Characterization of a Novel Lysogeny and Superinfection Immunity Mechanism In Arthrobacteriophages

ZACHARY J. MITCHELL

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Science in Microbiology and Immunology

Department of Biochemistry, Microbiology, and Immunology
Faculty of Medicine
University of Ottawa

Abstract

Viruses that replicate by infecting and killing bacteria, called bacteriophages (or just phages), are ubiquitous in nature and represent one of the largest reservoirs of genetic diversity on earth. A subset of phages have an alternative lifestyle where instead of infecting and immediately killing their bacterial host, they can integrate their genetic material with the host chromosome and become dormant in a process called lysogeny.

Most disease-causing bacteria are lysogens or polylysogens harboring one or more dormant phages. However, the systems diverse phages use to maintain stable lysogeny are not fully understood which limits our ability to harness their bacterial killing power in the treatment of disease. This is at a time when antimicrobial resistance is becoming widespread and new treatment options for bacterial disease, such as phages, are needed in clinics worldwide.

In this work, I characterize a brand new lysogen maintenance system in a group of phages that infect *Arthrobacter globiformis* which appear to use a DNA-binding protease to cleave the phage-encoded sigma factor responsible for reversing lysogeny and promoting bacterial killing. I also show how this system has naturally evolved in response to phage-phage competition.

Acknowledgements

To Mom and Dad for your unending love and support.

To Elisabeth for being a great sister and patiently awaiting a boat.

To Shannon for holding down the fort with me and sometimes doing my fair share of the dishes.

To Ben for making me laugh and cooking substantial meals with *lots* of butter.

To Garrett for being someone I can count on, anytime, anywhere.

To Martha, James, and all my family and friends in Ottawa for making me feel at home away from home.

To Adam for truly understanding me and fostering my love of research.

To Liz for being a voice of reason in the face of my many ridiculous ideas and scientific struggles.

To Kieran for helping me realize new ways of studying biology.

To Katia and Nicolas for your mentorship and friendship as I embarked on this journey. Also, to Katia for teaching me how to cook dumplings correctly.

To Leen, Katia (again), and the students of SEA-PHAGES for laying the groundwork of my project.

Table of Contents

List of figures

List of abbreviations

IR

immunity repressor, see chapter 1.4.1

SI

superinfection immunity, see chapter 1.4.2

A. glob

Arthrobacter globiformis

E. coli

Escherichia coli

bp

base pair, of DNA

SigAZ

AZ sigma factor, see chapter 1.5.1

ThbSP

Thebes protease, see chapter 1.5.2

pART3

plasmid used for gene expression in *Arthrobacter globiformis*

pGEX

plasmid used for inducible gene expression in *Escherichia coli*

EV

empty vector, an expression plasmid with no gene inserted, used as a control

WT

wild type, referring to a biological entity in its natural state, ‘free of mutation’

AZ cluster

group of genetically similar phages whose lysogeny maintenance system was studied in this thesis work, see chapter 1.5

Chapter 1: Introduction

1.1 Freewheeling, freeloading, and occasional friendship by viruses

Viruses represent nature's most prevalent symbionts. They exist in a discrete form as infectious particles called virions, each with the same fundamental parts: genetic information encoded in a DNA or RNA genome, and a protein coat called a capsid which surrounds and protects the genome.

The information in their genome is vital—a set of instructions for self-preservation. But they cannot carry out these instructions alone. For this, virions must transmit their genome to a compatible host cell through infection, where the virus, now in its indiscrete form, uses machinery and resources from the host to replicate[1].

The level of damage caused to the host cell by viral infection varies. At one extreme, the virus assumes a parasitic lifestyle, aggressively hijacking the host cell and using it to rapidly produce more virions, destroying it completely in the process. At the other, the virus assumes a mutualistic lifestyle, replicating passively as one with the host and even taking on some responsibility for its fitness. Each virus' lifestyle is a manifestation of its unique strategy for survival and can change from host to host.

1.2 What virus are *you* thinking about?

The set of all compatible hosts for a virus is known as its host range. Most viruses are highly specialized for their hosts and so often have a relatively narrow host range comprised of only one species, or sometimes even a single subspecies of organism[2].

Now a question for the reader: when you think about a virus, what comes to mind? Perhaps influenza A, measles, HIV, or COVID-19? These are all examples from a group of viruses whose collective host range includes humans, members of the animal kingdom and eukaryotic domain. However, there is a much larger group of viruses than those that infect humans or even animals. They are called bacteriophages (phages for short) and their collective host range includes single celled organisms of the domains bacteria and archaebacteria. They will be the subject of the remainder of this thesis.

1.3 Phages everywhere!

From the depths of the ocean to the ground we stand on, and even on and in our very bodies, bacteriophages exist in every ecological niche that bacteria do. It is predicted that there are an astounding 10^31 phage virions on earth at any given time[3]. And each day their host destructive replication is responsible for the death of up to 40 % of all bacteria, maintaining stability in microbial populations as well as balance in ecosystems and biogeochemical flux and worldwide[4,5,6].

Bacteria, however, are not simply bystander to phage infection, and have adapted various antiphage immune systems—restriction enzymes and CRISPR-Cas being the most well-known[7]. These apply strong selection for phage counteradaptations that bypass host immunity, kicking off an endless cycle of rapid phage-host coevolution. Couple this with individual phages' high host specificity across their immense collective population and it becomes evident why phages hold the largest reservoir of genetic biodiversity on earth. This diversity is so great that most coding genes in phages have little homology to any studied proteins, leaving us guessing at their function[1,8,9].

1.4 'Life' as a phage

To replicate, phages generally assume a host destructive lifestyle aptly named lysis. It occurs following infection, when the phage genome circularizes and undergoes replication, the viral capsid and other structural proteins are produced, and virions assemble with the newly replicated genomes, lysing the host cell and returning to the environment to infect again.

In addition to lysis, many phages can assume a passive lifestyle known as lysogeny when conditions for lysis are unfavourable, such as at cold temperature[10]. During lysogeny the phage genome, known as a prophage, resides within and replicates synchronously with the host bacterium. It is also transcriptionally active to an extent, producing proteins that can enhance host survival and in turn its own[11,12]. However, at any time, and in response to changing conditions like DNA damage in the host, the prophage can exit lysogeny and begin lytic replication[13].

The schematic in Figure 1A. illustrates these lifestyles through a tailed phage. The tail structure on the virion is common in addition to the capsid amongst phages, though other morphotypes exist[14]. The tail assists in binding to the host bacteria and injecting the genome during infection.

1.4.1 Do they have what it takes?

There are two fundamental requirements for stable lysogeny in phages:

A. A system to ensure fidelitous transmission of the prophage as the host replicates.

Across all domains of cellular life, the equal partitioning of genetic material during cell division is an essential process that occurs via a highly conserved mechanism[15]. It follows that prophage transmission relies on the same mechanism.

