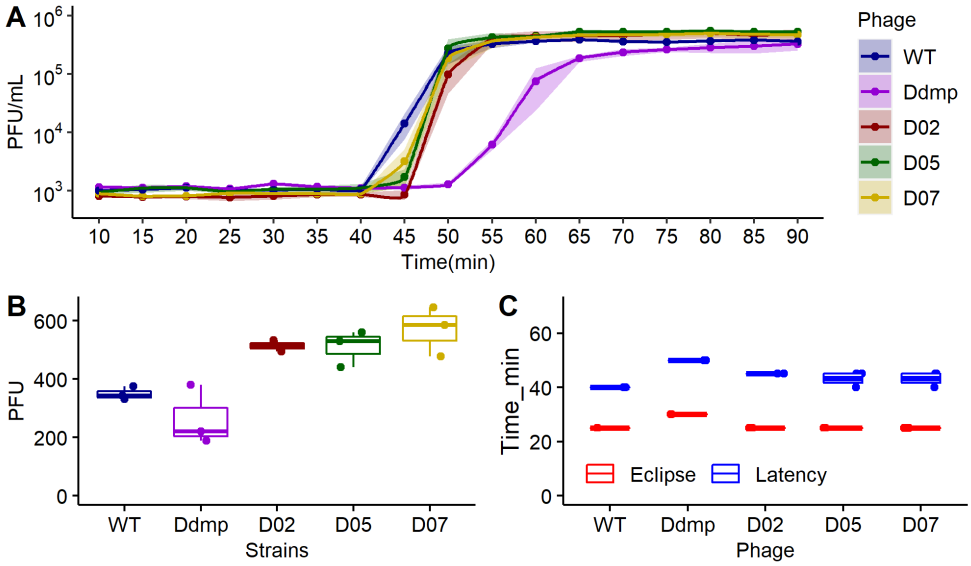


Figure 29. Infection kinetics of single-gene mutants of T5.
The plot indicates (A) One-Step-Growth (OSG) curve as the titer in PFU/mL versus time in minutes, after the addition of the phage dilution. (B) The burst size is depicted as PFU, it was obtained as a ratio between the averaged titers from the last and starting three time points. (C) The Eclipse and latent periods were calculated the last time point before a ten-fold rise of titer. Each data point is a mean and is surrounded by a colored ribbon as the s.e.m., both from of three experiments.

T5 DL, DR, and DLDR showed different phenotypes. T5 DL is one of the most affected: it produces a smaller burst, the eclipse and latency are delayed, on top of this, the phage release is slower than in the wild type. T5 DR showed similar burst size as the wild type with unaltered eclipse but with a delayed latency. Finally, T5 DLDR depicted an almost unaltered burst size, with longer eclipse and latent periods.

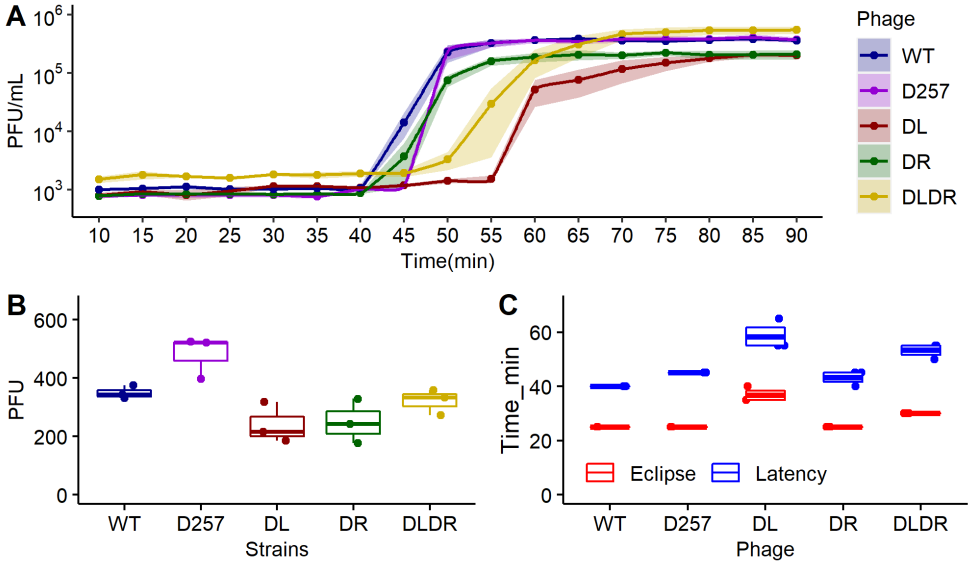


Figure 30. Infection kinetics of multiple-deletion mutants of T5.
The plots depict the (A) One step growth curve as the titer in PFU/mL versus time in minutes, (B) Burst size, (C) Latent and Eclipse periods. Data points and the shadowed ribbon are the mean and s.e.m of three experiments, respectively.

To test whether the differences were statistically significant, we made a comparison of the mutants and the wild type. The results indicate that the deletion of *02* leads to a higher burst size,

while the deletion of *dmp* leads concomitantly to a delay in the virus assembly and the release of the progeny (Table 5).

Table 5. Pairwise comparison of growth parameters.

Phage	Burst_size	Eclipse	Latency
Δ*dmp*	0.910	**$1.49x10^{-4}$**	**$6.61x10^{-3}$**
Δ*02*	**0.036**	1.00	0.417
Δ*05*	0.316	1.00	0.842
Δ*07*	0.066	1.00	0.842
D257	0.563	1.00	0.417
DL	0.746	**$4.86x10^{-10}$**	**$4.63x10^{-6}$**
DR	0.824	1.00	0.842
DLDR	0.999	**$1.49x10^{-4}$**	**$3.07x10^{-4}$**

Tukey's HSD post-hoc test for a comparison between mutants versus the wild type strains regarding the Burst size, Eclipse, and Latent periods obtained from the One Step Growth assays (Figure 29, Figure 30).

Certainly, some of the mutations alter the normal viral cycle, so, to assess the impact of these deletions on the mutant virulence we endeavored to evaluate the effects of the deletions on the phage virulence.

2.3.2. Deletions of pre-early genes impact the phage virulence

The virulence of a phage reflexes its killing capacity and is mainly affected by the phage fitness, the burst size, eclipse, latency time, efficiency of plating, and the titer. (Storms et al., 2020). One proxy to phage virulence is to study the reduction of the bacterial population when infected by the phage. The reduction in the bacterial population might be studied on solid media or in liquid (Konopacki et al., 2020; Storms et al., 2020).

Two models were recently proposed to evaluate the phage virulence on a specific host: the PhageScore (Konopacki et al., 2020) and the Phage Virulence Index (Storms et al., 2020). Both rely on the optical density reduction curve of a bacterial culture in liquid media infected at different MOIs to give the virulence a numeric value. The PhageScore is based on bacterial biomass concentration changes related to changes in the area under the growth curves (AUC) to gauge the lytic potential (Konopacki et al., 2020). The Phage virulence index (PVI) compares the optical density reduction curves to the uninfected control through the ratio between the AUCs (Storms et al., 2020). The fundamental difference between both methods is that PhageScore is a parametric index, since it makes some assumptions about the mathematical rules describing phage-bacteria interaction, which renders the estimations more resilient to outliers, while PVI is non-parametric and determines the AUC by the trapezoidal method of integration. We made use of the PVI since we have no robust data to support any theoretical model describing the bacterial biomass behavior over time during the infection of T5 mutant phages, and it would be wiser to explore them with no assumption to rely on a parametric method in the future.

We carried out a preliminary analysis and evaluated the bacterial optical density at different times in liquid media and good aeration in erlenmeyers at an MOI of 0.01. The bacterial cultures of *E. coli* strain F were inoculated at an OD_{600} of 0.1 and OD_{600} measurements were started when they reached an OD_{600} of 0.2. We added the phages when they reached an OD_{600} of 0.4.

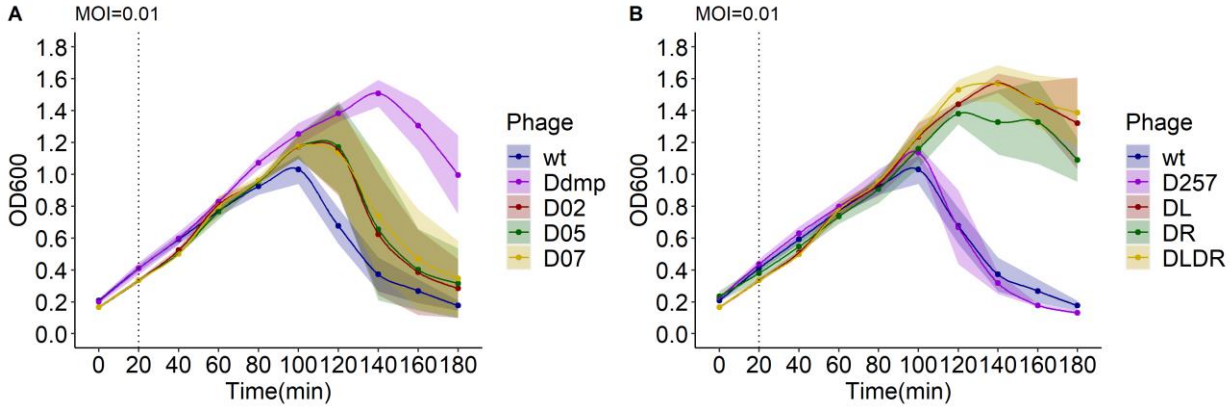


Figure 31. Optical density reduction curves from infections in liquid medium.
The optical density of bacterial cultures infected by T5 (A) single-deletion or (B) multiple-deletion mutants. Each time point and the surrounding ribbon is the average and s.e.m., respectively, from three replicas. Data collected every 20 minutes from cultures incubated in Erlenmeyer.

Wild-type T5 starts depleting the bacterial growth at 80 min and the OD_{600} stabilizes at 120-140 minutes. The culture infected by T5 Δ*dmp* started to decline at 120 min, while the cultures infected by T5 Δ*02*, Δ*05*, or Δ*07* started to decline at 100 minutes and stabilized between 120-140 minutes, as for the wild type (Figure 32, A).

On the other hand, bacterial cultures infected by T5 strains with multiple mutations on genes *02*, *05*, and *07* (T5 D257) behave similarly as cultures infected by the wild type. Bacterial cultures infected with T5 DL (similar to T5 D257 but lacking *dmp* and *03*) started to fall with a delay of more than 40 minutes regarding the wild-type cultures, similarly to the mutants DR and DLDR (Figure 32, B). A comparison between the curves from T5 Δ*dmp* and T5 DL suggests that the delay in the T5 DL might be mainly due to the deletion of *dmp*. The delay in the reduction of bacterial cultures turbidity might be an effect at the population level of phage kinetic characteristics, such as the latency and the burst size, tested through the OSG curves (Section 2.3.1).

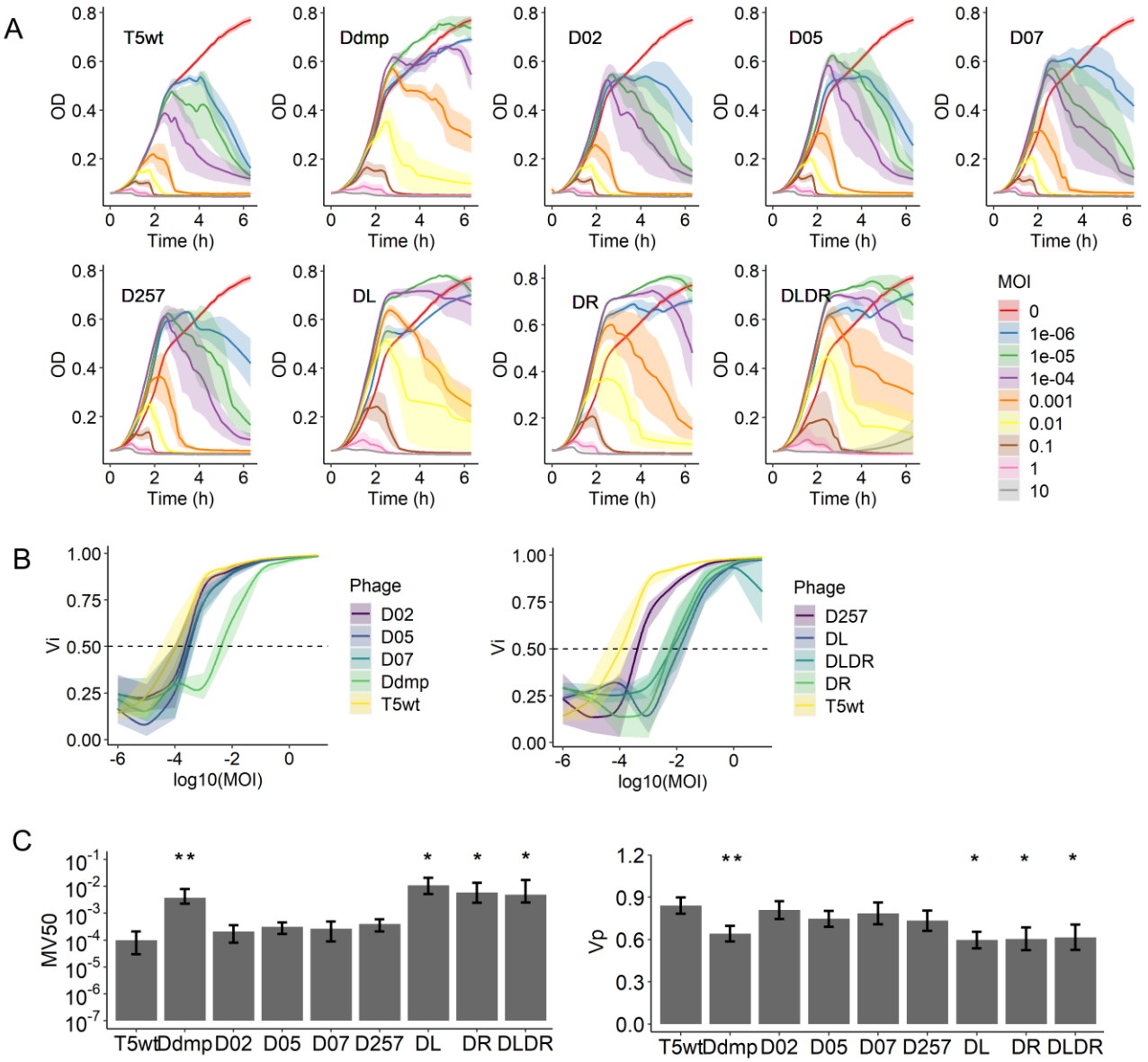


Figure 32. Virulence of mutants by deletion in T5.
(A) Bacterial reduction curves of cultures infected with T5 mutants. (B) Local virulence curves represent the Area under the curve (AUC) of bacterial reduction curves versus the logarithm of multiplicity of infection (MOI). Lines and colored ribbons indicate the mean and mean ± s.e.m., respectively, from three replicas. (C) MV50 and virulence index (Vp) as mean ± s.e.m. *p*-values indicated in Table 6. * p <0.05, ** p < 0.01.

To determine the PVI as Vp, we carried out an infection of bacterial cultures at different MOIs, from10^{-6} to 10^{1}, and followed the OD_{600} for up to six hours in a multi-plate reader, measuring every ten minutes and starting with a culture in exponential phase. The maximum doubling time for *E. coli* strain F was 20±0.6 minutes.

The most affected single-gene-deleted phage strain was T5 Δ*dmp*, which, at MOIs lower than 10^{-3} could not lyse its host after six hours, while the cultures challenged with strains T5 Δ*02*, Δ*05*, Δ*07*, and D257 behave similarly as those with the wild type. The multiple deletion mutants T5 DL, DR, and DLDR, showed the same effect as the deletion of *dmp*: below 10^{-3}, which probably infected not fast enough to keep the host from reaching the stationary phase. Puzzlingly, T5 DR produced a sudden

fall in the bacterial turbidity at approximately four hours, after the host had reached stationary phase (Figure 32, A).

The local virulence is represented as the AUC from the bacterial reduction curves, ranging from 0 to 1. Zero represents no virulence, while one represents the maximum theoretical virulence, which in empiric terms might be translated as an immediate bacterial lysis. From this curve, it is possible to deduce two parameters: MV_{50} (the MOI that corresponds to half the maximal virulence) and Vp (Figure 32, B).

The MV_{50} and Virulence Index show that all the mutants devoid of *dmp* (Δ*dmp*, DL, and DLDR) exhibited a reduced virulence index compared to wild-type, which corresponds to a 50- to 100-fold increase in MV_{50} (Figure 32, C and Table 6). These data are in agreement with the delayed kinetics of infection described above for these mutants. For all the other single-gene deletions besides those on *02*, *05*, and *07*, notwithstanding their delay in the latency, there is no difference between them and the wild-type strain regarding the virulence. Less expectedly, mutant T5 DR was as impaired as the mutants devoid of *dmp* (Figure 32, C and Table 6), suggesting that some gene(s) in the right arm of the FST promote(s) phage virulence.

Taken together, these results indicate that in addition to the essential genes *A1* and *A2*, *dmp* and one or more genes located on the right hand of the FST facilitate infection of *E. coli* by phage T5.

Table 6. Dunn's-test from comparison of phage virulence [a].

	MV50	Virulence Index
Kruskal-Wallis test	0.0100	0.01265
T5 Δdmp	0.0140	0.0159
T5 DL	0.0047	0.0040
T5 DR	0.0108	0.0108
T5 DLDR	0.0123	0.0108

[a] comparisons listed only for $p < 0.05$

2.3.3. Cytological analysis of infected bacteria

During the infection by a virulent phage, bacterial metabolism is redirected towards the synthesis and assembly of the viral progeny. Inevitably and concomitantly, such a detour in the metabolism lead to morphological changes of the infected cells. Therefore, to further explore the mechanism of action of the different phage effectors deployed pre-early on, and their effects later during the infectious process, we studied the differential effects of the constructed mutants on the bacterial morphology.

Some gene products were linked in the past to host DNA degradation (A1), nucleotides dephosphorylation (Dmp), replication halting (unknown), to study the host cell biology during

infection, bacteria harvested in exponential phase of the growth were infected by T5 or its mutants and fixed at different time intervals post infection. We evaluated the cell morphology of the infected cells by phase-contrast (PC) and fluorescence microscopy, using the fluorescent dyes DAPI and FM 4-64. DAPI (358/461 nm) binds strongly to AT-rich DNA regions and reflex the DNA content of the cell, while FM4-64 (515/640 nm) is a lipophilic dye that reveals the cell membranes (Figure 33).

Infection by the wild-type phage leads to a decline in the DAPI staining of the cells in the first ten minutes, that can be seen as a decrease in the average intensity of the DNA (WT-T20 and WT-T10 versus WT-T0, Figure 34) at 10 and 20 minutes. The decrease in the DAPI fluorescence is followed by an increase of its intensity at 30 minutes, linked to the phage DNA replication. The aspect of the DAPI staining was diffuse suggesting there is no confinement of the replicating T5 DNA in the bacterial cell (WT-T30 in Figure 34 and Section 7.3, Figure 65) as seen for other phages (Kraemer et al., 2012; Malone et al., 2020; Mendoza et al., 2020). In parallel, a loss of refringence of the cells and lack of staining with FM4-64 at 30 minutes reveal the onset of the lysis. Remarkably, through phase contrast we could observe the apparition of dots with no preferential location in the cell.

Amber mutants T5 *amA1* and T5 *amA2* led to no lysis after 30 minutes of infection but displayed two different phenotypes. Cells infected by T5 *amA1* displayed higher DAPI fluorescence intensity at all intervals than all the tested strains, there is however a decrease of DAPI intensity in the first 10 minutes, which stagnates until 30 minutes. The higher DAPI intensity of cells infected by the *A1*⁻ strain versus the wild type phage, at T0, suggests that the DNA degradation takes place rapidly before the fixation of the cells in the wild type. As a consequence, the imaging of infected bacteria at T0 does not reflect the real initial bacterial DNA content. On the other hand, the decrease in the DAPI fluorescence intensity from cells infected by T5 *amA1* hints that probably *A1* is not the only gene associated to the host DNA degradation, all this regarding that only one phage per million reverts to a wild type genotype (Section 2.1.4, Figure 20). Cells contained, all the same, black dots as when infected by the wild-type phage.

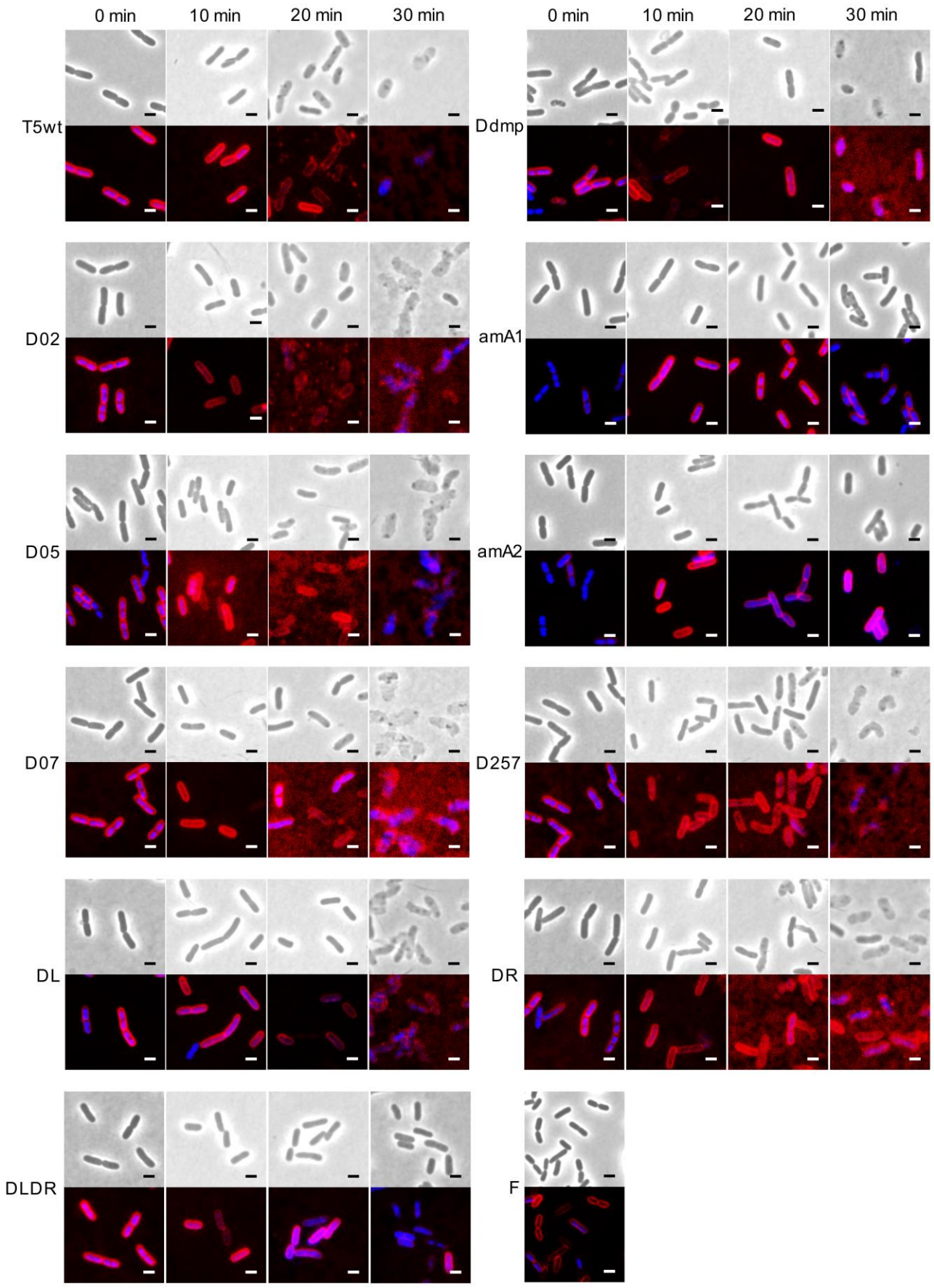


Figure 33. Microscopy images of bacteria infected by T5 mutants.
Bacterial cultures were infected at exponential phase at MOI of 10, incubated and fixed at different intervals. They were imaged by fluorescence microscopy after staining with dyes DAPI and FM4-64. (Scale bar, 2 μm).

The amber mutant on *A2* depicted a similar pattern of infection. Although the DAPI intensity reached lowest levels at 10 minutes and kept low, with no positive slope at 30 minutes due to the absence of phage replication, it did not disappear completely. This suggest that either not all the DNA was degraded, or some cells continued to replicate.

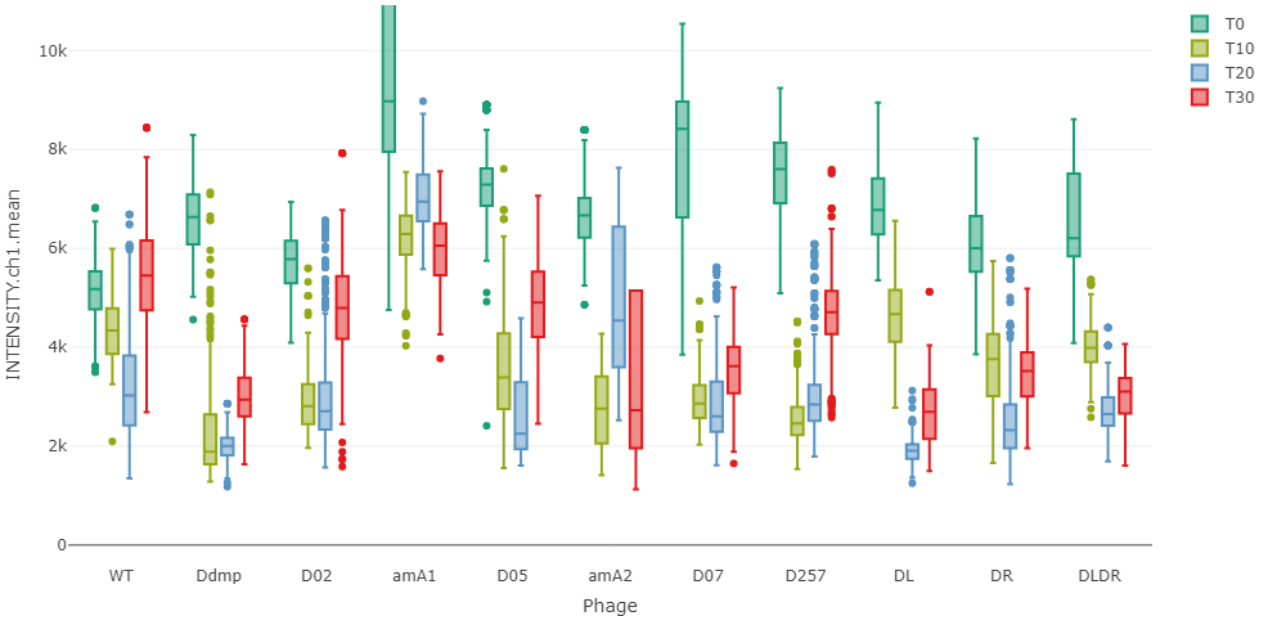


Figure 34. Fluorescence intensity of cells infected by T5 mutants.
The mean intensity of DAPI dye from cells infected by mutant phages over time. The intensity is correlated to the concentration of DAPI associated to the DNA.

For all the other mutants besides that on *A1* or on*A2*, cell DNA content goes down to the lowest levels in the first 10 or 20 minutes to go upwards near the end of the infection. Remarkably, the rise of DAPI intensity in cells infected by the less-virulent mutants Ddmp, DL, DR, and DLDR, denotes a lower amount of DNA synthesized near the end of the infection, which might explain the delay in the eclipse/latency.

All the strains tested apart from T5 DLDR, led to the production of the black foci, as seen under PC. T5 DLDR bears both DL and DR deletions and yet, these strains did lead to the formation of black foci in the cells at 30 minutes. It remains to be determined if such black foci appear afterwards and what is their function.

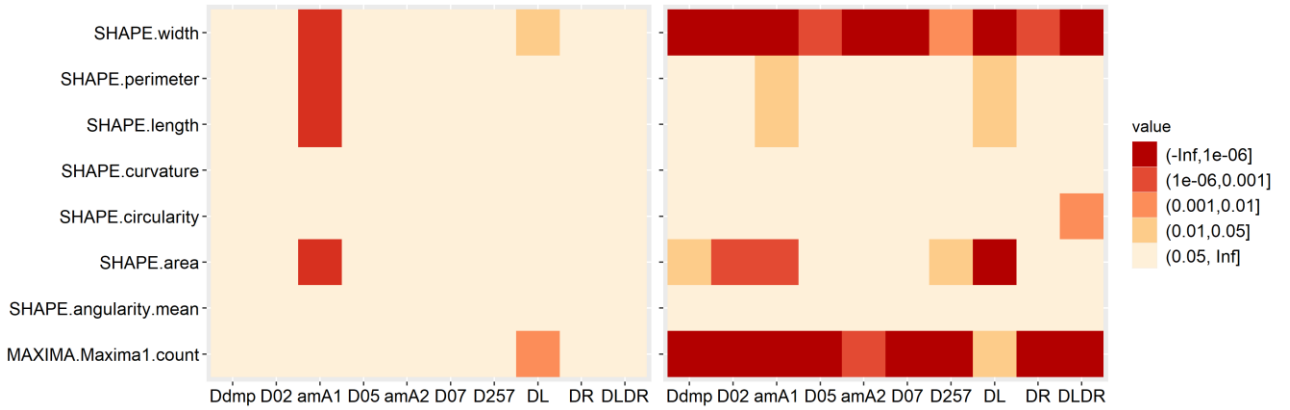


Figure 35. Comparison of shape parameters of T5-infected cells at T0 and T10.
Phase contrast images were treated to measure the dimensions of cells under infection at two time points: (left) 0 minutes and (right) 10 minutes. The shape parameters were compared to those of cells infected by the wild-type virus. The colors of cells are linked to the *p* value obtained from a Tukey *post hoc* analysis.

Shape parameters of the infected cells also presented some differences were also altered in the first 20 minutes of infection compared to the wild type. Thus, at time 0 minutes, we could see that only the cells infected by T5 *amA1* showed a significant difference to those with the wild type, regarding the width, length, perimeter, and area. This suggests that A1 absence leads to an increase of the cell size, probably without replication, and thereby A1 is probably not the only one but the most important gene to halt bacterial development in the first stages of the infection (Figure 35).

2.4. Functional exploration of T5 pre-early genes

2.4.1. Bioinformatic analysis of ORFan pre-early genes

Many functions were attributed to the pre-early proteins. However, most of the pre-early genes encoding these proteins are ORFan genes.

Therefore, our first approach to suggest a function for each was a bioinformatic analysis to look for distant homology through HHPred (Zimmermann et al., 2018), transmembrane domains through TMHMM (Krogh et al., 2001), signal peptides SignalP (Dyrløv Bendtsen et al., 2004), or test whether they might be anti-CRISPR proteins with PaCRISPR (Wang et al., 2020). The most relevant results are depicted in the Table 7.

Dmp HMM profile is similar to those of phosphatases, to which the poly nucleotide kinase (PNK) from the phage T4 (HHPred Probability: 96%). T4 PNK is a bifunctional protein with 5' kinase/3' Phosphatase, in which the 3' phosphatase activity does not require ATP. The phosphatase bears a DxDxT box present in many phosphotransferases (Wang and Shuman, 2002). The Dmp Asp24 (D24) is the most conserved residue and is included in a DxDxT motif as well (Figure 36). These results are in agreement with the previously demonstrated monophosphatase activity of the Dmp protein (Mozer and Warner, 1977)

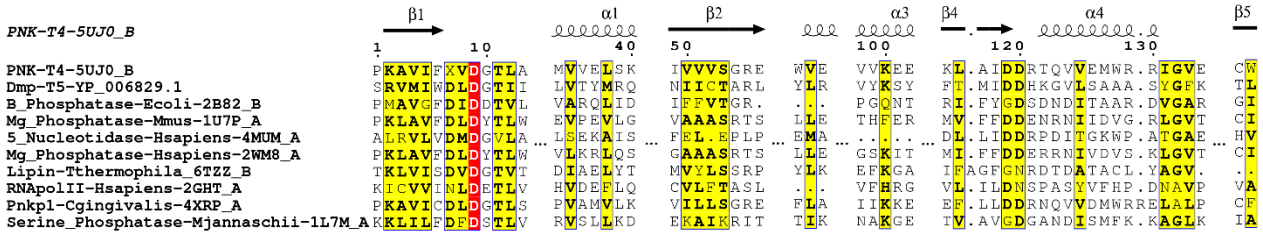


Figure 36: Alignment of Dmp with other phage-encoded similar proteins.
HMM-HMM profile comparison Dmp and proteins with known activity. Conserved residues are highlighted in red, while less conserved ones are in yellow.

Gp2 HMM profile match to several members of the poly(ADP-ribose) polymerase-1 (PARP-1) family (HHPred probability of 95). PARP-1 binds to nicked DNA, which activates its activity to catalyze the poly ADP-ribosylation of acceptor proteins (using NAD+ as a substrate) that are involved in DNA repair (Tao et al., 2008). This hit suggests that Gp02, like the dispensable early proteins Alt, ModA, and ModB of T4 may function as an ADP-ribosyl transferase.

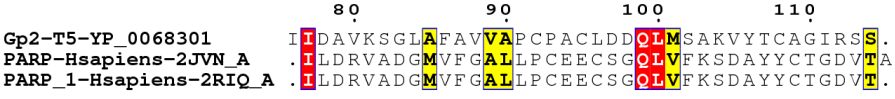


Figure 37. Amino acid sequence alignment of Gp2 and PARP enzymes.
HMM-HMM profile comparison of Gp2 and the domains A and D from PARP enzymes. Conserved residues are highlighted in red, while less conserved ones are in yellow.

HHPred searches on A1 hit several DNases (probabilities > 98%): the 3'-5' proof-reading exonuclease domain of the DNA polymerase from *Pyrococcus abyssi*, as well as Mre11 and SbcD exonucleases involved in the repair of DNA double-strand breaks (Figure 38). Although this hit covers a short region of homology, these results strongly suggest that A1 is a DNase. The DNase activity of A1 was confirmed *in vitro* using purified A1 protein (thesis Zangelmi, 2018).

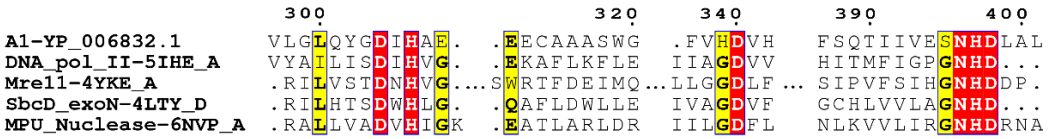


Figure 38. Protein sequence alignment of A1 and selected enzymes with known activity.
DNases with similar HMM profiles were aligned along with A1. Conserved residues that are essential for the DNase activity of DNAPol (5HIE), Mre11 (4YKE), SbcD (4LTY) and MPU (6NPV) are present in A1 (in red).

The protein A2 HHM profile hit a DNA-binding 3-helical bundle fold conserved in many DNA binding proteins, among them, the protein Timeless and transcription factor PAX (Figure 39). Timeless is a protein that recognize G-quadruplex (G4) secondary structures of DNA formed during replication for them to be repaired by the helicase DDX1 (Lerner et al., 2020). HHpred hits reinforce the prediction of a DNA-binding activity but does not suggests any specific function for A2.

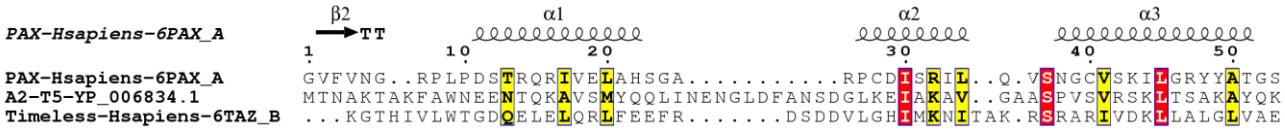


Figure 39. A2 alignment with DNA binding proteins.
A2 exhibit an HMM profile similar to that of several 3-helix DNA binding domains (Probabilities > 95 %). Conserved and similar residues are highlighted in red and yellow.

A search for transmembrane domains (TM) reveals that Gp5 and Gp13 are predicted to be membrane proteins with two TM each. Gp5 N- and C-terminus might be at the cytoplasmic side (Figure 40), while N- and C- extremities of Gp13 might be in opposite sense (Figure 41 A). Moreover, the first TM of Gp13 is likely to be a signal peptide according to SignalP (Table 7, Figure 41 B). Gp5 HMM profile is not significantly similar to proteins with known activity. On the contrary, Gp13 HMM profile is similar to some reductases (Probability 93%) (Figure 42)

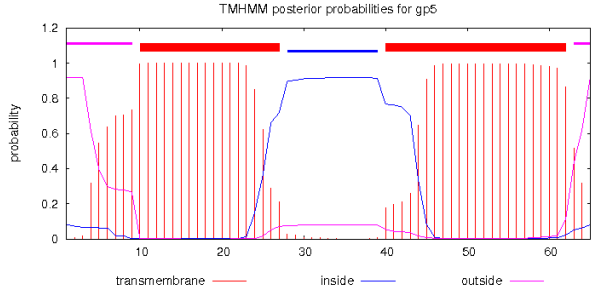


Figure 40. Transmembrane domains prediction for Gp5.
Gp5 is predicted to have two transmembrane domains. Gp5 N- and C-terminus are probably at the cytoplasm.

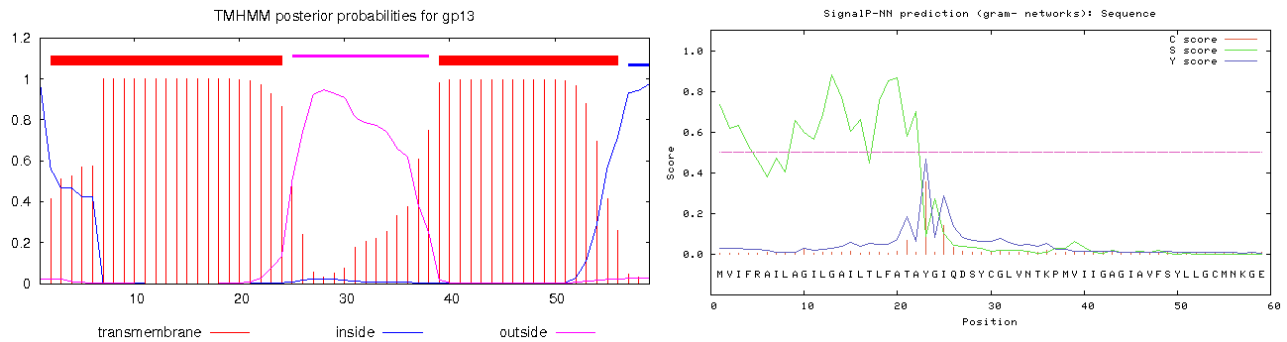


Figure 41. Transmembrane domains and signal peptide prediction for Gp13.
(Left) Gp13 is predicted to have two transmembrane. (Right) Gp13 probably bear a signal peptide at the N-terminus.

HegG exhibits strong similarity (HHPred probability >99%) with known endonucleases like the I-HmuI HNH homing endonuclease of bacteriophage SPO1, which structure has been solved (Shen et al., 2004). These proteins are site-specific DNA endonucleases that initiate DNA mobility by introducing double strand breaks at defined positions in genomes lacking the endonuclease gene

We also tested the pre-early protein sequences to look for CRISPR/Cas inhibitors. PaCRISPR is a machine learning based predictor of anti-CRISPR proteins (Wang et al., 2020) that was trained with known anti-CRISPR peptides with less than 70% sequence identity with these peptides. PaCRISPR extract four properties from the input peptides: Position-Specific Scoring Matrix (PSSM) to detect distant sequence similarities, and other three out of it, DPC-PSSM to detect local sequence-order effect, PSSM-AC to evaluate intra-vector correlations, and RPSSM to explore local sequence order effect based on a reduced initial PSSM matrix.

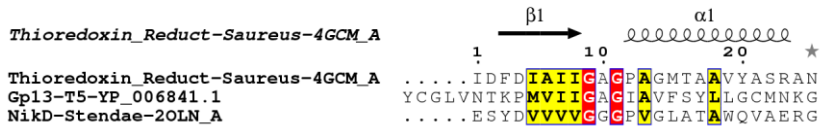


Figure 42.Alignment of Gp13 sequence to proteins with known activity.
Gp13 distant homology was detected through HHPred. Conserved residues are highlighted in red and less well conserved residues in yellow.

The conclusion of these *in silico* analyses remain limited: Dmp is confirmed as a phosphatase, Gp2 is predicted to be an ADP-ribosyltransferase, A1 a DNase, A2 a DNA binding protein, Gp5 and Gp13 are probably membrane proteins and Gp13 bears a signal peptide. Regarding the other peptides encoded at the FST-DNA, they lack significant homology to proteins with known functions. Therefore, we explored the function of these ORFan pre-early genes by expressing them in the host, out of the viral context.

.

Table 7. Bioinformatic analysis of putative pre-early proteins.

Protein	MW (KDa)	pI Compute_pI	Putative function Literature	Transmembrane Domain [a] TMHMM	Signal Peptide [b] SignalP	Anti-CRISPR [c] PaCRISPR	Homology detection structure prediction [d] HHPred
Dmp	28.2	5.27	dNMPase [e] (Mozer et al., 1977)			0.796	Phosphatase (1U7P_C, 97, 8.4e-4) Polynucleotide kinase (5UJ0_A, 96, 0.027)
Gp2	14.2	5.47				0.546	ADP-ribosylation (2JVN_A, 95, 0.014) Zn-binding domain (2RIQ_A, 95, 0.014)
Gp3	10.7	4.71				0.704	
A1	61.5	6.12	Host DNA degradation and Second Transfer (Lanni, 1969). DNase (thesis Zangelmi, 2018)				Exonuclease (5IHE_A, 99, 6.2e-9) Nuclease (4YKE_A 99, 3.2e-9) Exonuclease (4LTY_D, 99, 1.7e-10) Nuclease (6NVP_A, 99, 7e-10)
Gp5	7.2	9.10		0.081 (10:27, 40:62)			
A2	14.3	8.78	Second Transfer (Lanni, 1969)				DNA-binding protein (6TAZ_B, 98, 3.9e-4) Homeobox protein Pax-6 (6PAX_A, 95, 0.48)
Gp7	9.1	9.05					
Gp8	8.0	10.58					
Gp9	3.9	9.78					
HegG	20.0	9.43					HNH endonuclease (1U3E_M, 99, 4.2e-13)
Gp11	37.0	5.32				0.589	
Gp12	8.6	5.61				0.559	
Gp13	6.2	7.78		0.978 (1:24, 39:56)	0.462 (1:21)		Thioredoxin reductase (4GCM_A, 93, 0.1) Oxidoreductase (2OLN_A, 93, 0.11)
Gp14	8.1	10.00					
Gp15	13.0	5.82				0.789	
Gp16	9.6	9.41		0.077 (4:26)	0.256 (1:29)		
Gp17	7.7	4.39				0.514	

[a] Likelihood of N on the cytosolic side (start:end, start2:end2). [b] Likelihood (start:end). [c] Only peptides with a score > 0.5 are shown.
[d] Results shown as: Activity of the match (PDB accession, Probability, E-value). Only results with P > 93 are cited.
[e] deoxynucleotide monophosphatase

2.4.2. Ectopic expression uncovers toxic pre-early proteins

As a second approach in understanding the function of pre-early proteins, we systematically studied the impact of viral gene ectopic expression on bacterial growth and morphology. All the annotated pre-early genes (17) along with genes *A2* plus *A1* were cloned into plasmid pBAD24 under the control of the arabinose inducible promoter. Cloning of genes *A1* and *08* required numerous attempts and was only successful when glucose was added to the selective medium, hinting at a potential toxicity of these gene products in the cell.

Twenty strains were compared in total: one bearing the empty plasmid, 17 with each of the annotated pre-early genes, one bearing pBAD-A1A2, and one last without pBAD24-derived plasmids. We incubated them under the same conditions (Figure 43). Briefly, we transformed electrocompetent *E. coli* F cells with each plasmid, incubated for 30 minutes in LB glucose to repress the gene expression, then, resuspended in LB glucose with antibiotics and grown overnight at 37 °C under agitation at 220 RPM. With the overnight preculture, we started a new culture at OD_{600}~0.1, incubated until the exponential phase. Then, we followed the bacterial growth by measuring the OD_{600} (Figure 44), we tested the survival of the recipient cells (Figure 45), and observed them under fluorescence microscopy (Figure 46) as well.