A minority of plasmid-like prophages transmit by encoding their own partitioning system to disseminate a copy of the phage genome to each daughter cell[16]. The remainder encode an integrase to recombine into the host chromosome and let the host do this for them[17].

B. A system to repress the expression of host-toxic phage proteins and those involved in lysogeny exit, maintaining lysogeny.

In contrast to genome partitioning, there are a plethora of possible mechanisms for gene regulation to maintain lysogeny. Despite the immense diversity phages possess, only two systems for lysogen maintenance have been characterized.

The first and most well-known uses a protein called an immunity repressor (IR). IRs function by binding to the prophage genome to either stop the elongation of lytic transcripts or create DNA structures that prevent transcription initiation at lytic promotors. These structures can simultaneously enhance transcription of the IR itself to strengthen lysogeny. *Mycobacterium* and *E. coli* phages use this system with lambda phage as the classical example[18,19].

The second, much more recently discovered system is used by SPbeta-like *Bacillus* phages. These phages co-opt a host RNAse that selectively cleaves their lytic transcripts to maintain lysogeny[20].

1.4.2 Phage match (what is superinfection immunity?)

Based on the central dogma, any possible lysogen maintenance system should function through some specific feature of phage nucleic acid. This is what gives rise to the broadly observed phenomenon known as homotypic superinfection immunity (SI)[21].

To break this down, SI is when the presence of a prophage confers the lysogenic host immunity against secondary infection (superinfection) by another phage. Homotypic indicates that the superinfecting phage is genetically like the prophage. This means that the superinfecting phage possesses the same nucleic acid feature through which the prophage maintains lysogeny. Because of this it is susceptible to have its lytic replication repressed by the same mechanism as the prophage, which is what results in immunity for the host.

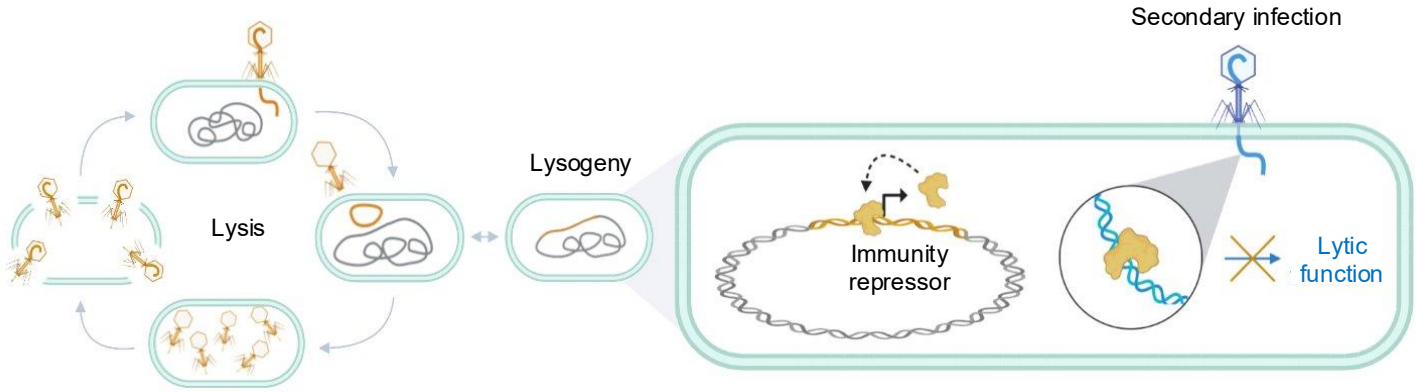
SI is one of the fitness benefits that prophages afford their host. The schematic in Figure 1A. illustrates an example of homotypic SI through an integrating prophage that maintains lysogeny using an IR.

1.5 Phages in Ottawa

Through the SEA-PHAGES course, undergraduate students at uOttawa have been isolating and characterizing novel Actinobacteriophages from environmental samples for the past half-decade, which now form a biobank of over 400 unique phages that are catalogued on the online database phagesdb[22]. Of the phages from this local collection, 51 have had their genomes sequenced.

Following sequencing, phages are assigned to phylogenetic groups called clusters with one another if they are genetically (and thus proteomically) similar to facilitate further

A.



B.

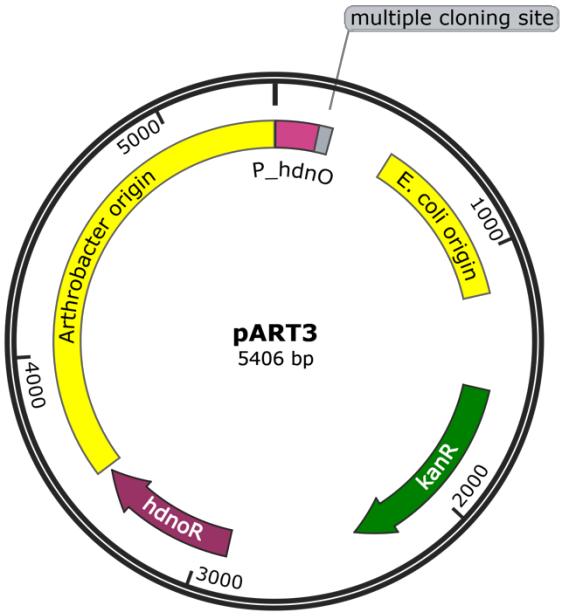


Figure 1. Introductory schematics. A. Phage lifestyles and an example of superinfection immunity, created with BioRender (not to scale). **B.** Plasmid map of pART3 made in SnapGene.

study. At the time of this writing, phages are clustered together if they share over 25 % proteomic equivalence, a combinatorial metric that considers gene content similarity across genomes, and amino acid similarity within gene products[23]. New clusters are named in the order A, B, C… AA, AB, etcetera. Groups within clusters, called subclusters are formed among phages that share over 60 % proteomic equivalence and named as follows: A1, A2, A3….

Of the 51 sequenced uOttawa phages, a large proportion—19 phages—which infect *Arthrobacter globiformis* (*A. glob*) are members of the AZ cluster. This poises our lab especially well to perform representative phenotypic assays with this cluster. There are also a total of 85* AZ phages with sequences available on phagesdb which enables robust bioinformatic analysis of the cluster through multiple sequence alignment (*at the time of this writing).

From a subset of the uOttawa AZ phages, a previous student, Katia Koziel Ly, isolated culturable lysogens and showed that they provided varying levels of homotypic SI against other cluster members. Their sequences reveal that although they encode an integrase for prophage transmission, none of the AZs have homologs of known lysogen maintenance systems, implying that they use a completely novel system.

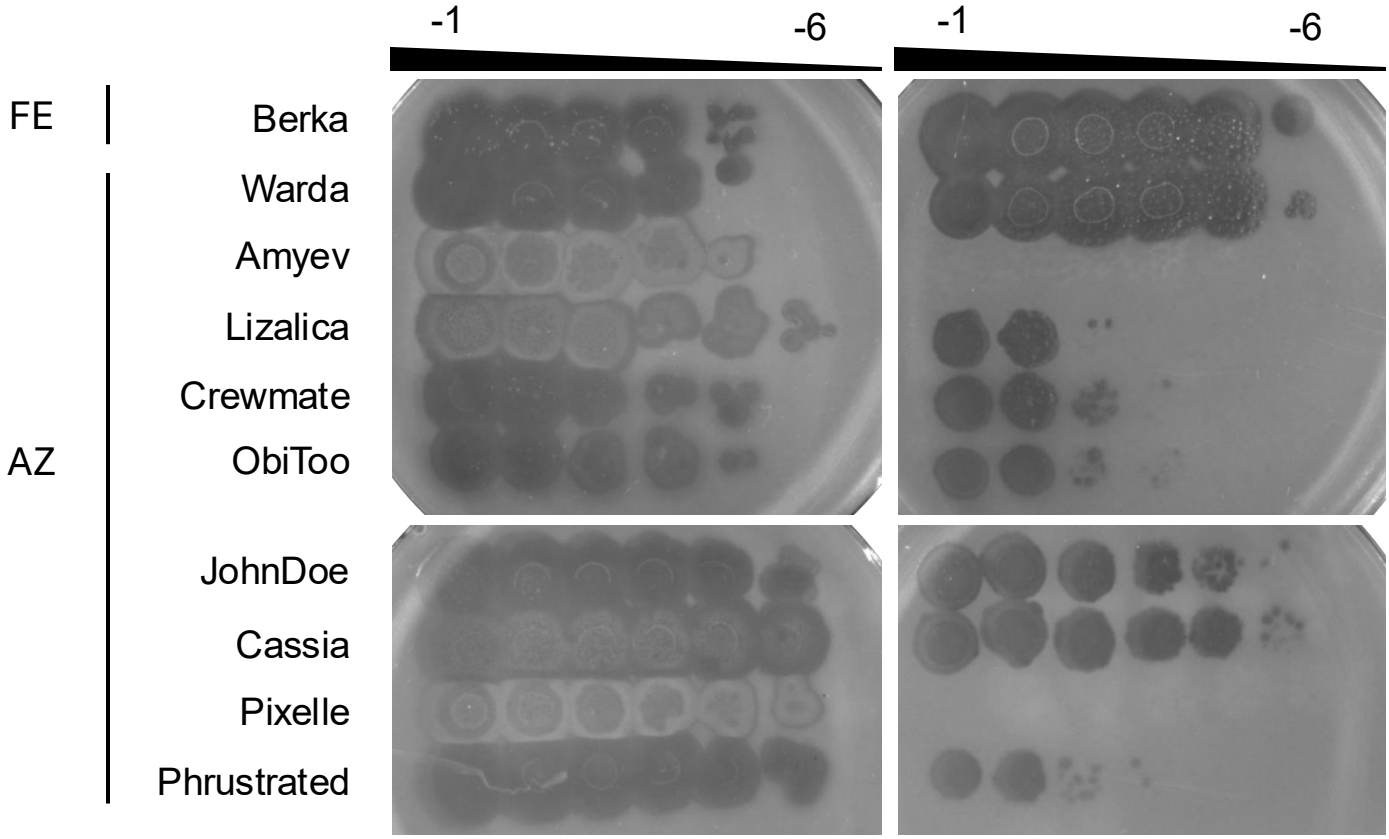
Of the lysogens Katia tested, Amyev and Pixelle were among the most susceptible to SI escape by other phages. Figure 2A. shows an example of this on the Pixelle lysogen. This raised the question of whether the phages which appear to partially escape SI (such as Lizalica, Crewmate, ObiToo, and Phrustrated), represent a population of genetic mutants from the parent lysate.

Leen Madani, another previous student, answered it by isolating and sequencing five independently amplified genetic mutants of ObiToo and Crewmate that escape superinfection immunity on Amyev and Pixelle lysogens. Figure 2B. shows an example of some of her escape mutants on Pixelle.

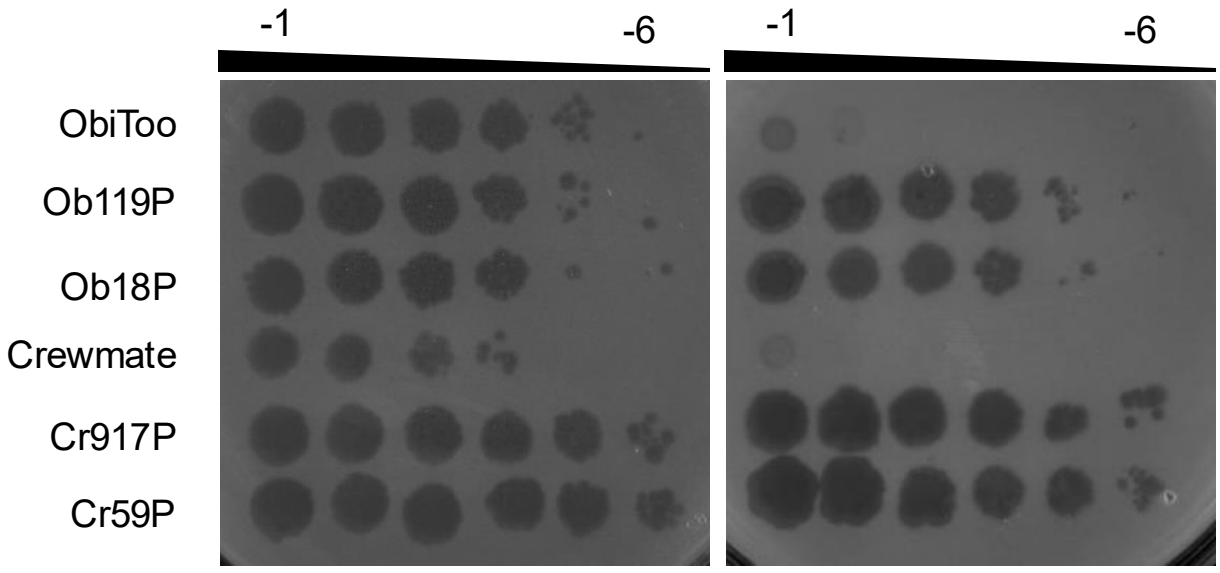
In all five of Leen’s sequenced mutants, single nucleotide substitutions were found in the promotor upstream of gene 50, and 4 of 5 were contained in a single 8 bp palindrome, defining this as an important locus for SI and lysogen maintenance (Figure 1C.). Although large deletions appeared in the three Crewmate mutants, these deletions were also found in the parent lysates which, without the single nucleotide substitutions, did not demonstrate SI escape. In ObiToo escape mutants, only the single nucleotide substitutions were present.