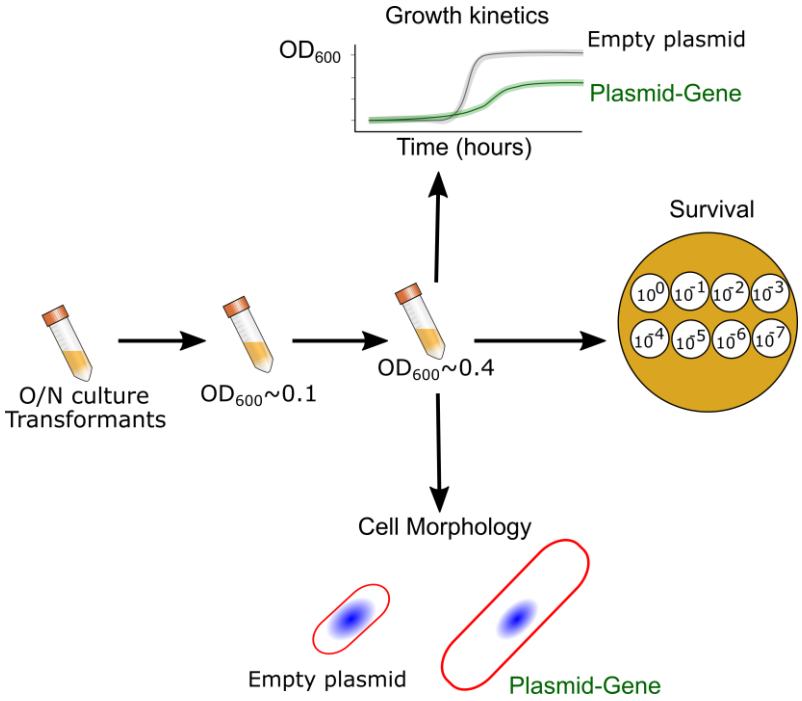


Figure 43. Workflow scheme for the ectopic expression of pre-early genes.
Bacteria were transformed with plasmids to code for T5 pre-early genes. These cells were incubated 16 hours in presence of glucose 0.4% and Ampicillin 200 µg/mL. Fresh cultures were started from them and incubated until exponential phase. Then, these cultures were distributed in microplates to follow the bacterial growth. Serial dilutions were plated in different media to test the survival of bacterial strains after induction. Finally, cell morphologies were studied under fluorescence microscopy.

Bacterial cultures expressing genes *A1*, *08*, *13*, and *A1-A2* displayed strong deviations from the behavior of the bacteria bearing the empty plasmid, pBAD24 (Figure 44, A). The expression of the other genes led to no remarkably effect on the culture's behavior (Figure 44, B, C, and D). A1 is probably a DNase and displayed a phenotype that goes along with such putative activity. The production of A1 plus A2 by the cells induced a disturbance of the normal growth curve as well, but in a lower degree than with A1 alone. The products of genes *08* and *13* also perturbed bacterial growth, hence, while Gp8 seemed to slow down bacterial growth pre-early in the incubation, Gp13 halted bacterial growth suddenly, in a similar fashion as by A1. In all cases, we found a reduction of the titer of survivals after the expression of the genes over a long period of time (16 hours).

It is rather expected that a lower culture turbidity will go along with lower bacterial titers, on the contrary, cultures with rising turbidity are expected to yield higher bacterial titers. The expression of *dmp*, *hegG*, *11*, or *15*, reduces the titers of surviving bacteria after >12 hours of induction, but makes no difference regarding the turbidity of cultures in the short term. It is worth noting that gene *11* in these assays is an *E. coli*-optimized synthetic gene; the reason for this is that we were unable to clone the gene after many attempts and despite the absence of restriction sites in the original gene sequence.

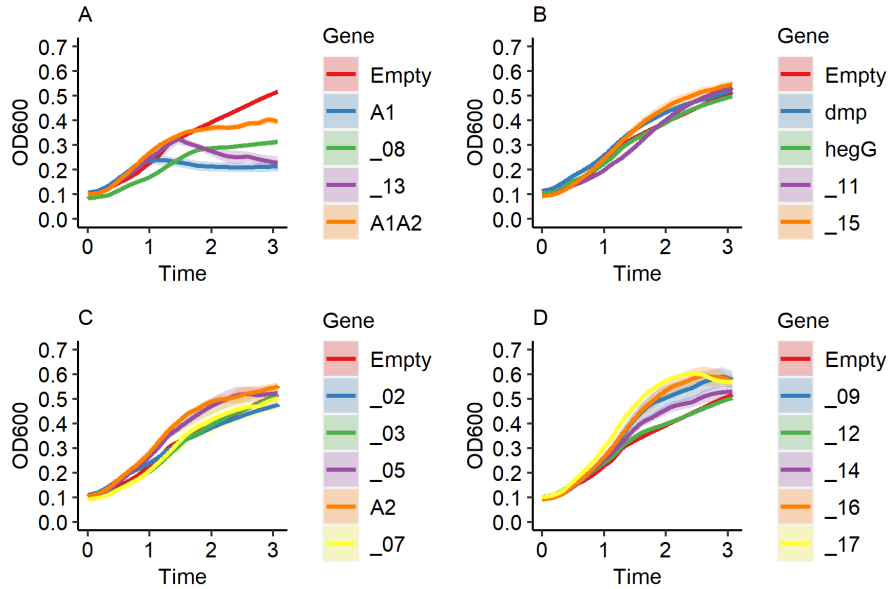


Figure 44. Bacterial growth curves for strains expressing T5 pre-early genes.
Bacterial cultures were started at OD_{600}~0.1. (A) Genes whose expression perturbs the bacterial growth, survival and morphology, (B) genes without an effect on bacterial growth but which strong survival and morphology perturbation, (C) genes with only morphology perturbations, and (D) genes with no negative effect on the bacterial growth. *n* = 4. The cultures were diluted in LB broth (glucose 0.4 % ampicillin 200 μg/mL).

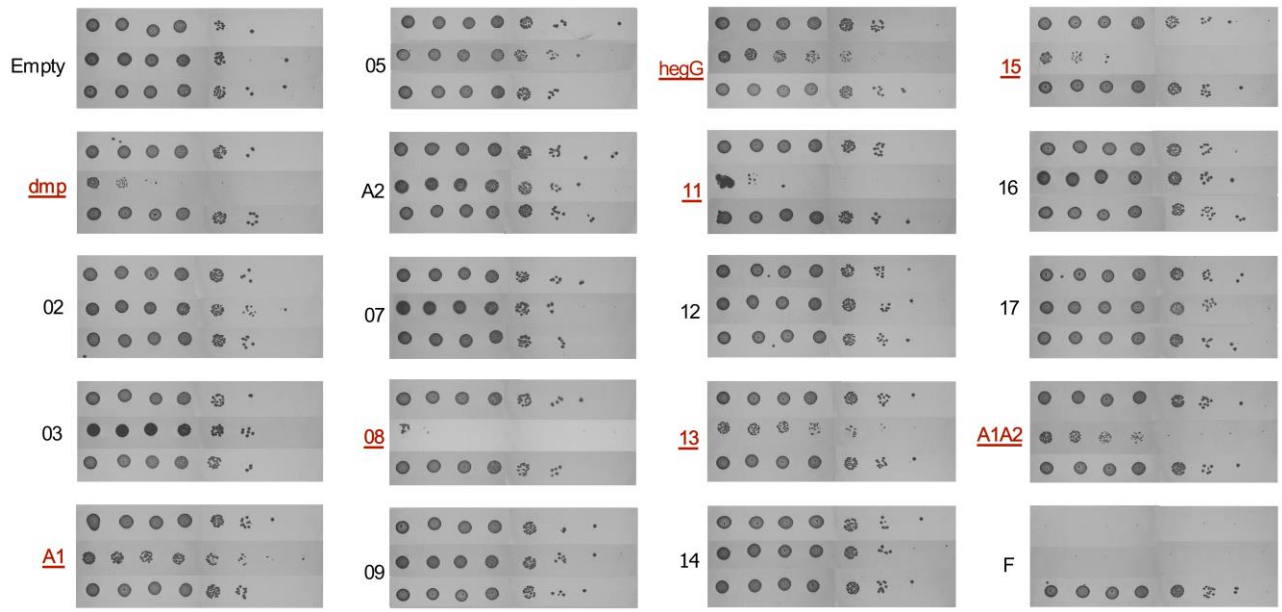


Figure 45. Survival of bacterial strains expressing T5 pre-early genes
E. coli F bacteria transformed with a plasmid bearing the named gene were serially ten-fold diluted. The highest to lowest dilutions were spotted from left to right, respectively. The serial dilutions were dropped onto plates under (above) repression of gene expression by glucose 0.4% plus ampicillin 200 μg/mL, (middle) induction of gene expression by arabinose 0.4 % plus ampicillin 200 μg/mL, (below) repression of gene expression by glucose 0.4% but no selection with antibiotics. The experiment was carried out twice with similar results.

These results indicate that some pre-early genes, even though they appear as dispensable for infection, are toxic for the host under the conditions of ectopic expression.

2.4.3. Alterations in cell morphology linked to pre-early gene ectopic expression

To better understand the impact of individual pre-early genes on bacteria, we observed the cell morphology by microscopy under the conditions of ectopic expression. First, we noted that the bacteria bearing the empty plasmid (Figure 46, Empty) were different from bacteria bearing no plasmids (Figure 46, F). Indeed, they had a higher width, length, therefore a larger area and perimeter, and slightly more rounded (not shown). Thus, we considered this difference and compared the strains bearing the plasmids with pre-early genes against the bacteria transformed with the empty plasmid. The cells expressing the different pre-early genes showed a variety of phenotypes as shown by the phase contrast (PC) and fluorescence microscopy using the two dyes DAPI and FM4-64 for staining the bacterial DNA and membranes respectively (Figure 46). We evaluated the cell parameters with MicrobeJ and made a pairwise comparison to screen those who were most affected depending on the expressed gene (Figure 47). A heatmap of DAPI-dye foci distribution across the cells with an orthogonal coordinate system is also depicted in the annexes (Section 7.3, Figure 66).

.

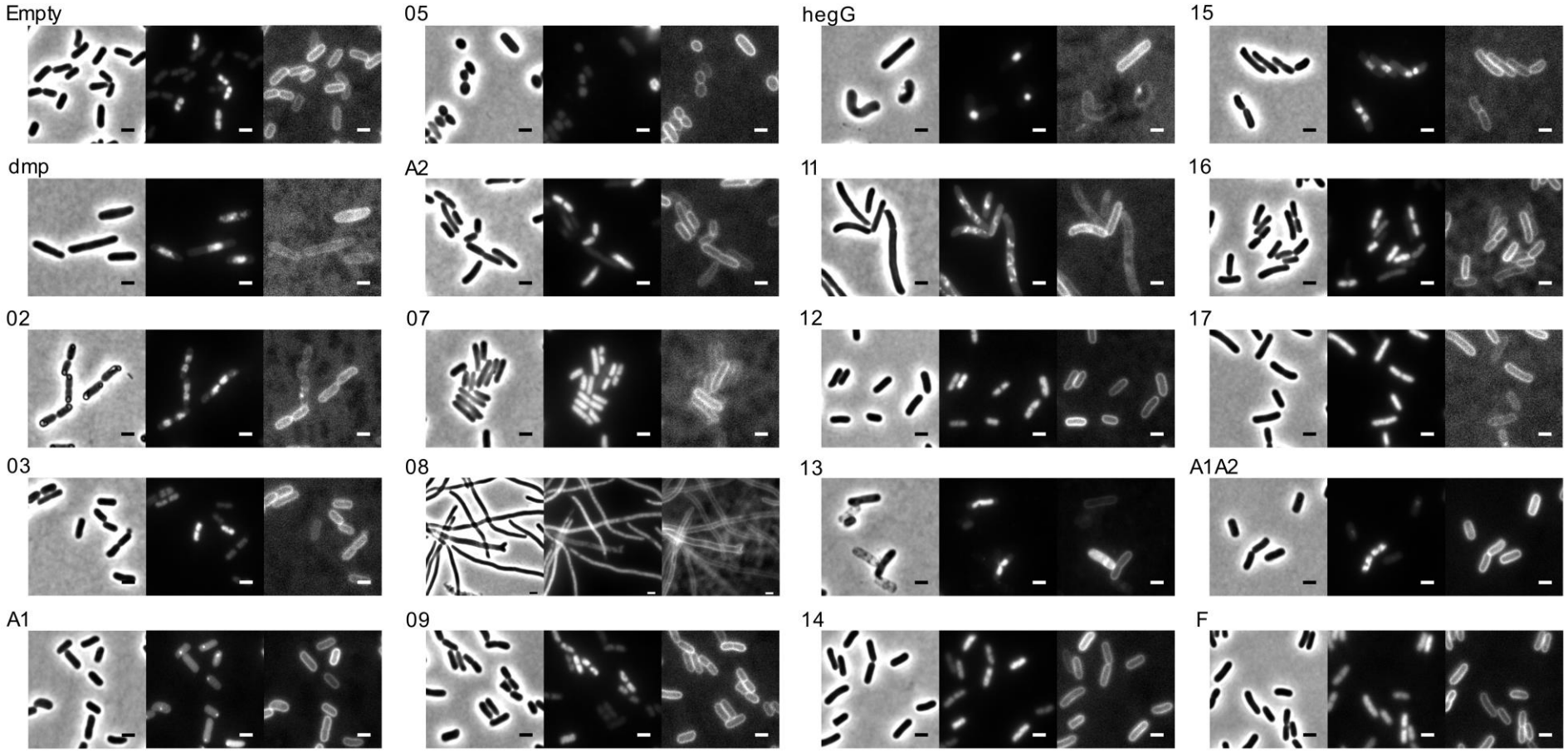


Figure 46. Cells expressing T5 pre-early gene products display different morphologies.
Fluorescence and contrast microscopy from cells expressing phage pre-early genes show different effects on their morphology and staining properties. Cell membranes were stained with FM4-64 and DNA with DAPI. (Scale bar, 2 μm). Phase contrast (left), DAPI (middle), FM4-64 staining (right) images are depicted for each overexpressed gene.

The expression of the gene *dmp* leads to an important elongation of the bacterial cells (Figure 46, dmp, left) that goes along with a cloudy nucleoid (Figure 46, dmp, middle), concentrated towards the center of the cell with a lower DAPI dye intensity (Figure 47, dmp). Dmp is a 5'-deoxynucleoside monophosphatase, with lower K_m for dAMP and dCMP, than dTMP, dUMP, or dGMP. Its effects might be linked to the phosphatase activity (Mozer and Warner, 1977), but the toxicity of the gene product had never been described before.

T5 gene *02* expression leads to significant effects on the morphology (cells become thinner and slightly longer). Moreover, the formation of highly refringent formations is visible by PC in the cells (Figure 46, 02, left). These refringent formations locate mainly at the poles and are not made up by lipids nor DNA (Figure 46, 02, middle and right), suggesting they contain mainly proteins, most probably Gp2. *E. coli* cells produce similarly refringent bodies when they are exposed to different classes of antibiotics, or when they are infected by certain phages, notably LM33_P1 or 536_P (Dufour et al., 2017).

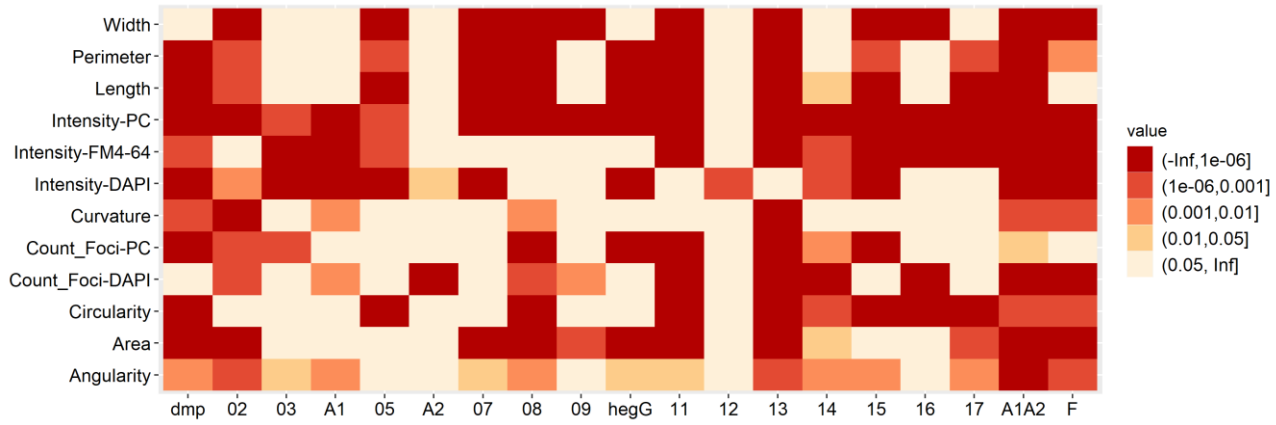


Figure 47. Pairwise comparison of parameters from cells expressing phage pre-early gene products.
Cell profiles were compared to the host bearing the empty plasmid. The cell color is given by the range of the *p*-Value from a pairwise comparison through a Tukey Honest Significant Test (CI=99.99%). Cells expressing pre-early T5 genes were studied through fluorescence microscopy. Segmentation and parameter extraction were performed with MicrobeJ.

A1 product affects few shape parameters: Area, length, perimeter, and width are the same as for the empty plasmid (Figure 46, A1, left and Figure 47, A1). However, it does affect the characteristics of bacterial DNA. *A1* ectopic expression drastically lowers the DAPI staining of the cells, and mainly what can be seen corresponds to a unique dot with no preferential position across the cell (Figure 46, A1, middle). This reinforces the hypothesis that A1 is a DNase capable of degrading the host DNA, probably leaving part of it unaffected, e.g., plasmids. Furthermore, the infection by *A1*$^-$ phages prompt the cells to grow bigger, while *A1* expression alone led to no changes in bacterial shape, therefore, the DNA digestion might halt multiplication fast enough to keep the bacterial shape intact, and quicker than the other effectors encoded at the FST altogether.

Bacteria producing Gp5 acquire rounded shapes (Figure 46, 05, left) in which the length, width, and perimeter are affected (Figure 47, 05). In parallel, DAPI intensity is remarkably lower than in bacteria bearing the empty plasmid (Figure 47, 05, middle). There are three main scenarios in which *E. coli* rods become round, such as inhibition of Mre, Mrd, or RodZ proteins. The MreBCD and PBP/RodA complex maintains cell width, while RodZ is the primary determinant of the cell length, more specifically RodZ and MreB ratio is crucial for the rod shape of the cells. Phage T7 Gp0.6 and Gp0.4 target MreB and their heterologous expression in *E. coli* alters the morphology of the cells (Molshanski-Mor et al., 2014). T5 Gp5 bears putative transmembrane domains and maybe interacts with bacterial membrane proteins that determine the cell shape.

The gene product of *A2* affects only the distribution of the DNA through the cells, hence, the nucleoid has a similar intensity as for the empty plasmid but is less defined (Figure 46, A2, middle). This suggest that A2 might interact with the bacterial genome or its organization but excerpts no effects over the bacterial growth.

Gp7-producing cells are mostly agglomerated (Figure 46, 07). At first sight, they look not so different to control cells. However, the aggregation might complicate and eventually skew the segmentation process. It is not possible to draw some hypothesis for putative mechanisms.

The expression of the gene *08* in *E. coli* cells dramatically perturbs their morphology: Cells become long filaments (Figure 46, 08, left), with no specific distribution of the DNA. Similar effects can be triggered in *E. coli* when there is no production of FstZ, a protein that is essential for cell division, or when this protein is inhibited (Margalit et al., 2004).

hegG gene holds homology to homing endonucleases, its expression is linked to perturbations on the cell shape (Figure 46, hegG, left) that were significative for the perimeter and length, due to curved cells. The nucleoids show a well-defined compaction towards the center of the cell (Figure 46, hegG, middle) and lower DAPI intensity (Figure 47, hegG).

Cells that express gene *11* are altered in most shape parameters but curvature, the DAPI intensity is not altered (Figure 47, 11) yet the nucleoid seems completely reorganized (Figure 46, 11, middle) with DAPI-related foci located all along the cells. These phenotypes suggest Gp11 might interact with the nucleoid, but it is hard to establish thus far whether the toxicity of the protein is linked to a specific bacterial process.

Gp13 is the only protein capable to lyse the cells by independent from the other viral factors. This can be observed by phase contrast microscopy (Figure 46, 13, left), some cells loose refringence suggesting lysis. Although most shape parameters seem to be altered, the automatic analysis of cells

had included lysed and non-lysed cells. Apparently, the nucleoid does not display any distinctive characteristic (Figure 46, 13, middle).

The last gene tested individually, 15, triggered yet another phenotype on the cells in which it was expressed: partial septation of cells with the nucleoid distributed mainly in the center of the cells (Figure 46, 15, middle. It seems that in them, the cell division is not completed, nor the growth, since the cells do not adopt a filamentous form as for Gp8 either.

We also tested simultaneously the expression of the probably only two essential T5 pre-early genes: *A1* plus *A2*. Their expression does not lead to the formation of single DAPI-related foci in the cells and in many cells, but not all, an apparently normal nucleoid can be seen (Figure 46, A1A2, middle). This add up to the hypothesis that A2 might bind to the DNA and partially protect it from the activity of A1.

In conclusion, we could observe that gene products A1, Gp8, HegG, Gp11, Gp13, and Gp15, were lethal for the cells when produced them over a long period. These proteins disturb the morphology of the cells suggesting some functions. Thus, A1 is a DNase *in vitro* (thesis Zangelmi, 2018), and probably degrades the DNA *in vivo* as well. Gp8 probably interferes with the cell division and the morphology of the cells recall inhibitory effects on FstZ (Margalit et al., 2004). HegG coding gene was annotated as an HNH endonuclease and probably targets the DNA, but probably not degrading it. Gp11 is lethal for *E. coli* but the cellular morphology suggests no particular mechanism. Gp13 has transmembrane domains, probably a signal peptide at its N-terminus and was the only product capable to lyse bacterial cells. And finally, Gp15, for which the cellular morphology entices a DNA-related mechanism.

2.4.4. Gp13 is a membrane protein

The product of gene *13* was the only phage pre-early protein capable to lyse the cells when it was ectopically produced independent from the context of the phage infection. The Gp13 protein is predicted to contain a N-terminal peptide signal and two putative transmembrane domains. To further investigate the properties of this protein, we cloned the gene *13* into the plasmid pET28 to overproduce the protein in *E. coli* cells. Three constructions were made: one to produce the protein without His-tag to evaluate its toxicity for the host, and the other two with a 6xHistidine tag fused to either the N terminus or the C terminus of Gp13. The constructions were called pET13, pET13HisN, and pET13HisC, respectively.

We transformed *E coli* BL21 cells with one of the three constructions, or the empty plasmid and started an overnight preculture at 37°C, 220 rpm shaking in LB broth supplemented with kanamycin 50 µg/ml glucose 0.5 %. Then, a fresh culture (150 mL) was started by inoculation with the overnight preculture. This culture was grown in the same conditions and was split in two flasks once the cultures reached an OD_{600} of 0.5. IPTG 0.4 mM was added in one flask or glucose 0.5 % in the second flask to induce or to repress gene *13* expression, respectively. Finally, we followed the optical density of the cultures every 20 minutes until 240 minutes.

Gp13 and the tagged versions stopped the bacterial growth 20 minutes after their production was induced. In addition, the induction of the empty plasmid also led to some perturbations on the normal growth of the cells but not at the same degree as those produced by Gp13. In contrast, the lack of inducer (IPTG) on the media was marked by growth curves with the same characteristics among empty plasmid and all the constructions. This implies that the observed effects were due to the overproduction of Gp13 (Figure 48).

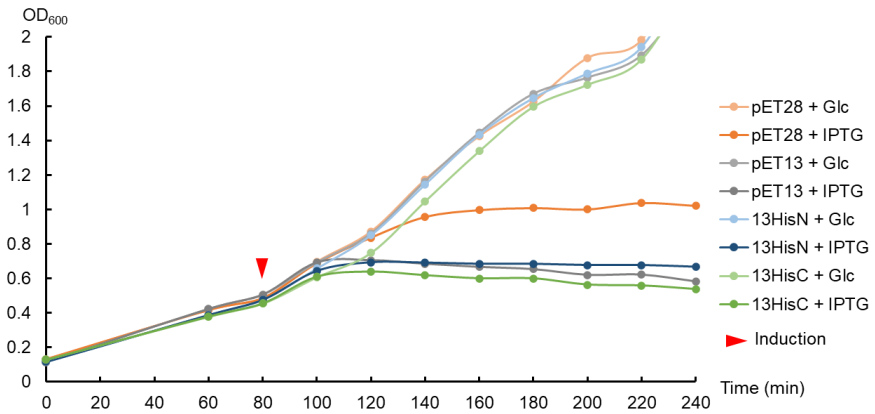


Figure 48. Ectopic expression of the gene 13 and its tagged version.
Cells were transformed with plasmids bearing the phage pre-early gene *13* bearing no poly histidine tags or either one at the N terminus, or at the C terminus. Half of the culture was separated and induced at 80 minutes (OD_{600}~0.5), and the OD_{600} was followed accordingly.

We harvested the cells by centrifugation and resuspended the pellets in Phosphate-Buffered Saline (PBS 1x) and adjusted all samples at the same OD_{600}. After the disruption of the cells by sonication we separated the lysate in different fractions: the total cellular content (T) was obtained after cell disruption. This total fraction was centrifuged at low speed (8,000 x g) to sediment the few unbroken cells and possible insoluble inclusion bodies of Gp13 proteins that could arise from gene 13 overexpression. The resulting supernatant was then fractionated by ultracentrifugation (at 100,000 x g) into the supernatant containing the soluble cytoplasmic material (S) and the bacterial membrane (M) recovered in the pellet. The proteins from these different fractions were separated by Tris-Tricine SDS-PAGE, which are optimal for the resolution for small proteins, transferred proteins onto a nitrocellulose membrane and the presence of His-tagged Gp13 proteins was verified by western blot using anti-Histidine antibodies (Figure 49).

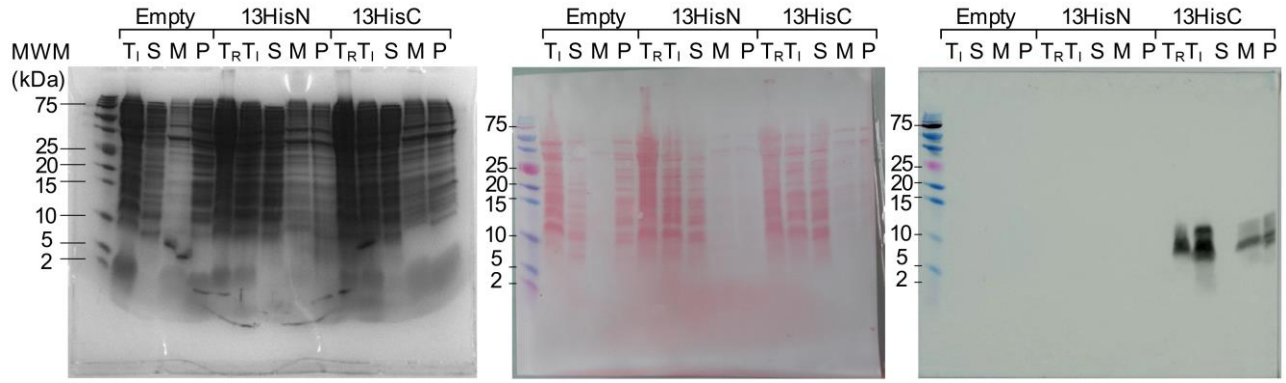


Figure 49. Protein electrophoresis of cell extracts producing Gp13.
Cells bearing the plasmids pET28, pET13HisN, and pET13HisC, were lysed and their cellular fractions analyzed by (A) Tris/Tricine SDS-PAGE and Coomassie Blue staining (B) Ponceau S staining after transfer on a nitrocellulose membrane and (C) western-blot revealed by using anti-His antibodies. The fractions included the total lysate from repression conditions (T_R), from induced conditions (T_I), and from this last, the soluble (S), membranes (M), and insoluble fractions (P) were migrated as well.

For the two plasmids pET13HisN and pET13HisC, the Tris/tricine gels displayed similar protein patterns under conditions of repression or induction of gene *13* expression. Thus, we could not observe the overproduction of Gp13 (MW= 6.2 kDa), neither in the total fraction, nor in the soluble or membrane fractions, either on gels stained with Coomasie blue or on the nitrocellulose membrane stained with Ponceau S. However, despite the absence of a visible overproduced band in the gel, anti-His antibodies revealed Gp13HisC in the following fractions: total repressed (T_R) and induced (T_I), membranes (M) and Pellet (P). As both Gp13HisN and Gp13HisC exhibit the same toxicity as the untagged protein for the overexpressing cells (Figure 48), both recombinant proteins are probably correctly folded upon their synthesis. Gp13HisC was not detected in the soluble fraction, but in the membrane fraction and the expected size (≈ 7.9 kDa) and also in the insoluble material sedimented at low speed, probably as inclusion bodies. This observation is in agreement with the bioinformatic

analyses predicting that Gp13 contains a peptide signal and two transmembrane domains (Figure 41) in Gp13HisN could not be detected by the anti-His antibodies, whatever the cellular fraction, which could be due to the cleavage of its N terminal region upon processing of the peptide signal. It is however worth noting that Gp13HisC can be detected under repression conditions, which implies that low levels of the protein might not be perturbing for the cell, while high concentrations lead to cell lysis.

This preliminary biochemical characterization of Gp13 confirms that it is a membrane protein. Further studies are required to understand its role during the first steps of T5 infection.

2.4.5. Revisiting the T5 resistance to EcoRI

Many phages deploy mechanisms to avoid restriction of their genomes by their hosts (Section 1.1.3.3) and T5 might not be the exception: indeed (i) T5 genome contains no EcoRI restriction site within the First-Step Transfer locus (FST) (Davison and Brunel, 1979b), while there are six EcoRI sites at the SST DNA, (ii) three mutants of T5 that carry an EcoRI site in the FST were isolated and characterized in the late 1970s and their growth is restricted in strains producing EcoRI (Brunel and Davison, 1979). These results suggested that there might be some pre-early proteins that inactivate host restriction enzymes (RE). One possible manner to test this possibility was to compare the efficiency of plating of the mutants deleted in the different pre-early genes in bacteria producing or not certain restriction enzymes. Thereby, we carried out the infection of three *E. coli* F strains: one with no plasmids, another transformed with pJJC1418, or pJJCC1428. The plasmids pJJC1418 and pJJC1428 carry the EcoRI Restriction/Methylation system: pJJC1418 codes for both the nuclease and the methylase of EcoRI system (R+/M+), while pJJC1428 codes only for the methylase (R-/M+) (Kulakauskas et al., 1995).

We designed two different experiments: the first was to evaluate whether the presence of EcoRI restriction sites at different locations across the FST locus makes a difference in terms of efficiency of infection, the second was to screen for EcoRI inhibitors among the genes mutated at the FST. Thus, the first deals with the target and the second with the inhibitor of EcoRI.

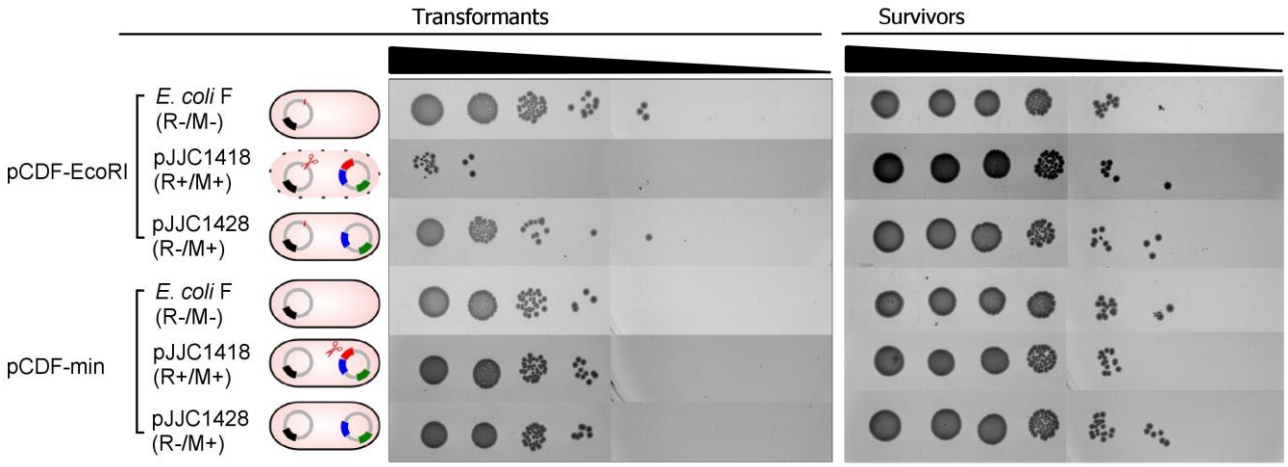


Figure 50. Transformation of cells encoding the EcoRI Restriction Methylation system.
Three strains of *E. coli* were tested: *E. coli* F that lacks the EcoRI restriction system, *E. coli* F(pJJC1418) that carries the EcoRI nuclease and methylase genes (R+/M+), or *E. coli* F(pJJC1428) bearing an EcoRI methylase gene only (R-/M+). These strains were transformed either with plasmid carrying an EcoRI restriction site (pCDF-EcoRI) or not (pCDF-min). Transformants were serially diluted and dropped onto plates with two antibiotics (Ampicillin-IPTG and Streptomycin) or only one (ampicillin-IPTG). Ampicillin 100 μg/mL and IPTG 0.5 mM support the replication of plasmids pJJC1418 and pJJC1428. Streptomycin 25 μg/mL selects pCDF-EcoRI or pCDF-min transformed bacteria. Survivors were resistant to ampicillin for *E. coli* F(pJJC1418) or *E. coli* F(pJJC1428) or no antibiotic for *E. coli* F. Transformants were counted as the resistant to ampicillin plus streptomycin for *E. coli* F(pJJC1418) and *E. coli* F(pJJC1428), transformed with pCDF-EcoRI or pCDF-min; or resistant to streptomycin only for *E. coli* F transformed with pCDF-EcoRI or pCDF-min.

To evaluate the efficacy of plasmid-encoded EcoRI in *E. coli* strain F, we evaluated the efficiency of transformation (EOP) with plasmids bearing EcoRI sites. Cells possessing the plasmid coding for both methylase and restrictase (R+/M+) were poorly transformable with pCDF-EcoRI, which contains an EcoRI restriction site (Figure 51), while cells lacking the EcoRI restrictase or R-/M+ cells were similarly transformable with the pCDF vectors regardless of the presence of restriction sites. In all cases, titer of survivors was in the same order of magnitude. Survivors were titrated in absence of antibiotic used for the selection of pCDF-transformed bacteria. In conclusion, only the presence of an EcoRI nuclease was not compatible with plasmids bearing EcoRI restriction sites.

Having established that R+/M+ strain F(pJJC1418) is capable of restricting the transformation of plasmid with EcoRI site, we then studied the restriction sensitivity of T5 when restriction sites were introduced at the FST. A previous study reported the isolation of four T5 mutants with EcoRI restriction sites at the FST by random mutagenesis, three of them were EcoRI-sensitive (*ris*) strains (T5 *ris1*, *ris2*, and *ris3*), with the restriction sites located between *A1* and *A2*, or upstream from *A2*, between positions 3000-4500 in the T5 genome. The fourth strain (T5 *ris4*) had an EcoRI site that was located at the position 9945. Only ris4 was resistant to EcoRI restriction; however, no other study placed EcoRI sites among genes 8 to 17. Therefore, I tried to replicate the test by placing one EcoRI site either between genes *A1* and *05*, or besides gene 17. The first T5 mutant was named T5 *risA105* (Figure 51, A), and the second T5 *ris16IS* (Figure 51, B).

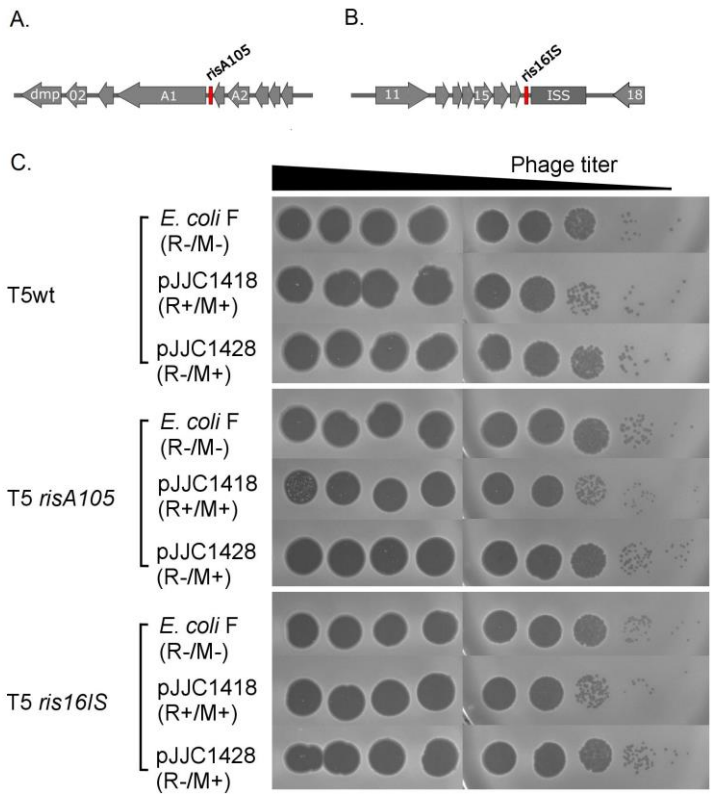


Figure 51. Effects of EcoRI restriction of T5 mutants carrying EcoRI sites in the FST
Phage T5 mutants were generated that bear EcoRI sites (A) between genes *A1* and *05* (T5 *risA105*), or (B) between gene 17 and the ISS (T5 *ris16IS*). (C) Wild-type and mutant phages were titrated on *E. coli* strains either devoid of EcoRI system (R-/M-), or bearing the EcoRI restrictase and methylase (R+/M+, encoded in pJJC1418), or bearing the EcoRI methylase only (R-/M+, encoded in pJJC1428). The top and the base agar contained Ampicillin 100 µg/mL and IPTG 0.5 mM.

T5wt, *risA105*, and *ris16IS* titers were almost the same on the three bacterial strains, with a slight decrease in their titer against *E. coli* F pJJC1418 (R+/M+), suggesting that strains with EcoRI sites in the FST are not susceptible to restriction by the EcoRI restriction enzyme.

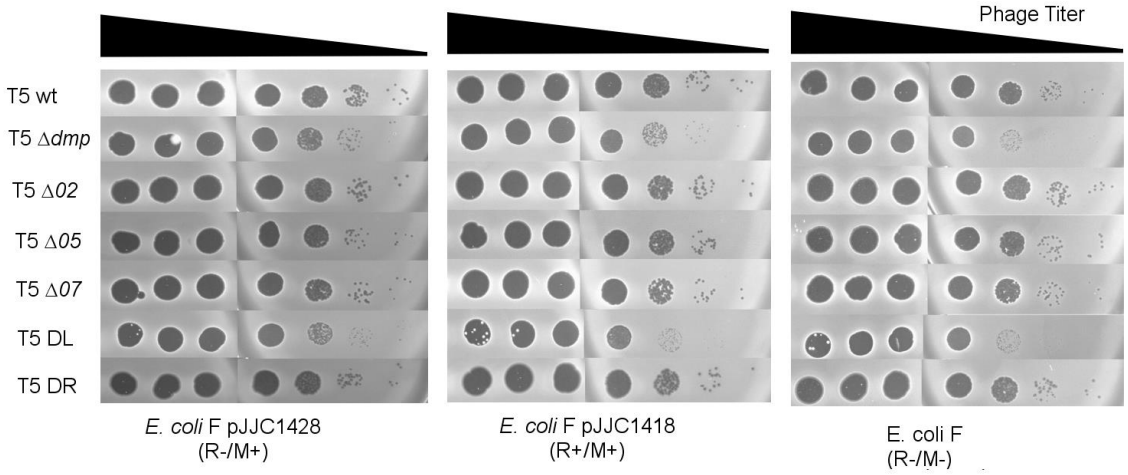


Figure 52. Effects of EcoRI restriction on T5 mutants.
T5 mutants were titrated on *E coli* strains coding for (left) EcoRI methylase, (center) EcoRI nuclease and methylase, or (right) no EcoRI system.

In the study of (Brunel and Davison, 1979), when titrating *ris1*, *ris2*, and *ris3* against *E. coli* EcoRI^{+} strains, they found efficiencies of plating of 10^{-2}, compared to 10^{-1} for the T5wt. Although there is a tenfold difference in the efficiency, variances were not included in the report. They tested the plating ability of *ris* mutants on *E. coli* K-12 strains transformed with high-copy-number ColE1-derivative plasmid NTP14, while in our assays we relied on *E. coli* strain F, transformed with a high-copy-number pBR322-derivative plasmid, pJJC1418 and pJJC1428. While both plasmids are similar, the genetic background is not. It remains to be tested whether different recipient strains with the same plasmids will yield different results as well.

In the last test we tried to screen for RE inhibitors by measuring the efficiency of plating of T5 mutants on the three *E. coli* strains used before: *E. coli* F (R-/M-), *E. coli* F pJJC1418 (R+/M+), and *E. coli* F pJJC1428 (R-/M+). It is worth noting that all T5 mutants were amplified in *E. coli* F (R-/M-).

The wild-type phage titer was equivalent in all three *E. coli* strains as expected. The same applied for mutants on genes *02*, *05*, *07*, and T5 DR. On the contrary, *dmp*$^{-}$ mutants showed slightly lower titers compared to all the other strains. The gene *dmp* codes for a phosphatase and it might be rather improbably that it is also acting as a nuclease inhibitor, unless both functions are indirectly linked, in any case further analysis will be needed.

2.5. Defining the Injection stop signal essential for infection

In parallel to the functional studies conducted on the T5 pre-early genes, we endeavored to start the identification of the injection stop signal. The ISS constitutes the non-coding region located between the pre-early genes and the rest of the phage genome (SST DNA). To better understand the interplay between this stop signal and the mechanism by which A1 and A2 resume the DNA transfer we attempted to define whether part of this non-coding region is actually essential for T5 infection. To this end, we constructed T5 mutants that will hopefully allow further investigations of the two-step injection process.

At the onset of infection, immediately after binding of T5 to its receptor, the T5 DNA delivery takes place in two steps. It has been proposed that the 8% of the genome injected during the first step ends at the contiguous non-coding region between genes *17* and *18* and that this 1.5-kb locus probably contains an injection stop signal (ISS).

We could delete almost all the pre-early genes with the exception of the essential genes *A1* and *A2*, and the genes *08* and *09* (9.2% of the genome). And the mutants obtained were less virulent but still infectious. These deletions show that the number of T5 FST genes needed for the two-step transfer and host takeover is probably limited and might be reduced to (*A1* and *A2*). Both gene products are necessary for the second DNA transfer as well and might interact with the ISS or with a complex formed between the ISS and host factors to restart the DNA delivery.

To determine whether the entire intergenic region between *17* and *18* is essential for T5 infection, I tried to construct deletion mutants on different segments. Namely, we divided the ISS region in three sections and tried to delete all but one at a time. Thus, we designed the following deletion mutants T5 ΔISS::1 (T5 Δ[9,736..10,936]), T5 ΔISS::2 (T5 Δ[9,204..9,735] Δ[10,414..10,936]), T5 ΔISS::3 (T5 Δ[9,204..10,391]), and T5 ΔISS (T5 Δ[9,204..10,936]) (Figure 53).

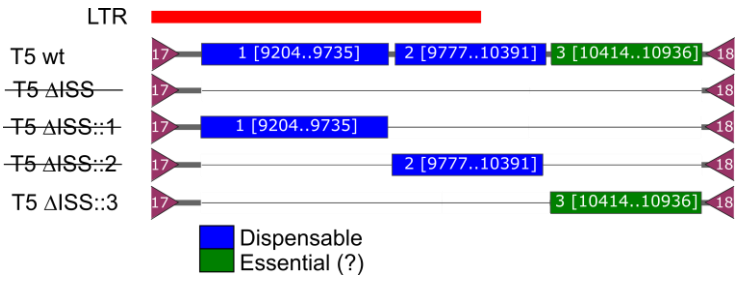


Figure 53. Reduction of the Injection Stop signal.
Partial deletions of the intergenic region between *17* and *18* were tried. Mutant names from strains that could not be isolated were stroke-through. Sections of the locus that were deleted are highlighted in blue, and the section that could not be deleted is highlighted in green. The red bar indicates the region comprised at the left terminal repeat (LTR).

The only mutant recovered was T5 Δ[9,204..10,391], also named T5 ΔISS::3, suggesting the left region [9,204..10,391] of the so-called ISS is dispensable for infection. We sequenced the intergenic

region left by Sanger sequencing (Figure 54). It is worth to emphasize that during the construction the T5 DR mutant in which the genes *hegG* to *17* were deleted, an additional deletion was introduced alongside (Section 6.2.5). It was Δ[9,875..10,340]. T5 DR mutant confirmed that the ISS can be largely truncated between Δ[9,875..10,390] without impairing T5 viability.

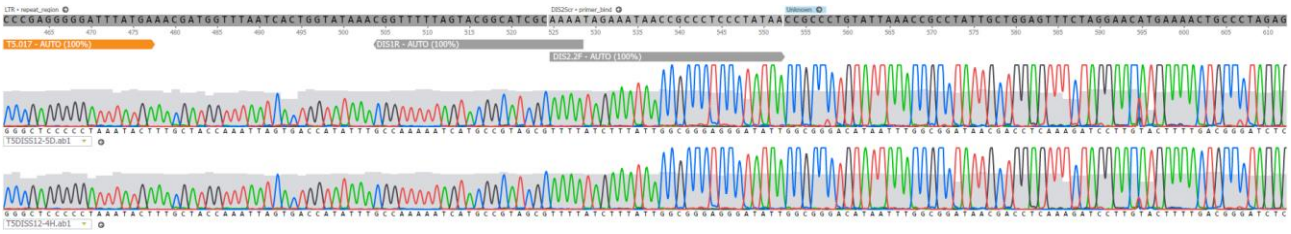


Figure 54. Sequencing of the T5 ΔISS::3 LTR.
The left terminal repeats from two strains of T5 ΔISS::3 was amplified by PCR and sequenced by Sanger method.