From these data, I theorized that the genes downstream of the promotor defined by the SI escape mutants might encode the lysogen maintenance system in AZ phages and act by binding to the 8 bp palindromic site.

A.


-1
-6
-1
-6
FE
Berka
AZ
Warda
Amyev
Lizalica
Crewmate
ObiToo
JohnDoe
Cassia
Pixelle
Phrustrated
A. glob 2880
Pixelle lysogen

B.


-1
-6
-1
-6
ObiToo
Ob119P
Ob18P
Crewmate
Cr917P
Cr59P
A. glob 2880
Pixelle lysogen

C.

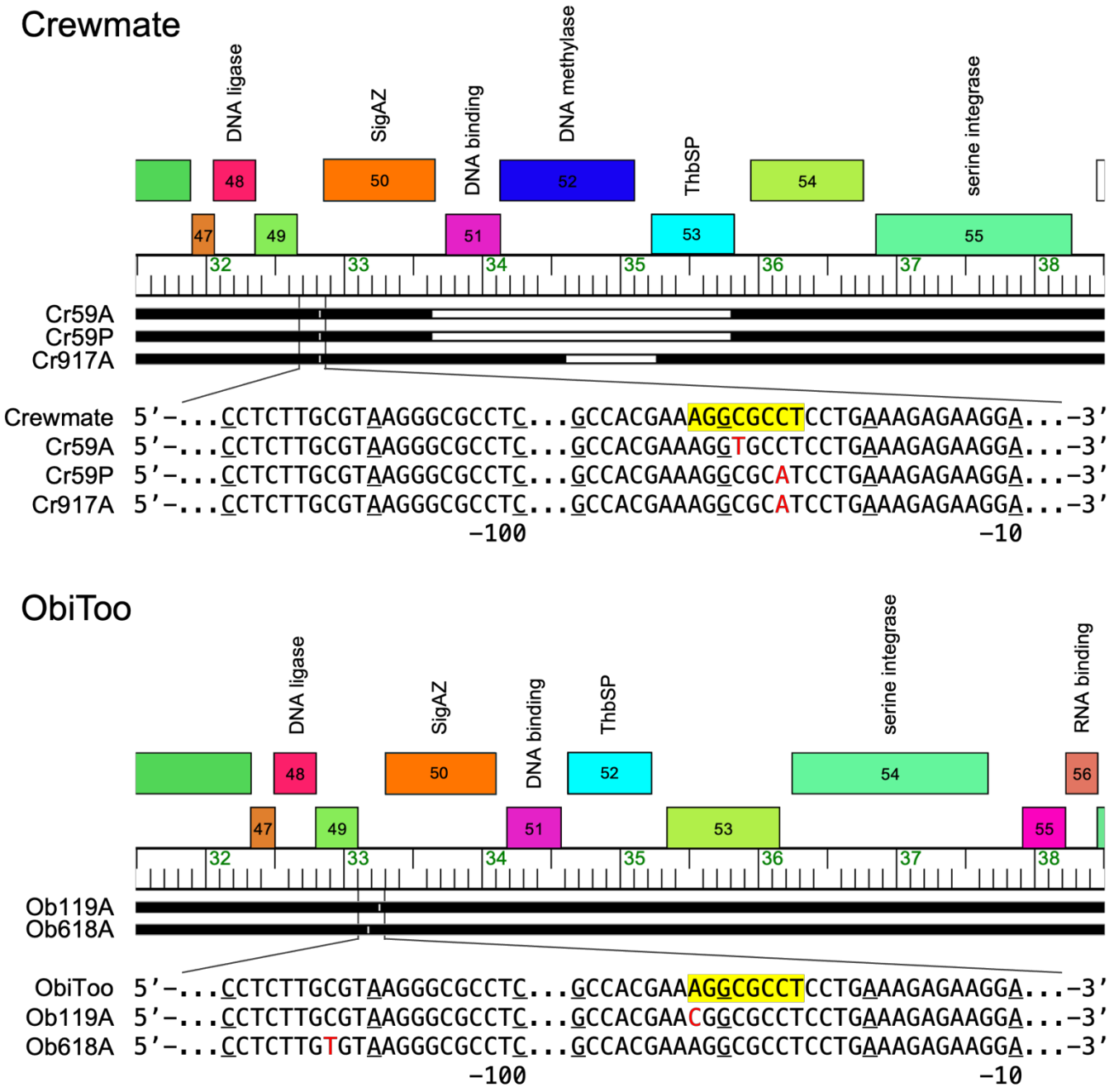


Figure 2. Superinfection immunity escape mutants arise in the promotor upstream of a phage encoded sigma factor. A. Many AZ phages escape SI on Pixelle. Phage dilutions from 10e-1 to 10e-6 spotted on either Pixelle lysogen or WT *A. glob* 2880 as a control. The FE phage Berka is genetically unrelated to the AZs. **B.** Independently amplified mutants of ObiToo and Crewmate demonstrating SI escape on Pixelle. As above. **C.** Five independent mutants of ObiToo and Crewmate define a consensus locus for SI escape. Sequence alignments of ObiToo and Crewmate SI escape mutants from SnapGene overlayed with with the WT gene maps from Phamerator. Single nucleotide substitutions in the promotor upstream of SigAZ are indicated in red. The 8 bp palindromic sequence is highlighted in yellow.

Upon further examination, only two of these genes were well conserved across AZ phages, making them the most promising candidates.

1.5.1 Sigma AZ

The gene immediately downstream of the promotor defined by the SI escape mutants (gene 50 in ObiToo and Crewmate) is fully conserved in AZ phages and bears strong homology to numerous bacterial sigma factors (>99.9% probability, HHPRED[24]). I will refer to it as sigma AZ (SigAZ) from now on.

In bacteria, sigma factors are the exchangeable subunit of RNA polymerase. They direct transcription at various promotors by binding the -35 and -10 motifs, recruiting the RNA polymerase core enzyme and melting the DNA duplex to initiate transcription. Each bacteria encodes a set of different sigma factors that they use to regulate different cellular processes and responses to environmental stress[25].

Sigma factors have been previously found encoded by phages that infect spore forming bacteria. In these instances, they influence the host's commitment to sporulation favouring phage replication[26].

1.5.2 Thebes protease

Gene 53 in Crewmate (and 52 in ObiToo) which I will refer to as Thebes protease (ThbSP) from now on, is homologous to Spartan protease (SPRTN) in humans (99.6% confidence, HHPRED[24]).

SPRTN is a zinc metalloprotease that is activated by dimerizing and binding DNA, where it cleaves DNA-protein complexes to initiate DNA repair[27]. There are also homologs in bacteria, but they are not characterized.

ThbSP is conserved in all AZ phages except members of the newly formed AZ5 subcluster that infect *Curtobacterium*.

1.6 Allow me to introduce our host

Arthrobacter globiformis and other members of its genus are soil bacteria that have been near exclusively studied in the context of their versatile metabolic systems. They have been shown to degrade many toxic xenobiotics which allows them to grow in places many other life forms cannot[28,29].

Among the long list of unorthodox carbon sources *Arthrobacters* can degrade is nicotine. It has been shown that production of first enzyme in this degradation pathway, 6-hydroxy-nicotine oxidase (6HdnO), is regulated by a repressor (6HdnoR) which is inactivated upon sensing nicotine[30,31].

This led Sandu et. al. to engineer a nicotine inducible plasmid (pART3) for use in *Arthrobacter* that encodes 6HdnoR and uses the 6HdnO promotor/operator to drive gene expression (Figure 1B.)[32].

1.7 Aims

With this base of knowledge and using pART3 as a tool, I set out to characterize the roles of SigAZ and ThbSP in the AZ phages' novel lysogen maintenance system as follows:

Aim 1. Assess the roles of SigAZ and ThbSP in lysogen maintenance and SI by expressing them in the host during infection.

Aim 2. Raise an antibody and confirm the expression of SigAZ *in vivo* during lysogeny by western blot.

Aim 3. Assess the proteolytic activity of ThbSP *in vitro* and verify if it or SigAZ binds the 8 bp palindromic site defined by the mutants using gel shift.

Chapter 2: Materials and methods

2.1 Culture of *Arthrobacter*

Frozen stocks of *Arthrobacter* were struck on PY plates and grown at 30 ˚C for 36-48 h.

Colonies were inoculated into PY liquid and grown to saturation at 30 ˚C with shaking (~48 h) before being removed to room temperature and used downstream within three days.

2.2 Preparation of electrocompetent cells from *Arthrobacter globiformis*

A new liquid culture of *A. glob* 2880 or 2979 was inoculated in PY at 1 % from a freshly saturated culture and grown at 30 ˚C with shaking. When it reached an OD600 of 0.3, it was dosed with 30 µg/mL ampicillin and returned to 30 ˚C with shaking for 30 mins.

The culture was then removed to chill on ice for 10 mins before being pelleted at 2000 xg for 15 mins in a 4 ˚C centrifuge. After discarding the supernatant, the cells were placed on ice and resuspended in 1/2 the initial culture volume of ice-cold wash buffer (10% glycerol, 0.5 M sorbitol, mqH_2O to volume and sterilize) then pelleted again. This process was repeated for a total of three resuspensions in wash buffer.

After discarding the final wash, the cells were resuspended in 1/100 th the initial culture volume of ice-cold electroporation buffer (same as wash buffer except without sorbitol) and distributed into 50 µL aliquots in tubes chilled on ice. These aliquots were flash frozen and stored at -80 ˚C for later use[33].

2.3 Plasmid DNA preparation

Frozen stock of plasmid-harbouring *E. coli* DH5alpha DE3 was inoculated into LB with appropriate selection (100 µg/mL ampicillin or 50 µg/mL kanamycin) and grown overnight at 37 ˚C with shaking.

5 mL of overnight culture was pelleted, and the plasmid was extracted using the EZ-10 Spin Column Plasmid DNA Miniprep kit (Bio Basic), eluted in mqH_2O, quantified by NanoDrop, and frozen at -20 ˚C for later use.

2.4 PCR

50 µL reactions were prepared in Phusion HF buffer (Thermo Scientific) with, 2 µM dNTPs, 2 % DMSO, 0.5 mM each of forward and reverse primer, mqH_2O to 48 µL, 1 µL template DNA, and 1 µL Phusion DNA polymerase. Once prepared, reactions were incubated at 95 ˚C for 2 min, then 35 cycles of 10 s denaturing at 95 ˚C, 30 s annealing at 60 ˚C, 20 s/kb elongation at 72 ˚C, finished by 5 mins at 72 ˚C then held at 4 ˚C. PCR products were frozen at -20 or used downstream immediately.

2.4.1 Template preparation

pAO1 screening:

Mixed colonies were picked into 15 mM NaOH and heated at 95 ˚C for 10 mins then used as template.

Phage for cloning:

NaOH was added to high titer phage lysate to final concentration of 15 mM then the mixture was heated at 95 ˚C for 10 mins and used as template.

2.4.2 Primer design

Primers for pAO1 screening were designed with Primer3[34].

hdnO FWD:

TAAGGTCGGGTTCTGTGGAC

hdnO REV:

CGAAAACGCTGGGGTAGATG

gntR FWD:

GCCATTGCCAAGATCGTCAT

gntR REV:

ATCGTCATGGAAATTGCGCA

coxE FWD:

AGGTCCCTAAAAGCTGCCAT

coxE REV:

AGACAATCTGAGGAGAGGCG

Primers for Crewmate gene inserts used in cloning were designed with SnapGene.

pART3-*thbSP* FWD:

GATTGTCAGCAGACAAAGGAGTTGGAAATGACCGCAAACTTCAAGACCCG

pART3-*thbSP* REV:

GTGTCTAGACTGCAGGGTACCACTAGTCTCACGCCTCGACCATCTGGG

pART3-*sigAZ* FWD:

GATTGTCAGCAGACAAAGGAGTTGGAAATGAGCTTCAACAACGTCCTCG

pART3-*sigAZ* REV:

GTGTCTAGACTGCAGGGTACCACTAGTCTCACGCGACGGCCAGAGC

pGEX-*sigAZ*$_{1\text{-}348}$ FWD:

CCGGAATTCATGAGCTTCAACAACGTCCTCG

pGEX-*sigAZ*$_{1\text{-}348}$ REV:

CCGGAATTCTCAGGCGTCGCGGCTCGGGTA

pGEX-*sigAZ*$_{348\text{-}573}$ FWD:

CCGGGATCCGCCTTCACGGTCCCGGCC

pGEX-*sigAZ*$_{348\text{-}573}$ REV:

CCGGAATTCTCAGGCGTCCCAGATCGGC

2.5 DNA gel electrophoresis

DNA samples mixed 5:1 with dye (5 mM Tris pH 8, 10 mM EDTA, 5 % glycerol, 0.005 % bromophenol blue, mqH_2O to volume) were loaded in agarose gels (1-1.2 % agarose, TAE to volume, melt, allow to cool, 5 µL/100mL EtBr, cast) and run at 100 V for 1 h in TAE (40 mM Tris, 1 mM EDTA, 1.142 mL/L acetic acid, mqH_2O to volume). After the completion of a run, gels were visualized by UV irradiation.