To assess the impact of the deleted region on the kinetics of T5 infection, we compared the one-step growth curves of the T5 ΔISS::3.and of wild type T5. The results of the deletion Δ[9,204..10,391] were slightly smaller burst sizes, with a delay in the release of the viral particles, with no changes in the eclipse compared to the wild type. Thus, clearly this deletion in the intergenic region does not affect T5 infectivity in our conditions of infection.

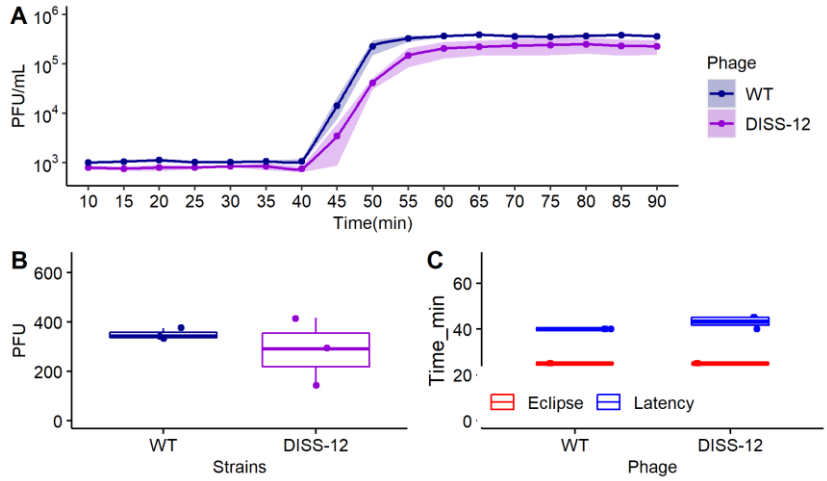


Figure 55. Growth parameters of mutant T5 ΔISS::3.
The one-step-growth experiment results show that the burst size of the strain T5 ΔISS::3 is slightly lower than that of the wild type. Three replicas were tested for each strain.

Finally, we have tried to further reduce the intergenic region remaining in T5 ΔISS::3. down to 150 bp smaller segments but we could not recover any mutant with the region "3" [10,414..10,936] reduced. This first approach of reducing the essential intergenic region between the FST and SST DNA to a maximum of 522 bp will now help us to investigate the mechanisms controlling the stop and restart process of the DNA two-step transfer.

3. Discussion

Bacteriophage T5 is an attractive model to study the early stages of infection because all the pre-early genes cluster together at the First-step-transfer (FST) locus and are expressed separately from the rest of the genome. Thus, T5 offers a unique opportunity to study the synergy of a few proteins in the full takeover of bacteria. Nevertheless, the actual T5 host takeover strategy still remains poorly understood since most of the pre-early genes are ORFan genes with no detectable homology to genes outside the *Tequintavirus* genus. They belong to the vast reservoir of genes with unknown functions that has been designated as the "biological dark matter" of the universe, and unraveling their function is of major interest to improve our understanding of the detailed molecular mechanism of phage infection. Although each of these pre-early genes is conserved in a variable degree in T5-like viral genomes, they tend to be clustered in a similar manner. Moreover, among the 17 genes annotated at the FST DNA, all the T5-like viruses described to date bear at least one copy of *dmp*, *A1*, and *A2*. The function of these three genes was studied in the past (more than 50 years ago), based on the characterization on *amber* mutants obtained by random mutagenesis. *dmp* was shown to be dispensable for the viral cycle and the nucleotide phosphatase activity of the Dmp protein was demonstrated. *A1* and *A2* were identified as genes essential for infection: they are required to resume the DNA delivery after the pause and expression of the FST genes. In addition, A1 was identified as an essential factor required for the early and massive host chromosome digestion and A2 as a putative DNA-binding protein. On the other hand, there is still very scarce information available today on the other 14 pre-early genes and some functions attributed to the FST, such as the host transcription and replication shut off or the inactivation of bacterial nucleases (EcoRI and RecBCD among others), remain unlinked to any pre-early gene.

During this work, we focused on the functional exploration of the T5 pre-early genes through reverse genetics and ectopic expression.

3.1. Reverse genetics

Genome engineering on the FST locus aimed at deleting pre-early genes to verify which are essential for the host takeover and to gauge the impact of the deletion on the viral cycle. For T5 genome editing, we tested two mutagenesis methods to introduce mutations – by using a plasmid template for homologous recombination (HR) or bacterial retro-elements (retrons) – combined with two enrichment methods to facilitate the recovery of mutants: counterselection with the CRISPR-Cas9 system or the dilution-amplification-screening (DAS) of phage lysates.

Retrons were used for the gene editing of *E. coli* to record analogically encoded information on the state of cells (Farzadfard and Lu, 2014a; Simon et al., 2018, 2019). This system had never been tested for phage genome editing. We have shown that retrons can be used to introduce modifications on the T5 genome through the construction of a T5 *amA1* strain. Such strategy might be helpful to target loci difficult to clone (e.g., bearing a strong promoter) in a high copy number plasmid.

We assessed the sensitivity of T5 infection to CRISPR/Cas9 (Type II-A) by targeting different sequences across the FST locus. with sgRNA-guided Cas9. We found that interference was not uniform. Most of the spacers targeting the left half of the FST locus lowered the T5 efficiency of plating by 10 to 100 times, while those targeting the right half of the FST locus were ineffective as they did not significantly interfere with the wild-type T5 infection. A somewhat similar pattern was previously observed when T5 was targeted by CRISPR/Cas (type I-E) from *E. coli*. Only a small number of spacers, those targeting the left half FST genes, were effective to restrict T5 infection, while no interference was observed against early and late genes (Strotskaya et al., 2017). Our current knowledge of the function of pre-early genes does not allow us to predict that one of them could code for a pan-anti-CRISPR mechanism, produced immediately after the FST and capable of preventing CRISPR interference. Hence, T5 might circumvent CRISPR/Cas systems in a passive way i.e., by protecting its DNA.

Regarding other phage models, T4 DNA is partially protected from CRISPR/Cas9 through the hydroxymethylation of cytosines plus glycosylation (Tao et al., 2017). T5 might not rely on such strategy, as no base modification was reported so far. Thus, there might be a generalized way, whereby T5 manages to escape CRISPR/Cas nucleases. Jumbo phages, such as ΦKZ, are inherently resistant to different types of CRISPR/Cas systems because they produce a "nucleus-like" proteinaceous shell confining and protecting the viral DNA from nucleases (Malone et al., 2020; Mendoza et al., 2020). However, whether phage T5 uses a similar mechanism is unknown. Further studies are necessary to understand how T5 resists CRISPR/Cas interference and to determine if its resistance mechanism stems from its two-step mode of infection, as it was previously suggested (Strotskaya et al., 2017).

Phage genome engineering without the counterselection provided by CRISPR/Cas9 can be a daunting task. We developed two alternatives to engineering the genome of T5 on regions in which CRISPR/Cas9 is poorly effective: DAS and blue/white screening. DAS consists in diluting the lysates to rise the proportion of mutants in some pools, amplify the phages to ease the recovery, and screen

the pools in which the mutants were enriched. Blue/white screening consists in the insertion a marker gene (*lacZ*α) along with the desired mutation in the genome to facilitate the mutant's identification. We used the DAS method to construct most of the mutants used in this study and tested the Blue/White screening as a proof of concept. The DAS strategy has certainly allowed obtaining most mutants constructed by homologous recombination throughout this work. However, it has its limitations, such as the high number of PCR tests needed to identify the pools containing the intended mutants. A genetic marker might work in phages as an analogous of the blue/white screening applied in plasmid screening. A color marker can facilitate the identification of mutants and, consequently, enable T5 gene editing.

Changes introduced in the genome of T5 unveil the degree of flexibility in the packaged DNA. The strain T5 st0 (AY692264.1) lacks 7% of the genome present in T5 wt (NC_005859.1), here we have deleted 9.2 % of the genome (T5 DLDR) and inserted 1 kb (T5 PNmC) of a heterologous sequence. It would be interesting to see the limits of such flexibility in the future.

To set up the conditions for gene editing by HR-CRISPR/Cas9 or HR-DAS, we obtained the mutants T5 *amA1*, T5 *amA2*, and T5 *amA1A2*. Historically, mutants in *A1* and *A2* were obtained by random mutagenesis. Here, we introduced the mutations at specific places near the start codon of the ORFs to get new amber mutants.

Prior to our systematic reverse genetic analysis of the pre-early genes, we evaluated the conservation of pre-early genes in T5-virus genomes and realized that *dmp*, *A1*, *A2*, and *07* are conserved among all of them, followed by *02*, *11*, and *14*. After the successful deletion of *02*, *dmp*, and *07*, we suspected that none of the other poorly conserved genes would be essential either. Finally, we confirmed that at least 13 out of the 17 annotated pre-early genes are dispensable under laboratory infection conditions. We could not delete *08* and *09* due to the impossibility to clone some of their flanking regions, specifically, the 30 bp upstream of *09* and the 30 bp downstream of *08* (Figure 64). The region around *08-09* and between *11-17* is difficult to clone as already seen in the past (Brunel et al., 1979; Heusterspreute et al., 1987; Wang et al., 2005). Mutagenesis of these regions might be facilitated by using retrons in the future since they require the cloning of much shorter homologous sequences. Although the predicted promoter for *08-09* is located upstream *hegG* and was deleted in the phages T5 DR and T5 DLDR, we cannot totally exclude that *08* and/or *09* are expressed in these mutants. Proteomic analysis of the pre-early gene products during infection with T5 DLDR would help answer this question. Without this information, we cannot conclude with

complete confidence that pre-early genes *08* and *09* dispensable for infection. Thus, the 13 deleted genes probably play auxiliary roles during the infection process that were explored afterward.

By whole genome sequencing, we have confirmed the absence of reads mapping to the 13 genes deleted in the respective strains (T5 DL, DR, and DLDR), which suggests that the modifications made at the LTR appear at the RTR as well, maybe due to an interaction between the terminal repeats during the viral cycle, particularly during T5 DNA replication. However, the mechanism of replication of T5 DNA remains poorly investigated (Everett, 1981) as well as how the direct terminal repeats are introduced in every single T5 genome. We found no other mutation in non-targeted pre-early genes. On the SST locus, T5 DLDR incorporated six mutations, two in common with its parental strain T5 DL and none with the non-related strain T5 DR. Further investigations would test whether specific mutations systematically appear at the SST locus upon mutagenesis of the FST locus.

Our phenotypic analysis of the mutant strains included the infection kinetics through the One-step growth (OSG) curve assay, virulence, and observation of bacterial morphology upon infection. The virulence in T5 was evaluated with the PVI through two parameters: $V_{p,}$ a dimensionless index that conveys the virulence of phage and MV_{50}, representing the titer of phage needed to equal half the maximal virulence, conceived as an analogous to the LD_{50} (for lethal dosis) (Storms et al., 2020).

T5 infection of *E. coli* strain F yielded about 350 PFUs per infected bacteria, after 25 minutes of eclipse and 40 min of latency. T5 virulence on *E. coli* F (in LB with Ca^{2+}-Mg^{2+}, V_{T5}: 0.8, MV_{50}: 10^{-4}) was consistently higher than T5's on *E. coli* MG1655 (in TSB broth, without Ca^{2+}-Mg^{2+}, V_{T5}: 0.17, MV_{50}: 0.1) described in the past (Storms et al., 2020). It is worth to mention that calcium and magnesium ions are more important for T5 than for any other T type phage (Adams, 1949). In addition, this difference may also be due to a difference in the two strains, F being an optimal fast adsorbing host strain for T5 isolated in the 1950s. Taken together, these phenotypic studies have highlighted that, in addition to the essential genes *A1* and *A2*, the dispensable *dmp* gene together with one or more genes located on the right hand of the FST facilitates infection of *E. coli* by phage T5. We cannot exclude that these genes become crucial under stressing environmental growth conditions.

Regarding the microscopic characteristics of the infected bacteria, we could notice the degradation of the host DNA at 10 minutes of infection, after which the DNA staining intensity rises, probably revealing the phage DNA replication. Few changes were observed in phase contrast before the lysis, as seen before (Boyd, 1949). However, at 30 minutes, we could notice lysis along with aggregates in the cells that might be made up by proteins since they were not stained by DAPI (DNA)

nor FM4-64 (lipids). The effects of the single and multiple deletions will be discussed below in the later section on the different pre-early genes.

3.2. How the ectopic expression of T5 pre-early genes impacts the host cell

The recombinant expression, or ectopic expression, of T5 pre-early genes in the host *E. coli* strain F was used as the second approach to explore the role of these genes in the host takeover. This strategy was used in the past to discover genes involved in the host takeover of *E. coli* or *P. aeruginosa* phages (Molshanski-Mor et al., 2014; Wagemans et al., 2015). In this study, we have measured the growth-inhibitory effects of pre-early T5 proteins on the *E. coli* F and observed their impact on the cell morphology over a short time (40 min to 3 h). On top of this, we assessed bacterial survival over a long time (16 h). The production of some T5 pre-early proteins led to lethal effects on the host or distortion of the normal morphology and might potentially reveal new mechanisms of action for T5 pre-early genes. A comparison of the effects will be clustered and discussed by gene.

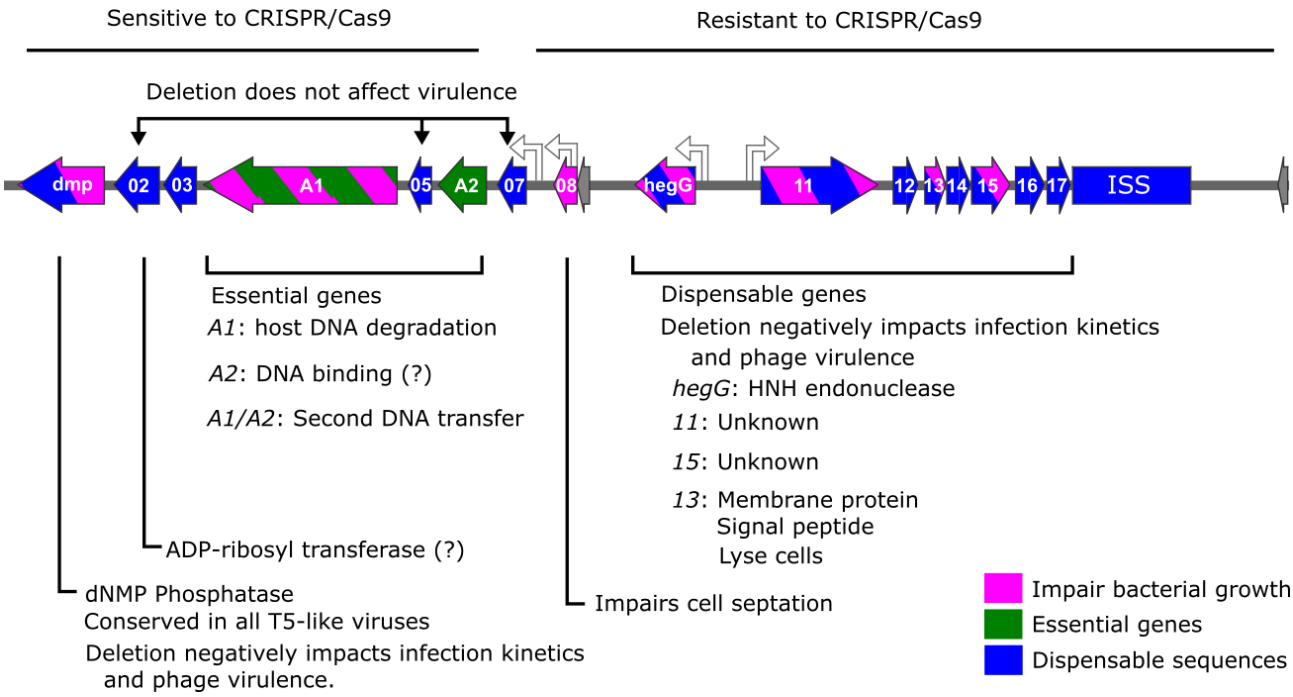


Figure 56. Characteristics of T5 pre-early genes.
Genes *dmp* to *A2* lay in a region sensitive to interference with CRISPR/Cas9, while the rest of the FST-DNA seems resistant to CRISPR/Cas9 interference, or the cleavage by Cas9 does not interfere with T5 infection. The 17 genes were classified into: essential or dispensable for infection, and capable or not to impair bacterial growth when ectopically expressed. Most relevant results from bioinformatic analysis on each gene product are cited.

3.3. Towards the understanding of Bacteriophage T5 pre-early genes

While some phages, such as T4 or T7, depend on the nucleotides from the host DNA degradation as precursors for their own DNA replication (Krüger and Schroeder, 1981; Kutter et al., 2018; Miller et al., 2003), T5 does not reuse the pool of host-derived nucleotides. Moreover, T5 excretes the nucleotides outside the cell after transforming some of them into nucleosides by the phosphatase Dmp (Mozer and Warner, 1977). In fact, a T5 *dmp*⁻ mutant strain reutilizes the nucleotides from the degradation of the host DNA for its DNA replication while the host DNA degradation slows down (Mozer et al., 1977). In our study, the deletion of *dmp* led to a delay in the assembly and release of the viral progeny as seen before (Mozer et al., 1977); it was also associated with a decrease in the burst size, which probably also affected the phage virulence. Among all the studied genes, *dmp* deletion led to the most negative effects for the phage. Bacteria infected by T5 Δ*dmp* reached their lowest DNA staining 10 minutes later than when infected with the wild type. This suggests that Dmp activity might play an important role in linking the host DNA degradation with the phage DNA replication.

In our study, deletion of the genes *02*, *05*, *07*, and all of them (D257), led to larger burst sizes along with a 5-min delay in the release of phage virions. Bacteriophages with shorter latent periods display smaller burst sizes, and both seem to be dependent on the bacterial density and resource availability, as demonstrated in the phage BF64 (Abedon et al., 2001, 2003). Thus, shorter latent periods, accompanied by smaller burst sizes, seem to render the phage specialized in high-density bacterial populations, but that advantage is lost when the specialized phage faces a low-density bacterial population. Phage T7 strains evolved to specialize in high-density cultures. Larger bursts go along with longer latencies (Heineman and Bull, 2007). Nevertheless, the interplay between both parameters might become more complex. In contrast, T4 mutants in the holin *t* gene do not consistently release more phages (Dressman and Drake, 1999). Therefore, delayed lysis does not necessarily result in larger burst sizes.

T5 gene *02* might code for an ADP ribosyl transferase, as suggested by its remote similarity to human Poly-ADP Ribosyl polymerase (PARP) enzymes. Phages producing no RNAP tend to code for factors that subvert and redirect the transcription by the host RNAP. Since no T5 RNAP has been identified until now, it is likely that T5 carries its own mechanism to control the host RNAP. Like T5, the T4 genome does not code for an RNAP: instead, it encodes three dispensable ADP-ribosyl transferases, Alt, ModA, and ModB, which change the specificity of the host RNA polymerase to different promoters. T4 *alt* gene deletion does not change the T4 burst size (Alawneh et al., 2016),

while a double mutant *alt*⁻ *mod*⁻ produces smaller plaques (Nechaev and Severinov, 2003). ModA and ModB tend to aggregate at concentrations above 1 mg/mL and, like Alt, are capable of autoribosylation (Tiemann et al., 1999). In T5, two studies have investigated viral proteins that can be copurified with the bacterial RNAP. T5 Gp26 (18 kDa) is an early protein that binds the RNAP. A Gp26-negative mutant exhibits a lowered burst size (Klimuk et al., 2020). Two pre-early T5 peptides were found to coprecipitate with the RNAP in the absence of a second transfer: a 60-kDa (A1) and 11-kDa polypeptides (McCorquodale et al., 1981). The 11-KDa polypeptide might be one of the small pre-early proteins predicted as soluble: Gp2 (14 kDa), Gp3 (11 kDa), Gp7 (9 kDa), Gp12 (8.6 kDa), Gp14 (8.1 kDa), Gp15 (13 kDa), Gp16 (9.6 kDa). Upon expression of gene *02* in the absence of other viral genes, bacteria harbored a refringent spot by phase contrast, reminiscent of inclusion bodies. Interestingly, it was shown that mutants in *A2* (that encodes a 14-kDa protein) lack ~15-kDa pre-early proteins altogether (McCorquodale and Lanni, 1970; Snyder and Benzinger, 1981), which has rendered Gp2 "invisible" in the past. This suggests that during the infection, Gp2 is stable or soluble only in the presence of A2. Future work will investigate whether T5 Gp2 (14 kDa) binds to the host RNAP along with A1, whether it displays ADP-ribosyl transferase activity and whether it interacts with other viral or host factors.

T5 gene *03* product shows no homology with known proteins. We did not construct a single deletion mutant T5D03, but its deletion in T5 DL strain did not trigger significant perturbations to the kinetic parameters of infection regarding T5 DL and T5 Δ*dmp*. Gp3 is not lethal for the host, does not alter in a significant manner the host morphology. Similarly, Gp9, Gp12, Gp14, Gp16, Gp17 presented no challenge for the cells in which they were produced.

The deletion of gene *05* induces no changes in the eclipse but lengthens the latency compared to other mutants with similar phenotypes, Δ*02* and Δ*07*. T5 Gp5 alters the morphology of the cells rendering them spherical but without being lethal for them. Some molecules can alter the rod to a spherical shape of *E. coli* cells: nigericin, monensin, mecillinam, and A22. Nigericin and monensin are ionophores, mecillinam is an extended-spectrum antibiotic that binds to PBP2 (Nonejuie et al., 2013), while A22 binds MreB inhibiting its polymerization (Shi et al., 2018). Thus, Gp5 might be a membrane protein that acts either as a transporter that depolarizes the cytoplasmic membrane or by altering the integrity or biosynthesis of the cell wall or interacting with the cytoskeleton-like protein MreB.

T5 *amA1* strains do not degrade the host DNA (Lanni, 1969) but stop the host DNA synthesis (Mozer et al., 1977). A1 was suggested to be a multimeric polypeptide (Beckman et al., 1971) needed for host DNA degradation (Lanni, 1969), second step transfer (Lanni, 1968), protein synthesis shut-

off (McCorquodale and Lanni, 1970), and capable of associating to the *E. coli* RNAP (Klimuk et al., 2020; McCorquodale et al., 1981) or to T5 A2 (Beckman et al., 1971). Two peptides with similar weight to A1 can be found during infection: at the estimated 65 kDa and 64 kDa (McCorquodale and Lanni, 1970). It was demonstrated that the recombinant 64-kDa version of A1 produced in *E. coli* has indeed a DNase activity *in vitro* (thesis Zangelmi, 2018). Our study has allowed exploring the A1 activity *in vivo*. Using microscopy and DAPI staining, we have seen that *E. coli* DNA amount remained constant when infected by an *A1*⁻ strain, probing that the host DNA is not degraded or replicated either. Moreover, the production of A1 in *E. coli* cells strongly reduces the bacterial DNA content in the absence of other viral factors, suggesting that the A1 DNase activity probably suffices to digest the host DNA. However, the presence of well-defined foci in A1-producing or T5wt-infected cells suggests that A1 degrades the host DNA up to a certain point. In any case, it remains to be investigated how is the T5 DNA protected from the degradation by A1 and whether and how is the A1 DNase activity essential for the second-step DNA transfer.

A2 amino acid sequence was elucidated by protein sequencing (Fox et al., 1982; Snyder, 1991) and confirmed later by the T5 genome sequencing. It was suggested that A2 is a dimeric DNA-binding protein (Snyder and Benzinger, 1981) loosely associated with the membrane (Snyder, 1984) and capable of associating with A1 (Beckman et al., 1971). The only phenotype of a T5 *amA2* strain was the absence of a second DNA transfer (Lanni, 1968). Cells infected by T5 *amA2* were reported to lack at least two peptides of 14 kDa and 30 kDa, respectively, where the first corresponds to A2 (Snyder and Benzinger, 1981). Regarding the 30-kDa protein identity, we can rule out the 28-kDa protein Dmp since T5 *amA2* induces dNMPase activity (Mozer et al., 1977). We hypothesize that it could be the 37-kDa protein Gp11. This leads to suspect that A2 might be necessary for the expression of gene *11*. In our study, the DNA in T5 *amA2*-infected cells was degraded and not replicated anymore due to the absence of the second DNA transfer. The overexpression of the *A2* in *E. coli* did not alter in a detectable way cell morphology or survival, suggesting that it does not interfere with the membrane homeostasis or the bacterial metabolism. It seems more probable that A2 interacts mainly with other phage factors.

Gp13 might play a role linked to the host membrane. It was the only gene product capable of lysing the cells upon ectopic expression. We produced the protein, tagged in either one of both extremities, and found that the N-terminus tag was lost, probably due to proteolytic processing. The C-ter tagged protein was not found in solution but associated with detergent-solubilized membranes and the high and low-speed centrifugation pellet, which supports that Gp13 is a membrane protein

with a cleavable peptide signal. Questions about the function of this protein at the very early stages of infection remain open.

Mutants T5 *A1*$^{-}$ or *A2*$^{-}$ strains are unable to resume DNA injection after the FST (Lanni, 1968). Phages carrying mutations in the three genes *A1*, *A2*, and *dmp*, were shown to halt the host DNA synthesis (Mozer et al., 1977). This finding implies that some other pre-early protein(s) is (are) capable stopping host DNA replication. In our study, we found that HegG, Gp8, Gp11, Gp13, and Gp15 production killed the host. HegG is predicted to be a homing endonuclease. Gp8 inhibited the cell division, but probably not the DNA replication since the DNA was visible all along the cells. Gp13 is a membrane protein. Gp11 might not be expressed when A2 is mutated, as discussed above. Thus, unless *11* to *17* are transcribed in the same operon, Gp15 is the most probable protein to be responsible for stopping the DNA replication.

HegG is a putative HNH homing endonuclease, an HMM-HMM profile search (Zimmermann et al., 2018) output that it is similar to I-HmuI from the bacteriophage SPO1. I-HmuI structure was solved, and its function was characterized (Shen et al., 2004). The HegG production promoted an alteration of the nucleoid morphology, which let to suppose that HegG might be a DNA interacting protein.

Taken together, this study has highlighted pre-early proteins from T5 that might play a significant role in the host takeover. A further functional and structural characterization of the gene products can help to elucidate the *E. coli* proteins targeted by T5 pre-early proteins and open new routes for a better understanding of the host takeover.

A pending great interrogation remains which are the factors that control the two-step transfer of T5 DNA. The DNA delivery halts at the same position independently of the phage gene transcription and resumes under the action of A1 and A2. The Injection Stop Signal (ISS) was suggested as the locus laying in the interphase between the bacterial cell and the bacteriophage (Heusterspreute et al., 1987) during the first step transfer, i.e., the locus at which the DNA injection stops. This region locates between genes 17 and 18. It is made up of 1,777 bp representing a length of 569 nm (linear DNA-B), much longer than the 25-nm thickness of the *E. coli* envelope. This intergenic region presents some heterogeneities: two-thirds make part of the LTR and are rich in direct and inverted repeats, while the remaining one third is not full of such sequences. The ISS was annotated between 9,140 and 10,149 in T5 wt (NCBI: NC_005859.1 and T5 st0 NCBI: AY692264.1) corresponding to the two thirds repetition-rich fragment, as proposed in the past (Davison, 2015; Heusterspreute et al., 1987). In this study, we show that the proposed locus for the ISS is dispensable

under laboratory conditions. We make two hypotheses: either the ISS is not an essential element for T5 infection, or the actual (and probably essential) ISS is located elsewhere in the genome, outside the terminal repeat.

4. Perspectives

4.1. Role of pre-early genes in the host takeover

We applied two approaches to identify essential T5 pre-early genes and explore their functions: reverse genetics and recombinant expression. The reverse genetics consisted in deleting dispensable genes and the effects of such deletions on phage virulence and kinetics. We could delete 9.3 % of the genome and concluded, therefore, that four or less out of 17 genes are essential for T5 infection. It is therefore possible that some of the essential pre-early proteins are multi-tasking or maybe moonlighting proteins. The functional and structural characterization of proteins A1 and A2, individually and together, is underway in our laboratory.

A1 is the most important protein for the host genome degradation, however, it remains to be elucidated whether it is the only pre-early protein holding a DNase activity, other gene products were predicted as endonucleases, as HegG, which might play an ancillary role during the host DNA degradation. This hypothesis can be tested determining the HegG activity *in vitro*. On top of a nuclease activity, A1 coprecipitates with the *E. coli* RNA polymerase (RNAP) (Klimuk et al., 2020; Szabo and Moyer, 1975) but it is not clear whether the RNAP and A1 interact directly and what might be the effects of such interaction on the RNAP. Again, biochemical analysis of A1 will help shed light on this matter.

The recombinant expression of the phage T5 pre-early genes in the host is a bacteriophage genome-driven screening to explore the function of pre-early proteins on the host takeover. Some T5 gene products impaired the bacterial multiplication. The targets of bactericidal T5 pre-early proteins can inspire the design of antibacterial agents as it was done for *Staphylococcus* in the past (Liu et al., 2004). Different bacterial target-driven approaches can be applied to complement the preliminary results obtained in our study: Identification of protein partners by pull-down assays followed by mass spectrometry, bacterial or yeast two-hybrid systems, or toxicity-abolishing screening of bacterial genes (Wan et al., 2021). We carried out the ectopic expression of phage pre-early genes in only one condition, 37 °C and LB broth. It might be useful to compare different temperatures of incubation and media. For example, the phage T4 protein Ndd (D2b) recognizes the *E. coli* HU (heat unstable) binding sites. Ndd activity could be found thanks to the incubation of cells recombinantly expressing *ndd* at 42°C (Bouet et al., 1998).

T5 gene *dmp* is present in all the T5-like viruses; moreover, among all the deleted genes, it was the only affecting the virion assembly. Mutants lacking *dmp* showed slightly lower titers in EcoRI$^+$ *E. coli* strains. Dmp function might be important not just for the dephosphorylation of nucleotides.

Previous studies reported that T5 Dmp reaches its maximum activity 5 minutes after the infection; then, it is rapidly inactivated (Mozer et al., 1977). The peak of activity coincides with the appearing of the deoxynucleotide kinase activity and DNA synthesis (Mozer et al., 1977). Because Dmp is apparently stable under laboratory conditions (Mozer and Warner, 1977), its activity might be inhibited, or Dmp might be degraded by another protein, probably encoded in the SST-DNA. The research for Dmp inhibitors can help elucidate the interactions between pre-early proteins and the rest of the phage proteins.

Data from the 1970s suggests that the production of A2 might be necessary for the presence of Gp2 and Gp11. There might therefore be a stepwise regulation of the transcription of the FST locus. Transcriptomic analysis of cells infected with wild-type and mutant T5 strains should be performed to test this hypothesis.

4.2. Identity and function of the ISS

It was proposed that the ISS interacts with A1/A2 and some bacterial protein for the second transfer of the DNA (Heusterspreute et al., 1987). Another scenario might be possible, involving the tail tape-measure protein with a muralytic activity (Boulanger et al., 2008) that is ejected from the tail tube and probably injected into the host immediately (Arnaud et al., 2017) after the irreversible binding of T5 to its receptor FhuA. This primary protein ejection certainly triggers the opening of the capsid and DNA delivery. Whether this protein is capable of transiently interact with the ISS until the action of A1 and A2 needs to be tested.

The reduction of the non-coding region between genes *17* and *18* confirms that not all the region is essential but does not prove that the remaining region contains the information for stopping the DNA transfer. Therefore, the function of the putative ISS must be confirmed through other approaches. Since genes *hegG* to *17* are dispensable for infection, one could introduce the putative ISS before them and test whether the injection stops at that point (Figure 57). These experiments should be carried out preferably with a modified T5 *amA1A2* mutant because it does not degrade the DNA nor reverts. The elimination of the phage particle with its SST DNA can be done by using a blender, allowing the sequencing of the FST DNA injected in the host. We have constructed the plasmid pUChegGter2 for the HR with T5 *amA1A2*. The selection of a T5 mutant with a shifted ISS to carry out such experiment.

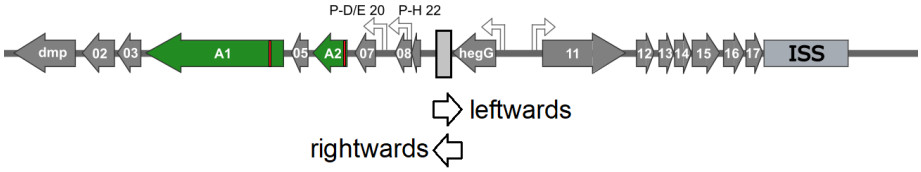


Figure 57. Proposed insertion of the putative Injection Stop Signal at the left terminal repeat.
DNA delivery can be studied in a T5 *amA1A2*, by inserting the putative ISS 4000 kb before the supposed ISS locus, in different orientations. A blender experiment can help identify the size of the transferred DNA, confirm the function of the sequence inserted, and suggest about the freedom of movement of the interacting protein.

A blender experiment could also answer other questions about the freedom of movement of the interacting proteins. On the one hand, if the orientation of the "new ISS" matters for the first delivery, it would suggest that the interacting proteins are restrained (membrane or tail-phage proteins). On the other hand, if the orientation of the new ISS does not affect the delivery, it would mean the interacting proteins might have some freedom of movement (soluble proteins).

An alternative hypothesis for the function of the intergenic region between genes 17 and 18 is that it contains a packaging signal. The mechanism of T5 DNA packaging remains poorly understood. We made a preliminary analysis from the NGS data to assess the left and right termini positions and

see if these termini were displaced after gene deletions. The left terminus was identified and constant in all cases, but the right terminus remains undefined. Left and right termini of deletion mutants must be tested in the future with the help of a single-strand-interruption-less mutant of T5 (Section 7.5.2, Table 12, Figure 67).

5. Conclusion

The virulent bacteriophage T5 uses an atypical strategy to take over its host by producing 17 pre-early proteins before the entire genome enters the cell. These proteins set up favorable conditions for phage propagation: they destroy the host genome, shut-off the host DNA replication, and neutralize the hostile host nucleases.

During this work, we have developed tools for gene editing in T5, applied them to construct deletion mutants with the aim of identifying which of the pre-early genes are essential for the infection. Surprisingly, we found that 13 of the 17 pre-early genes are dispensable for infection under laboratory conditions, suggesting that the host takeover probably relies on few multifunctional proteins and others with ancillary functions.

We also started exploring the function of all the annotated pre-early genes through expression in the host out of the viral context. Toxicity and cell morphology assays suggest that the viral gene products target DNA metabolism, host cell division, and the cell membrane. This exploration of the main events of T5 infection now calls for further functional and structural investigations of the pre-early proteins, which are crucial for host neutralization and thus might eventually inspire the design of new antimicrobial drugs.

6. Materials and Methods

6.1. Strains and culture conditions

6.1.1. Bacterial strains

The *Escherichia coli* strains utilized in this study were: DH5alpha for cloning, strain F for all the manipulations of T5 (e.g., phage stock production, titration) unless otherwise specified, CR63 (Appleyard et al., 1956) as an amber suppressive strain ($_{CUA}$tRNA^{Ser}), XL1-Blue (*phuA*$^{+}$ strain, sensitive to phage T5, from Stratagene) for infections with the phage T5 *lacZ*α, and K-12 sub strain. MG1655 for infections with T5 PNmC.

All mutants obtained in this study descend from Bacteriophage T5 (GenBank: NC_005859).

Unless otherwise stated, bacterial strains were incubated at 37 °C. Final antibiotic concentrations: 200 μg/mL ampicillin (Amp), 100 μg/mL kanamycin (Km), 100 μg/mL streptomycin (Sm), 25 μg/mL chloramphenicol (Cm). Other compounds used were arabinose 0.4 %, glucose 0.4 %, IPTG 0.5 mM , X-Gal 0.6 mg/mL.

Media: LB broth (Sigma #L3022), LB agar (Sigma #L2897), and M9.

100 mL	M9 salt solution (10X)	Na2HPO4	33.7 mM
		KH2PO4	22.0 mM
		NaCl	8.55 mM
		NH4Cl	9.35 mM
20 mL	20% glucose	Glucose	0.4 %
10 mL	20% casaminoacids	Casaminoacids	0.2 %
1 mL	1 M MgSO4	MgSO4	1 mM
0.3 mL	1 M CaCl2	CaCl2	0.3 mM
1 mL	Biotin (1 mg/mL)	Biotin	1 µg
1 mL	Thiamin (1 mg/mL)	Thiamin	1 µg
10 mL	Trace elements solution (100X)	Trace elements	1X
15 g	Agar		

Stocks:

M9 salt solution 10x	Na_2HPO_4-$2H_2O$	75.2 g/L	
	KH_2PO_4	30 g/L	
	NaCl	5 g/L	
	NH_4Cl	5 g/L	
Glucose 20%	Glucose	200 g/L	
$MgSO_4$	$MgSO_4$ $6H_2O$	228.5 g/L	1 M
$CaCl_2$	$CaCl_2$ $2H_2O$	147.0 g/L	1 M
Biotin	Biotin	50mg/50mL	
Thiamin	Thiamin	50 mg/50 mL	
100X trace elements solution	EDTA	5 g/L	13.4 mM
	$FeCl_3$	0.50 g/L	3.1 mM
	$ZnSO_4$-$7H_20$	177 mg/L	0.62 mM
	$CuCl_2$-$2H_2O$	13 mg/L	76 μM
	$CoSO_4$	9.3 mg/L	42 μM
	H_3BO_3	10 mg/L	162 μM
	$MnCl_2$-$2H_2O$	1.3 mg/L	8.1 μM

6.1.2. Plasmids

Plasmids were constructed by Golden Gate assembly (Weber et al., 2011) or by homologous recombination (HR) (Jacobus and Gross, 2015).

PCR amplifications were carried out from 10 ng of DNA or directly from a phage stock. The inserts were amplified with a cycling protocol: at 98°C, then, 30 cycles at 98 °C 30 sec, 55 °C 30 sec, 72 °C 30 sec/Kb, finally 5 minutes at 72 °C. The mix was prepared with Phusion polymerase (Thermo #F530S) 1U, 10uM of each primer, Quick change reactions were carried out as described before through amplifications in separated tubes with one primer in each, followed by an annealing step, digestion with DpnI (Thermo #FD1703), and transformation.

For Golden Gate cloning, a mix was prepared with 200 ng of plasmid DNA, insert/plasmid ratio 3:1, 1 U of BsaI (NEB #R3535) and 1 U of T4 ligase (Thermo #EL0014), incubated for 30 cycles at 37 °C and 22 °C, 5 min each, then 2 hours at 16°C. An aliquot was transformed in *E. coli* DH5alpha. To facilitate cloning of viral DNA, the multiple-cloning site in pUC19 was replaced by two BsaI sites, this by PCR amplification with primers pUCFPhos/pUCR and subsequent re-ligation of the linearized plasmid, generating pUC-GG. In parallel, the plasmid pBAD-GG was assembled from pBAD24, with the primers pBADFPhos/pBADR. The BsaI sites on the Ampicillin cassette (QC1589pUC/QC1608pUC) and on the terminator rrnB (QCpBADF/R) of pUC19 or BsaI were erased by Quick Change (Deng and Nickoloff, 1992). iPCR and ligation to replace the MCS (pBADFPhos/pBADR) (Figure 58).

To construct pUCA1am, the insert was amplified from T5 with primers GGA1TF/GGA1TR and cloned into pUC-GG, then, a point mutation was introduced by Quick Change with primers QCA11/2. pUCA2am insert was amplified with gg_pUC_A2_F/R and modified with primers QCamA2F/R.

For single gene deletions, to minimize polar effects, we kept the start codon and the last ten codons. To provide a template for homologous recombination during infection, DNA fragments carrying 300-500 bp on both sides of the targeted sequence were amplified and cloned into pUC-GG, all of them with BsaI restriction sites, by a cyclic ligation-restriction (Figure 58).

For HR cloning, the primers contain 20-bp overhangs that matched the target plasmid. The inserts were amplified with the PCR protocol cited above. Between 100-200 ng of DNA was used for transformation.

For the phage infection interference assays, plasmid pAC9 was constructed by homologous-recombination mediated cloning of the *cas9* gene was amplified with primers Cas9F/Cas9R and cloned under the control of the P_{BAD} promoter and the transcriptional regulator *araC* gene (from the plasmid pBAD24), this amplified with primers BC9GF/BC9GR; the origin of replication (p15A), from

the plasmid pCas9 was copied with primers ACYCoriF/ACYCoriR. This module was combined with the kanamycin cassette and the multicloning site for the constitutive transcription of a sgRNA from the plasmid psgRNAcos with primers psgF/psgR (Figure 58).

The plasmids pJJC1418 and pJJC1428 were kindly provided by M.A. Petit (INRAE), psgRNAcos by D. Bikard (Institut Pasteur).

pAC9 sgRNA-coding plasmids were constructed by cloning spacer oligonucleotides into the plasmid pAC9. 24-bp pairs of complementary oligonucleotides were aligned at a 0.1°C/s, starting from 98° to 20°C. The dsDNA formed was cloned into pAC9 by Golden Gate cloning.

For ectopic expression of viral genes, T5 pre-early genes were amplified and cloned in pBAD-GG. Primers follow the notation [gene]ATG/[gene]Stop e.g., *dmp* (dmpATG/dmpStop), *02* (02ATG/02Stop). All the T5 pre-early genes were cloned in pBADGG, except *11*, the wild type gene (Gene ID: 2777589) contains no BsaI site, which was also demonstrated by sequencing the PCR product, nevertheless, it was not possible to clone this gene. A codon-optimized synthetic version of *11* was ordered to IDT DNA technologies to be cloned into pBADGG. The insert for the pBADA1A2 was amplified from the phage T5 D05. Transformants bearing pBADA1A2 were plated and screened using M9-Glucose medium instead of LB, plus Ampicillin 200 ug/mL.

The sequence of sites *parS*(T1) and *parS*(P1), genes *mCherry* (AY678264, Excitation/Emission peaks of 587/610 nm for the encoded protein), *lacZ*α, and *11* optimized, as well as plasmids pFH2973 (Nielsen et al., 2006) and pJJC1418 (Kulakauskas et al., 1995) are cited in the Section 7.2. The sequence of primers for the construction of plasmids is detailed in the Section 7.6.