2.6 Plasmid assembly

Each PCR amplified gene insert and digested plasmid backbone, purified by gel extraction (Zymoclean Gel DNA Recovery Kit, Zymo Research) were mixed in a 2:1 molar ratio as determined from NanoDrop concentrations. Then, 5 µL of this mixture was added to 5 µL HiFi assembly mix (NEB) on ice and incubated at 50 ˚C for 30 mins, followed by immediate transformation or freezing at -20 ˚C for later use.

2.6.1 Backbone digestion

Digestion reactions were prepared with up to 10 µg plasmid backbone DNA in CutSmart buffer (NEB) and incubated for 1 h at 37 ˚C then purified immediately or frozen at -20 ˚C for later use.

BamH1 and Spe1 (NEB) were used to digest pART3 for *A. glob* expression plasmids.

BamH1 and EcoR1 (NEB) were used to digest pGEX-6P1 for *E. coli* expression plasmids.

2.7 Transformation of *Escherichia coli*

Up to 5 µL of plasmid DNA was mixed with freshly thawed, chemically competent *E. coli* (DH5alpha DE3 or Rosetta DE3), incubated on ice for 20 mins, heat shocked at 42 ˚C and recovered in 1 mL LB at 37 ˚C with shaking for 2 h. After recovery, cells were plated on LB with appropriate selection (50 µg/mL kanamycin, 100 µg/mL ampicillin, and/or 24 µg/mL chloramphenicol).

2.8 Electroporation of *Arthrobacter globiformis*

Aliquots of electrocompetent *A. glob* were thawed and gently mixed with 100-1000 ng of plasmid DNA, up to 5 µL, on ice, transferred to a 1 mm electroporation cuvette, shocked at 1500 V with 25 µF capacitance and 400 ohms resistance, then recovered in 1 mL of PY at 30 ˚C with shaking for 8h. After recovery, cells were plated on PY with appropriate selection (100 µg/mL kanamycin for *A. glob* 2979, 50 µg/mL for *A. glob* 2880. In downstream liquid cultures and top agar, 20 µg/mL kanamycin selection was used for *A. glob* 2979 and 10 µg/mL was used for *A. glob* 2880).

2.9 Phage amplification

Phage from frozen stocks was picked into phage buffer to create a phage base. Tenfold serial dilutions were performed from the base in phage buffer. Each dilution was added to 0.5 mL of saturated host culture, mixed with 4 mL of molten PYCa top agar and plated on PY. Once set, the plates were incubated at 30 ˚C for 24-48 h, saving the phage base at 4 ˚C during this time. If no webbed plates were achieved, the process was repeated, performing more targeted dilutions on the base.

Any webbed plates were flooded twice in succession for ~3 hours at 30 ˚C with 8 mL phage buffer each time, harvesting and filtering after each flood, and pooling the second filtrate with with the first. Pooled, filtered phage lysates were titered by spot assay and high titer lysates (>1e9 pfu/mL) were stored at 4 ˚C for downstream use.

2.10 Phage spot assays

0.5 mL of a saturated host culture was mixed with 4 mL of molten PYCa top agar containing appropriate selection and induction, if necessary (10 µg/mL kanamycin, 0.05 % nicotine), and plated on PY. Once set, tenfold serial dilutions of 5e8 pfu/mL phage lysates (if the titer was known), performed in phage buffer, were deposited onto the plate in 5 µL spots. After the spots dried the plate was incubated at 30 ˚C for 24-48 h and imaged thereafter.

2.11 Protein gel electrophoresis

Samples were loaded in polyacrylamide gels (Resolving: 12.5 % acrylamide, 0.104 % bis-acrylamide, 375 mM Tris pH 8.8, mqH_2O to volume, degas 5 mins, 0.05 % APS, 0.5 µL/mL TEMED, cast; Stacking: 5 % acrylamide, 0.13% bis-acrylamide, 125 mM Tris pH

6.8, mqH_2O to volume, 0.1 % APS, 1 µL/mL TEMED, cast) and run at 25 mA, to a maximum of 200 V for 105-135 mins in TG-SDS (50 mM Tris, 384 mM glycine, 0.1 % SDS, mqH_2O to volume).

After the completion of a run, gels were used downstream in western blotting immediately or visualized by staining for 15 mins with Coomassie (5 g/L Coomassie blue R250, 45 % methanol, 10 % acetic acid, mqH_2O to volume) and destaining overnight (25 % methanol, 7 % acetic acid, mqH_2O to volume).

2.11.1 Sample preparation

Peak fraction determination:

10 µL of eluted protein fractions were mixed 1:1 with 2x sample buffer (80 mM Tris pH 6.8, 10 mM EDTA, 2% SDS, 10 % glycerol, 0.005 % bromophenol blue, mqH_2O to volume, 10 % BME, 1 mM PMSF) before loading 15 µL.

Protein quantification:

20 µL of purified protein was mixed 1:1 with 2x sample buffer and from this, 8-10 twofold serial dilutions were performed in 1x sample buffer (2x sample buffer mixed 1:1 with mqH_2O). This process was repeated for a 10 mg/mL BSA standard and 10 µL of each dilution was loaded.

Western blotting:

Cell pellets for were mechanically lysed by bead beating for 2 mins in 300 µL of 1x sample buffer (2x sample buffer mixed 1:1 with mqH_2O) before loading 20 µL.

2.12 Protein purification

Frozen induced cell pellets were thawed and resuspended in 5 mL/g lysis buffer (1.8 mM KH_2PO_4, 2.7 mM KCl, 10 mM Na_2HPO_4, 500 mM NaCl, 1 mM EDTA, 1 mM EGTA, 15 mM DTT, 1 mM PMSF, 0.5 % Triton X-100) on ice. Once resuspended, lysozyme was added at 200 µg/mL and the mixture was incubated on ice for 15 mins before sonicating thrice for 45 s at 50 % power.

The insoluble portion was pelleted in a 4 ˚C centrifuge at 40000 xg for 20 mins then the supernatant was mixed with glutathione-agarose beads (Sigma-Aldrich) prewashed in bead buffer (same as lysis buffer except only 1 mM DTT, 0.01 mM PMSF, and with 0.1 % NP40 instead of Triton X-100) and incubated at 4 ˚C with gentle rocking for 2 h.

After 2 h of binding, the beads were pelleted at 500 xg for 2 mins and as much supernatant as possible was removed without disturbing the beads. Then the beads were transferred to a column at 4 ˚C and washed with 15 volumes cold bead buffer, 5 volumes cold column buffer (same as lysis buffer except only 1 mM DTT, 0.01 mM

PMSF, and without Triton X-100), and 5 volumes cold pre-elution buffer (50 mM Tris pH 8.1, 500 mM NaCl, mqH_2O to volume, 1 mM DTT, 0.01 mM PMSF).

The column was then eluted in 23-1 mL fractions using cold elution buffer (same as pre-elution buffer but with 5 mM reduced glutathione), and the peak fractions, determined by protein gel electrophoresis followed by Coomassie staining while the fractions remained at 4 ˚C, were pooled.

Pooled fractions were dialyzed thrice in 1 L dialysis buffer (1.8 mM KH_2PO_4, 2.7 mM KCl, 10 mM Na_2HPO_4, 500 mM NaCl, 15 % glycerol) at 4 ˚C for 3-12 h each time. After dialysis, purified proteins were aliquoted on ice, flash frozen, and stored at -80 for later use.

2.12.1 Protein expression

A new liquid culture of *E. coli* Rosetta DE3 harbouring pGEX-*sigAZ*$_{1\text{-}348}$ or -*sigAZ*$_{348\text{-}573}$ was inoculated at 0.1 % from an overnight culture into LB with appropriate selection (100 µg/mL ampicillin, 24 µg/mL chloramphenicol) and grown at 37 ˚C with shaking. When it reached an OD600 of 0.5, expression of the GST-tagged N-terminal fragment of SigAZ was induced by adding 2 µM IPTG and returning the culture to shaking at 30 ˚C for 4 h.

After 4 h the cells were pelleted, the supernatant discarded, and the pellet flash frozen and stored at -80 ˚C for later protein purification.

2.13 Antibody generation

A total of 1.5 mg of purified GST-SigAZ$_{1\text{-}116}$ and -SigAZ$_{116\text{-}191}$, quantified by comparison to a standard on a Coomassie stained gel, were mixed in equal parts by mass, adding Pen/Strep and PBS to 750 µL (1.8 mM KH_2PO_4, 2.7 mM KCl, 10 mM Na_2HPO_4, 137 mM NaCl, 10 µg/mL penicillin, 10 µg/mL streptomycin). This was then combined 1:1 with adjuvant and emulsified for injection (Freund’s complete, first injection; Freund’s incomplete, second and third injections; Millipore sigma).

Every month, for a total of three times, two intramuscular and four subcutaneous injections were administered to a rabbit (uOttawa animal care facility).

Three and a half weeks after the final injection, serum was extracted from the blood, aliquoted, flash frozen and stored at -80 for later use

2.13.1 Ethics statement

This study was performed in strict accordance with standards for animal care and use outlined in the Canadian Council on Animal Care (CCAC) policies and guidelines. The University of Ottawa holds a certificate of Good Animal Practice with the CCAC and is a registered research facility under the Province of Ontario's Animals for Research Act.

The animal use protocol (BMIe-3469) was approved by the University of Ottawa Animal Care Committee. All antigen injections and blood collection were administered under general anesthesia to minimize pain and suffering. Rabbits were first sedated with injectable sedatives, butorphanol and midazolam and induced into general anesthesia with inhaled isoflurane from a precision vapourizer.

2.14 Western blotting

After performing protein gel electrophoresis, proteins were fixed on a nitrocellulose membrane by transferring at 60 V for 90 mins in transfer buffer (20 mM Tris, 150 mM glycine, 20 % methanol, mqH_2O to volume) at 4 ˚C.

After transferring, the membrane was blocked twice in 4 % milk TBST (20 mM Tris, 150 mM NaCl, 0.1 % Tween-20, mqH_2O to volume) for 15 mins each time at room temperature, then incubated overnight at 4 ˚C in fresh 4 % milk TBST with 1:2500 anti-SigAZ antibody, using gentle rocking throughout.

The next day, the membrane was washed thrice in TBST at room temperature for 10 mins each time, incubated for 30 mins in TBST with 1:5000 anti-rabbit-HRP antibody (BioRad), then washed thrice more in TBST as before, using gentle rocking throughout.

Immediately thereafter, the blot was visualized with Clarity Western ECL Substrate (BioRad).

2.14.1 Sampling

For verifying pART3 induction:

New liquid cultures of *A. glob* 2880 and 2979 harbouring pART3-EV or -*sigAZ* were inoculated in PY with appropriate selection (10 or 20 µg/mL kanamycin) at 0.1 % and grown overnight at 30 ˚C with shaking.

The next day, exponentially growing cultures were split and induced with either 0.01 or 0.05 % nicotine or left uninduced as a control. After 5 hours, 2 mL samples of the cells were harvested by centrifugation, flash frozen, and stored at -80 ˚C for later use.

For assessing SigAZ cleavage in lysogens:

A 1 mL sample from a freshly saturated culture of Crewmate lysogen, as well as one from a culture of *A. glob* 2880 as a control, were harvested by centrifugation, flash frozen, and stored at -80 ˚C for later use.

The *A. glob* 2880 pART3-EV and -*sigAZ* controls were the same as the uninduced samples used for verifying pART3 induction.

2.15 Common recipes

2.15.1 PY

Plates:

15 g/L peptone, 1 g/L yeast extract, 15 g/L agar, mqH_2O to volume and sterilize. After sterilization aseptically add 0.1 % dextrose, 10 µg/mL cycloheximide, cast.

Liquid:

Same as plates except without agar.

Top agar:

Same as plates except with only 4 g/L agar and with 4.5 mM $CaCl_2$ added after sterilization.

2.15.2 LB

Plates:

10 g/L peptone, 10 g/L NaCl, 5 g/L yeast extract, 15 g/L agar, mqH_2O to volume and sterilize, cast.

Liquid:

Same as plates except without agar.

2.15.3 Phage buffer

10 mM Tris pH 7.5, 68 mM NaCl, 10 mM $MgSO_4$, 1 mM CaCl, 10 % glycerol, 0.002 % Tween-20, mqH_2O to volume and sterilize.

2.16 Structural modeling

Biomolecular structural modeling was performed using AlphaFold3 through the Google DeepMind server[35]. The model in Figure 7. was examined and recoloured in the PyMOL Molecular Graphics System, Version 2.5.4 Schrödinger LLC.

2.17 Generation of consensus sequences

Multiple sequence alignment was performed using MUSCLE with default settings through the EMBL-EBI Job Dispatcher[36,37]. Aligned sequence files in ClustalW format were opened in Jalview v2.11.4.1 to generate consensus sequences[38].