Table 8. Plasmids used in this study

Plasmids	Source	Reference
pCas9	Addgene #42876	(Jiang et al., 2013)
pUC19	Lab stock	Genbank: M77789.2
pBAD24	Lab stock	(Guzman et al., 1995)
pFF745	Addgene #61450	(Farzadfard and Lu, 2014b)
psgRNA	David Bikard (Institut Pasteur)	(Cui et al., 2018)
pFH2973	Frederic Boccard (I2BC)	(Nielsen et al., 2006)
pJJC1418	Marie-Agnès Petit (INRA)	(Kulakauskas et al., 1995)
pJJC1428	Marie-Agnès Petit (INRA)	(Kulakauskas et al., 1995)
From pUC19	Primer Set 1	Primer Set 2
pUCDdmp	DdmpF+dmpR2	dmpF2+DdmpR
pUC02	FD533/52-LR + FC1965/46-LR	
pUCD02	D02F + D02R	
pUC02parST1	02parST1G_F + 02parST1G_R	
pUCA1T	GGA1TF + GGA1TR	
pUCA1am	QCA11	QCA12
pUC05	05F +05R	
pUCD05	D05F + D05R	
pUCEcoA105	QCEcoA105F	QCEcoA105R
pUCNotA105	NotA105F + NotA105R	
pUC05-lacZa	pUC05-F + pUC05-R	LacZalpha-ATG + LacZalpha-Stop
pUC05-lacZa2	QCLacZSD1 + QCLacZSD2	
pUCA2	ggpUCA2F+ggpUCA2R	
pUCA2am	QCamA2F	QCamA2R
pUCD07	D07F + 07R3	07F3 + D07R
pUCDFSTR	DFST-R + FST-R	FST-F + DFST-R
pUCISSinv	FC8563inv+FC11322inv	
pUCEco16IS	QCISS-1	QCISS-2
pUC17parSP1	17parSPIC_F + 17parSPIC_R	
pUCDIS12	DIS2.2F + DIS1R	
pUCDIS13	DIS1F + DIS3.2R	
pUCDIS23	DISF3 + DIS2R	
pUCDIS123	DISF3 + DIS1R	
pUCPNmC	mCF + T5.1518	T5.151+mCR
From pAC9		
pAC_dmp	gdmpS + gdmpA	
pAC_02	gRNA-02S + gRNA-02A	
pAC_A1T	gRNAA1Ts + gRNAA1TA	
pAC_05	gRNA05S + gRNA05A	
pAC_07inv	g07invS + g07invA	
pAC_10	g10S + g10A	
pAC_10.4	g10.4S + g10.4A	
pAC_10.5	g10.5S + g10.5A	
pAC_10.6	g10.6S + g10.6A	
pAC_14.3	g14.3S + g14.3A	
pAC_14.4	g14.4S + g14.3A	
pAC_14.5	g14.5S + g14.5A	
pAC_14.7	g14.7S + g14.7A	
pAC_14.6	g14.6S + g14.6A	
pAC_IS3	gIS3S + gIS3A	
pAC_IS6	gIS6S + gIS6A	
pAC_IS5	gIS5S + gIS5A	
pAC_IS4	gIS4S + gIS4A	
pAC_IS7	gIS7S + gIS7A	
From pBAD24		Comments
pBADdmp	dmpATG + dmpStop	

pBAD_02		From Léo Zangelmi
pBAD_03	03ATG + 03Stop	
pBADA1long		From Madalena Renouard
pBAD_05		From Léo Zangelmi
pBADA2	A2ATG + A2Stop	
pBAD_07		From Léo Zangelmi
pBAD_08	08ATG + 08Stop	
pBAD_09	09ATG + 09Stop	
pBADhegG	hegGATG + hegGStop	
pBAD_11		Synthetic gene cloned
pBAD_12	12ATG + 12Stop	
pBAD_13	13ATG + 13Stop	
pBAD_14	14ATG + 14Stop	
pBAD_15	15ATG + 15Stop	
pBAD_16	16ATG + 16Stop	
pBAD_17	17ATG + 17Stop	
pBADA1A2	A2ATG + A1Stop	

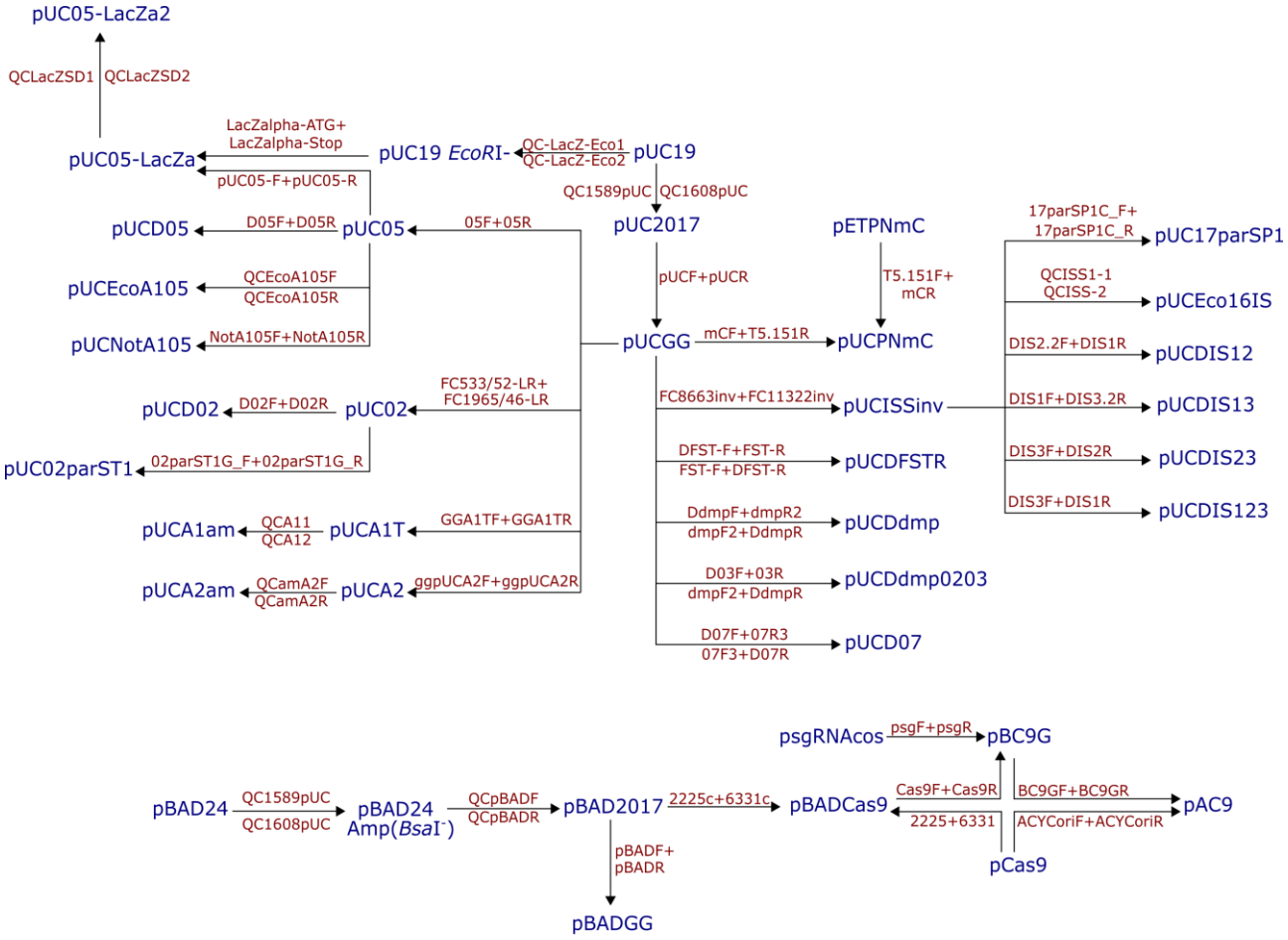


Figure 58: Pedigree of plasmids constructed.
Plasmids constructed (blue) were assembled using the primers (red) cited on each side of the arrows. Primers in the same side of an arrow were used in the same reaction.

6.1.3. Phage strains

Genes *dmp*, *02*, *05*, *07*, as well as the combinations *dmp-02-03* were deleted by homologous recombination between the wild type phage and plasmids pUCDdmp, pUCD02, pUCD05, pUCD07, and pUCDdmp0203, respectively (Table 9, Figure 59), during the infection of the recipient cell, *E. coli* strain F.

The strain T5 DR (Δ*hegG* Δ*11-17*) comes from the HR between pUCDFSTR and T5 *ris16IS*. The counterselection was carried out in F pJJC1318.

All the mutants were screened by PCR and sent to be sequenced by Sanger method whenever positive. The sequence of primers for the mutant screening is detailed in the Section 7.6.

Table 9. Phage mutants obtained during this study.

Name	Phage	Genotype[a]
T5 Δ*dmp*		Δ[153..856]
T5 Δ*02*		Δ[971..1332]
T5 *amA1* strain T28		*A1* T28Stop (Δ[3,293..3,296]::[ctag])
T5 *amA1* strain SS84		*A1* S84Stop, S85Stop (Δ[3,122..3,127]::[ctacta])
T5 *amA1A2*	T5 *amA1* *amA2*	*A1* T28Stop (Δ[3,293..3,296]::[ctag]) *A2* S37Stop (Δ[4,023..4,026]::[ccta])
T5 Δ*05*		Δ[3,505..3,668]
T5 *amA2*		*A2* S37Stop (Δ[4,023..4,026]::[ccta])
T5 Δ*07*		Δ[4,262..4,479]
T5 D257	T5 Δ*02* Δ*05* Δ*07*	Δ[971..1,332] Δ[3,505..3,668] Δ[4,262..4,479]
T5 DL	T5 Δ*dmp0203* Δ*05* Δ*07*	Δ[153..1,647]::[catt] Δ[3,505..3,668] Δ[4,262..4,479]
T5 DR	T5 Δ*hegG* Δ*11-17*	Δ[5,393..9,161] Δ[9,875..10,340]
T5 DLDR	T5 Δ*dmp0203* Δ*05* Δ*07* Δ*hegG* Δ*11-17*	Δ[153..1,647]::[catt] Δ[3,505..3,668] Δ[4,262..4,479] Δ[5,393..9,161]
T5 DISS-3		Δ[9,204..10,391]
T5 *lacZ*α		Δ[3,464..3,449]::*lacZ*α
T5 PNmC	T5 *pb10Nter-mcherry*	Δ[10,3638..103,928]::*mcherry*
T5 PNmC 02parST1	T5 *pb10Nter-mcherry* *02parS*(T1)	Δ[10,3638..103,928]::*mcherry* [10,366..10,367]::*parS*(T1)
T5 *risA105*		Δ[3,446..3,451]::[gaattc]
T5 *ris16IS*		Δ[9,233..9,238]::[gaattc]
T5 NotA105		Δ[3,446..3,456]::[gcggccgc]

[a] The coordinates consider Genbank: NC_005859.1 as the reference.

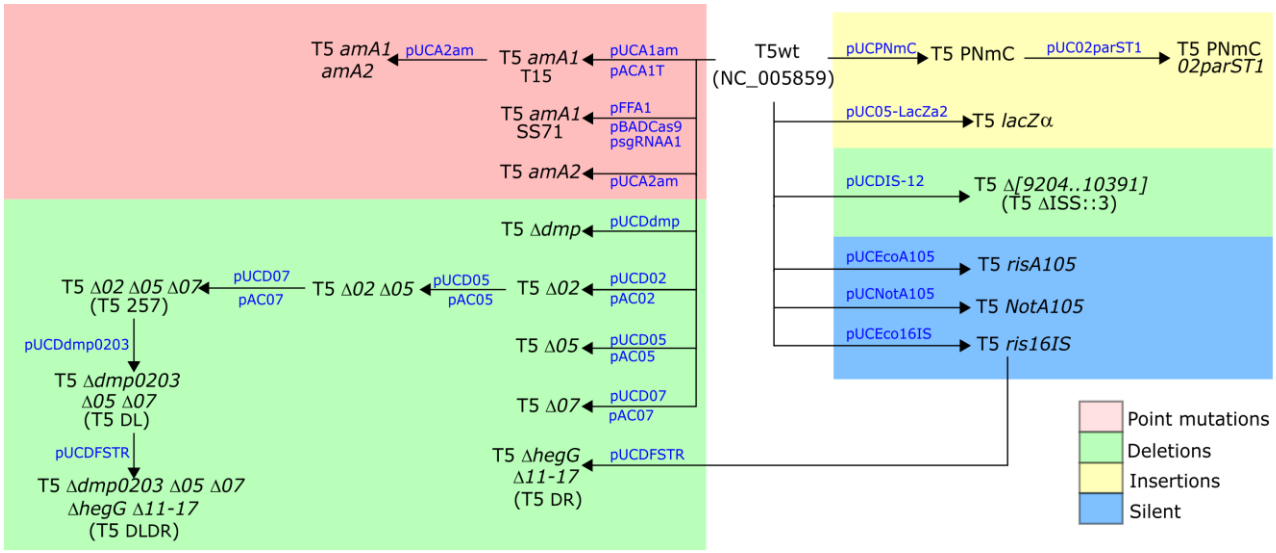


Figure 59. Pedigree of phages obtained during this study.
The phages are sorted according to their parental strain and grouped by type of modification. Template plasmids and CRISPR-Cas9-coding ones are highlighted in blue, on opposite sides of the lines. Strain names are in parenthesis.

6.1.4. Scripts used in this study

All the routines used in this study were deposited in the repository PhageT5 function of each is detailed in the sections where they were applied.

Repository: https://github.com/lumramirezch/PhageT5

Directories:

- `geneconservation/` used in section 2.2.1 to determine the conservation of T5 genes in phages from the same clade.
- `imageJprocessing/` used in sections 2.3.3 and 2.4.3 to analyze cell morphology parameters.
- `phagevir/`: used in section 2.3.2 to determine the phage virulence index and MV_{50}.

6.2. Deletion of phage T5 pre-early genes

6.2.1. T5 pre-early genes conserved among tequintaviruses

A list of taxa identification number (taxid) for genomes in a specific taxa was downloaded from the NCBI taxonomy browser. The taxid was input to the script OrthoFinder_script.sh, which retrieved the gene annotation files (protein.faa) from the NCBI genome database via Entrez and uniform the fasta (protein.faa) headers. From the fasta files, it constructs orthogroups with OrthoFinder (Emms and Kelly, 2019), using diamond as the sequence aligner. In the example, the option -a indicates the path to the directory containing the taxid list:

```
$ ./Orthofinder_script.sh -a "path/to/file/"
```

The list of genes was counted and plotted with the script Orthologs_Plot.py.

6.2.2. Site Directed mutagenesis

The mutation was designed and cloned flanked by 300 to 500 bp of homologous regions. In the pUC19 derivative plasmid Golden Gate compatible pUCGG.

Phage site directed mutagenesis is carried out through homologous recombination between the phage genome and the cloned template during the infection. The host bearing the template construction was incubated from an overnight preculture (OD_{600}~2.5) diluted at OD_{600}~0.1 in LB (Amp, Ca^{2+}, Mg^{2+}); then, incubated at 37°C in 220 rpm shaking until it reaches mid-log phase (OD_{600}~0.4) when it was infected at a MOI 10 with the parental phage (see phages pedigree table) and incubated for 5 minutes. The culture was washed three times and finally resuspended up to the volume with LB (Ca^{2+}, Mg^{2+}). It was then incubated until complete lysis (OD_{600}~0.1). The crude lysate was centrifuged at 8500 rpm for 3 min and the supernatant filtered with 0.2 μM nitrocelulose syringe filters.

6.2.3. Mutants screening with CRISPR-Cas9

The plasmids to obtain the different phages are listed in ST 1. *E. coli* CR63 was transformed either with pUCA1am (T5 *amA1*) or pUCA2am (T5 *amA2*), *E. coli* XL-1 Blue with pUC05-lacZa2 (T5 lacZa), while *E. coli* F was transformed with pUCPNmC (T5 PNmC).

The cited host were incubated from a single colony overnight at 37 °C. Host cultures were started at OD_{600} ~0.1 in LB broth (ampicillin 200 ug/mL, 1mM of Ca^{2+} and Mg^{2+}) from the pre-culture and incubated at 37 °C until they reached OD_{600} 0.4-0.5. The cultures were infected at MOI 10, incubated 10 minutes at 37 °C. The cells were washed twice through centrifugation at 5,000 x g for 3 min and resuspension with LB broth. Cells were harvested, resuspended at half of the initial volume, and incubated at 37 °C for 90 min. 0.2% chloroform was added to the culture, incubated 10 min at

37 °C. The lysate was centrifuged at 5,000 x g 3 min, and the supernatant was separated and stored at 4 °C.

6.2.4. Mutants screening by dilution

The lysate obtained from HR was filtered with 0.45-μm diameter nitrocellulose filters. Serial dilutions were made to adjust the lysate concentration to 10^8-10^{10} (10 mL each), with LB (CaCl2 1mM, MgSO4 1mM).

A host culture in stationary phase (OD_{600} 0.4-0.5) was distributed in a 96-well flat bottom polystyrene plate (Corning), 100 uL per well. 50 uL of the lysate, from the mutagenesis, serial dilutions (10 to 100 PFU) were added to each well and incubated for 3 hours at 37 °C and 220 rpm shaking. 20 uL of each well in the same row were pooled and tested by PCR.

6.2.5. Whole genome sequencing of T5 mutants

Phage DNA was purified by Phenol-Chloroform extraction, precipitated and washed with ethanol 70 % and resuspended. The samples were sent to the I2BC NGS platform for whole genome sequencing.

The samples (T5wt, T5DL, T5DR, T5DLDR) were prepared with TruSeq genomic, NextSeq 500/550 Mid Output Kit v2, with 80 cycles. Between 6-14 million reads were gathered for each sample.

To map the newly formed sequences as well as to see the deleted regions, the left terminal repeat was modified according to the results of Sanger sequencing of each mutant. The last terminal repeat was left with no modifications.

Short reads were mapped to each reference genome using Bowtie2 in default mode, SAMtools, BEDTools, as described in (Hung and Weng, 2017) in a Windows Subsystem Linux (WSL). The results were visualized with the Integrative Genomics Viewer (IGV) version 2.8.9.

6.3. Characterization of T5 mutants

6.3.1. One step growth curve

Cultures with the host (*E. coli* strain F) were incubated from OD_{600}~0.1 in LB (1mM of Ca^{2+} and Mg^{2+}) over 1 hour until they attained OD_{600}~0.4. Then, 9.9 mL of the culture was infected with 100 uL a phage dilution at 1x10^7 PFU/mL and replaced in incubation for 5 minutes at 37 °C 220 rpm shaking. After 5 min 100 uL was taken from the infected bacterial culture to 9.9 mL of prewarmed LB (1mM of Ca^{2+} and Mg^{2+}) medium (flask A). The serial dilution was carried out in a second flask with 9.9 mL of LB (1mM of Ca^{2+} and Mg^{2+}, flask C). For strains with a slow rise in titers during the infection, namely T5 Ddmp, DL, and DLDR, 1 mL of the flask A was diluted in 9 mL of prewarmed media (flask B), and again, 1 mL was transferred to 9 mL of prewarmed media (flask C). This allowed a better titration between the onset of the lysis and until stabilization of titer of released phages.

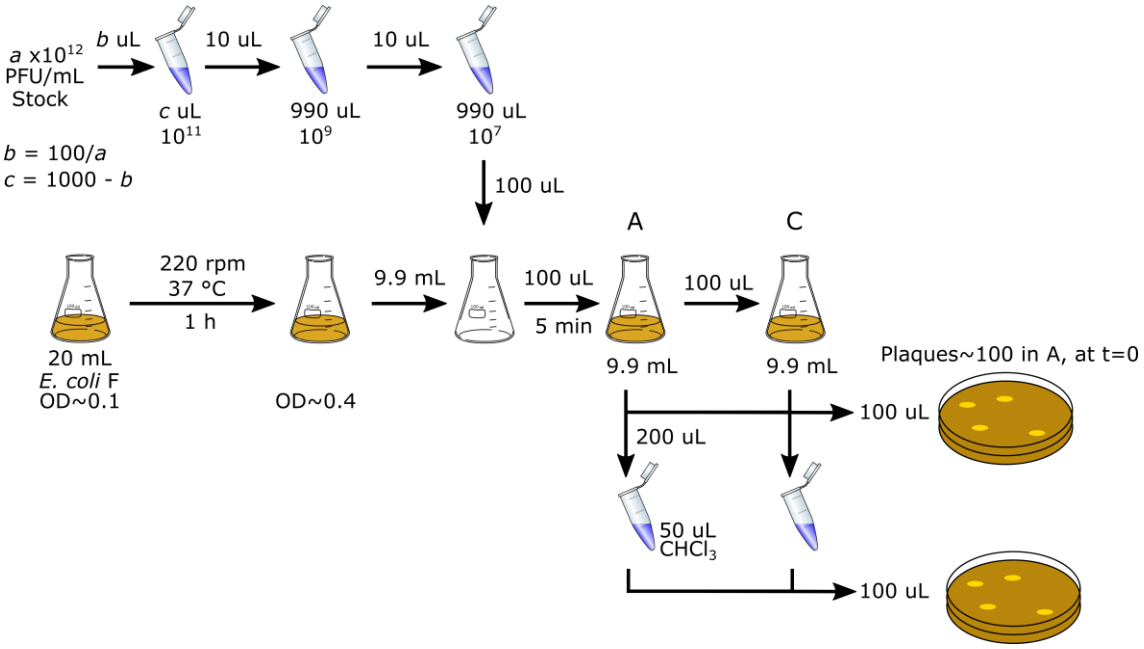


Figure 60. One Step growth protocol used in this study.
The stock of phage was diluted to 10^7 PFU/mL. 100 uL of diluted stock was added to 9.9 mL of a bacterial culture in exponential phase and incubated for 5 min at 37°C 220 rpm shaking. The mix was serially diluted twice 100 times (flasks A and C). Both flasks were placed in incubation at 37 °C 220 rpm for up to 90 minutes. 300-uL aliquots were taken every 5 min from the flasks A and C. 200 µL were added 50 µL of $CHCl_3$ to stop the infection, while 100 µL were mixed with top agar plus a 100 uL of the host culture and immediately plated. The aliquot on chloroform were plated in the same manner afterwards. Adapted from (Kropinski, 2018).

The flasks A and C (and B if present) were incubated at 37 °C 220 rpm of shaking. Every 5 minutes two aliquots were taken: 200 µL for the first 30 min and 100 µL until 90 minutes. First, 200 µL of each flask was transferred to a 1.5-mL tube with 50 µL of Chloroform to stop the reaction, count non-adsorbed phages and determine the onset of fully assembly of viral particles (aka end of eclipse period), for which 100-µL aliquot of the mix with chloroform was incubated for 40 min, then, it was

plated along with 100 µL of the host at OD_{600}~1.0 in molten top agar over a LB agar petri plate. Second, 100 µL of the cultures from flasks were taken and immediately poured along with the host on the overlay agar on top of LB-agar petri plates (Figure 60).

Plates were incubated overnight at 37 °C. The day after, those with 30 to 300 plaques were considered. The burst size equals the average of the three final titers over the three first. Eclipse and latent time correspond to the last time point before a tenfold rise in the titer.

6.3.2. Virulence index

Stocks of phages were adjusted to 10^9 PFU/mL with LB (Ca^{2+}, Mg^{2+}, 1 mM each). Serial dilutions from 10^9 to 10^2 were prepared and distributed in a 96-well plate. 100 µL of each dilution was dispensed in each well of a 96-well plate. A pre-culture of *E. coli* strain F (OD_{600}~0.4) was adjusted to an OD_{600}~0.1 and distributed in each well. The plate was placed into a multi-plate reader to evaluate the OD_{600} for 3 hours every 10 minutes.

Immediately after the onset of the infection at different MOIs, the Optical Density (OD) was measured every 10 minutes for each MOI, and for up to 6 hours; the results were plotted as OD versus Time (h) grouped by phage, these are the bacterial reduction curves. As explained in (Storms et al., 2020), the bacterial reduction curves were then used to infer three sets of parameters, for each phage:

- Local virulence curve (v_i versus $\log_{10}$ (MOI))
- MOI for a local virulence of 0.5 (MV_{50})
- Virulence Index (V_p)

The local virulence (v_i) curve equals to the area under the bacterial reduction curves for each MOI (A_i in which i corresponds to each MOI), divided by that of the uninfected control (A_0):

$$v_i = 1 - \frac{A_i}{A_0} \qquad (1)$$

The area under the curve (AUC) was calculated with the MESS::auc library in R from splines joining the time points. It computes the AUC through linear interpolation for two vectors (joining points), the x values (Time (h)) and the y values (OD). The MOI required to produce a local virulence of 0.5, considering a maximum virulence of 1 (MV_{50}), corresponds to the MOI at the intersection v_i~0.5 on the local virulence curve (v_i versus $\log_{10}$(MOI)).

$$MV_{50} \equiv MOI|_{v_i=0.5} \qquad (2)$$

The virulence index (V_p) is the AUC of the virulence curve (v_i), obtained from all the tested MOIs. In this case, from 10^{-6} to 1; divided by the maximum theoretical virulence:

$$A_P = \int_{-6}^{1} v_i d(\log MOI) \quad (3)$$

$$A_{max} = \int_{-6}^{1} v_i d(\log MOI) = 7, \text{ being } v_i = 1 \text{ for all MOIs.} \quad (4)$$

$$V_p = \frac{A_P}{A_{max}} \quad (5)$$

6.3.3. Fluorescence microscopy

For the analysis of bacterial cells infected with the different mutants, overnight pre-cultures of *E. coli* F were diluted to OD_{600}~0.1 in LB (1mM of Ca^{2+} and Mg^{2+}) and incubated for 1 hour at 37 °C 220 rpm of shaking until it reached exponential phase (OD_{600}~0.4). Then, 25 mL of the culture were aliquoted on 50-mL Falcon tubes and chilled on ice for 15 minutes. The cells were harvested by centrifugation then resuspended to adjust the OD_{600} to 1. The resuspended cells were infected at MOI~10, gently mixed, and incubated on ice for 5 min.

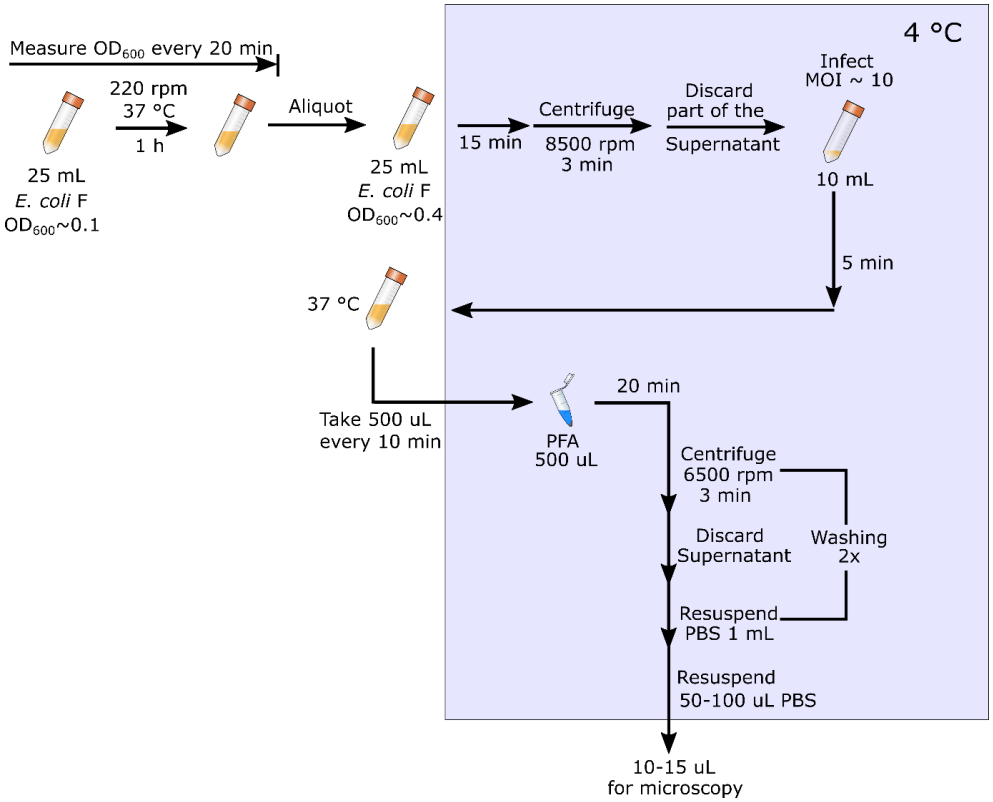


Figure 61. Fluorescence microscopy of infected bacteria.
Cells were prepared following the protocol exposed in the figure. Cell cultures of *E. coli* F were infected during the exponential phase at MOI of 10, incubated for 15 min at 4 °C, centrifuged and resuspended with PBS, fixed with PFA solution, washed and resuspended with PBS 1x. Cells were observed by phase contrast and fluorescence microscopy with appropriate filters.

To trigger the infection, the cultures were set at 37 °C and 220 rpm of shaking. Aliquots of 500 μL were taken every 10 minutes (0, 10, 20, 30 minutes were tested) and added to a 1.5-mL tube with 500 μL of cold PFA solution to fix the bacterial cells. The mix of aliquot plus PFA was gently mixed

and set on ice for 20 minutes. After this time, the fixed cells were harvested by centrifugation then, resuspended with 1 mL of PBS 1x. Process that was repeated twice before the final resuspension of the cells on 50 to 100 μL of PBS. Finally, 10 to 15 μL of the fixed washed infected cells were mixed with DAPI (Thermo #62248) and FM4-64 (Thermo #T13320) dye solutions (Figure 61).

The stained cells were mounted with agarose pads. Agarose pads were prepared by resuspending 1% agarose (Sigma #A9539) with PBS 1x, melting and pouring onto glass slides (Menzel-Gläser; Thermo Scientific) and overlaid with a coverslip (Menzel-Gläser; Thermo Scientific). Contrast and fluorescence microscopy were carried out immediately after montage. Images of DAPI, FM4-64, and phase contrast were obtained with a Zeiss Axio Observer Z1 microscope.

Solutions used in the process had the following composition:

- PFA: Formaldehyde 5%, glutaraldehyde 0.06%, phosphate buffer solution (PBS 1x).
- GTE/DAPI: Glucose 0.9 %, Tris 20 mM, EDTA 10 mM pH 8, DAPI 1mg/mL.

6.4. Ectopic expression of phage genes

6.4.1. Cell growth parameters

Bacterial cells were transformed with the plasmids in which T5 genes were cloned. The cells were incubated for 16 h at 37 °C in LB broth with ampicillin 200 μg/mL, 0.5 % glucose. After it, new cultures were started from the pre-cultures at OD_{600}~0.1 in LB broth with ampicillin 200 μg/mL. The culture was incubated for 1 hours until it achieved OD_{600}~0.4. Then, the cultures were used for three tests: growth (Figure 62, 1), survival (Figure 62, 2), and microscopy (Figure 62, 3) after induction (or repression) of the gene expression.

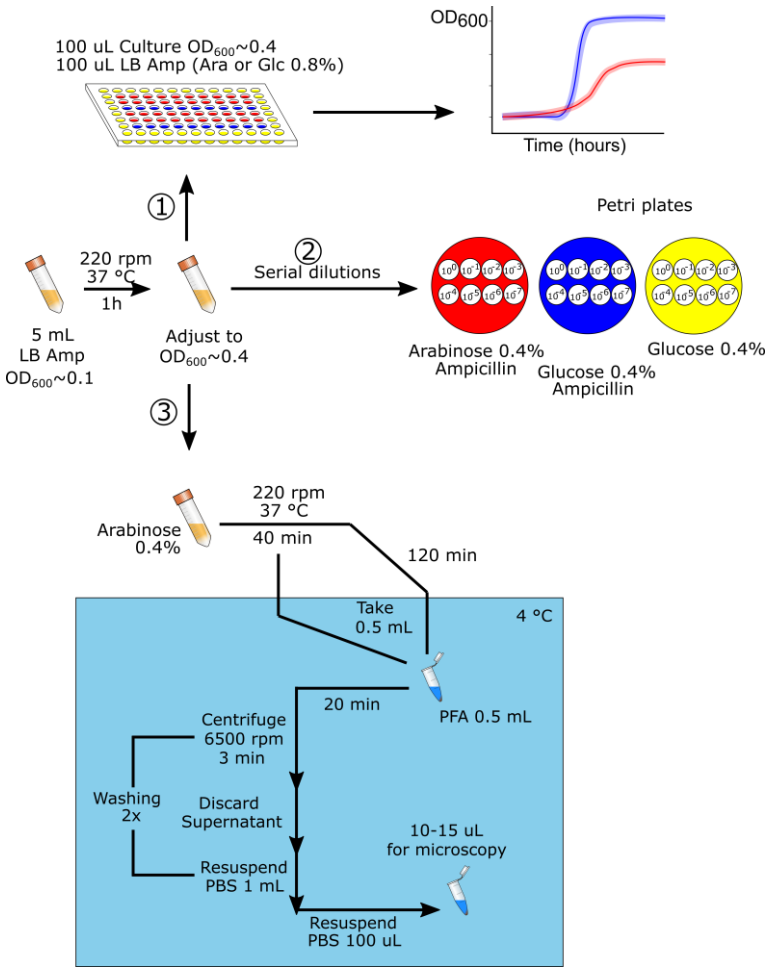


Figure 62. Protocol for ectopic expression.
Plasmids were transformed by electroporation. The day after a fresh culture was started and incubated until exponential phase. Samples were taken to measure (1) Optical density of the cultures after induction, for up to three hours, (2) serial dilutions were spotted onto agar LB with antibiotics and glucose or arabinose, or glucose without antibiotics. (3) gene expression was induced in the cells with arabinose, cells incubated for up to 120 min then fixated to be observed by fluorescence microscopy using DAPI or FM4-64 dyes.

Cultures were distributed in a 96-well plate (100 μL of the culture per well plus 100 of LB with ampicillin 200 μg/mL and either arabinose 0.4 % or glucose 0.4 %) to measure the OD_{600} in a multiplate reader (Figure 62, 1) after induction or repression of the gene expression.

Serial dilutions in LB ampicillin 200 μg/mL were made from the cultures in exponential phase (Figure 62, 2). 5-μL aliquots of each dilution were spotted onto LB agar with ampicillin 200 μg/mL and arabinose 0.4 %, ampicillin 200 ug/mL and glucose 0.4 %, or glucose 0.4 % with no antibiotics.

Finally, the cultures in exponential phase were induced with arabinose 0.4 % and incubated at 37 °C 220 rpm of shaking (Figure 62, 3). 500-μL aliquots were taken after 40 minutes and 120 minutes to be fixed with the same volume of PFA solution and incubated at 4°C. After 20 minutes, cells were centrifuged and resuspended in PBS buffer twice, discarding the supernatant after each centrifugation. Cells were resuspended in 50 μL PBS and conserved at 4 °C until observation by fluorescence microscopy.

Montage from microscopy images were split into single-channel .tiff files with the imageJ macro Split_channels.ijm. Then, merged, cropped, and a scale bar (2 μm) was drawn through the macro Merge_channels.ijm. The analysis of the images was carried out with the ImageJ plugin MicrobeJ 5.13m (2)-beta version (Ducret et al., 2016). The generated dataset was analyzed with the R package to compare cells parameters.

6.5. Functional exploration of pre-early genes

6.5.1. Production of Gp13

E. coli BL21 cells were transformed by electroporation with the constructions pET28, pET-13, pET-13HisN, and pET-13HisC. Being the empty plasmid, the plasmid with the gene *13*, and the fusion to produce Gp13 tagged with 6xHistidine at the N terminus or at the C terminus, respectively. The strains were incubated over 16 hours at 37°C 220 rpm of shaking, in LB broth added with kanamycin 50 ug/mL and glucose 0.5 %.

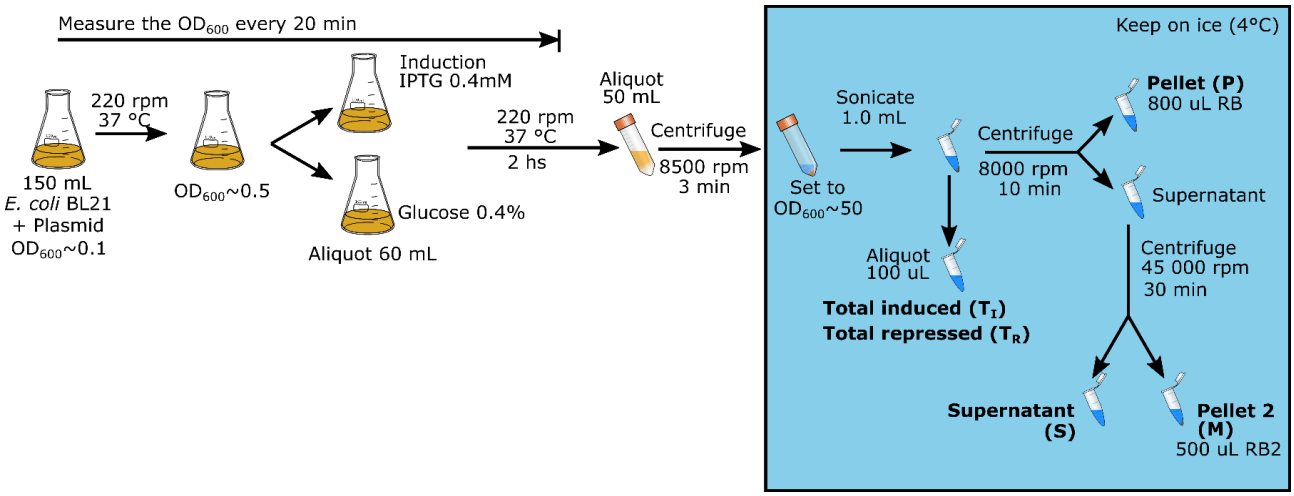


Figure 63. Production of Gp13.
Cells were transformed with plasmids pET28, pET-13, pET-13HisN, or pET-13HisC, then incubated overnight at 37 °C. They were used to start a fresh culture at OD_{600} of 0.1. Once the cultures reached exponential phase, they were aliquoted in two: one aliquot was induced with IPTG 0.4 mM and the other repressed with Glucose 0.4%. The incubation continued for 2 hs. The cells were harvested by centrifugation, resuspended in Resuspension Buffer (RB), sonicated and aliquoted. One aliquot of each induced and repressed sonicated cells was used as a control (T_I and T_R, respectively). The total extract of cells was fractionated by centrifugation at low speed to separate the pellet (P), high speed to separate the membranes (M), and left the supernatant (S) with soluble proteins. All the fractions were migrated in Tris/Tricine SDS-PAGE.

Fresh cultures were started from the transformed cells, they were incubated in LB kanamycin 50 μg/mL at 37 °C 220 rpm and the OD_{600} was evaluated every 20 min. Once the cultures reached OD_{600}, they were split in two aliquots. One culture was induced by the addition of IPTG 0.4 μM, while the other was added with glucose 0.5%. The cultures were incubated for up to 90 min, during which the OD_{600} was controlled every 20 minutes. All cultures were centrifuged at 8500 rpm for 3 min, the supernatant was discarded, and the pellet stored at -20 °C.

The pellets were resuspended to OD_{600}~50 with resuspension buffer (RB), then, sonicated in 4-min cycles (4-sec pulses and 2-sec rest). One aliquot was taken as a control for the total fraction (T_I for the induced cells and T_R for cells incubated in repression) and kept on ice. Sonicated cells were centrifuged at 8000 rpm for 10 min to separate the supernatant. The pellet was resuspended with 800 μL RB and kept on ice (fraction P1). 500 μL of the supernatant was centrifuged at 100 000 g for

30 min, 4 °C to sediment membranes. The supernatant was separated (fraction S) and the pellet of membranes (M) was resuspended with 500 μL of RB2.

Aliquots of 60 μL were mixed with 20 μL of Sampling Buffer 4x and vortexed. The samples were evaporated for 95 °C for 5 min, chilled, then, 15 μL were taken for electrophoresis in BioRad Mini-Protean® Tris-Tricine Precast Gels 16.5%. The samples were migrated at 50 V for 2 hours.

Another gel was run in parallel to transfer the proteins to a membrane and detect them by Western Blot. This gel was transferred to a nitrocellulose membrane (pore size: 0.45 μm) in 1L 1X Transfer Buffer for 45min at 90V. The membrane was stained with a Ponceau Red solution.

The membrane was blocked with 20 mL PBST (5% Milk) for 1h. It was rinsed with PBST, once with 10 mL then 3 times with 5 mL. The membrane was incubated at room temperature with a Peroxydase-coupled Anti-His6 antibody (Roche #11965085001) at 1/1000 dilution in PBST for 1 h. The membrane was rinsed again with PBST, once with 10 mL, then 3 times with 15 mL during 5 min each.

The peroxidase activity was detected with Supersignal WestPicoChemiluminescent substrate (Thermo #34080). Capture the luminescence with GE Healthcare LAS 500 (exposure: 1 min).

Solutions:

- Detergent stock: n-Octyl β-D-Glucopyranoside 100 g/L, diluted with water.
- Resuspension buffer (RB): 250 mM NaCl, 50 mM Tris pH 7.4, 1mM EDTA: 2.5 mL 5M NaCl, 2.5mL 1M Tris pH 7.4, 1mM EDTA, 45 mL Water.
- Resuspension buffer 2 (RB2): RB 9 mL + Detergent stock 1mL.
- Sampling Buffer 4x (BioRad Cat. 161-0739): 200mM Tris-HCl pH 6.8, 40% glycerol, 2% SDS, 0.04% Coomassie Blue G-250, 2% β-Mercaptoethanol.

7. Annexes

7.1. T5 pre-early genes conserved among *Tequintavirus* members

Table 10: Conservation of T5 genes among phages from the genus *Tequintavirus*.
Percentage of virus bearing at least one T5 gene. Known essential genes were highlighted in green.

locus	gene	Total	locus	gene	Total	locus	gene	Total	locus	gene	Total
T5.001	*dmp*	100	T5.041	*C1*	100	T5.083		99	**T5.126**	***D11***	**100**
T5.002		97	T5.042		99	T5.084		99	**T5.127**	***D12***	**100**
T5.003		4	T5.043	*dnk*	100	T5.085		99	**T5.128**	***D13***	**100**
T5.004	***A1***	**100**	T5.044		29	T5.086		96	**T5.129**	***D14***	**100**
T5.005		65	T5.045		93	T5.087		99	**T5.130**	***D15***	**100**
T5.006	***A2***	**100**	T5.046	*hegF*	88	T5.088		99	T5.131	*dut*	100
T5.007		99	T5.047		99	T5.089		99	T5.132		85
T5.008		83	T5.048		98	T5.090		99	T5.133	*ltf*	96
T5.009		9	T5.049		99	T5.091	*rnh*	100	T5.135		99
T5.010	*hegG*	17	T5.050		81	T5.092	*thy*	33	**T5.137**	***D17***	**100**
T5.011		90	T5.052		99	T5.093	*frd*	100	**T5.138**	***D16***	**100**
T5.012		73	T5.053		84	T5.094	*nrdB*	100	T5.139		98
T5.013		68	T5.054		99	T5.096	*nrdA*	100	**T5.140**	***D18-19***	**100**
T5.014		96	T5.055		95	T5.097		97	T5.141		4
T5.015		86	T5.056		98	T5.098	*phoH*	100	T5.142		96
T5.016		2	T5.057		5	T5.099	*nrdD*	100	T5.143		99
T5.017		55	T5.058		97	T5.100		33	T5.144		99
T5.018		43	T5.059		85	T5.101		93	**T5.145**	***N4***	**100**
T5.019		99	T5.060		92	T5.102		96	T5.146		87
T5.020		93	T5.061	*hegD*	24	T5.103		3	T5.147		99
T5.021		88	T5.062		95	T5.104		78	T5.148		99
T5.022		25	T5.063		27	T5.105		99	**T5.149**	***D20-21***	**100**
T5.023		89	T5.065		86	T5.106		95	T5.150		99
T5.024		33	T5.066		16	T5.107		95	T5.151		99
T5.025		99	T5.067		37	T5.108		99	T5.152		99
T5.026		98	T5.068		97	T5.110		80	T5.153		97
T5.027		91	T5.069		41	**T5.111**	***D2***	**100**	T5.154		34
T5.028		99	T5.070		58	T5.112		91	T5.155		99
T5.029		99	T5.071		84	**T5.113**	***D3***	**100**	T5.156		98
T5.030		99	T5.072		95	T5.114		99	**T5.157**	***oad***	**98**
T5.031		21	T5.073		98	T5.115		95	T5.158	*lpp*	6
T5.032		98	T5.074		41	T5.116	*ligA*	99	T5.159		27
T5.033		99	T5.075	*hegA*	6	T5.117	*ligB*	100	T5.160		75
T5.034		96	T5.076		59	**T5.118**	***D5***	**100**	T5.161		65
T5.035		99	T5.077		37	**T5.119**	***D6***	**99**	T5.162		65
T5.036		38	T5.078		37	T5.121	*pri*	100			
T5.037		99	T5.079		2	**T5.122**	***pol***	**100**			
T5.038		99	T5.080		99	T5.123		79			
T5.039		77	T5.081		99	**T5.124**	***D10***	**100**			
T5.040	*lys*	100	T5.082		98	T5.125		77			

7.2. Uncloned regions

Here we present a summary of the regions in the FST that could be amplified but not cloned, or those that could be so. The difference uncloned versus cloned matches the intergenic regions *hegG-11*, *11-12*, *12-13*, and the 30 bp upstream and downstream *08* or *09*.

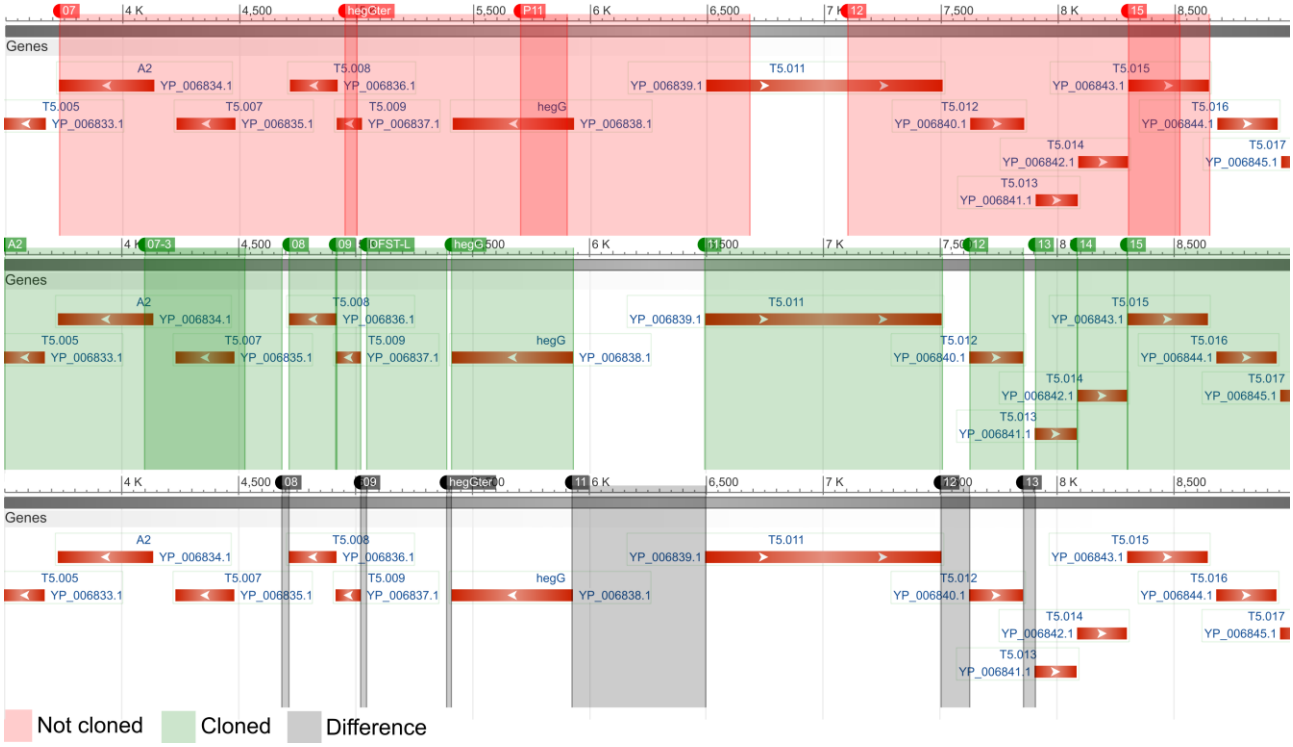


Figure 64. Summary of the cloned regions of the FST.
Above, in red, the regions of the FST that were amplified but could not be cloned after different attempts (n=3). Middle, in green, the regions of the FST that were amplified and could be cloned either in a high or in a middle-copy-number plasmid. Below, in gray, the difference between the uncloned and cloned regions: 30 bp upstream *09*, 30 bp downstream *08*, and intergenic regions *hegG-11*, *11-12*, *12-13*.

7.3. Distribution of DAPI-dye foci in bacterial cells

Fluorescent microscopy images were analyzed by the MicrobeJ plugin of ImageJ (Section 6.3.3). DAPI dye foci in detected cells were counted and a two dimensional histogram constructed to depict their distribution across an "average" cell.

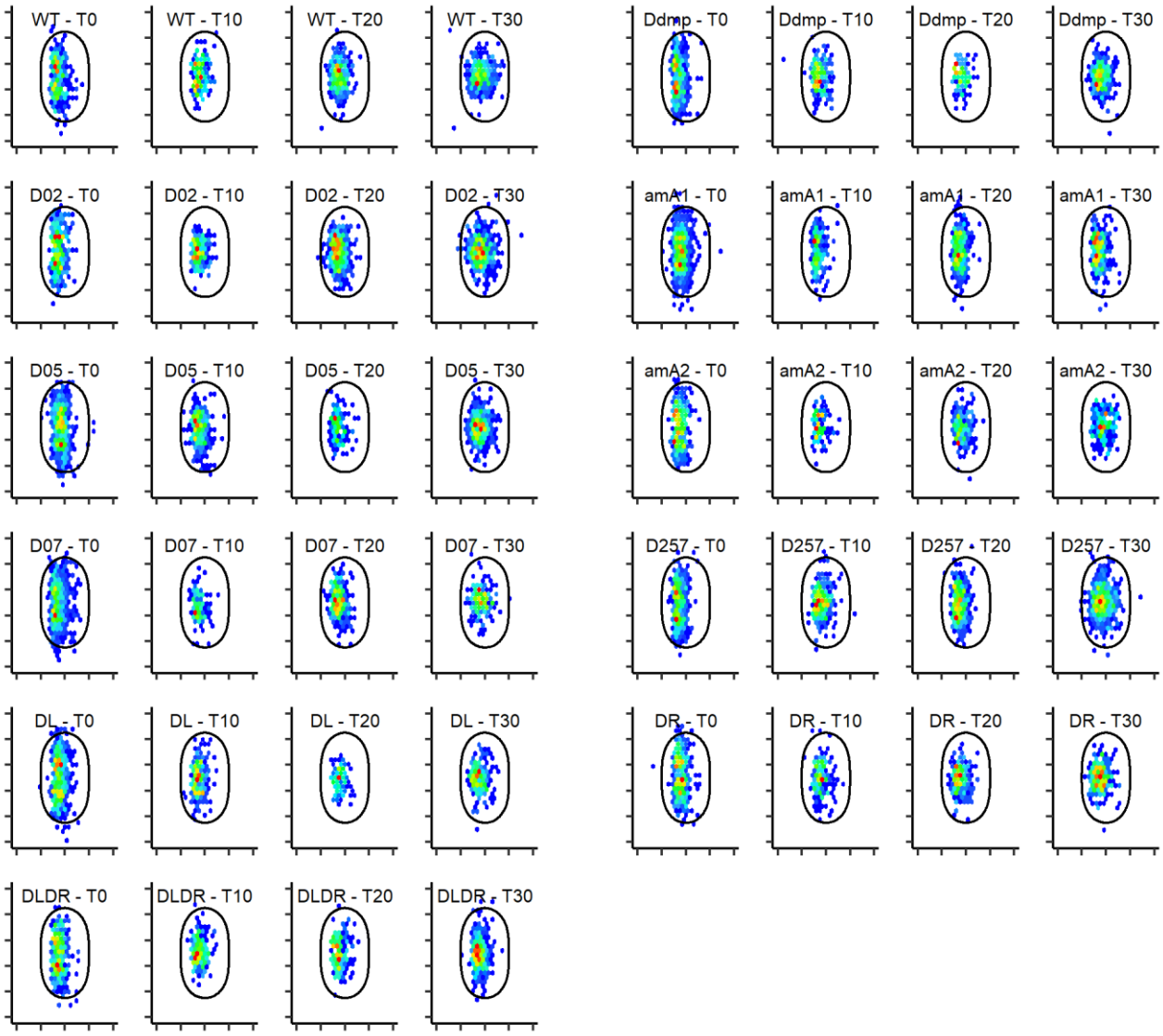


Figure 65. Distribution of foci in cells infected by mutants of T5.
Escherichia coli F cells were infected and followed for up to 30 minutes and evaluated every 10 minutes. The ellipses represent a bacterial cell over which it was plotted a hexagonal heatmap of 2D bin counts of foci detected with DAPI staining. The foci were set on an orthogonal coordinate system and were colored based on the crowdedness. Segmentation and foci detection were carried out with the MicrobeJ plugin for ImageJ. Every diagram is labeled indicating the mutant strain and the time of infection in minutes, for instance, WT – T0 corresponds to the T5 wild-type strain at time zero.

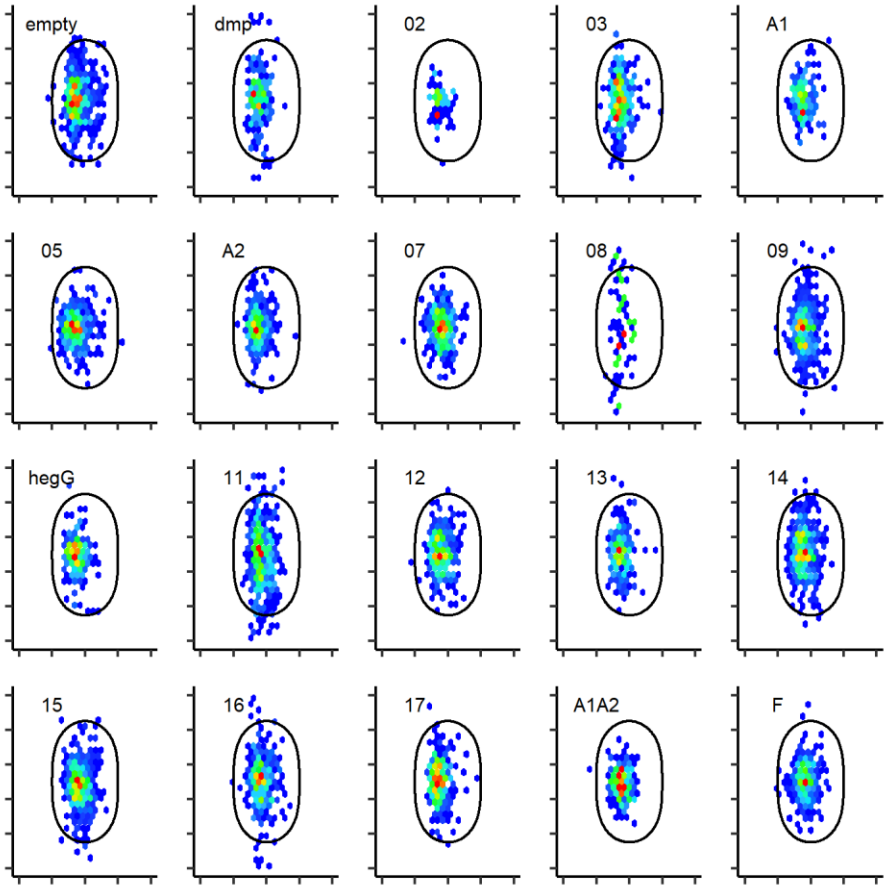


Figure 66. Distribution of DAPI dye foci in cells that express early T5 genes.
Escherichia coli F cells expressing T5 early genes were evaluated after 140 minutes of expression. The ellipses represent a bacterial cell over which it was plotted a hexagonal heatmap of 2D bin counts of foci detected with DAPI staining. Segmentation and foci detection were carried out with the MicrobeJ plugin for ImageJ. Every diagram is labeled indicating the mutant strain and the time of infection in minutes, for instance, WT – T0 corresponds to the T5 wild-type strain at time zero. The foci were set on an orthogonal coordinate system and were colored based on the crowdedness.

7.4. Sequences of interest

The heterologous genes inserted in the T5 genome are underlined and in upper case, their flanking sequences are in lower case.

>pUCPNmC
atggggattgattatagtggtctaaggaccatttttggtgaaaaactaccagaatctcatatcttctttgctacggttgctgcgcataaatatgttccta
gctatgcttttctgcgtagagaactagggctttcatctgcgcatactaaccgtaaagtatggaagaaatttgtagaggcttatggtaaagcaattcctcc
ggccccaccggctccacca
ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGA
TCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTC
CCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGC
GTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACT
TCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAA
GCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTC
AACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGC
TGTACAAGTCTTAA
tggaggctaataaatgacacaagctgctattgactataacaagttaaaat

The *mcherry* gene is highlighted in red and underlined.

>pUC05-LacZa2
tgttgcaaaggaggaaatcac
ATGACCATGATTACGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAGCTCAAACTCACTGGCCGTCGTTTTACAACGTCGTG
ACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTC
CCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGT
ACAATCTGCTCTGATGCCGCATAG
attcaaggcgttatatttggcaatattgccaataacaaacgtttttaatccctttaatggagttttattattatggttatttccgcagaaaaacaaacc
gttatcctgaaa

The *lacZa* gene is highlighted in blue and underlined.

>Gene11_optimized
aaaggtctcatcacATGCGTTTATATAAGCCCGACAACGCAACGGTCCTGAAAGGGGCGCTTCGCAATCTTCTGGACGGAAGTCGCACCACATCGATTAA
GCATTTCATCAACAAAGCTGAAACCATTCACCAGAATTTTTTCAATGACTATGATGCTTACGACATGGATTCATTCTTACCACTGTATAAACAAGGAGCA
ACTATCATCCTTTACAACCGCGACTGTCACATTTTCACGGGACGCGGGGCCGGGCCAGATACAGACTTTATTGTTGCCCAAGGTAACAATATTGTCGCAT
ATCACAGCAATTCTGGTCGCCAAAGTTTTAATATTAATAACTTCGAACTGGTAGGTTTAGAACTTATTAACGGTTTGAGTAGTTTGGAGGGCTTCTTAGA
TGTTATCTGGAAATACTTATCAGCCGGCTGGCACCGCCCAGACTCTGACGGGTCTCAATATACCGCGGTCATGGCATTTGACGGAGATGGAAACTTTTTG
TTGCCGTCCGAACTGTATCCCTATGTGTACAATGAACTTGTGCAAGAGGCTGGCTTTATCTTTTGGAAAGATGAGAACCAGATCATCGATAACACGGAGA
AGCACGCTGTAAACAGTAATACACAGAGCGTTAACACTCAAATTACCGAGAACAAGGGGAAAAATGCTATGACAAAGATCGCTAACATTGTAGCCGCCAA
TAAATCGGCTGTAGTTAACGCAGCGAAACTTGAGGCGGGTAAGATTGCACTTACACAAATCACTAAAGTAGCGGCGAAGAAGGCTCCATTCATGATTAAG
GGGTACATCGATACACCGATTGGCCGCGTTGTAATCGCAAACCTGTTATCTGTCGCGGTCGACCAGTACGCCCCCTCCAACCAAAAGGCCAAAGCAGTCG
CCGGTGCGGCAATGGAAGCTGCCATGTTGGAGATGGTGCAGTCTTTCAACATTGCGGAAATGATTGACGAGATGGTTAAGGGAATCGATATCAGCACGTT
TACGGTAAACACAGAATCAGAATAAtttgtgagacctttt

The BsaI restriction sites are underlined.

>pJJ1418
CAGGGGAAGGACTGGATACAACTGACGGAGTGATACGGAATGACAGAGCTGGAAACACATTTGCTGAACGCCTTAGAGCAGCTGCAACAGGACTATATGC
AGCGGCTGAACGAATGGGAGAGCGCCTTCGTGGAATTGCAGAAGATGTTTTCGCTTACGCAACGGGACAACGCGATGCTGAACGAGCGGGTCATGCAGTT
GAGTCAGCAGGTGCAACACTTGAGCGAGCGGACAGAACGCTTGAGCCAGTTATACAGCGAGAACAGGAGATAAGGGATGAACGGCTGATGCAGGAGCGCG
AACATGCATTATCcctggagcgggaacgccagctggaaatacaggaacgcacactggatggtccttcgttgtcgctgtgatggcgaaactatgaaaaatg
gcaggttcggtggattttgacgggctaatgtggtctgcaccatctggttgcataggtattcatacggttaaaatttatcaggcgcgatcgcggcagtttt
tcgggtggtttgttgccatttttacctgtctgctgccgtgatcgcgatgaacgcgttttagcggtgcgtacaattaagggattatggtaaatcaaacgta
tgttaatctatcgacatatgtaactttataaaataacagtggaaacatggattcATGTCTAATAAAAAACAGTCAAATAGGCTAACTGAACAACATAAGT
TATCTCAAGGTGTAATTGGGATTTTTGGGGATTATGCAAAAGCTCATGATCTCGCTGTTGGTGAGGTTTCAAAATTAGTAAAGAAAGCTCTTAGCAACGA
ATACCCTCAATTATCATTTCGATATAGAGATAGTATAAAGAAAACAGAAATAAATGAAGCTTTAAAAAAAAATTGACCCTGATCTTGGCGGTACTTTATTT
GTTTCAAATTCCAGCATCAAACCTGATGGTGGAATTGTAGAGGTCAAAGATGATTATGGTGAATGGAGAGTTGTACTTGTTGCTGAAGCCAAACACCAAG
GTAAAGATATTATAAATATAAGGAATGGTTTGTTAGTTGGGAAAAGAGGAGATCAAGATTTAATGGCTGCTGGTAATGCTATCGAAAGATCTCATAAGAA
TATATCAGAGATAGCGAATTTTATGCTCTCTGAGAGCCACTTTCCTTACGTCCTTTTCTTAGAGGGGTCTAACTTTTTAACAGAAAATATCTCAATAACA
AGACCAGATGGAAGGGTTGTTAATCTTGAGTATAATTCTGGTATATTAAATAGGTTAGATCGACTAACTGCAGCTAATTATGGAATGCCTATAAATAGTA
ATCTATGTATTAACAAATTTGTAAATCATAAAGACAAAAGCATTATGCTACAAGCAGCATCTATATATACTCAAGGAGATGGGAGGGAGTGGGATTCGAA
AATCATGTTTGAAATAATGTTTGATATATCAACGACTTCGCTCAGAGTGTTGGGGCGTGACTTGTTTGAACAGCTTACATCTAAGTGAtatttttattt
taataaggttttaattaATGGCTAGAAATGCAACAAACAAGTTACTGCACAAAGCTAAAAAATCGAAAAGCGACGAATTTTACACTCAGTATTGTGATAT
TGAGAACGAACTGCAATACTACAGAGAGCACTTCTCTGATAAGGTTGTTTATTGCAATTGTGATGATCCTAGAGTAAGCAATTTCTTTAAATATTTTGCA
GTGAATTTTGATAATCTTGGCTTGAAAAAGTTAATAGCATCTTGCTATGTAGAGAATAAAGAAGGTTTTTCTAGTAGCGAAGCCGCGAAGAACGGATTTT
ACTATGAATATCATAAAGAAAATGGAAAGAAATTAGTTTTTGATGATATTAGTGTTTCTTCTTTCTGTGGCGATGGCGATTTTCGCAGTTCGGAGAGCAT
TGATCTGCTAAAAAAATCAGATATTGTTGTTACGAATCCTCCATTCTCGTTATTTAGAGAGTATCTTGATCAACTAATTAAGTATGATAAGAAATTCCTT
ATAATTGCTAATGTTAATTCAATAACATATAAAGAGGTGTTTAATCTAATTAAGGAAAATAAGATTTGGCTTGGGGTTCATCTCGGGAGAGGTGTTTCTG
GATTTATTGTTCCAGAGCATTATGAATTATATGGTACTGAGGCGAGAATTGATTCTAATGGTAATAGAATTATCTCGCCAAACAACTGCTTATGGCTAAC
TAACCTAGATGTCTTTATTAGGCATAAAGACTTGCCTCTTACAAGAAAATATTTTGGGAATGAAAGTTCATATCCAAAATATGATAATTATGATGCTATA
AATGTAAACAAAACAAAGGATATTCCATTAGATTACAATGGGGTTATGGGGGTTCCTATCACATTCTTGCATAAGTTTAACCCTGAGCAATTTGAGTTAA
TAAAATTTAGAAAGGGTGTTGATGAAAAAGATTTGTCTATAAATGGTAAATGCCCTTATTTCAGAATTTTGATAAAAAACAAACGATTACAAAAGTAA
ttgatgtttgttagtttttcttgagatcattagcttcgtcgtaagctt

Genes EcoRI nuclease (*ecoRIR*) and EcoRI methylase (*ecoRIM*) are highlighted in red and blue, respectively and underlined.

```
>pUC02parST1
atagaagactgagcgcgatcgaatgggatacggttagatttagccatgtttataactccaattagtttaataagggattt
TTTCACCACGCCAATTTCATGGTTATTTGTTTGATTATAAAGGAAAATTGAAAATTCTTTCACACTGAAATCACCACGCTTTTCAA
aatttatttgctagcttcaagacttacaacaaaagcgagatttgtttctttccatacgtcaggaatatcattttcctgta
```

The *parS* (T1) sequence is highlighted in green and underlined.

```
>pUC17parSP1
tagcaaaagatacacctacagatactgaggcgtatcaaaaagcataccaagaattattataatactctgcatatctcggt
GCACAATACCTGTACCGAGACGGTGAAATCGTGGCGATTTCACCTTGAAATTTTAGAGTAATTTACTTTAAAAACAGTCAGTTAATAGTGAAATTTGAAT
GGCGAAACAAATACAAACTTTCGTTGACAGATAGC
taaatgtatgttatcataactctatgagttaatagaattaaccatttagagttgccacacagaaggtgtggtggttcagt
```

The *parS* (P1) sequence is highlighted in green and underlined.

```
>pFH2973
cgcaacgcaattaatgtaagttagcgcgaattgatctggtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggcagccatcgg
aagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgctcaaggcgcactcccgttctggataatgtttttgcgccgacatcataa
cggttctggcaaatattctgaaatgagctgtttacaattaatcatccggctcgcattatgtgtggaattgtgagcggataacaatttcacacaggaaaca
gacc
ATGAGTAAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGTGATGTTAATGGGCACAAATTTTCTGTCAGTGGAGAGGGTG
AAGGTGATGCAACATACGGAAAACTTACCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCGACCCTGGTGACCACCCTGACCTGGGG
CGTGCAGTGCTTCGCGCGCTACCCGGATCACATGAAACGTCACGACTTTTTCAAATCTGCCATGCCGGAAGGTTATGTACAGGAACGTACCATCTTCTTC
AAAGATGACGGTAACTACAAAACCCGTGCTGAAGTGAAATTTGAAGGTGATACCCTGGTGAACCGTATCGAACTGAAAGGTATTGATTTTAAAGAAGATG
GTAACATCCTGGGCCATAAACTGGAATACAACTACAACTCTCATAACGTATACATCGTGGCAGACAAACAGAAAAACGGTATCAAAGTGAACTTCAAAAT
CCGTCATAACATCGAAGATGGTTCTGTTCAACTGGCAGACCATTATCAGCAGAACACCCCAATTGGCGATGGCCCGGTGCTGCTGCCGGACAACCATTAC
CTGTCCACGCAGTCTGCCCTTTCGAAAGATCCGAACGAAAAACGTGACCACATGGTGCTGCTGGAATTTGTGACCGCTGCCGGCATCACGCATGGCATGC
ATGAGCTGTATAAAGAATTCGAGCTC
ATGGTCGAGCAGGTATTCAAGTTATCAACTGGTCGCCAGGCCACATTTATTGAAGAGGTAATACCTCCGAACCAGGTAGAAAGCGATACCTTTGTTGATC
AGCATAACAACGGGCGTGACCAGGCATCTCTTACGCCAAAATCATTAAAAAGTATCCGAAGCACTATTAAGCATCAGCAATTTTACCCTGCAATAGGTGT
TAGACGGGCTACAGGGAAAATTGAAATTTTGGATGGTTCCCGGCGTCGAGCTTCTGCCATCTTAGAGAACGTAGGGTTGCGGGTTTTAGTCACGGACCAG
GAGATCAGCGTTCAGGAAGCGCAAAATTTAGCGAAAGACGTTCAGACAGCATTGCAGCACAGCATTCGAGAAATAGGTCTGCGTTTGATGCGAATGAAAA
ATGATGGGATGAGTCAGAAGGATATTGCAGCCAAAGAAGGGCTGTCTCAGGCGAAGGTCACGCGTGCTCTCCAGGCAGCGAGTGCTCCGGAAGAATTAGT
CGCCCTTTTCCCTGTGCAGTCGGAATTAACCTTTTCGGACTACAAAACGCTTTGTGCTGTTGGCGACGAAATGGGGAACAAGAATTTAGAGTTTGATCAG
CTTATTCAAAACATATCCCCGGAAATAAACGACATCTTATCCATTGAAGAAATGGCCGAAGATGAAGTTAAAAATAAAATCCTGCGCTTGATAACAAAGG
AAGCCTCACTACTCACGGATAAAGGTTCTAAAGATAAGTCCGTAGTTACTGAATTATGGAAATTTGAGGACAAGGATCGCTTTGCAAGGAAGCGCGTGAA
AGGCCGTGCATTTTCTTATGAGTTTAATCGACTCTCAAAAGAGTTACAGGAAGAACTCGACAGGATGATTGGGCATATCCTTAGAAAGAGCCTCGATAAA
AAGCCGAAGCCTTAA
actttcgccattcaaatttcactattaacccggggatcctctagattaaggaggccat
ATGTCTAAAGGTGAAGAACTGTTTACGGGTGTTGTGCCGATCCTGGTGGAACTGGATGGTGATGTGAACGGTCATAAATTCTCTGTGTCTGGTGAAGGTG
AAGGTGATGCAACCTACGGTAAGCTGACCCTGAAATTTATCTGTACCACGGGTAAACTGCCGGTTCCATGGCCGACCCTGGTGACCACCTTCGCGTATGG
TCTGCAATGCTTTGCGCGCTACCCGGATCACATGAAACGTCACGACTTTTTCAAATCTGCCATGCCGGAAGGTTATGTACAGGAACGTACCATCTTCTTC
AAAGATGACGGTAACTACAAAACCCGTGCTGAAGTGAAATTTGAAGGTGATACCCTGGTGAACCGTATCGAACTGAAAGGTATTGATTTTAAAGAAGATG
GTAACATCCTGGGCCATAAACTGGAATACAACTACAACTCTCATAACGTATACATCACGGCAGACAAACAGAAAAACGGTATCAAAGCGAACTTCAAAAT
CCGTCATAACATCGAAGATGGTTCTGTTCAACTGGCAGACCATTATCAGCAGAACACCCCAATTGGCGATGGCCCGGTGCTGCTGCCGGACAACCATTAC
CTGTTCATCCAGTCTGCCCTTTCGAAAGATCCGAACGAAAAACGTGACCACATGGTGCTGCTGGAATTTGTGACCGCTGCCGGCATCACGCATGGCATGC
ATGAGCTGTATAAA
TCGTACGGCGGCGACGGTGATGTTAAAACATTTACCCTTAAATCAGGCGTGCAGGCTAGGTTTGTCAAAACAGTTGTGTTAAGCGGAGAAGTTGAATCAA
AAACGTTCGTTGATGCTTCGGTTAATGGACGTGATCAGACAATGCTTACGCGCGAGTCCGTCAGTGATATTTCCCGGACGATAAAGCTGCAGCAGTTCTT
TCCGGCTATTGGTCGGGAGGTTAATGGACTAATTGAGATATTAGATGGAACCCGACGTCGTGCTGCCTGCATCTTTAATAACGTTAAATTCGAAATTCTG
GTAACAAAAGATGATATCTCACTCGCCGATGCACGGCAGTTGGCGAAAGATATCCAGACTGCCAGAGAACATAGTCTTCGCGAGCTGGGGAAGCGACTCG
AAGTTACCTACGGAACCAGCATGACGAAAGAAGATATTGCGTTGAAAGAAAATCTCTCTCCGGCGAAAGTGACACGTGCGTTTCAGGCCGCAGCAGTGCC
AGACGAAATGGTTGCAGTGTTCCCGGTGATAAATGATATTTCGTTGTCAGATTATCAGTTTTTACTGAAACTGGCCGAAGAAGCAAACAACAAGCAAACA
TCGGTAACAGAGCTGATGGAAAAAGTTCAGCATCGGTTGAAGACCATGCCAGATTATCCGGCAATTGATAAAAGCAAAATTCTTGCGGTTATCCGGTCGG
AGAGCAAATTGCTGACAGCCCTCCCAACTAGAACGGTTCAAACAGAGAAGCTGAGAGAATTTTCAGATCGTAATCAGTTTGCCAGAAAGAAAACTGATCC
AAAGAAGCGACTTGTTGTTTATGAGTTTTCCCGTATTTCCGCTGAGGCACAGTCGGAGATTGATAAGGCAATAAAACGTATTCTGGAAAGACTTCCAGAA
TCAGGTGAGTAA
gggataaggatcaagcttggctgttttggcggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaacagaatt
tgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgag
agtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatct
```

The fusions *cfp-parB* (P1), *gfp-parB* (T1), are highlighted in blue and black, green, and red, respectively.

```
LOCUS       Exported                8670 bp ds-DNA     circular GCC 07-MAY-2020
DEFINITION  pAC9.
ACCESSION   .
VERSION     .
KEYWORDS    .
SOURCE      synthetic DNA construct
  ORGANISM  synthetic DNA construct
REFERENCE   1  (bases 1 to 8670)
  AUTHORS   Boulanger Lab
  TITLE     Direct Submission
  JOURNAL   Exported May 7, 2020 from SnapGene Viewer 4.1.0
            http://www.snapgene.com
COMMENT     This file was created using Genome Compiler.
FEATURES             Location/Qualifiers
     source          1..8670
                     /organism=" synthetic DNA construct"
                     /mol_type="genomic DNA"
     source          7301..7300
                     /organism="pAC9"
                     /mol_type="genomic DNA"
     terminator      50..136
                     /label="rrnB"
                     /note="rrnB T1 terminator; transcription terminator T1 from
                     the E. coli rrnBgene"
     misc_feature    151..374
                     /label="Min_lambda_cos"
                     /note="Minimal lambda cos sequence"
     terminator      complement(450..775)
                     /label="rrnB"
                     /locus_tag=""
                     /product=""
     CDS             complement(903..5012)
                     /codon_start=1
                     /label="cas9"
                     /translation="MMDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIK
                     KNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE
                     SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI
                     KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR
                     LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
                     QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALV
                     RQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL
                     LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR
                     GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
                     YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECF
                     DSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE
                     RLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
                     MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
                     RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL
                     YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNV
                     PSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK
                     HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAY
                     LNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFK
                     TEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS
                     KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLG
                     ITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGN
                     ELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA
                     DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEV
                     LDATLIHQSITGLYETRIDLSQLGGD"
     promoter        complement(5055..5083)
                     /label="PBAD_promoter"
                     /locus_tag=""
                     /product=""
     misc_feature    complement(5312..5327)
                     /label="araO2"
     misc_feature    complement(5159..5170)
                     /label="araO1"
     promoter        complement(5178..5206)
                     /label="Pc_promoter"
     misc_feature    complement(5115..5128)
                     /label="CAP"
     misc_feature    complement(5080..5118)
                     /label="AraI1I2"
     CDS             5357..6235
                     /label="AraC"
                     /locus_tag=""
                     /translation="MQYGQLVSSLNGGSMKSMAEAQNDPLLPGYSFNAHLVAGLTPIE
                     ANGYLDFFIDRPLGMKGYILNLTIRGQGVVKNQGREFVCRPGDILLFPPGEIHHYGRH
                     PEAREWYHQWVYFRPRAYWHEWLNWPSIFANTGFFRPDEAHQPHFSDLFGQIINAGQG
                     EGRYSELLAINLLEQLLLRRMEAINESLHPPMDNRVREACQYISDHLADSNFDIASVA
```

```
                     QHVCLSPSRLSHLFRQQLGISVLSWREDQRISQAKLLLSTTRMPIATVGRNVGFDDQL
                     YFSRVFKKCTGASPSEFRAGCEEKVNDVAVKLS*"
     rep_origin      6339..7251
                     /label="p15A_ori"
     terminator      7339..7433
                     /label="t0"
                     /note="lambda t0 terminator; transcription terminator fromp
                     hage lambda"
     CDS             complement(7464..8258)
                     /codon_start=1
                     /label="KanR"
                     /translation="MIEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRP
                     VLFVKTDLSGALNELQDEAARLSWLATTGVPCAAVLDVVTEAGRDWLLLGEVPGQDLLS
                     SHLAPAEKVSIMADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQ
                     GLAPAELFARLKARMPDGEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIA
                     LATRDIAEELGGEWADRFLVLYGIAAPDSQRIAFYRLLDEFF"
     CDS             8486..8501
                     /codon_start=1
                     /product="Cas9 (Csn1) endonuclease from the Streptococcus"
                     /label="Cas9"
                     /note="Cas9; generates RNA-guided double strand breaks in D
                     NA"
                     /translation="ARR*L"
     promoter        8529..8557
                     /label="P_BBaJ23119"
                     /note="from BioBricks"
     misc_RNA        8564..8583
                     /label="spacer"
                     /note="guide RNA"
     misc_RNA        8584..8666
                     /label="gRNA_scaffold"
                     /note="gRNA scaffold; guide RNA scaffold for the CRISPR/Cas
                     9 system"
     CDS             8667..1
                     /codon_start=1
                     /locus_tag="b0345"
                     /product="lactose-inducible lac operon transcriptionalrepre
                     ssor"
                     /label="lacI"
                     /note="lacI; transcriptional repressor of the lac operon"
                     /db_xref="GI:49176012; ASAP:ABE-0001189; UniProtKB/Swiss-Pr
                     ot:P03023; EcoGene:EG10525; GeneID:945007"
                     /protein_id="NP_414879.3"
                     /translation="G"
ORIGIN
        1 gtgcagcgcg atcgtaatca ggatcccatg gtacgcgtgc tagaggcatc aaataaaacg
       61 aaaggctcag tcgaaagact gggcctttcg ttttatctgt tgtttgtcgg tgaacgctct
      121 cctgagtagg acaaatccgc cgccctagac cctccacgca cgttgtgata tgtagatgat
      181 aatcattatc actttacggg tcctttccgg tgatccgaca ggttacgggg cggcgacctc
      241 gcgggttttc gctatttatg aaaattttcc ggtttaaggc gtttccgttc ttcttcgtca
      301 taacttaatg tttttattta aaatacccctc tgaaaagaaa ggaaacgaca ggtgctgaaa
      361 gcgaggcttt ttggggcgtt cggctgcggc gagcggtatc agctcactca aaggcggtaa
      421 gggttattgt ctcatgagcg gatacatatt tgaatgtatt tagaaaaata aacaaaagag
      481 tttgtagaaa cgcaaaaagg ccatccgtca ggatggcctt ctgcttaatt tgatgcctgg
      541 cagtttatgg cgggcgtcct gcccgccacc ctccgggccg ttgcttcgca acgttcaaat
      601 ccgctcccgg cggatttgtc ctactcagga gagcgttcac cgacaaacaa cagataaaac
      661 gaaaggccca gtctttcgac tgagcctttc gttttatttg atgcctggca gttccctact
      721 ctcgcatggg gagaccccac actaccatcg gcgctacggc gtttcacttc tgagttcggc
      781 atggggtcag gtgggaccac cgcgctactg ccgccaggca aattctgttt tatcagaccg
      841 cttctgcgtt ctgatttaat ctgtatcagg ctgaaaatct tctctcatcc gccaaaacag
      901 cctcagtcac ctcctagctg actcaaatca atgcgtgttt cataaagacc agtgatggat
      961 tgatggataa gagtggcatc taaaacttct tttgtagacg tatatcgttt acgatcaatt
     1021 gttgtatcaa aatatttaaa agcagcggga gctccaagat tcgtcaacgt aaataaatga
     1081 ataatatttt ctgcttgttc acgtattggt ttgtctctat gtttgttata tgcactaaga
     1141 actttatcta aattggcatc tgctaaaata acacgcttag aaaattcact gatttgctca
     1201 ataatctcat ctaaataatg cttatgctgc tccacaaaca attgtttttg ttcgttatct
     1261 tctggactac ccttcaactt ttcataatga ctagctaaat ataaaaaatt cacatatttg
     1321 cttggcagag ccagctcatt tccttttttgt aattctccgg cactagccag catccgttta
     1381 cgaccgtttt ctaactcaaa aagactatat ttaggtagtt taatgattaa gtctttttta
     1441 acttccttat atcctttagc ttctaaaaag tcaatcggat ttttttcaaa ggaacttctt
     1501 tccataattg tgatccctag taactcttta acggatttta acttcttcga tttccctttt
     1561 tccaccttag caaccactag gactgaataa gctaccgttg gactatcaaa accaccatat
     1621 ttttttggat cccagtcttt tttacgagca ataagcttgt ccgaatttct ttttggtaaa
     1681 attgactcct tggagaatcc gcctgtctgt acttctgttt tcttgacaat attgacttgg
     1741 ggcatggaca atactttgcg cactgtggca aaatctcgcc ctttatccca gacaatttct
     1801 ccagtttccc cattagtttc gattagaggg cgtttgcgaa tctctccatt tgcaagtgta
     1861 atttctgttt tgaagaagtt catgatatta gagtaaaaga aatattttgc ggttgctttg
     1921 cctatttctt gctcagactt agcaatcatt ttacgaacat cataaacttt ataatcacca
     1981 tagacaaact ccgattcaag ttttggatat ttcttaatca aagcagttcc aacgacggca
     2041 tttagatacg catcatgggc atgatggtaa ttgttaatct cacgtacttt atagaattgg
```

2101 aaatcttttc ggaagtcaga aactaattta gattttaagg taatcacttt aacctctcga
2161 ataagtttat cattttcatc gtatttagta ttcatgcgac tatccaaaat ttgtgccaca
2221 tgcttagtga tttggcgagt ttcaaccaat tggcgtttga taaaaccagc tttatcaagt
2281 tcactcaaac ctccacgttc agctttcgtt aaattatcaa acttacgttg agtgattaac
2341 ttggcgttta gaagttgtct ccaatagttt ttcatctttt tgactacttc ttcacttgga
2401 acgttatccg atttaccacg atttttatca gaacgcgtta agaccttatt gtctattgaa
2461 tcgtctttaa ggaaactttg tggaacaatg tgatcgacat cataatcact taaacgatta
2521 atatctaatt cttggtccac atacatgtct cttccatttt ggagataata gagatagagc
2581 ttttcatttt gcaattgagt attttcaaca ggatgctctt taagaatctg acttcctaat
2641 tctttgatac cttcttcgat tcgtttcata cgctctcgcg aatttttctg gccctttga
2701 gttgtctgat tttcacgtgc catttcaata acgatatttt ctggcttatg ccgccccatt
2761 actttgacca attcatcaac aacttttaca gtctgtaaaa taccttttt aatagcaggg
2821 ctaccagcta aatttgcaat atgttcatgt aaactatcgc cttgtccaga cacttgtgct
2881 ttttgaatgt cttctttaaa tgtcaaacta tcatcatgga tcagctgcat aaaattgcga
2941 ttggcaaaac catctgattt caaaaaatct aatattgttt tgccagattg cttatcccta
3001 ataccattaa tcaattttcg agacaaacgt ccccaaccag tataacggcg acgtttaagc
3061 tgtttcatca ccttatcatc aaagaggtga gcatatgttt taagtctttc ctcaatcatc
3121 tccctatctt caaataaggt caatgttaaa acaatatcct ctaagatatc ttcattttct
3181 tcattatcca aaaaatcttt atctttaata atttttagca aatcatggta ggtacctaat
3241 gaagcattaa atctatcttc aactcctgaa atttcaacac tatcaaaaca ttctattttt
3301 ttgaaataat cttcttttaa ttgcttaacg gttacttttc gatttgtttt gaagagtaaa
3361 tcaacaatgg ctttcttctg ttcacctgaa agaaatgctg gttttcgcat tccttcagta
3421 acatatttga cctttgtcaa ttcgttataa accgtaaaat actcataaag caaactatgt
3481 tttggtagta ctttttcatt tggaagattt ttatcaaagt ttgtcatgcg ttcaataaat
3541 gattgagctg aagcaccttt atcgacaact tcttcaaaat tccatggggt aattgtttct
3601 tcagacttcc gagtcatcca tgcaaaacga ctattgccac gcgccaatgg accaacataa
3661 taaggaattc gaaaagtcaa gatttttca atcttctcac gattgtcttt taaaaatgga
3721 taaaagtctt cttgtcttct caaaatagca tgcagctcac ccaagtgaat ttgatgggga
3781 atagagccgt tgtcaaaggt ccgttgcttg cgcagcaaat cttcacgatt tagtttcacc
3841 aataattcct cagtaccatc cattttttct aaaattggtt tgataaattt ataaaattct
3901 tcttggctag ctcccccatc aatataacct gcatatccgt tttttgattg atcaaaaaag
3961 atttctttat acttttctgg aagttgttgt cgaactaaag cttttaaaag agtcaagtct
4021 tgatgatgtt catcgtagcg tttaatcatt gaagctgata ggggagcctt agttatttca
4081 gtatttactc ttaggatatc tgaaagtaaa atagcatctg ataaattctt agctgccaaa
4141 aacaaatcag catattgatc tccaatttgc gccaataaat tatctaaatc atcatcgtaa
4201 gtatcttttg aaagctgtaa tttagcatct tctgccaaat caaaatttga tttaaaatta
4261 ggggtcaaac ccaatgacaa agcaatgaga ttcccaaata agccattttt cttctcaccg
4321 gggagctgag caatgagatt ttctaatcgt cttgatttac tcaatcgtgc agaaagaatc
4381 gctttagcat ctactccact tgcgttaata gggttttctt caaataattg attgtaggtt
4441 tgtaccaact ggataaatag tttgtccaca tcactattat caggatttaa atctccctca
4501 atcaaaaaat gaccacgaaa cttaatcata tgcgctaagg ccaaatagat taagcgcaaa
4561 tccgctttat cagtagaatc taccaatttt tttcgcagat gatagatagt tggatatttc
4621 tcatgataag caacttcatc tactatattt ccaaaaatag gatgacgttc atgcttcttg
4681 tcttcttcca ccaaaaaaga ctcttcaagt cgatgaaaga aactatcatc tactttcgcc
4741 atctcatttg aaaaaatctc ctgtagataa caaatacgat tcttccgacg tgtatacctt
4801 ctacgagctg tccgtttgag acgagtcgct tccgctgtct ctccactgtc aaataaaaga
4861 gcccctataa gatttttttt gatactgtgg cggtctgtat ttcccagaac cttgaacttt
4921 ttagacggaa ccttatattc atcagtgatc accgcccatc cgacgctatt tgtgccgata
4981 tctaagccta ttgagtattt cttatccatc atggtgaatt cctcctgcta gcccaaaaaa
5041 acgggtatgg agaaacagta gagagttgcg ataaaaagcg tcaggtagga tccgctaatc
5101 ttatggataa aaatgctatg gcatagcaaa gtgtgacgcc gtgcaaataa tcaatgtgga
5161 cttttctgcc gtgattatag acacttttgt tacgcgtttt tgtcatggct ttggtcccgc
5221 tttgttacag aatgctttta ataagcgggg ttaccggttt ggttagcgag aagagccagt
5281 aaaagacgca gtgacggcaa tgtctgatgc aatatggaca attggtttct tctctgaatg
5341 gcgggagtat gaaaagtatg gctgaagcgc aaaatgatcc cctgctgccg ggatactcgt
5401 ttaatgccca tctggtggcg ggtttaacgc cgattgaggc caacggttat ctcgattttt
5461 ttatcgaccg accgctggga atgaaaggtt atattctcaa tctcaccatt cgcggtcagg
5521 gggtggtgaa aaatcaggga cgagaatttg tttgccgacc gggtgatatt ttgctgttcc
5581 cgccaggaga gattcatcac tacggtcgtc atccggaggc tcgcgaatgg tatcaccagt
5641 gggtttactt tcgtccgcgc gcctactggc atgaatggct taactggccg tcaatatttg
5701 ccaatacggg gttctttcgc ccggatgaag cgcaccagcc gcatttcagc gacctgtttg
5761 ggcaaatcat taacgccggg caaggggaag ggcgctattc ggagctgctg gcgataaatc
5821 tgcttgagca attgttactg cggcgcatgg aagcgattaa cgagtcgctc catccaccga
5881 tggataatcg ggtacgcgag gcttgtcagt acatcagcga tcacctggca gacagcaatt
5941 ttgatatcgc cagcgtcgca cagcatgttt gcttgtcgcc gtcgcgtctg tcacatcttt
6001 tccgccagca gttagggatt agcgtcttaa gctggcgcga ggaccaacgt atcagccagg
6061 cgaagctgct tttgagcacc acccggatgc ctatcgccac cgtcggtcgc aatgttggtt
6121 ttgacgatca actctatttc tcgcgggtat ttaaaaaatg caccggggcc agcccgagcg
6181 agttccgtgc cggttgtgaa gaaaaagtga atgatgtagc cgtcaagttg tcataattgg
6241 taacgaatca gacaattgac ggcttgacgg agtagcatag ggtttgcaga atccctgctt
6301 cgtgccagga accgtaaaaa ggccgcgttg ctggcgttgc gctagcggag tgtatactgg
6361 cttactatgt tggcactgat gagggtgtca gtgaagtgct tcatgtggca ggagaaaaaa
6421 ggctgcaccg gtgcgtcagc agaatatgtg atacaggata tattccgctt cctcgctcac
6481 tgactcgcta cgctcggtcg ttcgactgcg gcgagcggaa atggcttacg aacggggcgg
6541 agatttcctg gaagatgcca ggaagatact taacagggaa gtgagagggc cgcggcaaag
6601 ccgtttttcc ataggctccg ccccccctgac aagcatcacg aaatctgacg ctcaaatcag
6661 tggtggcgaa acccgacagg actataaaga taccaggcgt ttcccccctgg cggctccctc
6721 gtgcgctctc ctgttcctgc ctttcggttt accggtgtca ttccgctgtt atggccgcgt
6781 ttgtctcatt ccacgcctga cactcagttc cgggtaggca gttcgctcca agctggactg
6841 tatgcacgaa ccccccgttc agtccgaccg ctgcgcctta tccggtaact atcgtcttga

```
6901 gtccaacccg gaaagacatg caaaagcacc actggcagca gccactggta attgatttag
6961 aggagttagt cttgaagtca tgcgccggtt aaggctaaac tgaaaggaca agttttggtg
7021 actgcgctcc tccaagccag ttacctcggt tcaaagagtt ggtagctcag agaaccttcg
7081 aaaaaccgcc ctgcaaggcg gttttttcgt tttcagagca agagattacg cgcagaccaa
7141 aacgatctca agaagatcat cttattaatc agataaaata tttctagatt tcagtgcaat
7201 ttatctcttc aaatgtagca cctgaagtca gccccatacg atataagttg tgaagatcct
7261 ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg
7321 gtcatgacta gtgcttggat tctcaccaat aaaaaacgcc cggcggcaac cgagcgttct
7381 gaacaaatcc agatggagtt ctgaggtcat tactggatct atcaacagga gtccaagcga
7441 gctctcgaac cccagagtcc cgctcagaag aactcgtcaa gaaggcgata gaaggcgatg
7501 cgctgcgaat cgggagcggc gataccgtaa agcacgagga agcggtcagc ccattcgccg
7561 ccaagctctt cagcaatatc acgggtagcc aacgctatgt cctgatagcg gtccgccaca
7621 cccagccggc cacagtcgat gaatccagaa aagcggccat tttccaccat gatattcggc
7681 aagcaggcat cgccatgggt cacgacgaga tcctcgccgt cgggcatgcg cgccttgagc
7741 ctggcgaaca gttcggctgg cgcgagcccc tgatgctctt cgtccagatc atcctgatcg
7801 acaagaccgg cttccatccg agtacgtgct cgctcgatgc gatgtttcgc ttggtggtcg
7861 aatgggcagg tagccggatc aagcgtatgc agccgccgca ttgcatcagc catgatggat
7921 actttctcgg caggagcaag gtgagatgac aggagatcct gccccggcac ttcgcccaat
7981 agcagccagt cccttcccgc ttcagtgaca acgtcgagca cagctgcgca aggaacgccc
8041 gtcgtggcca gccacgatag ccgcgctgcc tcgtcctgca gttcattcag ggcaccggac
8101 aggtcggtct tgacaaaaag aaccgggcgc ccctgcgctg acagccggaa cacggcggca
8161 tcagagcagc cgattgtctg ttgtgcccag tcatagccga atagcctctc cacccaagcg
8221 gccggagaac ctgcgtgcaa tccatcttgt tcaatcatgc gaaacgatcc tcatcctgtc
8281 tcttgatcag atcttgatcc cctgcgccat cagatccttg gcggcaagaa agccatccag
8341 tttactttgc agggcttccc aaccttacca gagggcgccc cagctggcaa ttccgacgtc
8401 taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt
8461 cgtcttcggg cccattaagt tctgtgctag gaggtgactg aagtatattt taggaattct
8521 aaagatcttt gacagctagc tcagtcctag gtataatact agttgagacc agtctaggtc
8581 tcggttttag agctagaaat agcaagttaa aataaggcta gtccgttatc aacttgaaaa
8641 agtggcaccg agtcggtgct ttttttggta//
```

7.5. Whole genome sequencing

7.5.1. Differences between the T5wt and the NCBI deposited sequence.

Table 11. Off-target mutations in the T5 mutants

SNP	Phages[a]	Protein	Annotation	Function
T21158C	DL, DLDR	E198G	T5.045	
[27492]::[TTT]	WT, DL, DR, DLDR	G27492SE	T5.055	
C30586T	WT, DL, DR, DLDR		tRNA	
T36543C	WT, DL, DR, DLDR	I173V	hegA	
G39965C	WT, DL, DR, DLDR	N257K	T5.080	
C40360T	DR	D126N	T5.080	
G43860T	DLDR	T94N	T5.085	
G52707-	WT, DL, DR, DLDR	L43NPN...	T5.097	
(Poor quality 55120-55142)	WT, DL, DR, DLDR		nrdD	
(Poor quality 68763-68778)	WT, DL, DR, DLDR		D6	
C79465A	DLDR	T565K	D13	
GGC85893CAG	WT, DL, DR, DLDR	PG53AA	T5.134	
CG85929GC	WT, DL, DR, DLDR	VD41LE	T5.134	
AG86341-	WT, DL, DR, DLDR	L45SRG...	T5.134	
T98285C	WT, DL, DR, DLDR	D22G	T5.144	
T103331A	WT, DL, DR, DLDR	I100N	T5.150	

[a] WT, DL, DR, and DLDR stand for T5 wt, T5 DL, T5 DR, and T5 DLDR, respectively.

Only non-synonym mutations were cited. A sudden fall in the coverage depth was found at 9668, 22256, 39666, and 78493. They are probably due to single-strand interruptions in the T5 DNA (Bujard, 1969). Exact positions in the Table 12.

7.5.2. Phage termini analysis with PhageTerm

T5 genome has long exact direct repeat ends but the mechanism by which these long repeats are generated or the DNA is packaged into the virions remains unknown. One way to explore these mechanisms is through Next Generation Sequencing (Garneau et al., 2017). We have sequenced the genome of deletion mutants to confirm our constructions, nevertheless, the sequencing raw output data can also be used to explore other questions about the T5 infection mechanism, in particular, on the packaging of T5 DNA.

Gene deletions could either lead to the packaging of the same amount of DNA (121750 bp) or shorten the size of DNA packaged (<121750 bp). In the first case, we could expect headful packaging mechanism, while in the second we can expect that there is a *pac* signal contiguous to the direct terminal repeats as in T3 or T7 (Fujisawa and Morita, 1997).

Table 12. DNA termini found with PhageTerm

Strand	T5wt	T5DL	T5DR	T5DLDR	Single-chain interruptions
Positive [a] (Left end)	2	2	2	2 [b]	
	111533	111533	111533	111533	
Negative [a] (Right end)	9748	9748		9748	9748
	22336	22336	22336	22336	22336
	39746	39746	39746	39746	39746
	78573	78573	78573	78573	78573
	121279	121279	121279		121279

[a] only results with *p-value* < 10^{-3} and T > 0.5 were included ; [b] *p-value* > 10^{-3}

We tested the Galaxy-based tool PhageTerm (Garneau et al., 2017) to determine the left and right ends of the packaged DNA from the strains T5wt, T5 DL, T5 DR, T5 DLDR, in which we had deleted 0, 3, 7, 9.2 % of the genome. PhageTerm detects biases in the number of reads across the genome to predict the terminal ends relying on high-throughput sequencing of random fragmented DNA. Thus, short reads covering the termini tend to be more abundant than those covering the rest of the genome (Garneau et al., 2017).

The results (Table 12, Figure 67) show that the starting position remains unaltered, it corresponds to the position 2 or 111533. The right terminal repeat (RTR) starts at 111533. On the contrary, the Right end of the packaged DNA coincides with the DNA single-chain interruptions of T5. Moreover, coverage pattern at the right terminus in T5 is similar to that of phages with 5' cohesive ends, e.g., lambda. Illumina libraries were made by fragmentation of the DNA at random positions along the genome and probably T5 DNA single-strand interruptions are fragile points in which the DNA is frequently fragmented during the preparation of these libraries, which makes it impossible

to ascertain the phage right terminus position. To allow the observation of the right terminus in future experiments, it will be necessary to test nick-less T5 DNA

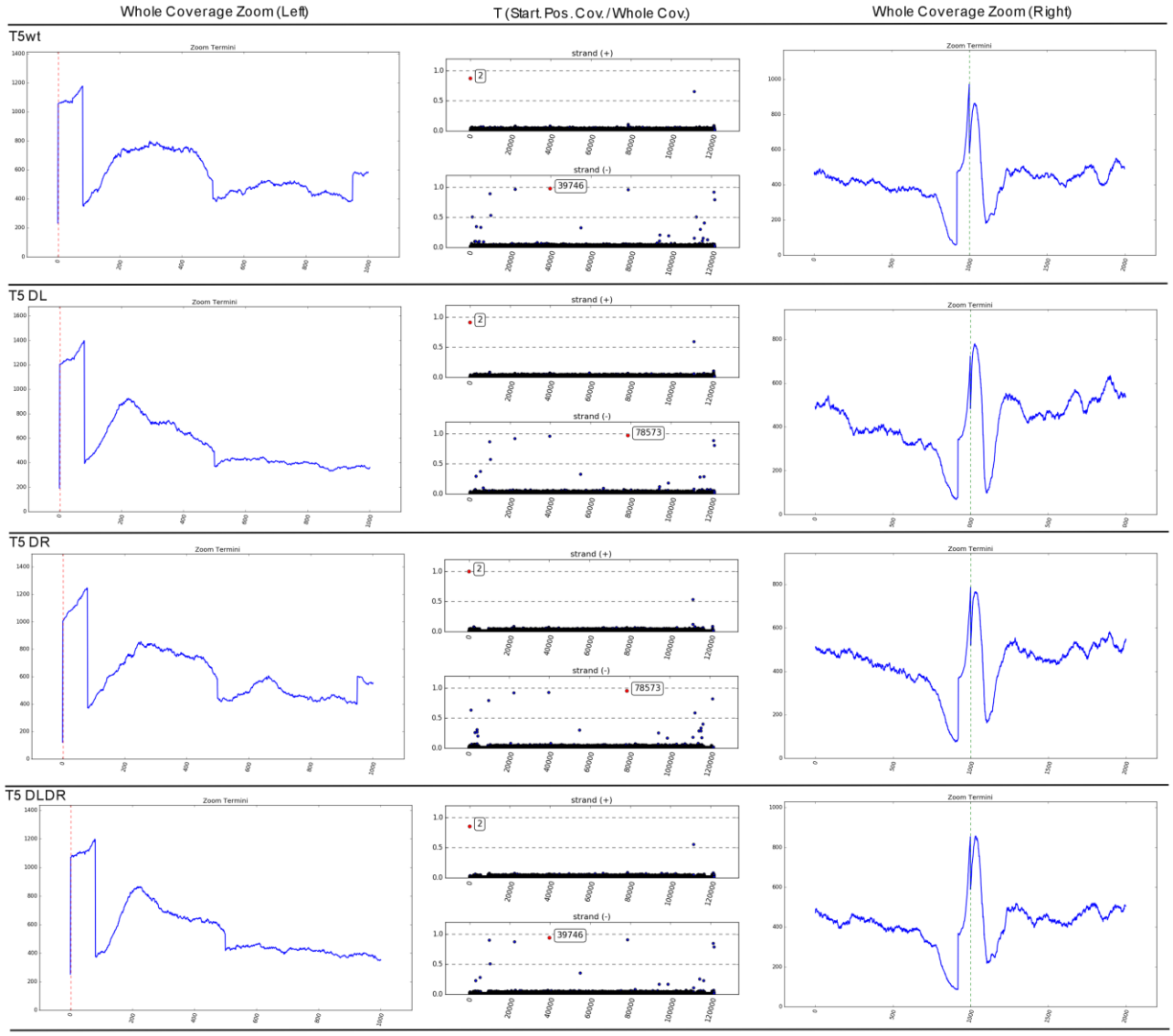


Figure 67. Phage termini in deletion mutants of T5
Short reads obtained by Illumina sequencing of bacteriophages T5wt, T5 DL, T5 DR, and T5 DLDR, were analyzed with the Galaxy-based program PhageTerm (galaxy.pasteur.fr) (Garneau et al., 2017). On the left, the coverage of the left terminus. In the middle, coverage versus position in the genome for (above) the positive strand and (below) the negative strand. On the right, the coverage of the right terminus.

7.6. Oligonucleotides used in this study

Mutant	Primer	Sequence
pAC9	2225	tttttgggctaacaggaggaattaaccATGGATAAGAAATACTCAATAGGCTTAGATATCGGCAC
	6331	gaaaatcttctctcatccgccaaaacaGCCTCAGTCACCTCCTAGCTGACTCAAATCAATG
	2225c	CATGGTTAATTCCTCCTGTTAGCCCAAAAA
	6331c	GGCTGTTTTGGCGGATGAGAGAAGATTTT
	Cas9F	gccttttttacggttcctggcACGAAGCAGGGATTCTGCAA
	Cas9R	cagctcactcaaaggcggtaAGGGTTATTGTCTCATGAGCGG
	psgF	TACCGCCTTTGAGTGAGCTG
	psgR	GCCAGGAACCGTAAAAAGGC
	QCpBADF	GTGTGGGGTCACCCCATGCGAG
	QCpBADR	CTCGCATGGGGTGACCCCACAC
	ACYCoriF	GTAAAAAGGCCGCGTTGCTGGCGTTGCGCTAGCGGAGTGTATACTGGC
	ACYCoriR	CCGTAGAAAAGATCAAAGGATCTTCACAACTTATATCGTATGGGGCTGACTTCAGG
	BC9GF	GAAGATCCTTTGATCTTTTCTACGGGGTCTGA
	BC9GR	AACGCCAGCAACGCGGC
pUCGG	pUCF	[Phos]-aaaaggtctcattttCTGGCCGTCGTTTTACAACG
	pUCR	aaaaggtctcaggggTCATGGTCATAGCTGTTTCCTGT
T5 Δdmp	gdmpS	tagtGGGAAATATGCGGGAAATTAC
	gdmpA	aaacGTAATTTCCCGCATATTTCCC
	DdmpF	aaaaggtctcacattTTGTTACGTTCTCCATTTGAGG
	DdmpR	aaaaggtctcaaatgCAATATATTGAGAAATTAAAAGTTGCGTAATAATTAAAG
	dmpF2	aaaaggtctcaaaaaGCGGGAGTTTTCGGAATGACT
	dmpR2	aaaaggtctcaccccCGAGGCCGGTCACGTTATTG
	dmp-scr	GGAAATATGCGGGAAATTACGcta
	DdmpScr	CTTTAATTATTACGCAACTTTTAATTTCTCAATATATTGcatt
T5 Δ*02*	gRNA-02S	tagtCGCTCAGTCTTCTATTGTTC
	gRNA-02A	aaacgaACAATAGAAGACTGAGCG
	FC533/52	atgaccatgattacgccCTATCACCACGACCGCGCAA
	FC1965/46	gtaaaacgacggccagtTTGAATGCGGCGTACACGGT
	D02F	aaaggtctcaTTTATCGCTGATGCGCTAGCGTAT
	D02R	aaaggtctcaTAAACATGTTTATAACTCCAATTAGTTTAATAAG
T5 *amA1* SS82	gRNA-A1S	tagtGGCGGGGATTGTTTCTAGCA
	gRNA-A1A	aaacTGCTAGAAACAATCCCCGCC
	3087	aaaaggtctcaGTAGACGGttgAGCAATGGGAAGTAAAACAAGGTAAACGGAATTCAGGAAAACAGACAGTAACTCA
	3127	aaaaggtctcaCTACTAAACAATCCCCGCCGCTTCTGGGCGCTCATCGTTTGAATTCAGGAAACCCGTTTTTTCTG
	3106	CCCATTGCTCAACCGTCTACT
T5 *amA1* T28	gRNAA1TS	tagtACGTCTGCGCGCTACTAAAT
	gRNAA1TA	aaacATTTAGTAGCGCGCAGACGT
	GGA1TF	aaaaggtctcaccccAGACCTTCAAGATTCAGCGCG
	GGA1TR	aaaaggtctcaaaaaCACTGGAAAGTGCAAAATTGGAAAC
	QCA11	ATATCGTCGCAAACTTCCAActtctAGGCGCGCAGACGTTTACCGTA
	QCA12	TACGGTAAACGTCTGCGCGCCTagaagTTGGAAGTTTGCGACGATAT
	2772	AGACCTTCAAGATTCAGCGCG
	A1Scr	CGGTAAACGTCTGCGCGCCTagaag
T5 Δ*05*	gRNA05S	aaacAGAATGAGGGATTCAGTAGG
	gRNA05A	tagtCCTACTGAATCCCTCATTCT
	05F	aaaaggtctcaaaaaACCGCAAAATTCGCTTGGAA
	05R	aaaaggtctcaccccCTAACCAGCAAATCAGCGCC
	D05F	aaaggtctcacaccCATTTTAAAACTCCAGTCAAAGGG
	D05R	aaaggtctcaGGTGTTAGTTCTAGCCTTATGCCTTT
T5 *amA2* S37	ggpUCA2F	aaaaggtctcaccccCGGCGGCAATAAAACAAATC
	ggpUCA2R	aaaaggtctcaaaaaCAGGTAATGCGGAACAATCC
	QCamA2F	ATCTCTTTTAAACCGTCctaATTTGCGAAATCTAAACC
	QCamA2R	GGTTTAGATTTCGCAAATtagGACGGTTTAAAAGAGAT
	4050.2	GAAAACGGTTTAGATTTCGCAAAgtag

Mutant	Primer	Sequence
T5 Δ*07*	07F-3	aaaaggtctcaccccTTCCAAGCGAATTTTGCGGT
	07R-3	aaaaggtctcaaaaaCCCAGTTCGATGAGAGCGAT
	D07F	aaaggtctcaCAGCGTAACGAAAAATGGGAAGC
	D07R	aaaggtctcagctgCATAATATAAACTCCAGTTTATTAAGGGG
	gRNA07S	tagtAGTAACGAAAGGGATAAAAA
	gRNA07A	aaactttTTATCCCTTTCGTTACT
T5 DR	FST-F	aaaaggtctcaaaaaTACCCAAAAATGGCGCAACC
	FST-R	aaaaggtctcaccccGCACGATTCGCCTTGACAAA
	DFST-F	aaaaggtctcacgagGATGGTTTAATCACTGGTATAAACGG
	DFST-R	aaaaggtctcaCTCGTTTAGTCTTGATTTTTAAAAGTCAATATC
	DFSTScr	GTTTATACCAGTGATTAAACCATCctcg
T5 DL	03F	aaaaggtctcaaaaaCATTAGTCGAGCGGTGCAAATAATG
	D03R	aaaaggtctcaAATGTTGTAAGTCTTGAAGCTAGCAAATAA
	Ddmp0203-Scr	GCCGCTATTACAGGAGCTAAATTAAATGaatg
T5 ΔISS::3	FC8663inv	agggttttcccagtcacgacgttgtaaaacgacggccagtGTAATGTTGAGGCTGCGAATGATT
	FC11322inv	acaatttcacacaggaaacagctatgaccatgattacgccCAGTTAGGGCATCACAAGGTAAAG
	DIS1F	aaaaggtctcaaaaaCAATTGCCCTGCGCTTTTGT
	DIS1R	aaaaggtctcattttGCGATGCCGTACTAAAAACCG
	DIS2.2F	aaaaggtctcaaaaaTAGAAATAACCGCCCTCCCTATAA
	DIS2R	aaaaggtctcattttATGGGGGCAAATGTCGGATT
	DIS3F	aaaaggtctcaaaaaACCACCACACCTTCTGTGTG
	DIS3.2R	aaaaggtctcattttTTATAGGGAGGGCGGTTATTTCTA
T5 *lacZ*α	05F	aaaaggtctcaaaaaACCGCAAAATTCGCTTGGAA
	05R	aaaaggtctcaccccCTAACCAGCAAATCAGCGCC
	pUC05-F	aaaaggtctcaTAGGCGTTATAATTTGCAACATTAATTTAAAGGCATAAGG
	pUC05-R	aaaaggtctcaATTCAAGGCGTTATATTTGGCAATATTGCCAATAAC
	LacZalpha-ATG	aaaaggtctcacctaATGACCATGATTACGCCAAGCTTGC
	LacZalpha-Stop	aaaaggtctcagaatCTATGCGGCATCAGAGCAGATTGTAC
	QC-LacZ-Eco1	GGATCCCCGGGTACCGAGCTCAAACTCACTGGCCGTCGTTTTACAA
	QC-LacZ-Eco2	TTGTAAAACGACGGCCAGTGAGTTTGAGCTCGGTACCCGGGGATCC
	QCLacZSD1	GCAAGCTTGGCGTAATCATGGTCATgtGattTccTccTTTGCAACATTAATTTAAAGGCATA
	QCLacZSD2	TATGCCTTTAAATTAATGTTGCAAAggAggAaatCacATGACCATGATTACGCCAAGCTTGC
	A2ATG	aaaaggtctcatcacATGACTAACGCTAAAACCGCAAAATTC
T5 PNmC	T5.151F	aaaaggtctcaccccTGACCATTCATGCTCTTAGCGA
	T5.151R	aaaaggtctcaaaaaTGCTTGATGGTGGTAATAATGCG
	mCF	aaaaggtctcaattaAGACTTGTACAGCTCGTCCAT
	mCR	aaaaggtctcataatGGAGGCTAATAAATGACACAAGC
	104870	GGATAACAGTTACGTGTGTGGC

Upper-case sequences match with the T5 genome or the plasmid pUC19.

Mutant	Primer	Sequence
pBAD derived	dmpATG	aaaaaagaattcaacATGAATCAAGTTAAAACGAATATTACCCGT
	dmpStop	aaaaaaaagcttTTACGCAACTTTTAATTTCTCAATATATTGTTG
	03ATG	aaaaggtctcatcacATGGCTATTAAAATTAATCTTCCCAGCAT
	03Stop	aaaaggtctcacaaaTTATTTGCTAGCTTCAAGACTTACAACAAA
	A2ATG	aaaaggtctcatcacATGACTAACGCTAAAACCGCAAAATTC
	A2Stop	aaaaggtctcaaattAAATTTATTCGCCTGCGGCTTG
	A1ATG	aaaaggtctcaaattACCGTTATCCTGAAAATGGCCG
	A1ATG2	aaaaggtctcaaatttTTTAATGGAGTTTTATTATTATGGTTATTTCCG
	A1Stop	aaaaggtctcacaaaTTACGCAAAGAATTTACCATTCTTAAAGT
	08ATG	aaaaggtctcatcacATGAAACAAACTTTATTGATCACTGGTAAAC
	08Stop	aaaaggtctcacaaaTCAAACTATTTTTTGCAGTTTTTTATGATTTTC
	09ATG	aaaaggtctcatcacATGACGAACTACACGGCGG
	09Stop	aaaaggtctcacaaaTCATATCGTCGATTCCTTATAATATATAAG
	hegGATG	aaaaggtctcatcacATGGGTCTATTTAAAGATATAGATTATTCGC
	hegGStop	aaaaggtctcacaaaTTAATTGACAATTCCAAATTCAGCTTTTTTAG
	11ATG	aaaaggtctcatcacATGCGGCTATATAAACCAGATAACG
	11Stop	aaaaggtctcacaaaTTACTCGCTTTCGGTGTTAACGG
	12ATG	aaaaggtctcatcacATGGTTGCTTATTCTTCTTCTGACG
	12Stop	aaaaggtctcacaaaTTACATTAAAACAATTTTGGCGGCTTGTT
	13ATG	aaaaggtctcatcacATGGTTATTTTTCGAGCTATTCTGGC
	13Stop	aaaaggtctcacaaaTTATTCTCCTTTATTCATGCAGCCC
	14ATG	aaaaggtctcatcacATGATCCGCAACGTTTCTCTTGC
	14Stop	aaaaggtctcacaaaTTATAACACTCCACGAAAAGCCG
	15ATG	aaaaggtctcatcacATGGTAATTTATGAAGGCAATCGCTTTG
	15Stop	aaaaggtctcacaaaTCATTGCGCCACCTTTCTGAAAG
	16ATG	aaaaggtctcatcacATGATTCCGCTGGTGGCGATA
	16Stop	aaaaggtctcacaaaTTAGTTGACTATCTCAAAGATAGCGTTC
	17ATG	aaaaggtctcatcacATGAGCAATAAAATTATTGTGACCAAAACTAC
	17Stop	aaaaggtctcacaaaTCATAAATCCCCCTCGGGTAATC
pET-13 derived	ATG-pET-13	aaaaggtctcaattcATGGTTATTTTTCGAGCTATTCTGGC
	Stop-pET-13-HisC	aaaaggtctcaaagcttagtggtggtggtggtggtgTTCTCCTTTATTCATGCAGCCCAG
	ATG-pET-13-HisN	aaaaggtctcaattcatgcaccaccaccaccaccacATGGTTATTTTTCGAGCTATTCTGGC
	Stop-pET-13	aaaaggtctcaaagcTTATTCTCCTTTATTCATGCAGCCC

Upper-case sequences match with the T5 genome.

7.7. Publication of results

The results obtained in this thesis will be submitted for publication in a peer-reviewed journal. The first report is presented in the following pages.

1. First report: to be submitted.

Strategies for bacteriophage T5 mutagenesis: expanding the toolbox for phage genome engineering. Ramirez L. M., Boulanger P. and Rossier O.

Contribution: Ramirez L. M. designed the research, performed all the experiments and wrote the first draft.

2. Second report: in preparation

Dissection of bacteriophage T5 pre-early genes uncovers the long elusive nuclease necessary for host takeover. Ramirez L.M., et al.

Contribution: Ramirez L. M. designed and performed most of the experiments, analyzed the data, and wrote part of the first draft.

Strategies for bacteriophage T5 mutagenesis: expanding the toolbox for phage genome engineering

Luis Ramirez-Chamorro[1], Pascale Boulanger[1] and Ombeline Rossier[1,*]

[1]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell, 91198 Gif-sur-Yvette, France

[*]To whom correspondence should be addressed.
Tel: (+33) 1 69 82 61 21; Email: ombeline.rossier@universite-paris-saclay.fr

Keywords: bacteriophages, genome engineering, CRISPR-Cas, retrons, Dilution-Amplification-Screening

ABSTRACT

Phage genome editing is crucial to uncover the molecular mechanisms of virus infection and to engineer bacteriophages with enhanced antibacterial properties. Phage genetic engineering rely mostly on homologous recombination (HR) assisted by the targeted elimination of wild-type phages by CRISPR/Cas nucleases. These strategies are often less effective in virulent bacteriophages with large genomes. T5 is a virulent phage that infects *Escherichia coli* and it can avoid CRISPR/Cas type I-E cleavage. We found that also CRISPR/Cas9 system (type II-F) had ununiform efficacies against T5, which impairs a reliable use of CRISPR-Cas-assisted counterselection in the gene editing of T5. We present here alternative strategies for the construction of mutants in T5. Bacterial retroelements (retrons) proved to be efficient for T5 gene editing by introducing point mutations in the essential gene *A1*. We set up a protocol based on dilution-amplification-screening (DAS) of phage pools for mutant enrichment that was used to introduce a conditional mutation in another essential gene (*A2*), insert a new gene (*lacZ*α) and construct a translational fusion of a late gene with a fluorescent protein (*pb10-mCherry*). The method might be applicable to other virulent phages that are naturally resistant to CRISPR/Cas nucleases.

INTRODUCTION

Bacteriophage genome engineering opens avenues to investigate the role of viral genes in infection and to generate bacteriophages with new properties. Until the late 1970s, genetic modifications were only possible through random mutagenesis (1). Recombinant DNA technology later allowed the first phage gene editing to be carried out in T4 (2). Continuous improvements of targeted mutagenesis methods were proposed to increase the recombination rate between a template and a bacteriophage, such as BRED (Bacteriophage Recombineering of Electroporated DNA) (3). Nevertheless, they depend on the compatibility of the recombinases to the phage to be engineered and on the phage genome size. Another progress consisted in the use of CRISPR-Cas nucleases, enzymes that can be targeted to precise sequences thanks to their association with so-called guide RNAs. CRISPR-Cas nucleases were successfully used to eliminate the wild-type phages that have not undergone HR and thereby enrich the phage population in mutant viruses. Both type I-E and type II-A CRISPR systems improved the recovery of T7 or T4 mutants by counterselection (4, 5), and the latter system was also used with *Lactococcus* or *Listeria* virulent phages (6, 7).

Despite these improvements in the reverse genetic toolbox, some phages are recalcitrant to nucleases, either by inhibition (Acr proteins protecting the phages from CRISPR-Cas) (8) or by genome shielding (9, 10). Alternatives to phage DNA modification in vivo is to “reboot” the phages. Following modification in vitro, the entire phage genome is transformed into a recipient bacterium to yield viral particles (11, 12). This bottom-up rebooting method requires a thorough optimization of transfection conditions and is very inefficient for phages with large genomes. Another avenue for rebooting could be the cell-free TXTL synthesis of phages in vitro: synthesis of T4 from the intact 170-kb genome T4 was recently reported, albeit with small yields (13). Therefore, no universal method seems to emerge for the editing of phages with large genomes.

Our laboratory studies bacteriophage T5, a virulent phage that infects *Escherichia coli* and has a linear 121-kb dsDNA genome. Most mutants in T5 were isolated in the past century after chemical mutagenesis and screening for amber mutations (14, 15). Since the 1980s, directed mutagenesis of T5 was also performed upon infection of host strains carrying a plasmid with the desired mutation. Usually ca. 500-bp homologous DNA is provided on each side of the mutation to favor HR. Screening of plaques was performed by hybridization of radiolabeled oligonucleotides in stringent conditions (16, 17) a rather cumbersome method. A recent report raised doubts about the possibility to use CRISPR-Cas for T5wt counterselection: T5 was reported to escape the CRISPR-Cas type I-E from *E. coli*: accordingly, although most of the spacers incorporated into CRISPR arrays map to pre-early genes, only few of them conferred protection against T5 (18).

In this report, we tested the use of a Cas nuclease normally not present in *E. coli*, i.e., the type II-A nuclease Cas9, for restricting T5 infection. We also describe an alternative method based on a scheme of dilutions for improved recovery of mutants. Finally, we test the use of retrons as another template for homologous recombination. These strategies likely constitute welcome additions to the toolbox for the phage genome engineering.

MATERIAL AND METHODS

Bacteria, phage and culture media

The *E. coli* bacterial strains used in this work are described in the supplementary data (ST 1). They were routinely grown at 37 °C in LB broth (10 g tryptone, 5 g yeast extract and 5 g NaCl per liter) supplemented with 1.5 % agar for solid media. When appropriate, antibiotics were used at the following concentrations: ampicillin, 200 μg/mL; chloramphenicol, 50 μg/mL kanamycin, 100 μg/mL. Prior to phage infections, cultures were supplemented with $CaCl_2$ (1 mM) and $MgCl_2$ (1 mM). T5 wild type (T5wt), T5 *lacZ*α, and T5 PNmC were propagated in *E. coli* F. T5 *amA1* SS84, T5 *amA1* T15, and T5 *amA2* S37 were amplified in *E. coli* CR63, which codes for an amber suppressor $_{CUA}tRNA^{Ser}$. Plaques were visualized on a double agar layer using molten LB agar (0.5 %) supplemented with the indicator bacteria and the phages. For the blue/white screening of lysis plaques the molten top agar also contained X-Gal (0.6 mg/mL) and isopropyl β-D-1-thiogalactopyranoside (IPTG, 3 mM). For visualization of lysis plaques, plates were incubated at 37 °C overnight and, when appropriate, a further incubation at 25 °C for 24 h enhanced the blue color of the plaques.

Plasmid constructions

Plasmid pAC9 was assembled from plasmids pCas9, pBAD24 and psgRNAcos (SF 1, ST 3). For sgRNA cloning, sense and anti-sense oligonucleotides (ST 2) were annealed using a temperature gradient from 98 °C to 20 °C at 0.1 °C/s, and then cloned by Golden-Gate method using the type II restriction enzyme BsaI into pAC9 (19).

Plasmid pFFA1, used for retron-mediated mutagenesis, was obtained by PCR from the plasmid pFF745 with primers 3087 and 3127, and subsequent religation in a Golden-Gate reaction (19).

Plasmids used for homologous recombination were assembled with primers cited in ST 3. In all cases, pUCGG was the recipient; this plasmid was obtained by PCR (pUCF/pUCR) with pUC19 as the template, followed by re-ligation. Inserts were amplified, purified and cloned in pUCGG by the Golden-Gate method (19). The QuikChange (QC) site-directed mutagenesis method was used to introduce specific mutations in the plasmids when needed (20). pUCA1am was constructed by cloning a PCR fragment from T5 (using primers GGA1TF/ GGA1TR) and introducing a stop codon (QCA11/QCA12) (20) at the *A1* codon T28. For pUCA2am, the amplified insert (ggpUCA2F/ggpUCA2R) was cloned in pUGG and the *A2* codon S37 was then changed to stop codon (QCamA2F/QCamA2F). Plasmid pUC05-LacZa2 was constructed in two steps: the first insert in pUCGG bears a T5 genome fragment comprising gene *05* flanked by parts of genes *A1* and *A2*, amplified by PCR (05F/05R); the resulting plasmid, pUC05, was amplified (pUC05-F/pUC05-R) to clone *lacZ*α (LacZalpha-ATG/LacZalpha-Stop) from pUC19. The *EcoR*I site in *lacZ*α was erased (QC-LacZ-Eco1) and a Shine Delgarno sequence added (QCLacZSD1/QCLacZSD2) upstream to allow translation. For pUCPNmC construction, two inserts were fused, one from T5 (T5.151F/mCR) and from pET-Nter-Cherry (mCF/T5.151R).

CRISPR-Cas9 interference assay

Each host (*E. coli* strains F or CR63) bearing a pAC9 with or without the spacer against T5wt was grown until an OD_{600}~0.4-0.5 was reached at 37 °C in LB supplemented with kanamycin, Ca^{2+} (1 mM) and Mg^{2+} (1 mM). 100 uL of serial dilutions of the phage and 100 uL of the host were added to the top agar of a double agar overlay assay in presence of kanamycin as well as arabinose (0.4%) to induce the expression of Cas9. To measure infection interference, the host and phage were *E. coli* F pAC9 and T5wt; for CRISPR-cas9 assisted counterselection of T5wt following mutagenesis, *E. coli* CR63 was infected with the lysate from HR.

Mutagenesis using plasmid template for homologous recombination

The template plasmids used for phage mutagenesis are listed in ST 1. *E. coli* CR63 was transformed either with pUCA1am (to generate T5 *amA1*) or pUCA2am (T5 *amA2*), *E. coli* XL-1 Blue with pUC05-lacZa2 (T5 *lacZ*α), while *E. coli* F was transformed with pUCPNmC (T5 PNmC).

Precultures of the cited hosts were started from a single colony inoculated in LB with ampicillin, shaken overnight at 37°C and diluted to OD_{600} ~0.1 in LB broth supplemented with ampicillin, Ca^{2+} and Mg^{2+}. Cultures were shaken at 37 °C until they reached OD_{600} 0.4-0.5, and were then infected with T5 wt at MOI 10.

After 10 min at 37 °C, cells were washed twice through centrifugation at 5000 x g for 3 min and resuspension with LB broth supplemented with ampicillin, Ca^{2+} and Mg^{2}+. Cells were harvested, resuspended at half of the initial volume, and incubated at 37 °C for 90 min. Then 0.2% chloroform was added to the culture. After 10 min at 37 °C, the lysate was centrifuged at 5000 x g 3 min, and the supernatant was separated and stored at 4 °C.

Mutagenesis using retron

E. coli cells CR63 carrying the plasmid pFFA1 were incubated overnight at 37 °C, shaking at 220 rpm. This preculture was used to initiate a 20-mL culture at OD_{600} 0.1 in LB broth with chloramphenicol and 1mM each of IPTG , $CaCl_2$ and $MgSO_4$. The culture was incubated at 37 °C *and* shaking at 220 rpm until OD_{600} of 0.4, and then infected with T5wt at MOI 10 and incubated at 37 °C while shaking.

After 10 min at 37 °C, cells were washed twice through centrifugation at 5000 x g, resuspended in half the initial volume with LB supplemented with chloramphenicol, IPTG, Ca^{2+} and Mg^{2+}. Cell cultures were treated from hereon as detailed for the mutagenesis by homologous recombination.

Mismatched amplification mutation assay (MAMA-PCR)

For the PCR screening of phages obtained after mutagenesis, we performed a mismatched amplification mutation assay (MAMA-PCR) (21). The sequence of primers for the detection of T5 *amA1* (2772/A1Scr), T5 *amA2* (05R/4050.2), T5 *lacZ*α (LacZalpha-Stop/A2ATG), Pb (mCF/104870) are shown in ST 3. PCR conditions were 98 °C for 3 min, 30 cycles at 55 °C (30 sec) and 72 °C (1 min), then 72 °C (5 min). PCR mix contained 10uM of each desalted primer (Eurofins), DMSO 1.5 %, 1 uL of filtered lysate, and Taq DNA Polymerase 0.05 U/uL with 1x Buffer with KCl (Thermo Scientific).

Dilution-Amplification-Screening (DAS)

The lysate obtained from the HR was filtered with 0.45-um diameter nitrocellulose filters. The lysates were first diluted: we made 10-fold serial dilutions to reach ca. 10-100 PFU/mL (usually up to a dilution factor of 10^8 to 10^{10}), of 10 mL volume each, with LB supplemented with $CaCl_2$ (1mM) and $MgSO_4$ (1mM). Fifty μL of each dilution were distributed in twelve wells of a 96-well sterile flat bottom polystyrene plate (Corning), one row per dilution factor. To amplify the phages in each well, one-hundred-μL aliquots of a CR63 bacterial culture in exponential phase were added. The plate was incubated for 3 hours at 37 °C 220 rpm shaking. For the first screening by MAMA-PCR, we pooled 20 μL from six wells in the same row. Positive pools corresponding to the highest phage dilutions were selected, and the six individual wells were tested individually for the second MAMA-PCR screening. Finally, 50 μL taken from a PCR- positive well were plated on a double agar to test 10-20 lysis plaques individually (SF 2).

Fluorescence microscopy

E. coli strain F was transformed with plasmid pFH2973 (22). The strain was incubated at 37 °C in LB supplemented with ampicillin, $CaCl_2$ and $MgCl_2$, until it reached an OD_{600} of 0.4. To synchronize the infections, the cells were chilled at 4°C for 15 min. Phage T5 PNmC was added to the culture at an MOI of 5 and incubated for 5 min at 4°C. The infected cells were washed by centrifugation at 5000 *x g* for 3 min and resuspension in LB. An aliquot of the culture was taken every 10 minutes and fixed with the same volume of PFA solution (Formaldehyde 5%, Glutaraldehyde 0.06%, diluted in PBS buffer) for 20 min. Fixed cells were washed in PBS buffer (NaCl 0.14M, KCl 2.68mM, Na_2HPO_4 6.46mM, KH_2PO_4 1.148mM) twice by centrifugation-resuspension. The cells were finally resuspended in PBS and stored at 4 °C until microscopy imaging.

Fixed cells were deposited on thin agarose pads (1% molecular-biology-grade agarose in PBS) molded on glass slide and observed under cover slips with a Zeiss Axio Observer.Z1/7 microscope using the ZEN software, version 3.0. The objective was Plan Apochromat 63x/ 1.40 Oil Ph 3 M27. The exposure time was 200 ms for mCherry mPlum HXP (Ex: 587 nm, Em: 610 nm) and 20 ms for phase contrast.

RESULTS

Restriction of T5 infection by CRISPR/Cas9 depends on the targeted locus

To test whether phage T5 infection is restricted by CRISPR-Cas9, we transformed *E. coli* F with a plasmid coding for the nuclease Cas9 and customizable sgRNAs. The sgRNAs matched different locations across the first 10 kb of the T5 phage genome. By double-layer agar method, we compared the titer of a T5 stock on two host strains: *E. coli* F expressing both Cas9 and each sgRNA, versus the same host lacking the sgRNA. The ratio of PFUs between the two hosts was considered as the efficiency of plating (EOP) and reflects the protective capacity of CRISPRs against phages (5, 18) for each sgRNA. Restriction of T5 infection varied across the tested region regardless of the DNA strand that was targeted. Spacers with the highest interference mapped to the genes *dmp*, *02*, *A1*, *05*, and *07* and lowered the titer of T5wt between 10 to 100 times. The other spacers did not provide protection as the titer of T5wt decreased less than ten times. Therefore, Cas9 can be efficiently used to restrict T5wt infection when it is targeted to a region within the first 4.5 kb of the T5 genome.

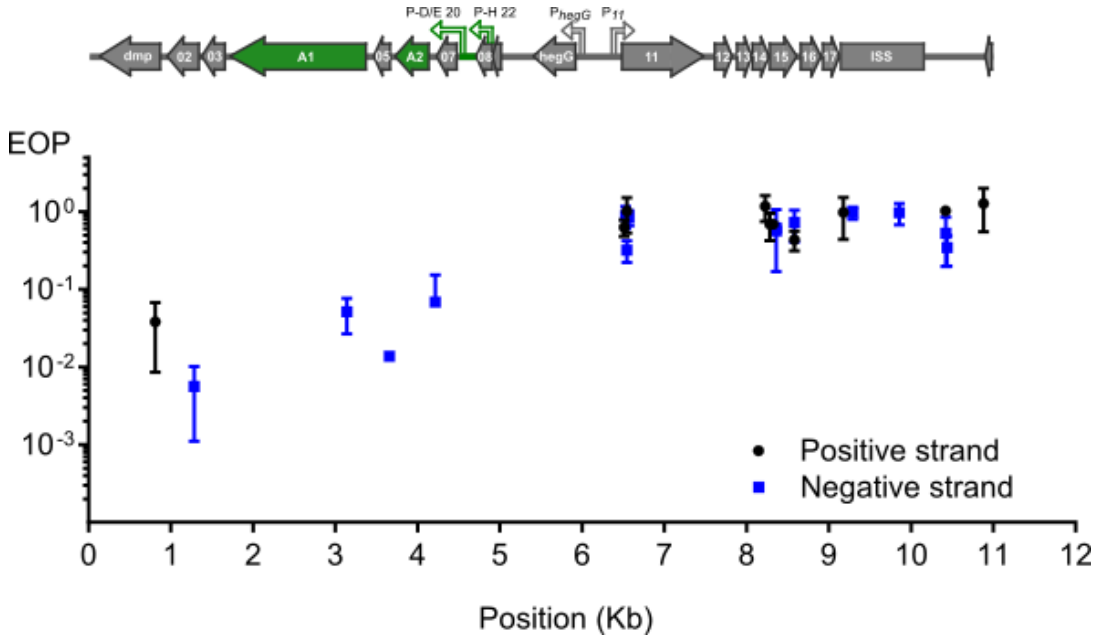


Figure 1. **Restriction of phage T5 infection by sgRNA-guided Cas9.** The restriction of T5 infection was measured as the efficiency of plating (EOP), calculated as the titer of T5 on sgRNA- and Cas9-containing hosts versus the titer on the same host lacking the sgRNA. Titers were determined by plaque assay. EOP using spacers matching the positive or negative strand were highlighted in black or blue, respectively. Each point represents the mean and s.e.m. from three replicas. In the genetic map of the first 11 kb from the 121-kb genome, green arrows represent genes *A1* and *A2* which are predicted to be essential based on the phenotype of amber mutations genetically mapped in the 1960s.

Bacteriophage T5 genome editing through HR plus CRISPR/Cas9

In order to introduce genetic modifications in T5, we focused on the region with effective spacers. In the late sixties, Lanni and coworkers described the isolation of T5 amber mutants that could only be propagated in the CR63 permissive strain (15, 23). Mutations were genetically mapped to *A1* or *A2* genes (Fig. 1) but no sequencing data were obtained to confirm this finding. Therefore, we decided to introduce a stop codon early on in the open reading frame of the gene *A1*, resulting in the synthesis of truncated A1 (27 amino acids instead of 556). A 1-kb fragment of the T5 genome was amplified and cloned into a pUC19 plasmid. This fragment comprises 500-bp on each side of the *A1* start codon in order to provide ample homologous sequences on either side of the desired mutation, and the codon for T28 on the insert was replaced with a stop codon. Phage mutagenesis was performed by allowing infection of *E. coli* CR63 carrying the resulting plasmid. Counterselection of the T5wt within the crude lysate was carried out by double agar overlay plaque assay onto *E. coli* CR63(pAC9-A1), a strain that expressed Cas9 and a sgRNA directed against the wild-type *A1* gene

(Fig. 2A). The resulting plaques were then picked and patched onto both *E. coli* strains F (non permissive) and CR63 (amber suppressive). We recovered six amber mutant plaques in 100 patched ones (Fig. 2B). Sequencing confirmed the introduction of the intended mutation in the T5amA1 phages (Fig. 2C). Thus, our results demonstrate that it is possible to use HR to introduce point mutations in the genome of T5 and to obtain a good yield in screening of the phage mutants with the assistance of CRISPRcas9 to introduce point mutations in the genome of T5.

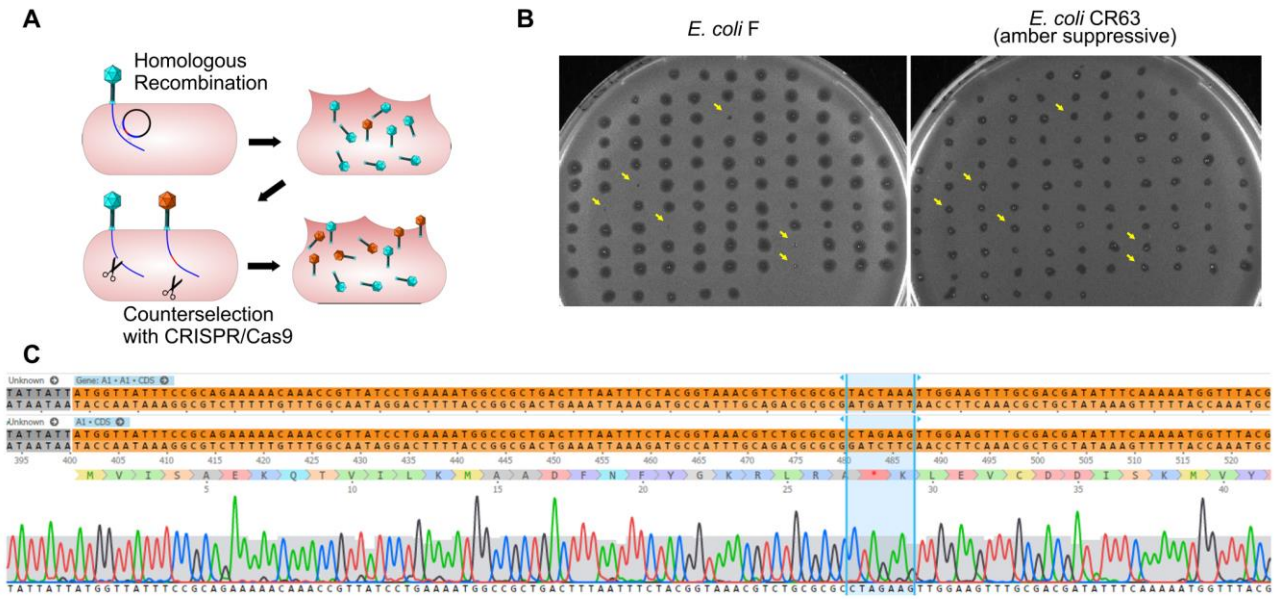


Figure 2. **Construction of phage mutants by homologous recombination and mutant enrichment using CRISPR/Cas9.** (A) Upon infection by wild-type phages of a host bearing a template plasmid, HR introduces the genetic modification in a subset of phages. In a second step, mutants among the progeny are enriched by counterselection onto a host producing the Cas9 nuclease targeted against the wild-type sequence. (B) Screening of the phages obtained after mutagenesis and counterselection. Lysis plaques were tested onto a non-permissive *E. coli* strain F and the amber suppressive strain CR63. Plaques with no lysis on *E. coli* F were considered positive and sequenced. (C) Sequencing results for one of the six amber mutants T5amA1 T28. Top lane corresponds to the T5 nucleotide sequence with the desired amber mutation within gene *A1* at position (3226..3332) of the T5 genome. Note that *A1* coding sequence is on the complementary strand. Second lane corresponds to amino acid residues 16 to 50 translated from the sequence above. Third lane shows sequencing results.

Bacteriophage genome engineering with retrons

We explored the use of alternative forms of DNA templates for phage genome engineering. Bacterial retroelements or retrons are chimeric RNA/DNA molecules composed of covalently linked ssRNA (msr) and ssDNA (msd). Both msr and msd are coded in the same cistron, altogether with a reverse transcriptase responsible for the partial reverse transcription and linking of the msr-msd molecules (24). These elements can provide ssDNA *in vivo* for gene modification, and were used for bacterial genome engineering in the past (25, 26). We tested whether retrons are also effective for phage genome engineering in combination with CRISPR-Cas9. For the retron template, we modified plasmid pFF745, which codes for the msr, msd and the reverse transcriptase in a polycistron controlled by an IPTG inducible promoter (25). Part of the msd sequence was replaced with a 75-bp segment of the phage *A1* gene centered on the serine codons for S84 and S85. These codons were substituted by stop codons, generating plasmid pFFA1. To introduce

the mutation *A1* S84Stop/S85Stop into the phage genome, we grew *E. coli* CR63(pFFA1) in the presence of IPTG and infected the bacteria with the wild-type phage. The crude lysate was plated onto *E. coli* CR63 bearing plasmid pAC9 and sgRNA matching the unmodified codons S84/S85 of T5 *A1*. For mutant screening, we picked 100 plaques and patched them onto plates with bacterial lawns of *E. coli* strains F and CR63. We recovered 2 out of 100 plaques that could only lyse the amber-suppressing host strain. Such mutants were designated T5amA1 SS84. Sanger sequencing results showed no other alteration in the targeted locus (Fig. 2). Our results suggest that retrons could be effectively used for phage genome engineering.

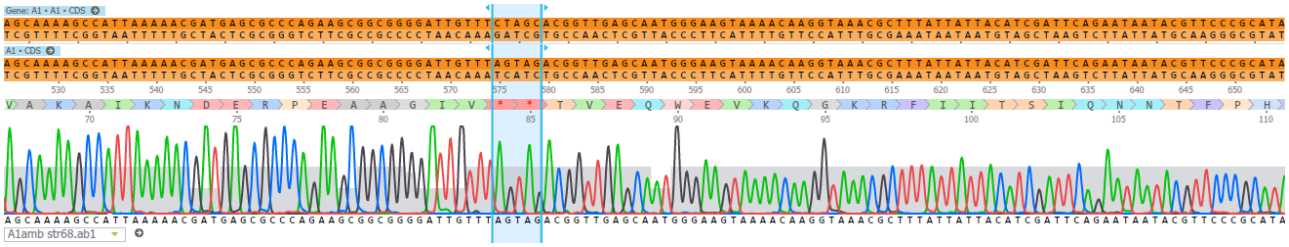


Figure 3. **Sequencing of the phage T5 amA1 SS84 constructed by retron-mediated recombination.** Purified amber mutant phages T5 amA1 SS84 obtained by retron-mediated mutagenesis were used as a template for PCR amplification. Amplicon sequences showed the presence of two stop codons instead of serine codons (S84 and S85) in the T5 *A1* gene.

Dilution-Amplification-Screening to isolate mutants from phage T5

As we have seen above, some regions of the T5 genome are not amenable to the CRISPR/Cas9 counterselection, a feature that enriches the phage progeny in mutants relative to wild type after mutagenesis. We reasoned that a similar enrichment could be achieved with a simple dilution-amplification screening (DAS) (Fig. 4A). As a proof of principle for this procedure, we sought to obtain a phage carrying an amber mutation in the essential gene *A2* at the codon S37. We allowed the recombination between the phage T5 and the plasmid bearing a mutated copy of the gene *A2* during infection. Following filtration to eliminate uninfected bacteria, the crude lysate was serially diluted from ca. 10^{10} PFU/mL down to 10^{2} PFU/mL. We distributed each dilution into several wells containing the amber-suppressive host and incubated for 3 h to amplify the phages (Fig. 4A and Fig. S2). We screened the wells by mismatched amplification mutation assay (MAMA)-PCR to detect the mutation and the positive wells were plated by a double agar layer method to obtain individual plaques. The plaques were picked and patched successively onto two plates with a non-permissive and a permissive host, respectively. The plaques that lysed only the permissive host were considered positive (Fig. 4B). We recovered 6 positives mutants from 50 tested, and they were all? confirmed by Sanger sequencing (Fig. 4C). Since our estimation of T5 recombination rates is around 0.25% (see below), our recovery rate of 12% indicates that we could enrich the mutant in some pools enough to simplify the recovery of T5 *amA2* S37 mutants. Hence, DAS is an efficient method for mutant isolation.

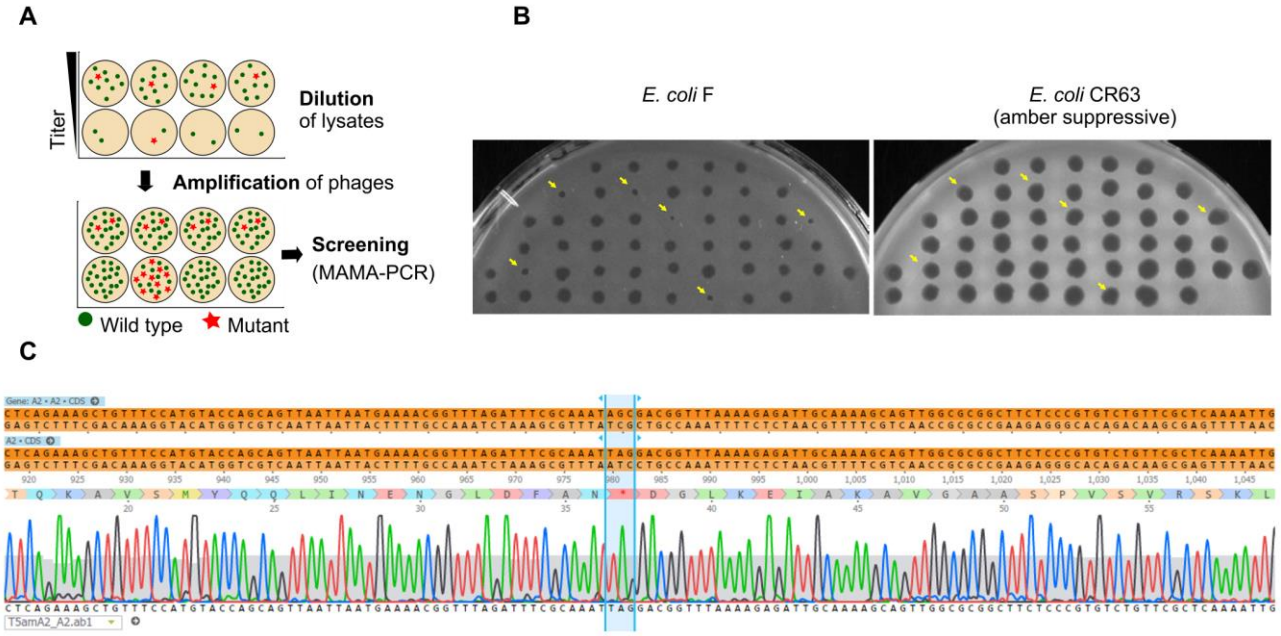


Figure 4. **Screening of T5 mutants by Dilution-Amplification Screening (DAS).** (A) DAS principle: following mutagenesis mutants are far less abundant than wild-type phages in the crude lysates. In the first step, dilutions of the lysates are performed to lower the phage titers and alter the proportion of mutants present in each pool (above, second row). In the second step, phages in the diluted pools are amplified by adding the host. While the mutants remain scant within amplified pools from initial high viral titers (below, first row), the modified phages become enriched in the progeny obtained with some of the highly diluted pools (below, second row). (B) Plaques obtained from a PCR-positive well were patched onto the non-permissive strain F and the amber-suppressive host CR63. Positive plaques are highlighted with an arrow. (C) Sequencing result shows the correct introduction of point mutations in the T5 genome.

Blue T5: Insertion of *lacZ*α gene into the T5 genome

Following the successful introduction of point mutations in the T5 genomes, we further tested whether insertion of exogenous genes was possible. To generate a T5 phage that produces blue lysis plaques, a 156-bp ORF encoding the fragment LacZα of the *E. coli* β-galactosidase was introduced by HR between the genes *05* and *A1* and DAS was used to recover mutants (Fig. 5A). The T5 *lacZ*α (Blue T5) plaques displayed a β-Gal activity (Fig. 5B) when plated in a soft agar overlay with the strain *E. coli* XL-1 Blue in the presence of X-Gal and IPTG. When we skipped the DAS procedure and relied only on the blue/white plaque screening, we were able to measure the rate of recombination in T5 after HR. We counted the number of blue lysis plaques obtained from the filtered crude lysate and found that 0.27 ± 0.01 % (n = 6) of the phages were recombinants, i.e., had integrated the *lacZ*α gene. In T4, the frequency of mutants was reported to be between 0.1-0.5 % (2), while in T3, it is comprised between 0.1-1% (27). Therefore, the HR in T5 is consistent with that of other phages from the T series.

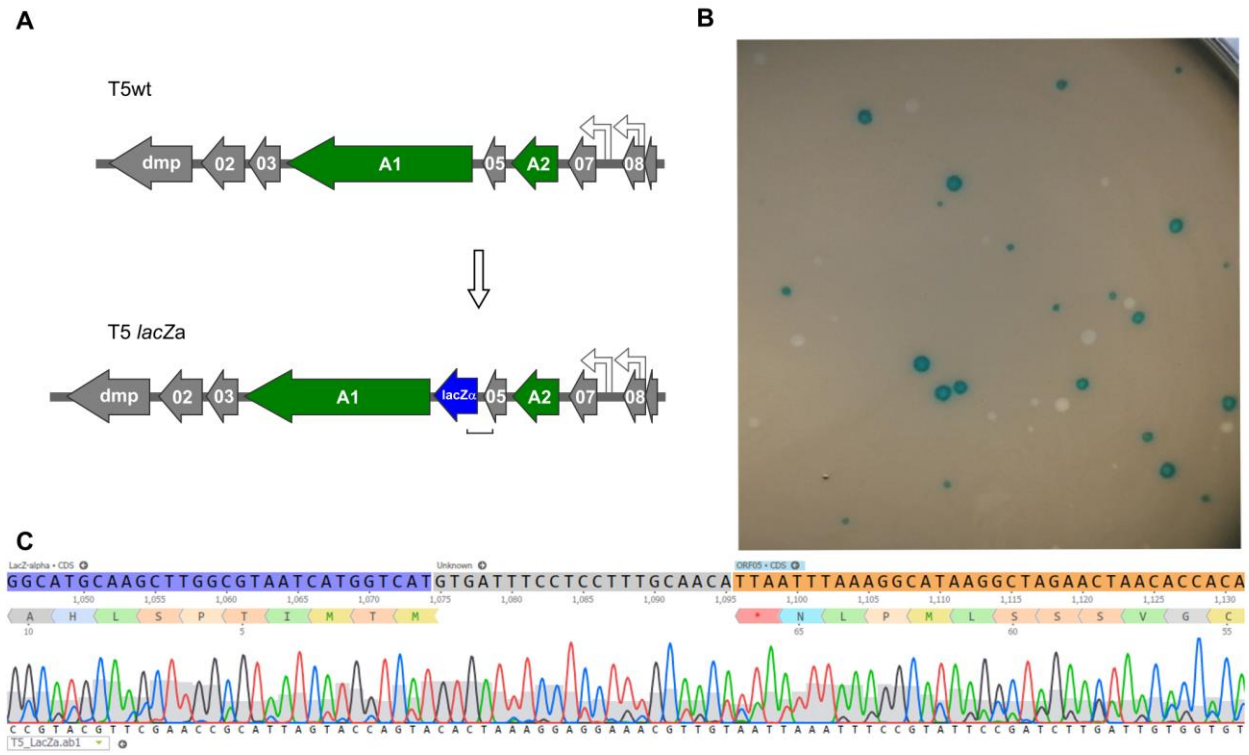


Figure 5: **Construction of a Blue T5 (T5 *lacZα*) strain.** (A) Genetic maps showing that the gene *lacZα* was introduced between genes *A1* and *05* in the genome of T5. (B) Plaques of T5 *lacZα* and T5wt after plating on a *E. coli* XL-1 Blue lawn supplemented with X-Gal and IPTG: only the mutant produced blue plaques. (C) Sequencing results of one of the blue plaques, showing the correct insertion of *lacZα* downstream of gene *05.*

Phages producing a chimeric fluorescent late protein

To test whether our mutagenesis protocol was applicable to another region of the T5 genome, we generated a gene fusion between part of gene *151* and a reporter gene encoding a fluorescent protein, The product of T5 gene *151*, the late protein pb10, is a two-domain protein that decorates the surface of the T5 capsid. Although dispensable for capsid assembly, it contributes to its stability under high temperature (17). The N-ter domain of pb10 that bind to specific sites on the capsid was fused to mCherry (Fig. 6). Mutant phages were recovered and could be used to monitor the timing of production of late proteins during infection by fluorescence microscopy. There was a rapid rise of red fluorescence inside the cells after around 20 min of infection (Fig. 6B), consistent with the expected late synthesis of the PNmC fusion.

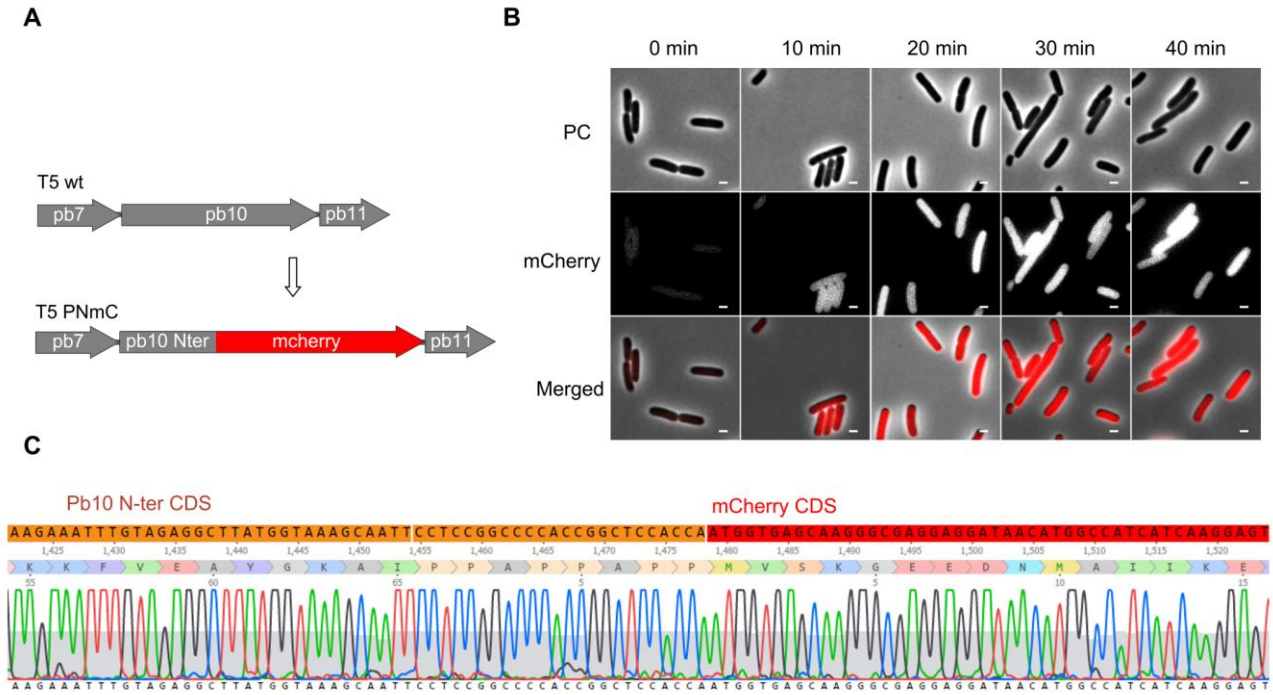


Figure 6: **Gene fusion encoding a chimeric protein between the first half of the late protein Pb10 and a fluorescent protein.** (A) Map of the fusion between the first half of *pb10* and *mcherry* (in red). (B) Fluorescent microscopy of *E. coli* F cells infected by T5 PNmC. The time after infection is indicated above. Top row, phase contrast (PC); middle row, mCherry fluorescence; bottom row, merged images. White bars, 2 μm. (C) Sequencing results of the pb10 gene in T5 PNmC showing the region where the Pb10 Nter coding sequence (in orange) merges with the *mcherry* ORF (in red).

DISCUSSION

In this work, we have edited the genome of bacteriophage T5 and could generate mutants carrying point mutations or insertions. Template DNA for mutagenesis was provided on a plasmid with ca. 500-bp homologous DNA on either side to facilitate homologous recombination, a classical method used for the genome engineering of many phages (28). Here, we also chose to test an alternative template DNA, i.e., bacterial retroelements called retrons. Retrons code for mixed, covalently linked molecules constituted by single-stranded RNA and DNA, msd-msr, which can provide ssDNA molecules *in vivo*. They were successfully used for gene editing of *E. coli* cells expressing the lambda recombinase Beta (24). Although the mechanism of mutagenesis by msd-msr molecules remains unknown, it has been suggested that the ssDNA produced from msd-msr cleavage replaces Okasaki fragments during replication leading to the introduction of modifications in the genome (25). Retrons were used in the past to store analogically encoded information on the bacterial DNA upon exposure of the cells to certain stimuli (25) or to continuously target specific sequences in the bacterial genome to drive evolution (26). To our knowledge, our report is the first example of phage genome editing using retrons. The small size of the homologous domain (75-bp) is incompatible with insertions of large sequences in the genome but was sufficient to successfully introduce point mutations in T5 (here two stop codons). Further work will test whether retrons can also be used to construct gene deletions in phages.

Reported phage homologous recombination rates vary widely from 10^{-10} to 10^{-4} depending on the phages and the genes that are targeted (28). These rates warrant efficient screening strategies of the phage progeny after mutagenesis. One common approach is the targeted elimination of wild-type phage using interference by CRISPR-Cas. Mutant phages are thus enriched within the phage population (4–7). Using Cas9 and sgRNAs targeting genes within the first 4.5 kb of T5 DNA, we were successful in isolating conditional mutants in the essential gene *A1*.

However, by targeting different locations across the pre-early gene locus with sgRNA-guided Cas9, we found that interference was not uniform, something similar to what was been seen in T4 (5). While T4 DNA is partially protected

315 through the modification of cytosines with 5-hydroxymethylation plus glycosylation, no modification of T5 DNA has been
316 reported so far. This resistance to CRISPR-Cas interference was not limited to Cas9 nuclease: indeed a recent study of
317 the impact of CRISPR/Cas type I-E against T5 showed that the few efficient spacers matched sequences that are on the
318 left half of T5 DNA but not the rest of the genome (18). These observations suggest that there might be a generalized
319 way, other than base modification, in which T5 manages to escape CRISPR/Cas nucleases. A similar observation in
320 jumbo phages led to the discovery of a viral proteinaceous shell that protects phage DNA from nucleases (9, 10).
321 However, whether phage T5 uses a similar mechanism is unknown.
322 For phages resistant to CRISPR-Cas interference, we have established an alternative enrichment method by Dilution,
323 Amplification, and Screening (DAS). The strategy relies on the dilution of the viral progeny into parallel pools, followed by
324 their amplification in the host. Screening of the most dilute pools occurs using a discriminative PCR. This method was
325 successful in isolating conditional as well as insertion mutants of bacteriophage T5. In this work we generated a
326 bacteriophage mutant useful to detect late viral gene expression in fluorescence microscopy of *E. coli* infected by T5. We
327 also used DAS to generate a phage carrying a 156-bp *lacZ*α gene among the early genes of T5. This phage was easily
328 trackable when infecting *E. coli* cells designed for alpha complementation of beta-galactosidase activity, since its blue
329 plaques are easily distinguished from colorless plaques. Using this phenotype, we were able to determine the
330 recombination efficiency of T5, which is similar to that of T3 (27) and T4 (2). Such an easily trackable phage could prove
331 useful in competition assays with other phages in the future. Finally, we believe that DAS can document that some genes
332 targeted for mutagenesis are essential for infection, as the phages carrying the mutation after HR will not propagate
333 during the amplification step and therefore will only be abundant in the least diluted pools.
334 Taken together, our work has shown that retrons and DAS were successful tools in the mutagenesis of T5, a phage
335 with a large genome mostly recalcitrant to CRISPR-Cas nucleases. These tools are likely welcome additions to the
336 biotechnological toolkit for phage genome editing.
337

SUPPORTING INFORMATION

ST 1: Plasmids, bacterial strains and bacteriophages used in this study.

Plasmids	Reference
pCas9	(29)
pUC19	Genbank: M77789.2
pBAD24	(30)
pFF745	(25)
psgRNAcos	(31)
pAC9	This study
pFFA1	This study
pUCamA1	This study
pUCamA2	This study
pUC05-LacZa2	This study
pET_Nter	(17)
pUC_PNmC	This study
Escherichia coli strains	
DH5alpha	New England Biolabs
CR63	(32)
F	(33)
MG1655	GenBank: U00096.3
XL-1 Blue	Stratagene
Bacteriophages	
T5 wt	GenBank: NC_005859.1
T5 amA1 strain SS84 (S84stop S85stop)	This study
T5 *amA1* strain T28 (T28stop)	This study
T5 *amA2* (S37stop)	This study
T5 *lacZ*α	This study
T5 PNmC	This study

ST 2: Oligonucleotides to generate the sgRNA for the CRISPR-Cas9 Infection Interference Assay

Name [a]	Plasmid	Sense (S)	Anti-Sense (A)	Strand	Locus	EOP
gdmp	pAC_dmp	tagtGGGAAATATGCGGGAAATTA	aaacGTAATTTCCCGCATATTTCC	Pos	829	$3.81x10^{-2}$
gRNA-02	pAC_02	tagtCGCTCAGTCTTCTATTGTTC	aaacGAACAATAGAAGACTGAGCG	Neg	1283	$5.61x10^{-3}$
gRNAA1T	pAC_A1T	tagtACGTCTGCGCGCTACTAAAT	aaacATTTAGTAGCGCGCAGACGT	Neg	3289	$4.22x10^{-1}$
gRNA05	pAC_05	aaacAGAATGAGGGATTCAGTAGG	tagtCCTACTGAATCCCTCATTCT	Neg	3658	$1.38x10^{-2}$
g07inv	pAC_07inv	tagtTATATCCCCTAGTAGTTCGG	aaacCCGAACTACTAGGGGATATA	Neg	4217	$6.29x10^{-2}$
g10	pAC_10	tagtGCAACCTGCTGGATGGTAGC	aaacGCTACCATCCAGCAGGTTGC	Pos	6563	$1.02x10^{+0}$
g10.4	pAC_10.4	tagtAGATAACGCAACCGTCTTAA	aaacTTAAGACGGTTGCGTTATCT	Pos	6512	$6.31x10^{-1}$
g10.5	pAC_10.5	tagtTTTAAGACGGTTGCGTTATC	aaacGATAACGCAACCGTCTTAAA	Neg	6532	$8.87x10^{-1}$
g10.6	pAC_10.6	tagtGCGCAAGGCGCCCTTTAAGA	aaacTCTTAAAGGGCGCCTTGCGC	Neg	6545	$3.21x10^{-1}$
g14.5	pAC_14.5	tagtTTTCGTGGAGTGTTATAAAA	aaacTTTTATAACACTCCACGAAA	Pos	8281	$6.89x10^{-1}$
g14.7	pAC_14.7	tagtCTTTGTTGCTATTTGCCGAC	aaacGTCGGCAAATAGCAACAAAG	Pos	8323	$6.90x10^{-1}$
g14.6	pAC_14.6	tagtAAGTAGTTTGCTAATAATCC	aaacGGATTATTAGCAAACTACTT	Neg	8364	$5.66x10^{-1}$
g14.3	pAC_14.3	tagtCGCATATATTCAGTCAACTG	aaacCAGTTGACTGAATATATGCG	Pos	8584	$4.37x10^{-1}$
g14.4	pAC_14.4	tagtGCCTTTCATTATCAGGCGCT	aaacAGCGCCTGATAATGAAAGGC	Neg	8584	$7.28x10^{-1}$
gIS3	pAC_IS3	tagtAACGGGGAGCTGATCCCCGT	aaacACGGGGATCAGCTCCCCGTT	Neg	9293	$9.73x10^{-1}$
gIS7	pAC_IS7	tagtGCACACTATAAAAATTTTTC	aaacGAAAAATTTTTATAGTGTGC	Neg	9859	$9.74x10^{-1}$
gIS6	pAC_IS6	tagtTGTATTAAACCGCCTATTGC	aaacGCAATAGGCGGTTTAATACA	Pos	10421	$1.03x10^{+0}$
gIS5	pAC_IS5	tagtCAGGGCGGTTATAGGGAGGG	aaacCCCTCCCTATAACCGCCCTG	Neg	10422	$5.25x10^{-1}$
gIS4	pAC_IS4	tagtATAGGCGGTTTAATACAGGG	aaacCCCTGTATTAAACCGCCTAT	Neg	10437	$3.45x10^{-1}$

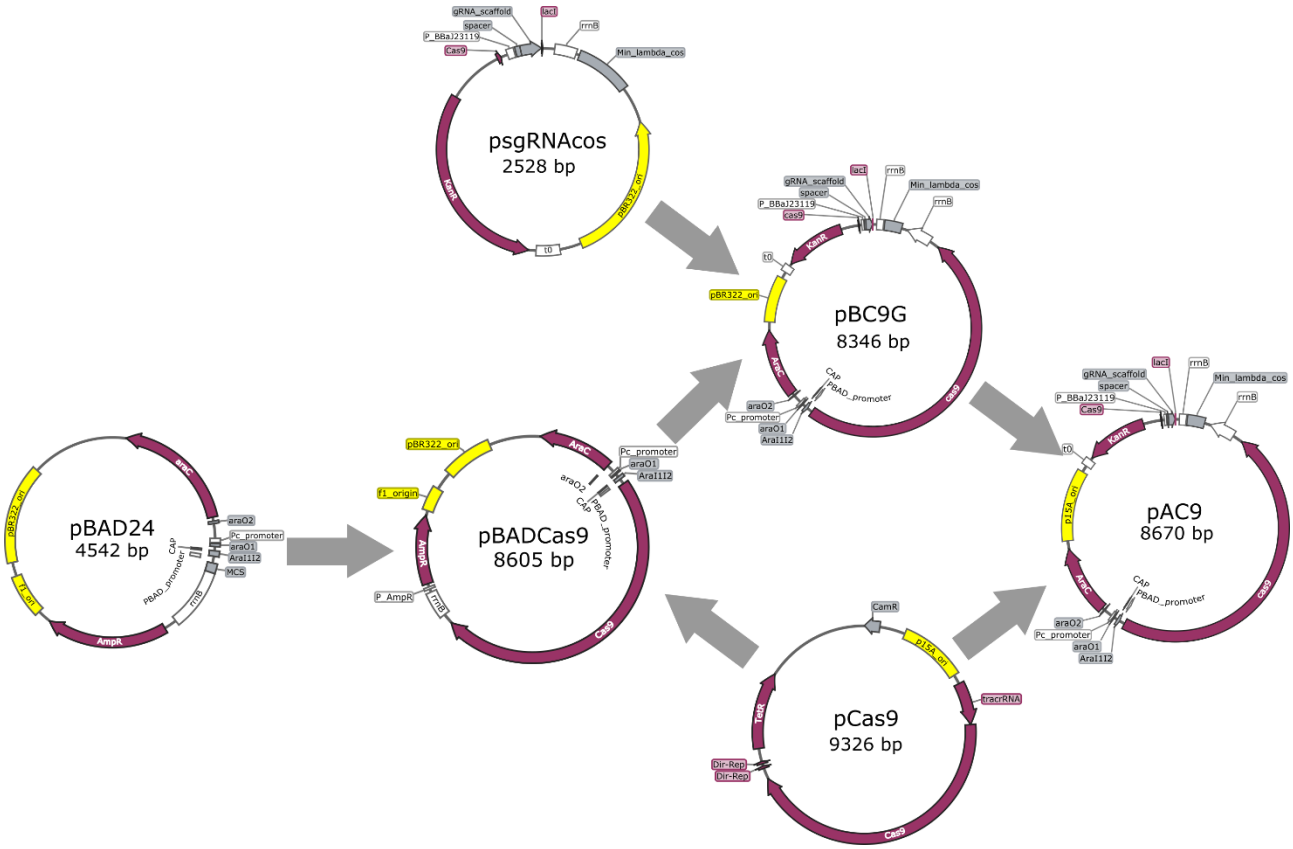
[a] Add the character S or A at the end of the primer to get the primer's name.

343 ST 3: Oligonucleotides used for cloning or mutagenesis

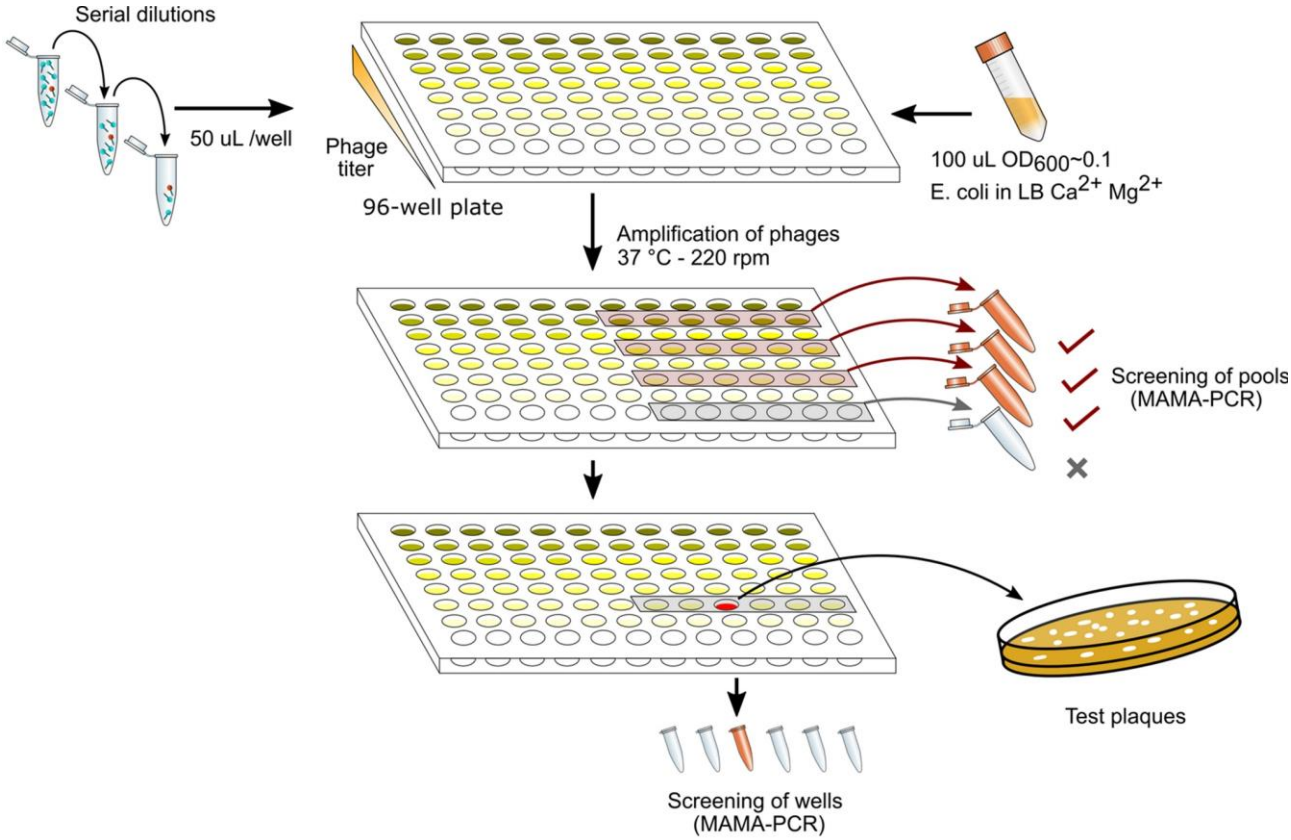
Construction	Primer	Sequence
pAC9	2225	tttttgggctaacaggaggaattaaccATGGATAAGAAATACTCAATAGGCTTAGATATCGGCAC
	6331	gaaaatcttctctcatccgccaaaacaGCCTCAGTCACCTCCTAGCTGACTCAAATCAATG
	2225c	CATGGTTAATTCCTCCTGTTAGCCCAAAAA
	6331c	GGCTGTTTTGGCGGATGAGAGAAGATTTT
	Cas9F	gccttttacggttcctggcACGAAGCAGGGATTCTGCAA
	Cas9R	cagctcactcaaaggcggtaAGGGTTATTGTCTCATGAGCGG
	psgF	TACCGCCTTTGAGTGAGCTG
	psgR	GCCAGGAACCGTAAAAAGGC
	QCpBADF	GTGTGGGGTCACCCCATGCGAG
	QCpBADR	CTCGCATGGGGTGACCCCACAC
	ACYCoriF	GTAAAAAGGCCGCGTTGCTGGCGTTGCGCTAGCGGAGTGTATACTGGC
	ACYCoriR	CCGTAGAAAAGATCAAAGGATCTTCACAACTTATATCGTATGGGGCTGACTTCAGG
	BC9GF	GAAGATCCTTTGATCTTTTCTACGGGGTCTGA
	BC9GR	AACGCCAGCAACGCGGC
	gRNAseq	TCGGTCTTGACAAAAAGAACCGGG
pUC19 [b]	pUCF	[Phos]-aaaaggtctcattttCTGGCCGTCGTTTTACAACG
	pUCR	aaaaggtctcaggggTCATGGTCATAGCTGTTTCCTGT
T5 *amA1*	3087	aaaaggtctcaGTAGACGGttgAGCAATGGGAAGTAAAACAAGGTAAACGGAATTCAGGAAAACAGACAGTAACTCA
SS71	3127	aaaaggtctcaCTACTAAACAATCCCCGCCGCTTCTGGGCGCTCATCGTTTGAATTCAGGAAACCCGTTTTTTCTG
	3106	CCCATTGCTCAACCGTCTACT
	gRNA-A1S	tagtGGCGGGGATTGTTTCTAGCA
	gRNA-A1A	aaacTGCTAGAAACAATCCCCGCC
T5 *amA1*	GGA1TF	aaaaggtctcaccccAGACCTTCAAGATTCAGCGCG
T15	GGA1TR	aaaaggtctcaaaaaCACTGGAAAGTGCAAAATTGGAAAC
	QCA11	ATATCGTCGCAAACTTCCAActtctAGGCGCGCAGACGTTTACCGTA
	QCA12	TACGGTAAACGTCTGCGCGCCTagaagTTGGAAGTTTGCGACGATAT
	2772	AGACCTTCAAGATTCAGCGCG
	A1Scr	CGGTAAACGTCTGCGCGCCTagaag
	gRNAA1TS	tagtACGTCTGCGCGCTACTAAAT
	gRNAA1TA	aaacATTTAGTAGCGCGCAGACGT
T5 *amA2*	gg pUC A2 F	aaaaggtctcaccccCGGCGGCAATAAAACAAATC
S37	gg pUC A2 R	aaaaggtctcaaaaaCAGGTAATGCGGAACAATCC
	QCamA2F	ATCTCTTTTAAACCGTCctaATTTGCGAAATCTAAACC
	QCamA2R	GGTTTAGATTTCGCAAATtagGACGGTTTAAAAGAGAT
	4050.2	GAAAACGGTTTAGATTTCGCAAAgtag
T5 *lacZ*α	05F	aaaaggtctcaaaaaACCGCAAAATTCGCTTGGAA
	05R	aaaaggtctcaccccCTAACCAGCAAATCAGCGCC
	pUC05-F	aaaaggtctcaTAGGCGTTATAATTTGCAACATTAATTTAAAGGCATAAGG
	pUC05-R	aaaaggtctcaATTCAAGGCGTTATATTTGGCAATATTGCCAATAAC
	LacZalpha-ATG	aaaaggtctcacctaATGACCATGATTACGCCAAGCTTGC
	LacZalpha-Stop	aaaaggtctcagaatCTATGCGGCATCAGAGCAGATTGTAC
	QC-LacZ-Eco1	GGATCCCCGGGTACCGAGCTCAAACTCACTGGCCGTCGTTTTACAA
	QC-LacZ-Eco2	TTGTAAAACGACGGCCAGTGAGTTTGAGCTCGGTACCCGGGGATCC
	QCLacZSD1	GCAAGCTTGGCGTAATCATGGTCATgtGattTccTccTTTGCAACATTAATTTAAAGGCATA
	QCLacZSD2	TATGCCTTTAAATTAATGTTGCAAAggAggAaatCacATGACCATGATTACGCCAAGCTTGC
	A2ATG	aaaaggtctcatcacATGACTAACGCTAAAACCGCAAAATTC
T5 PNmC	T5.151F	aaaaggtctcaccccTGACCATTCATGCTCTTAGCGA
	T5.151R	aaaaggtctcaaaaaTGCTTGATGGTGGTAATAATGCG
	mCF	aaaaggtctcaattaAGACTTGTACAGCTCGTCCAT
	mCR	aaaaggtctcataatGGAGGCTAATAAATGACACAAGC
	104870	GGATAACAGTTACGTGTGTGGC
T5amA1 SS84	3087	aaaaGGTCTCAGTAGACGGTTGAGCAATGGGAAGTAAAACAAGGTAAACGGAATTCAGGAAAACAGACAGTAACTCA
	3127	aaaaGGTCTCACTACTAAACAATCCCCGCCGCTTCTGGGCGCTCATCGTTTGAATTCAGGAAACCCGTTTTTTCTG
	3106	CCCATTGCTCAACCGTCTACT
	pFFseq	GTCGATTATAACCAAGAGAATGGAAC

344 Upper-case letters match with the [a] T5 genome or [b] plasmid pUC19.

345



Supplementary Figure 1. Construction of plasmid pAC9. The plasmid pBAD24 was amplified with primers 2225c/6331c to clone the gene *cas9* taken from the plasmid pCas9 with primers 2225/6331. pBADCas9, amplified with Cas9F/R, was fused with the sgRNA coding fragment of psgRNAcos. psgRNAcos was amplified with primers psgF/R. The origin of replication pBR322, from pBC9G, was replaced by the origin of p15A from pCas9 to obtain a mid-copy number plasmid: p15A was amplified with primers ACYCoriF/R, while the plasmid pBC9G was amplified with primers BC9GF/R.



Supplementary Figure 2: Phage mutant enrichment protocol by Dilution-Amplification-Screening (DAS).
The phage lysate obtained was filtered with 0.45-um diameter nitrocellulose filters. The lysates were first diluted: we made 10-fold serial dilutions to reach ca. 10-100 PFU/mL (usually up to a dilution factor of 10^8 to 10^{10}), of 10 mL volume each, with LB supplemented with $CaCl_2$ (1mM) and $MgSO_4$ (1mM). Fifty μL of each dilution were distributed in twelve wells of a 96-well sterile flat bottom polystyrene plate (Corning), one row per dilution factor. To amplify the phages in each well, one-hundred-μL aliquots of bacterial culture in exponential phase were added. The plate was incubated for 3 hours at 37 °C 220 rpm shaking. For the first screening by MAMA-PCR, we pooled 20 μL from six wells in the same row. Positive pools corresponding to the highest phage dilutions were selected, and the six individual wells were tested individually for the second PCR screening. Finally, 50 μL taken from a PCR- positive well were plated on a double agar to test 10-20 lysis plaques individually

ACKNOWLEDGEMENTS

We thank D. Bikard for his kind gift of plasmid psgRNAcos.

FUNDING

This work was supported by the Paraguayan Fellowship program BECAL-FRANCE01 to L.R.C., and the Agence Nationale pour la Recherche [ANR-17-CE11-0038 to P.B.].

CONFLICT OF INTEREST

The authors declare to have no conflict of interest.

REFERENCES

1. Drake,J.W. and Baltz,R.H. (1976) The Biochemistry of Mutagenesis. *Annual Review of Biochemistry*, **45**, 11–37.
2. Völker,T.A. and Showe,M.K. (1980) Induction of mutations in specific genes of bacteriophage T4 using cloned restriction fragments and marker rescue. *Molecular and General Genetics MGG*, **177**, 447–452.
3. Marinelli,L.J., Piuri,M., Swigoňová,Z., Balachandran,A., Oldfield,L.M., Kessel,J.C. van and Hatfull,G.F. (2008) BRED: A Simple and Powerful Tool for Constructing Mutant and Recombinant Bacteriophage Genomes. *PLOS ONE*, **3**, e3957.
4. Kiro,R., Shitrit,D. and Qimron,U. (2014) Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biology*, **11**, 42–44.
5. Tao,P., Wu,X., Tang,W.-C., Zhu,J. and Rao,V. (2017) Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. *ACS Synthetic Biology*, **6**, 1952–1961.
6. Hupfeld,M., Trasanidou,D., Ramazzini,L., Klumpp,J., Loessner,M.J. and Kilcher,S. (2018) A functional type II-A CRISPR–Cas system from Listeria enables efficient genome editing of large non-integrating bacteriophage. *Nucleic Acids Research*, **46**, 6920–6933.
7. Martel,B. and Moineau,S. (2014) CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. *Nucleic Acids Research*, **42**, 9504–9513.
8. Marino,N.D., Zhang,J.Y., Borges,A.L., Sousa,A.A., Leon,L.M., Rauch,B.J., Walton,R.T., Berry,J.D., Joung,J.K., Kleinstiver,B.P., *et al.* (2018) Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, **362**, 240–242.
9. Malone,L.M., Warring,S.L., Jackson,S.A., Warnecke,C., Gardner,P.P., Gumy,L.F. and Fineran,P.C. (2020) A jumbo phage that forms a nucleus-like structure evades CRISPR–Cas DNA targeting but is vulnerable to type III RNA-based immunity. *Nature Microbiology*, **5**, 48–55.
10. 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. (2020) A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. *Nature*, **577**, 244–248.
11. Ando,H., Lemire,S., Pires,D.P. and Lu,T.K. (2015) Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Systems*, **1**, 187–196.
12. Kilcher,S., Studer,P., Muessner,C., Klumpp,J. and Loessner,M.J. (2018) Cross-genus rebooting of custom-made, synthetic bacteriophage genomes in L-form bacteria. *Proceedings of the National Academy of Sciences*, **115**, 567–572.
13. Rustad,M., Eastlund,A., Jardine,P. and Noireaux,V. (2018) Cell-free TXTL synthesis of infectious bacteriophage T4 in a single test tube reaction. *Synthetic Biology*, **3**.
14. Hendrickson,H.E. and McCorquodale,D.J. (1971) Genetic and Physiological Studies of Bacteriophage T5 I. An Expanded Genetic Map of T5. *Journal of Virology*, **7**, 612–618.
15. Lanni,Y.T., Lanni,F. and Tevethia,M.J. (1966) Bacteriophage T5 Chromosome Fractionation: Genetic Specificity of a DNA Fragment. *Science*, **152**, 208–210.
16. Klimuk,E., Mekler,V., Lavysh,D., Serebryakova,M., Akulenko,N. and Severinov,K. (2020) Novel Escherichia coli RNA Polymerase Binding Protein Encoded by Bacteriophage T5. *Viruses*, **12**, 807.

17. Vernhes,E., Renouard,M., Gilquin,B., Cuniasse,P., Durand,D., England,P., Hoos,S., Huet,A., Conway,J.F., Glukhov,A., *et al.* (2017) High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid. *Scientific Reports*, **7**, 1–14.
18. Strotskaya,A., Savitskaya,E., Metlitskaya,A., Morozova,N., Datsenko,K.A., Semenova,E. and Severinov,K. (2017) The action of Escherichia coli CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Research*, **45**, 1946–1957.
19. Werner,S., Engler,C., Weber,E., Gruetzner,R. and Marillonnet,S. (2012) Fast track assembly of multigene constructs using Golden Gate cloning and the MoClo system. *Bioengineered*, **3**, 38–43.
20. Deng,W.P. and Nickoloff,J.A. (1992) Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Analytical Biochemistry*, **200**, 81–88.
21. Cha,R.S., Zarbl,H., Keohavong,P. and Thilly,W.G. (1992) Mismatch amplification mutation assay (MAMA): application to the c-H-ras gene. *Genome Research*, **2**, 14–20.
22. Nielsen,H.J., Ottesen,J.R., Youngren,B., Austin,S.J. and Hansen,F.G. (2006) The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. *Molecular Microbiology*, **62**, 331–338.
23. Lanni,Y.T. (1968) First-step-transfer deoxyribonucleic acid of bacteriophage T5. *Bacteriological Reviews*, **32**, 227–242.
24. Simon,A.J., Ellington,A.D. and Finkelstein,I.J. (2019) Retrons and their applications in genome engineering. *Nucleic Acids Research*, **47**, 11007–11019.
25. Farzadfard,F. and Lu,T.K. (2014) Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science*, **346**.
26. Simon,A.J., Morrow,B.R. and Ellington,A.D. (2018) Retroelement-Based Genome Editing and Evolution. *ACS Synthetic Biology*, **7**, 2600–2611.
27. Yehl,K., Lemire,S., Yang,A.C., Ando,H., Mimee,M., Torres,M.D.T., Fuente-Nunez,C. de la and Lu,T.K. (2019) Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell*, **179**, 459-469.e9.
28. Pires,D.P., Cleto,S., Sillankorva,S., Azeredo,J. and Lu,T.K. (2016) Genetically Engineered Phages: a Review of Advances over the Last Decade. *Microbiology and Molecular Biology Reviews*, **80**, 523–543.
29. Jiang,W., Bikard,D., Cox,D., Zhang,F. and Marraffini,L.A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, **31**, 233–239.
30. Guzman,L.M., Belin,D., Carson,M.J. and Beckwith,J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, **177**, 4121–4130.
31. Cui,L., Vigouroux,A., Rousset,F., Varet,H., Khanna,V. and Bikard,D. (2018) A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nature Communications*, **9**, 1–10.
32. Appleyard,R.K., McGregor,J.F. and Baird,K.M. (1956) Mutation to extended host range and the occurrence of phenotypic mixing in the temperate coliphage lambda. *Virology*, **2**, 565–574.
33. Lanni,Y.T. (1958) Lysis inhibition with a mutant of bacteriophage T5. *Virology*, **5**, 481–501.

8. References

Abedon, S.T., Herschler, T.D., and Stopar, D. (2001). Bacteriophage Latent-Period Evolution as a Response to Resource Availability. Appl. Environ. Microbiol. *67*, 4233–4241.

Abedon, S.T., Hyman, P., and Thomas, C. (2003). Experimental Examination of Bacteriophage Latent-Period Evolution as a Response to Bacterial Availability. Appl. Environ. Microbiol. *69*, 7499–7506.

Abelson, J., and Thomas, C.A. (1966). The anatomy of the T5 bacteriophage DNA molecule. J. Mol. Biol. *18*, 262–288.

Adams, M.H. (1949). The Calcium Requirement of Coliphage T5. J. Immunol. *62*, 505–516.

Alawneh, A.M., Qi, D., Yonesaki, T., and Otsuka, Y. (2016). An ADP-ribosyltransferase Alt of bacteriophage T4 negatively regulates the Escherichia coli MazF toxin of a toxin–antitoxin module. Mol. Microbiol. *99*, 188–198.

Appleyard, R.K., McGregor, J.F., and Baird, K.M. (1956). Mutation to extended host range and the occurrence of phenotypic mixing in the temperate coliphage lambda. Virology *2*, 565–574.

Arnaud, C.-A., Effantin, G., Vivès, C., Engilberge, S., Bacia, M., Boulanger, P., Girard, E., Schoehn, G., and Breyton, C. (2017). Bacteriophage T5 tail tube structure suggests a trigger mechanism for Siphoviridae DNA ejection. Nat. Commun. *8*.

Bandyopadhyay, P.K., Studier, F.W., Hamilton, D.L., and Yuan, R. (1985). Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. J. Mol. Biol. *182*, 567–578.

Beckman, L.D., Hoffman, M.S., and McCorquodale, D.J. (1971). Pre-early proteins of bacteriophage T5: Structure and function. J. Mol. Biol. *62*, 551–564.

Belley, A., Callejo, M., Arhin, F., Dehbi, M., Fadhil, I., Liu, J., McKay, G., Srikumar, R., Bauda, P., Ha, N., et al. (2006). Competition of bacteriophage polypeptides with native replicase proteins for binding to the DNA sliding clamp reveals a novel mechanism for DNA replication arrest in Staphylococcus aureus. Mol. Microbiol. *62*, 1132–1143.

Bergh, Ø., BØrsheim, K.Y., Bratbak, G., and Heldal, M. (1989). High abundance of viruses found in aquatic environments. Nature *340*, 467–468.

Blaisdell, P.W., and Warner, H.R. (1986). Cell-free transcription and translation of isolated restriction fragments localize bacteriophage T5 pre-early genes. J. Virol. *57*, 759–764.

Blinov, V.M., Koonin, E.V., Gorbalenya, A.E., Kaliman, A.V., and Kryukov, V.M. (1989). Two early genes of bacteriophage T5 encode proteins containing an NTP-binding sequence motif and probably involved in DNA replication, recombination and repair. FEBS Lett. *252*, 47–52.

Bonhivers, M., and Letellier, L. (1995). Calcium controls phage T5 infection at the level of the Escherichia coli cytoplasmic membrane. FEBS Lett. *374*, 169–173.

Bouet, J.-Y., Krisch, H.M., and Louarn, J.-M. (1998). Ndd, the Bacteriophage T4 Protein that disrupts the Escherichia coli Nucleoid, has a DNA Binding Activity. J. Bacteriol. *180*, 5227–5230.

Boulanger, P., Jacquot, P., Plançon, L., Chami, M., Engel, A., Parquet, C., Herbeuval, C., and Letellier, L. (2008). Phage T5 Straight Tail Fiber Is a Multifunctional Protein Acting as a Tape Measure and Carrying Fusogenic and Muralytic Activities. J. Biol. Chem. *283*, 13556–13564.

Boyd, J.S.K. (1949). Development of Bacteriophage in Escherichia coli B. Nature *164*, 874–875.

Brunel, F., and Davison, J. (1979). Restriction insensitivity in bacteriophage T5: III. Characterization of EcoRI-sensitive mutants by restriction analysis. J. Mol. Biol. *128*, 527–543.

Brunel, F., Davison, J., and Merchez, M. (1979). Cloning of bacteriophage T5 DNA fragments in plasmid pBR322 and bacteriophage λ gtwes. Gene *8*, 53–68.

Bujard, H. (1969). Location of single-strand interruption in the DNA of Bacteriophage T5. Proc. Natl. Acad. Sci. *62*, 1167.

Chernov, A.P., and Kaliman, A.V. (1987). [Various characteristics of the anti-restriction mechanism in bacteriophage T5]. Mol. Genet. Mikrobiol. Virusol. 14–19.

Chetverin, A.B., Ugarov, V.I., and Chetverina, H.V. (2019). Unsolved Puzzles of Qβ Replicase. Mol. Biol. *53*, 791–801.

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. Nat. Commun. *9*, 1–10.

Dame, R.T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Mol. Microbiol. *56*, 858–870.

Davidson, A.R., Lu, W.-T., Stanley, S.Y., Wang, J., Mejdani, M., Trost, C.N., Hicks, B.T., Lee, J., and Sontheimer, E.J. (2020). Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. Annu. Rev. Biochem. *89*, 309–332.

Davison, J. (2015). Pre-early functions of bacteriophage T5 and its relatives. Bacteriophage *5*, e1086500.

Davison, J., and Brunel, F. (1979a). Restriction insensitivity in bacteriophage T5 I. Genetic characterization of mutants sensitive to EcoRI restriction. J. Virol. *29*, 11–16.

Davison, J., and Brunel, F. (1979b). Restriction insensitivity in bacteriophage T5. II. Lack of EcoRI modification in T5+ and T5ris mutants. J. Virol. *29*, 17–20.

Davison, J., and Lafontaine, D. (1984). Direction of transcription in bacteriophage T5 first-step transfer DNA. J. Virol. *50*, 629–631.

Dendooven, T., and Lavigne, R. (2019). Dip-a-Dee-Doo-Dah: Bacteriophage-Mediated Rescoring of a Harmoniously Orchestrated RNA Metabolism. Annu. Rev. Virol. *6*, 199–213.

Deng, W.P., and Nickoloff, J.A. (1992). Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. Anal. Biochem. *200*, 81–88.

Depping, R., Lohaus, C., Meyer, H.E., and Rüger, W. (2005). The mono-ADP-ribosyltransferases Alt and ModB of bacteriophage T4: Target proteins identified. Biochem. Biophys. Res. Commun. *335*, 1217–1223.

Dillingham, M.S., and Kowalczykowski, S.C. (2008). RecBCD Enzyme and the Repair of Double-Stranded DNA Breaks. Microbiol. Mol. Biol. Rev. *72*, 642–671.

Dillon, S.C., and Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat. Rev. Microbiol. *8*, 185–195.

Dong, L., Guan, X., Li, N., Zhang, F., Zhu, Y., Ren, K., Yu, L., Zhou, F., Han, Z., Gao, N., et al. (2019). An anti-CRISPR protein disables type V Cas12a by acetylation. Nat. Struct. Mol. Biol. *26*, 308–314.

Dressman, H.K., and Drake, J.W. (1999). Lysis and Lysis Inhibition in Bacteriophage T4:rV Mutations Reside in the Holin t Gene. J. Bacteriol. *181*, 4391–4396.

Duckworth, D.H., and Dunn, G.B. (1976). Membrane protein biosynthesis in T5 bacteriophage-infected Escherichia coli. Arch. Biochem. Biophys. *172*, 319–328.

Ducret, A., Quardokus, E.M., and Brun, Y.V. (2016). MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat. Microbiol. *1*, 1–7.

Dufour, N., Delattre, R., Ricard, J.-D., and Debarbieux, L. (2017). The Lysis of Pathogenic Escherichia coli by Bacteriophages Releases Less Endotoxin Than by β-Lactams. Clin. Infect. Dis. *64*, 1582–1588.

Dunn, J.J., and Studier, F.W. (1973). T7 Early RNAs are Generated by Site-Specific Cleavages. Proc. Natl. Acad. Sci. U. S. A. *70*, 1559–1563.

Dyrløv Bendtsen, J., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved Prediction of Signal Peptides: SignalP 3.0. J. Mol. Biol. *340*, 783–795.

Effantin, G., Boulanger, P., Neumann, E., Letellier, L., and Conway, J.F. (2006). Bacteriophage T5 Structure Reveals Similarities with HK97 and T4 Suggesting Evolutionary Relationships. J. Mol. Biol. *361*, 993–1002.

Eisenberg, S., and Ascarelli, R. (1981). The A* protein of phi X174 is an inhibitor of DNA replication. Nucleic Acids Res. *9*, 1991–2002.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238.

Everett, R.D. (1981). DNA Replication of Bacteriophage T5. 3. Studies on the Structure of Concatemerie T5 DNA. J. Gen. Virol. 14.

Farzadfard, F., and Lu, T.K. (2014a). Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science *346*.

Farzadfard, F., and Lu, T.K. (2014b). Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. Science *346*, 1256272.

Fox, J.W., Barish, A., Snyder, C.E., and Benzinger, R. (1982). Amino terminal sequence of the bacteriophage T5-coded gene A2 protein. Biochem. Biophys. Res. Commun. *106*, 265–269.

Fujisawa, H., and Morita, M. (1997). Phage DNA packaging. Genes Cells *2*, 537–545.

Garneau, J.R., Depardieu, F., Fortier, L.-C., Bikard, D., and Monot, M. (2017). PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. Sci. Rep. *7*, 8292.

Gentz, R., and Bujard, H. (1985). Promoters recognized by Escherichia coli RNA polymerase selected by function: highly efficient promoters from bacteriophage T5. J. Bacteriol. *164*, 70–77.

Goff, C.G., and Setzer, J. (1980). ADP ribosylation of Escherichia coli RNA polymerase is nonessential for bacteriophage T4 development. J. Virol. *33*, 547–549.

Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., and Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. EMBO J. *34*, 169–183.

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J. Bacteriol. *177*, 4121–4130.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. *17*, 148.

Hansen, F.G., and Atlung, T. (2018). The DnaA Tale. Front. Microbiol. *9*.

Hartmann, R.K., Heinrich, J., Schlegl, J., and Schuster, H. (1995). Precursor of C4 antisense RNA of bacteriophages P1 and P7 is a substrate for RNase P of Escherichia coli. Proc. Natl. Acad. Sci. *92*, 5822–5826.

Häuser, R., Blasche, S., Dokland, T., Haggård-Ljungquist, E., von Brunn, A., Salas, M., Casjens, S., Molineux, I., and Uetz, P. (2012). Bacteriophage Protein–Protein Interactions. In Advances in Virus Research, (Elsevier), pp. 219–298.

Hayward, S.D., and Smith, M.G. (1973). The chromosome of bacteriophage T5: III. Patterns of transcription from the single-stranded DNA fragments. J. Mol. Biol. *80*, 345–359.

Heineman, R.H., and Bull, J.J. (2007). Testing Optimality with Experimental Evolution: Lysis Time in a Bacteriophage. Evolution *61*, 1695–1709.

Heller, K., and Braun, V. (1979). Accelerated Adsorption of Bacteriophage T5 to Escherichia coli F, Resulting from Reversible Tail Fiber-Lipopolysaccharide Binding. J. Bacteriol. *139*, 32–38.

Heller, K., and Braun, V. (1982). Polymannose O-antigens of Escherichia coli, the binding sites for the reversible adsorption of bacteriophage T5+ via the L-shaped tail fibers. J. Virol. *41*, 222–227.

Hendrickson, H.E., and Bujard, H. (1973). Structure and Function of the Genome of Coliphage T5. Eur. J. Biochem. *33*, 529–534.

Hendrickson, H.E., and McCorquodale, D.J. (1971a). Genetic and Physiological Studies of Bacteriophage T5 I. An Expanded Genetic Map of T5. J. Virol. *7*, 612–618.

Hendrickson, H.E., and McCorquodale, D.J. (1971b). Genetic and physiological studies of bacteriophage T5 2. The relationship between phage DNA synthesis and protein synthesis in T5-infected cells. Biochem. Biophys. Res. Commun. *43*, 735–740.

d'Herelle, F. (1917). On an invisible microbe antagonistic to dysentery bacilli. Note by M. F. d'Herelle, presented by M. Roux. Comptes Rendus Academie des Sciences 1917; 165:373–5. Bacteriophage *1*, 3–5.

Heusterspreute, M., Ha-Thi, V., Tournis-Gamble, S., and Davison, J. (1987). The first-step transfer-DNA injection-stop signal of bacteriophage T5. Gene *52*, 155–164.

Hilliard, J.J., Simon, L.D., Van Melderen, L., and Maurizi, M.R. (1998). PinA Inhibits ATP Hydrolysis and Energy-dependent Protein Degradation by Lon Protease. J. Biol. Chem. *273*, 524–527.

Ho, C.-H., Wang, H.-C., Ko, T.-P., Chang, Y.-C., and Wang, A.H.-J. (2014). The T4 Phage DNA Mimic Protein Arn Inhibits the DNA Binding Activity of the Bacterial Histone-like Protein H-NS. J. Biol. Chem. *289*, 27046–27054.

Hong, Y., and Black, L.W. (1993). An expression-packaging-processing vector which selects and maintains 7-kb DNA inserts in the blue T4 phage genome. Gene *136*, 193–198.

Howard-Varona, C., Hargreaves, K.R., Abedon, S.T., and Sullivan, M.B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. ISME J. *11*, 1511–1520.

Hung, J.-H., and Weng, Z. (2017). Mapping Short Sequence Reads to a Reference Genome. Cold Spring Harb. Protoc. *2017*, pdb.prot093161.

Hupfeld, M., Trasanidou, D., Ramazzini, L., Klumpp, J., Loessner, M.J., and Kilcher, S. (2018). A functional type II-A CRISPR–Cas system from Listeria enables efficient genome editing of large non-integrating bacteriophage. Nucleic Acids Res. *46*, 6920–6933.

Hyman, P., and Abedon, S.T. (2009). Practical Methods for Determining Phage Growth Parameters. In Bacteriophages, M.R.J. Clokie, and A.M. Kropinski, eds. (Totowa, NJ: Humana Press), pp. 175–202.

Ishihama, A. (2000). Functional Modulation of Escherichia Coli RNA Polymerase. Annu. Rev. Microbiol. *54*, 499–518.

Jacobus, A.P., and Gross, J. (2015). Optimal Cloning of PCR Fragments by Homologous Recombination in Escherichia coli. PLOS ONE *10*, e0119221.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233–239.

Kaczanowska, M., and Rydén-Aulin, M. (2007). Ribosome Biogenesis and the Translation Process in Escherichia coli. Microbiol. Mol. Biol. Rev. *71*, 477–494.

Kashlev, M., Nudler, E., Goldfarb, A., White, T., and Kutter, E. (1993). Bacteriophage T4 Alc protein: A transcription termination factor sensing local modification of DNA. Cell *75*, 147–154.

Kiro, R., Molshanski-Mor, S., Yosef, I., Milam, S.L., Erickson, H.P., and Qimron, U. (2013). Gene product 0.4 increases bacteriophage T7 competitiveness by inhibiting host cell division. Proc. Natl. Acad. Sci. *110*, 19549–19554.

Kiro, R., Shitrit, D., and Qimron, U. (2014). Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. RNA Biol. *11*, 42–44.

Klimuk, E., Akulenko, N., Makarova, K.S., Ceyssens, P.-J., Volchenkov, I., Lavigne, R., and Severinov, K. (2013). Host RNA polymerase inhibitors encoded by ϕKMV-like phages of pseudomonas. Virology *436*, 67–74.

Klimuk, E., Mekler, V., Lavysh, D., Serebryakova, M., Akulenko, N., and Severinov, K. (2020). Novel Escherichia coli RNA Polymerase Binding Protein Encoded by Bacteriophage T5. Viruses *12*, 807.

Konopacki, M., Grygorcewicz, B., Dołęgowska, B., Kordas, M., and Rakoczy, R. (2020). PhageScore: A simple method for comparative evaluation of bacteriophages lytic activity. Biochem. Eng. J. *161*, 107652.

Koonin, E.V. (2005). Orthologs, Paralogs, and Evolutionary Genomics. Annu. Rev. Genet. *39*, 309–338.

Kovalenko, A.O., Chernyshov, S.V., Kutyshenko, V.P., Molochkov, N.V., Prokhorov, D.A., Odinokova, I.V., and Mikoulinskaia, G.V. (2019). Investigation of the calcium-induced activation of the bacteriophage T5 peptidoglycan hydrolase promoting host cell lysis. Metallomics *11*, 799–809.

Kraemer, J.A., Erb, M.L., Waddling, C.A., Montabana, E.A., Zehr, E.A., Wang, H., Nguyen, K., Pham, D.S.L., Agard, D.A., and Pogliano, J. (2012). A Phage Tubulin Assembles Dynamic Filaments by an Atypical Mechanism to Center Viral DNA within the Host Cell. Cell *149*, 1488–1499.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L.L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. J. Mol. Biol. *305*, 567–580.

Kronheim, S., Daniel-Ivad, M., Duan, Z., Hwang, S., Wong, A.I., Mantel, I., Nodwell, J.R., and Maxwell, K.L. (2018). A chemical defence against phage infection. Nature *564*, 283–286.

Kropinski, A.M. (2018). Practical Advice on the One-Step Growth Curve. In Bacteriophages: Methods and Protocols, Volume 3, M.R.J. Clokie, A.M. Kropinski, and R. Lavigne, eds. (New York, NY: Springer New York), pp. 41–47.

Krüger, D.H., and Schroeder, C. (1981). Bacteriophage T3 and bacteriophage T7 virus-host cell interactions. Microbiol. Rev. *45*, 9–51.

Kulakauskas, S., Lubys, A., and Ehrlich, S.D. (1995). DNA restriction-modification systems mediate plasmid maintenance. J. Bacteriol. *177*, 3451–3454.

Kutter, E., Bryan, D., Ray, G., Brewster, E., Blasdel, B., and Guttman, B. (2018). From Host to Phage Metabolism: Hot Tales of Phage T4's Takeover of E. coli. Viruses *10*.

Lanni, Y.T. (1958). Lysis inhibition with a mutant of bacteriophage T5. Virology *5*, 481–501.

Lanni, Y.T. (1960). Invasion by bacteriophage T5: II. Dissociation of calcium-independent and calcium-dependent processes. Virology *10*, 514–529.

Lanni, Y.T. (1968). First-step-transfer deoxyribonucleic acid of bacteriophage T5. Bacteriol. Rev. *32*, 227–242.

Lanni, Y.T. (1969). Functions of two genes in the first-step-transfer DNA of bacteriophage T5. J. Mol. Biol. *44*, 173–183.

Lanni, Y.T., Lanni, F., and Tevethia, M.J. (1966). Bacteriophage T5 Chromosome Fractionation: Genetic Specificity of a DNA Fragment. Science *152*, 208–210.

Lemay, M.-L., Tremblay, D.M., and Moineau, S. (2017). Genome Engineering of Virulent Lactococcal Phages Using CRISPR-Cas9. ACS Synth. Biol. *6*, 1351–1358.

Lerner, L.K., Holzer, S., Kilkenny, M.L., Šviković, S., Murat, P., Schiavone, D., Eldridge, C.B., Bittleston, A., Maman, J.D., Branzei, D., et al. (2020). Timeless couples G-quadruplex detection with processing by DDX 11 helicase during DNA replication. EMBO J. *39*.

Liu, Q., and Richardson, C.C. (1993). Gene 5.5 protein of bacteriophage T7 inhibits the nucleoid protein H-NS of Escherichia coli. Proc. Natl. Acad. Sci. *90*, 1761–1765.

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N., Kwan, T., et al. (2004). Antimicrobial drug discovery through bacteriophage genomics. Nat. Biotechnol. *22*, 185–191.

Łoś, M., and Węgrzyn, G. (2012). Pseudolysogeny. In Advances in Virus Research, (Elsevier), pp. 339–349.

Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. *18*, 67–83.

Malone, L.M., Warring, S.L., Jackson, S.A., Warnecke, C., Gardner, P.P., Gumy, L.F., and Fineran, P.C. (2020). A jumbo phage that forms a nucleus-like structure evades CRISPR–Cas DNA targeting but is vulnerable to type III RNA-based immunity. Nat. Microbiol. *5*, 48–55.

Marchand, I., Nicholson, A.W., and Dreyfus, M. (2008). Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase: T7 kinase protects T7 RNAP mRNAs against RNase E. Mol. Microbiol. *42*, 767–776.

Margalit, D.N., Romberg, L., Mets, R.B., Hebert, A.M., Mitchison, T.J., Kirschner, M.W., and RayChaudhuri, D. (2004). Targeting cell division: Small-molecule inhibitors of FtsZ GTPase perturb cytokinetic ring assembly and induce bacterial lethality. Proc. Natl. Acad. Sci. *101*, 11821–11826.

Margolin, W. (2009). Sculpting the Bacterial Cell. Curr. Biol. *19*, R812–R822.

Martel, B., and Moineau, S. (2014). CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. Nucleic Acids Res. *42*, 9504–9513.

Martin, D.F., and Godson, G.N. (1975). Identification of a X174 coded protein involved in the shut-off of host dna replication. Biochem. Biophys. Res. Commun. *65*, 323–330.

Mattenberger, Y., Mattson, S., Métrailler, J., Silva, F., and Belin, D. (2011). 55.1, a gene of unknown function of phage T4, impacts on Escherichia coli folate metabolism and blocks DNA repair by the NER. Mol. Microbiol. *82*, 1406–1421.

Mattenberger, Y., Silva, F., and Belin, D. (2015). 55.2, a Phage T4 ORFan Gene, Encodes an Inhibitor of Escherichia coli Topoisomerase I and Increases Phage Fitness. PLOS ONE *10*, e0124309.

McCorquodale, D.J., and Buchanan, J.M. (1968). Patterns of Protein Synthesis in T5-infected Escherichia coli. J. Biol. Chem. *243*, 2550–2559.

McCorquodale, D.J., and Lanni, Y.T. (1970). Patterns of protein synthesis in Escherichia coli infected by amber mutants in the first-step-transfer DNA of T5. J. Mol. Biol. *48*, 133–143.

McCorquodale, D.J., and Warner, H.R. (1988). Bacteriophage T5 and Related Phages. In The Bacteriophages, R. Calendar, ed. (Boston, MA: Springer US), pp. 439–475.

McCorquodale, D.J., Shaw, A.R., Shaw, P.K., and Chinnadurai, G. (1977). Pre-early polypeptides of bacteriophages T5 and BF23. J. Virol. *22*, 480–488.

McCorquodale, D.J., Chen, C.W., Joseph, M.K., and Woychik, R. (1981). Modification of RNA polymerase from Escherichia coli by pre-early gene products of bacteriophage T5. J. Virol. *40*, 958–962.

McGinn, J., and Marraffini, L.A. (2019). Molecular mechanisms of CRISPR–Cas spacer acquisition. Nat. Rev. Microbiol. *17*, 7–12.

Mekler, V., Minakhin, L., Sheppard, C., Wigneshweraraj, S., and Severinov, K. (2011). Molecular Mechanism of Transcription Inhibition by Phage T7 gp2 Protein. J. Mol. Biol. *413*, 1016–1027.

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. (2020). A bacteriophage nucleus-like compartment shields DNA from CRISPR nucleases. Nature *577*, 244–248.

Mikoulinskaia, G.V., Odinokova, I.V., Zimin, A.A., Lysanskaya, V.Y., Feofanov, S.A., and Stepnaya, O.A. (2009). Identification and characterization of the metal ion-dependent l-alanoyl-d-glutamate peptidase encoded by bacteriophage T5. FEBS J. *276*, 7329–7342.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Rüger, W. (2003). Bacteriophage T4 Genome. Microbiol. Mol. Biol. Rev. *67*, 86.

Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., and Sorek, R. (2020a). Bacterial retrons function in anti-phage defense (Microbiology).

Millman, A., Melamed, S., Amitai, G., and Sorek, R. (2020b). Diversity and classification of cyclic-oligonucleotide-based anti-phage signalling systems. Nat. Microbiol.

Minakhin, L., and Severinov, K. (2005). Transcription regulation by bacteriophage T4 AsiA. Protein Expr. Purif. *41*, 1–8.

Molshanski-Mor, S., Yosef, I., Kiro, R., Edgar, R., Manor, M., Gershovits, M., Laserson, M., Pupko, T., and Qimron, U. (2014). Revealing bacterial targets of growth inhibitors encoded by bacteriophage T7. Proc. Natl. Acad. Sci. *111*, 18715–18720.

Mondigler, M., Holz, T., and Heller, K.J. (1996). Identification of the Receptor-Binding Regions of pb5 Proteins of Bacteriophages T5 and BF23. Virology *219*, 19–28.

Moyer, R.W., and Rothe, C.T. (1977). Role of the T5 gene D15 nuclease in the generation of nicked bacteriophage T5 DNA. J. Virol. *24*, 177–193.

Mozer, T.J., and Warner, H.R. (1977). Properties of Deoxynucleoside 5′-Monophosphatase Induced by Bacteriophage T5 After Infection of Escherichia coli. J. Virol. *24*, 7.

Mozer, T.J., Thompson, R.B., and Berget, S.M. (1977). Isolation and Characterization of a Bacteriophage T5 Mutant Deficient in Deoxynucleoside 5′-Monophosphatase Activity. J. Virol. *24*, 9.

Mruk, I., and Kobayashi, I. (2014). To be or not to be: regulation of restriction–modification systems and other toxin–antitoxin systems. Nucleic Acids Res. *42*, 70–86.

Nechaev, S., and Severinov, K. (2003). Bacteriophage-Induced Modifications of Host RNA Polymerase. Annu. Rev. Microbiol. *57*, 301–322.

Nechaev, S., Yuzenkova, Y., Niedziela-Majka, A., Heyduk, T., and Severinov, K. (2002). A Novel Bacteriophage-encoded RNA Polymerase Binding Protein Inhibits Transcription Initiation and Abolishes Transcription Termination by Host RNA Polymerase. J. Mol. Biol. *320*, 11–22.

Nielsen, H.J., Ottesen, J.R., Youngren, B., Austin, S.J., and Hansen, F.G. (2006). The Escherichia coli chromosome is organized with the left and right chromosome arms in separate cell halves. Mol. Microbiol. *62*, 331–338.

Nonejuie, P., Burkart, M., Pogliano, K., and Pogliano, J. (2013). Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules. Proc. Natl. Acad. Sci. *110*, 16169–16174.

Nudler, E., and Gottesman, M.E. (2002). Transcription termination and anti-termination in E. coli. Genes Cells *7*, 755–768.

O'Donnell, M. (2006). Replisome Architecture and Dynamics in *Escherichia coli*. J. Biol. Chem. *281*, 10653–10656.

Ofir, G., Melamed, S., Sberro, H., Mukamel, Z., Silverman, S., Yaakov, G., Doron, S., and Sorek, R. (2018). DISARM is a widespread bacterial defence system with broad anti-phage activities. Nat. Microbiol. *3*, 90–98.

Otsuka, Y., and Yonesaki, T. (2012). Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. Mol. Microbiol. *83*, 669–681.

Page, R., and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. Nat. Chem. Biol. *12*, 208–214.

Qi, D., Alawneh, A.M., Yonesaki, T., and Otsuka, Y. (2015). Rapid Degradation of Host mRNAs by Stimulation of RNase E Activity by Srd of Bacteriophage T4. Genetics *201*, 977–987.

Rifat, D., Wright, N.T., Varney, K.M., Weber, D.J., and Black, L.W. (2008). Restriction Endonuclease Inhibitor IPI* of Bacteriophage T4: A Novel Structure for a Dedicated Target. J. Mol. Biol. *375*, 720–734.

Roberts, R.J. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. *31*, 1805–1812.

Rosenkranz, H.S. (1973). RNA in Coliphage T5. Nature *242*, 327–329.

Rostøl, J.T., and Marraffini, L. (2019). (Ph)ighting Phages: How Bacteria Resist Their Parasites. Cell Host Microbe *25*, 184–194.

Roucourt, B., and Lavigne, R. (2009). The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. Environ. Microbiol. *11*, 2789–2805.

Sakaki, Y. (1974). Inactivation of the ATP-dependent DNase of Escherichia coli After Infection with Double-Stranded DNA Phages. J. Virol. *14*, 1611–1612.

Samson, J.E., Magadán, A.H., Sabri, M., and Moineau, S. (2013). Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. *11*, 675–687.

Sanders, G.M., Kassavetis, G.A., and Geiduschek, E.P. (1995). Rules governing the efficiency and polarity of loading a tracking clamp protein onto DNA: determinants of enhancement in bacteriophage T4 late transcription. EMBO J. *14*, 3966–3976.

Sarkar, P., Switzer, A., Peters, C., Pogliano, J., and Wigneshweraraj, S. (2017). Phenotypic consequences of RNA polymerase dysregulation in Escherichia coli. Nucleic Acids Res. *45*, 11131–11143.

Scheible, P.P., and Rhoades, M. (1975). Heteroduplex Mapping of Heat-Resistant Deletion Mutants of Bacteriophage T5. J. Virol. *15*, 1276–1280.

Shajani, Z., Sykes, M.T., and Williamson, J.R. (2011). Assembly of Bacterial Ribosomes. Annu. Rev. Biochem. *80*, 501–526.

Sharma, U.K., and Chatterji, D. (2008). Differential Mechanisms of Binding of Anti-Sigma Factors Escherichia coli Rsd and Bacteriophage T4 AsiA to E. coli RNA Polymerase Lead to Diverse Physiological Consequences. J. Bacteriol. *190*, 3434–3443.

Shaw, A.R., and Davison, J. (1979). Polarized Injection of the Bacteriophage T5 Chromosome. J. Virol. *30*, 933–935.

Shen, B.W., Landthaler, M., Shub, D.A., and Stoddard, B.L. (2004). DNA Binding and Cleavage by the HNH Homing Endonuclease I-HmuI. J. Mol. Biol. *342*, 43–56.

Shi, H., Bratton, B.P., Gitai, Z., and Huang, K.C. (2018). How to Build a Bacterial Cell: MreB as the Foreman of E. coli Construction. Cell *172*, 1294–1305.

Simon, A.J., Morrow, B.R., and Ellington, A.D. (2018). Retroelement-Based Genome Editing and Evolution. ACS Synth. Biol. *7*, 2600–2611.

Simon, A.J., Ellington, A.D., and Finkelstein, I.J. (2019). Retrons and their applications in genome engineering. Nucleic Acids Res. *47*, 11007–11019.

Singer, B.S., Gold, L., Gauss, P., and Doherty, D.H. (1982). Determination of the amount of homology required for recombination in bacteriophage T4. Cell *31*, 25–33.

Sirbasku, D., and Buchanan, M. (1970a). Patterns of Ribonucleic Acid Synthesis in T5-infected Escherichia coli: II-Separation of High Molecular Weight Ribonucleic Acid species by Disc Electrophoresis on acrylamide gel columns. J. Biol. Chem. *245*, 15.

Sirbasku, D.A., and Buchanan, J.M. (1970b). Patterns of Ribonucleic Acid Synthesis in T5-infected Escherichia coli III. Separation of low molecular weight Ribonucleic acid species by Disc Electrophoresis on Acrylamide gel columns. J. Biol. Chem. *245*, 2693–2703.

Snyder, C.E. (1984). Bacteriophage T5 gene A2 protein alters the outer membrane of Escherichia coli. J. Bacteriol. *160*, 1191–1195.

Snyder, C.E. (1991). Amino acid sequence of the bacteriophage T5 gene A2 protein. Biochem. Biophys. Res. Commun. *177*, 1240–1246.

Snyder, C.E., and Benzinger, R.H. (1981). Second-step transfer of bacteriophage T5 DNA: purification and characterization of the T5 gene A2 protein. J. Virol. *40*, 248–257.

Storms, Z.J., Teel, M.R., Mercurio, K., and Sauvageau, D. (2020). The Virulence Index: A Metric for Quantitative Analysis of Phage Virulence. PHAGE *1*, 27–36.

Strotskaya, A., Savitskaya, E., Metlitskaya, A., Morozova, N., Datsenko, K.A., Semenova, E., and Severinov, K. (2017). The action of Escherichia coli CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. Nucleic Acids Res. *45*, 1946–1957.

Sulakvelidze, A., Alavidze, Z., and Morris, J.G. (2001). Bacteriophage Therapy. Antimicrob. Agents Chemother. *45*, 649–659.

Suttle, C.A. (2007). Marine viruses — major players in the global ecosystem. Nat. Rev. Microbiol. *5*, 801–812.

Sváb, D., Falgenhauer, L., Rohde, M., Szabó, J., Chakraborty, T., and Tóth, I. (2018). Identification and Characterization of T5-Like Bacteriophages Representing Two Novel Subgroups from Food Products. Front. Microbiol. *9*.

Szabo, C., and Moyer, R.W. (1975). Purification and properties of a bacteriophage T5-modified form of Escherichia coli RNA polymerase. J. Virol. *15*, 1042–1046.

Szabo, C., Dharmgrongartama, B., and Moyer, R.W. (1975). Regulation of transcription in bacteriophage T5-infected Escherichia coli. Biochemistry *14*, 989–997.

Tabib-Salazar, A., Liu, B., Shadrin, A., Burchell, L., Wang, Z., Wang, Z., Goren, M.G., Yosef, I., Qimron, U., Severinov, K., et al. (2017). Full shut-off of Escherichia coli RNA-polymerase by T7 phage requires a small phage-encoded DNA-binding protein. Nucleic Acids Res. *45*, 7697–7707.

Tao, P., Wu, X., Tang, W.-C., Zhu, J., and Rao, V. (2017). Engineering of Bacteriophage T4 Genome Using CRISPR-Cas9. ACS Synth. Biol. *6*, 1952–1961.

Tao, Z., Gao, P., Hoffman, D.W., and Liu, H. (2008). Domain C of Human Poly(ADP-ribose) Polymerase-1 Is Important for Enzyme Activity and Contains a Novel Zinc-Ribbon Motif †,‡. Biochemistry *47*, 5804–5813.

Tejada-Arranz, A., de Crécy-Lagard, V., and de Reuse, H. (2020). Bacterial RNA Degradosomes: Molecular Machines under Tight Control. Trends Biochem. Sci. *45*, 42–57.

Tiemann, B., Depping, R., and Ruger, W. (1999). Overexpression, purification, and partial characterization o f ADP-Ribosyltransferases ModA and ModB of Bacteriophage T4. Gene Expr. 10.

Twort, F.W. (1915). An Investigation on the nature of ultra-microscopic viruses. The Lancet *186*, 1241–1243.

Van den Bossche, A., Hardwick, S.W., Ceyssens, P.-J., Hendrix, H., Voet, M., Dendooven, T., Bandyra, K.J., De Maeyer, M., Aertsen, A., Noben, J.-P., et al. (2016). Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome. ELife *5*.

Vernhes, E., Renouard, M., Gilquin, B., Cuniasse, P., Durand, D., England, P., Hoos, S., Huet, A., Conway, J.F., Glukhov, A., et al. (2017). High affinity anchoring of the decoration protein pb10 onto the bacteriophage T5 capsid. Sci. Rep. *7*, 1–14.

Völker, T.A., and Showe, M.K. (1980). Induction of mutations in specific genes of bacteriophage T4 using cloned restriction fragments and marker rescue. Mol. Gen. Genet. MGG *177*, 447–452.

Wagemans, J., Delattre, A.-S., Uytterhoeven, B., De Smet, J., Cenens, W., Aertsen, A., Ceyssens, P.-J., and Lavigne, R. (2015). Antibacterial phage ORFans of Pseudomonas aeruginosa phage LUZ24 reveal a novel MvaT inhibiting protein. Front. Microbiol. *6*.

Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M., Edwardson, J.M., and Dryden, D.T.F. (2002). Structure of Ocr from Bacteriophage T7, a Protein that Mimics B-Form DNA. Mol. Cell *9*, 187–194.

Wan, X., Hendrix, H., Skurnik, M., and Lavigne, R. (2021). Phage-based target discovery and its exploitation towards novel antibacterial molecules. Curr. Opin. Biotechnol. *68*, 1–7.

Wang, L.K., and Shuman, S. (2002). Mutational analysis defines the 5′-kinase and 3′-phosphatase active sites of T4 polynucleotide kinase. Nucleic Acids Res. *30*, 1073–1080.

Wang, J., Jiang, Y., Vincent, M., Sun, Y., Yu, H., Wang, J., Bao, Q., Kong, H., and Hu, S. (2005). Complete genome sequence of bacteriophage T5. Virology *332*, 45–65.

Wang, J., Dai, W., Li, J., Xie, R., Dunstan, R.A., Stubenrauch, C., Zhang, Y., and Lithgow, T. (2020). PaCRISPR: a server for predicting and visualizing anti-CRISPR proteins. Nucleic Acids Res. *48*, W348–W357.

Warner, H.R., Drong, R.F., and Berget, S.M. (1975). Early Events After Infection of Escherichia coli by Bacteriophage T5. I. Induction of a 5′-Nucleotidase Activity and Excretion of Free Bases. J. Virol. *15*, 8.

Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLOS ONE *6*, e16765.

Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. FEMS Microbiol. Rev. *28*, 127–181.

White, C.L., and Gober, J.W. (2012). MreB: pilot or passenger of cell wall synthesis? Trends Microbiol. *20*, 74–79.

Willis, L., and Huang, K.C. (2017). Sizing up the bacterial cell cycle. Nat. Rev. Microbiol. *15*, 606–620.

Wommack, K.E., and Colwell, R.R. (2000). Virioplankton: Viruses in Aquatic Ecosystems. Microbiol. Mol. Biol. Rev. *64*, 69–114.

Xiong, X., Wu, G., Wei, Y., Liu, L., Zhang, Y., Su, R., Jiang, X., Li, M., Gao, H., Tian, X., et al. (2020). SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities. Nat. Microbiol. 1–12.

Yang, H., Ma, Y., Wang, Y., Yang, H., Shen, W., and Chen, X. (2014). Transcription regulation mechanisms of bacteriophages: Recent advances and future prospects. Bioengineered *5*, 300–304.

Yano, S.T., and Rothman-Denes, L.B. (2011). A phage-encoded inhibitor of Escherichia coli DNA replication targets the DNA polymerase clamp loader. Mol. Microbiol. *79*, 1325–1338.

Yehl, K., Lemire, S., Yang, A.C., Ando, H., Mimee, M., Torres, M.D.T., Fuente-Nunez, C. de la, and Lu, T.K. (2019). Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. Cell *179*, 459-469.e9.

Zangelmi, L. (2018). Etude structurale et fonctionnelle de la protéine A1 du bactériophage T5 : une DNase octamérique originale. These de doctorat. Université Paris-Saclay (ComUE).

Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G., and Inouye, M. (2003). MazF Cleaves Cellular mRNAs Specifically at ACA to Block Protein Synthesis in Escherichia coli. Mol. Cell *12*, 913–923.

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A.N., and Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J. Mol. Biol. *430*, 2237–2243.

Zivanovic, Y., Confalonieri, F., Ponchon, L., Lurz, R., Chami, M., Flayhan, A., Renouard, M., Huet, A., Decottignies, P., Davidson, A.R., et al. (2014). Insights into Bacteriophage T5 Structure from Analysis of Its Morphogenesis Genes and Protein Components. J. Virol. *88*, 1162–1174.

Saved: 9/2/2021 09:35:00

université PARIS-SACLAY

Ecole doctorale n°569
Innovation Thérapeutique :
Du fondamental à l'appliqué (ITFA)

Titre : Exploration fonctionnelle des gènes précoces du bactériophage T5 dans la prise de contrôle de l'hôte : vers des nouvelles stratégies antibactériennes ?

Mots clés : bactériophages, gènes précoces, ingénierie des génomes.

Résumé : Les bactériophages (phages) sont les virus qui infectent les bactéries. Au début de l'infection, ils inactivent les défenses bactériennes et détournent les fonctions cellulaires à leur profit pour produire de nouveaux virions. Les stratégies qu'ils utilisent pour prendre le contrôle de leur hôte ont été étudiées chez très peu de phages. Une limite à ces études est due à la diversité des gènes précoces exprimés lors des premières étapes de l'infection, qui le plus souvent n'ont pas d'homologues détectables ni de fonction connue. Or, ils représentent une source importante de nouveaux gènes dont les produits sont essentiels pour la neutralisation de l'hôte. La découverte de leur fonction est donc cruciale pour décrypter les mécanismes de prise de contrôle de l'hôte et pour identifier les fonctions bactériennes ciblées par les protéines des phages.

T5 est un phage virulent qui infecte *Escherichia coli* en utilisant un mode d'injection de son génome en deux étapes. Il transfère d'abord 8 % de son ADN et l'injection s'arrête. Ce fragment, appelé locus FST (*First Step Transfer*), code pour 17 protéines précoces qui inhibent les mécanismes de défense bactérienne, induisent la dégradation de l'ADN de l'hôte, stoppent la synthèse des protéines de l'hôte et redémarrent le transfert de l'ADN viral. Le génome complet est alors injecté, permettant la réplication de l'ADN et l'assemblage des particules virales, puis la lyse de la bactérie. Seuls trois gènes ont été étudiés dans le passé : *dmp*, gène non essentiel pour l'infection codant pour une phosphatase ; *A1* et *A2*, deux gènes nécessaires pour la deuxième étape de transfert de l'ADN. *A1* est également impliquée dans la dégradation de l'ADN de l'hôte. Les 14 autres gènes précoces sont des gènes dits "ORFphelins" dont les fonctions sont inconnues. Ce mécanisme original d'injection de l'ADN en deux étapes, associé au regroupement des gènes précoces dans le locus FST, fait de T5 un modèle privilégié pour étudier les premiers événements de l'infection par les phages.

Cette thèse porte sur l'étude fonctionnelle du locus FST du phage T5. Pour comprendre le rôle des gènes précoces de T5, nous avons d'abord développé des outils de génétique inverse. Une méthode alternative à l'utilisation du système CRISPR-Cas a été mise au point avec T5, permettant de modifier les génomes de phages virulents qui "résistent" aux nucléases CRISPR-Cas. Ces stratégies ont permis d'obtenir des mutants portant des délétions simples ou multiples des gènes précoces et d'évaluer leur impact sur l'infection. La délétion de certains gènes a entraîné des modifications de la production virale, de la latence et de la virulence des mutants testés. La délétion du gène *dmp* est celle qui altère le plus l'infection, de manière comparable aux délétions multiples. Un phage T5 avec un locus FST minimal a pu être construit, suggérant que seuls quelques gènes précoces - dont *A1* et *A2* - sont nécessaires pour l'infection. Enfin, pour explorer la fonction des gènes précoces de T5, nous avons étudié l'impact de leur expression sur la bactérie *E. coli*, en dehors du contexte infectieux. Les produits de certains gènes sont toxiques et altèrent la croissance bactérienne, produisant des perturbations de la morphologie cellulaire ou du nucléoïde bactérien, voire induisant une lyse cellulaire. Ces résultats suggèrent que certaines protéines précoces interfèrent avec la division cellulaire, le métabolisme de l'ADN ou l'intégrité des membranes. Il est important de noter que la capacité de A1 à déclencher la destruction massive du chromosome bactérien a été observée en l'absence d'infection.

Cette exploration des premières étapes de l'infection par T5 ouvre maintenant la voie vers la caractérisation fonctionnelle et structurale des protéines précoces qui « domestiquent » l'hôte au profit de la production virale et pourraient déboucher sur la conception de nouveaux agents antimicrobiens.

Title : Functional exploration of bacteriophage T5 pre-early genes in the host takeover: a route towards new antibacterial strategies?

Keywords : Bacteriophages, pre-early genes, genome engineering.

Abstract : Bacteriophages (phages) are viruses that infect bacteria. In the early stages of infection, they defeat bacterial defenses and hijack cellular functions to produce new viral particles. Their strategies to take over host cell resources remain poorly understood and have only been studied in a limited number of phages. Importantly, phage genes expressed early during infection are highly diverse from one phage to another and most have no assigned function yet. They represent a library of novel genes whose products are crucial for host neutralization; therefore, elucidating their function is essential for deciphering the mechanisms of host takeover and identifying host factors targeted by phage proteins.

T5 is a virulent phage that infects *Escherichia coli* in two steps. First, it delivers 8 % of its linear DNA and the genome transfer stops. This fragment, called the First Step Transfer (FST) locus, encodes 17 pre-early proteins that inhibit bacterial defense mechanisms, promote host DNA degradation, stop host protein synthesis and promote the resumption of viral DNA delivery. Then, the rest of the genome enters, enabling DNA replication and viral particle assembly and subsequent release through lysis. Only three pre-early genes were studied in the past: *dmp* encodes a phosphatase that is dispensable for infection, *A1* and *A2* are two genes required for the second-step transfer of DNA, and *A1* is also involved in host DNA degradation. The remaining 14 pre-early genes are "ORFan" genes with unknown functions. Together with this original two-step DNA injection, the clustering of the pre-early genes makes T5 an attractive model to investigate the early events of infection.

This thesis focused on the functional investigation of the FST locus of T5. In order to uncover the role of pre-early genes in host takeover, we first developed reverse genetic tools for editing the T5 genome. Alternatives to the use of the CRISPR-Cas system were set up for T5 genome engineering, which include an attractive method to modify genomes of virulent phages that "resist" CRISPR-Cas nucleases. Through these strategies, single and multiple pre-early gene deletion mutants were obtained, that allowed to assess the significance of pre-early genes for infection. Deletion of some pre-early genes led to alterations in burst size, latency and virulence of the mutants tested. Among them, deletion of the *dmp* gene resulted in the strongest alteration of infection, similar to multiple deletion mutants. A bacteriophage with a minimal FST locus could be constructed, which suggests that only very few pre-early genes - including both *A1* and *A2* - are required for the host takeover process by T5. Finally, to further explore the function of all T5 pre-early genes, we investigated the impact of their expression on host cells in the absence of other viral factors. Some gene products were toxic for the host and altered bacterial growth, producing strong perturbations of the cell morphology or the nucleoid layout, or even inducing cell lysis. These results suggest that some pre-early proteins interfere with cell division, DNA metabolism or membrane integrity. Importantly, the ability of A1 to trigger the massive destruction of the bacterial chromosome in the absence of infection was unravelled.

This exploration of the main events of T5 infection now calls for further functional and structural investigations of the pre-early proteins, which are crucial for host neutralization and thus might eventually inspire the design of new antimicrobial drugs.

Université Paris-Saclay

Espace Technologique / Immeuble Discovery

Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France