2.18 Generation of phylogenetic trees

Multiple sequence alignment was performed using MUSCLE with default settings through MEGA11[39]. Aligned sequence files in MEGA format were used to generate phylogenetic trees by neighbour joining using the p-distance method and bootstrapping with 1000 replications.

2.19 Sequence availability

All AZ phage genome and protein sequences are available on phagesdb[22].

pART3 (DQ191048), pGEX-6P1 (U78872), and pAO1 (AJ507836) DNA sequences are available from GenBank.

Chapter: 3 Results

3.1 SigAZ drives lytic replication while ThbSP may be responsible for superinfection immunity and lysogen maintenance

Once I optimized a method to transform plasmid DNA into *A.glob* 2880, I tested whether the expression of SigAZ from pART3 might give immunity against AZ phage infection, assuming it functioned analogously to an IR. This assumption follows from the simplest hypothesis to explain the SI escape mutants whereby the mutations would prevent SigAZ from binding to and upregulating expression at its own promotor to bypass SI. However, as shown in Figure 3A., SigAZ does not give immunity against infection, instead increasing plaque size and infection by 10 to 1000-fold and behaving like a transcriptional regulator that drives lysis.

I followed up by testing in the same way whether ThbSP, encoded by the other conserved gene downstream of *sigAZ*, would have any impact on infection. Not only did its expression give immunity against WT Crewmate and ObiToo, but this immunity was bypassed by the Crewmate and ObiToo SI escape mutants, implicating ThbSP as responsible for SI at the palindromic site affected by the mutants (Figure 3B.). A critical caveat to this result is that it only occurred once despite numerous attempts to reproduce it.

3.2 pART3 is not a robust expression vector in *Arthrobacter globiformis* probably owing to low bioavailability of its inducer, nicotine

After struggling to reproduce the immunity phenotype demonstrated by ThbSP, I questioned whether poor induction and expression from pART3 was to blame. To confirm this, I assessed the expression of SigAZ homologs off pART3 in *A.glob* by western blot, probing with a polyclonal antibody I raised against two fragments of the SigAZ N-terminal domain from Crewmate. The result in Figure 4A highlights that pART3 is not substantially induced in *A. glob* 2880 or 2979 beyond a basal level of expression—even after prolonged exposure to a high dose of nicotine (5 h, 0.05 %).

It stands to reason that this basal level was adequate to repeatedly demonstrate the effect of SigAZ on infection through its likely amplificatory transcriptional impact, but inadequate to do the same for ThbSP whose effect depends upon meeting a greater expression threshold. This is supported by the presence of a conserved canonical promoter to drive ThbSP expression in AZ phages that is absent upstream of SigAZ (Figure 8A.).

Based on a report by Ganas and Brandsch, nicotine import into *Arthrobacter* occurs via facilitated diffusion through a yet unidentified transporter[40]. And for nicotine and its derivates to accumulate to detectable levels intracellularly, as would be required for pART3 induction, strains must actively catabolize nicotine.

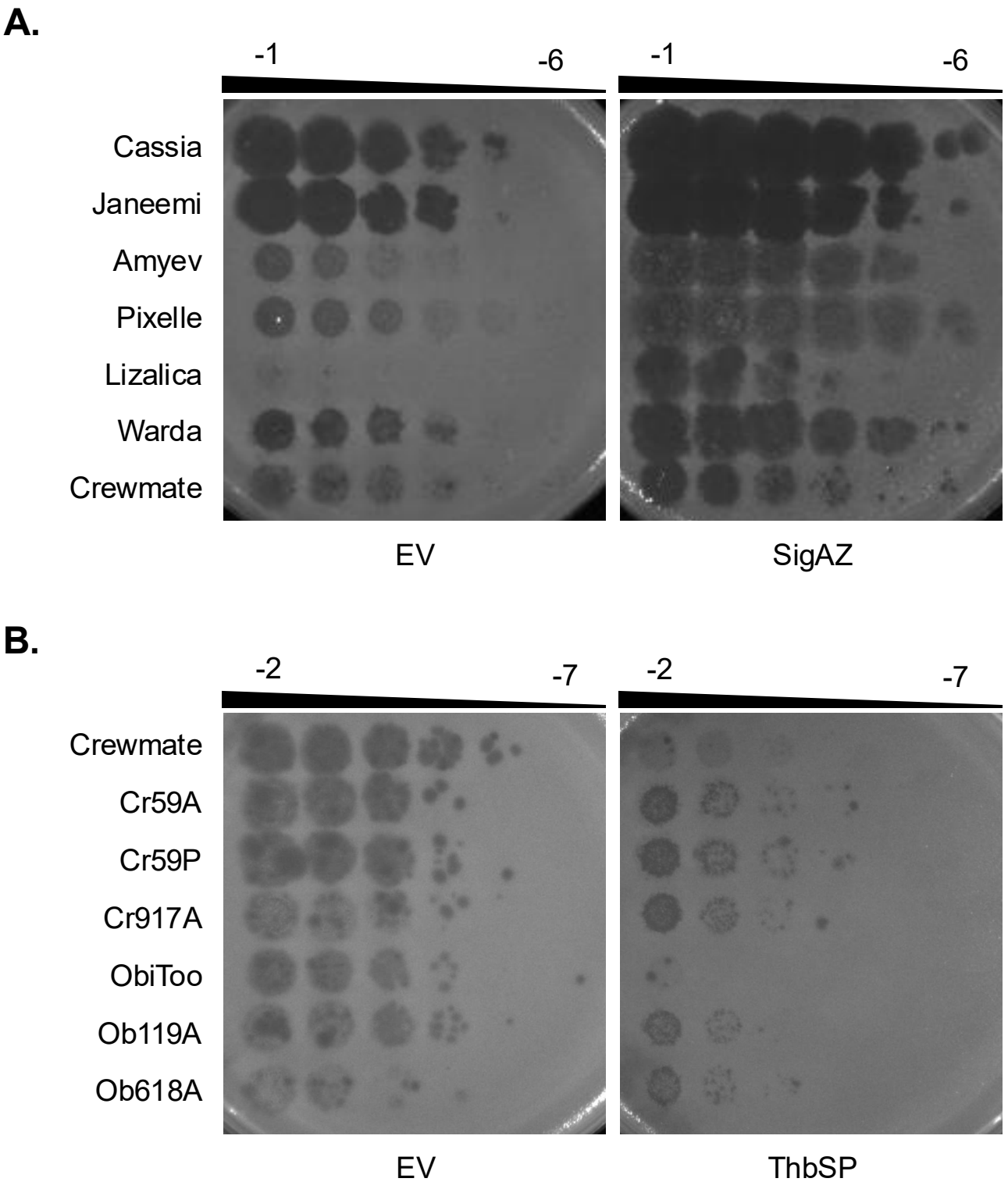
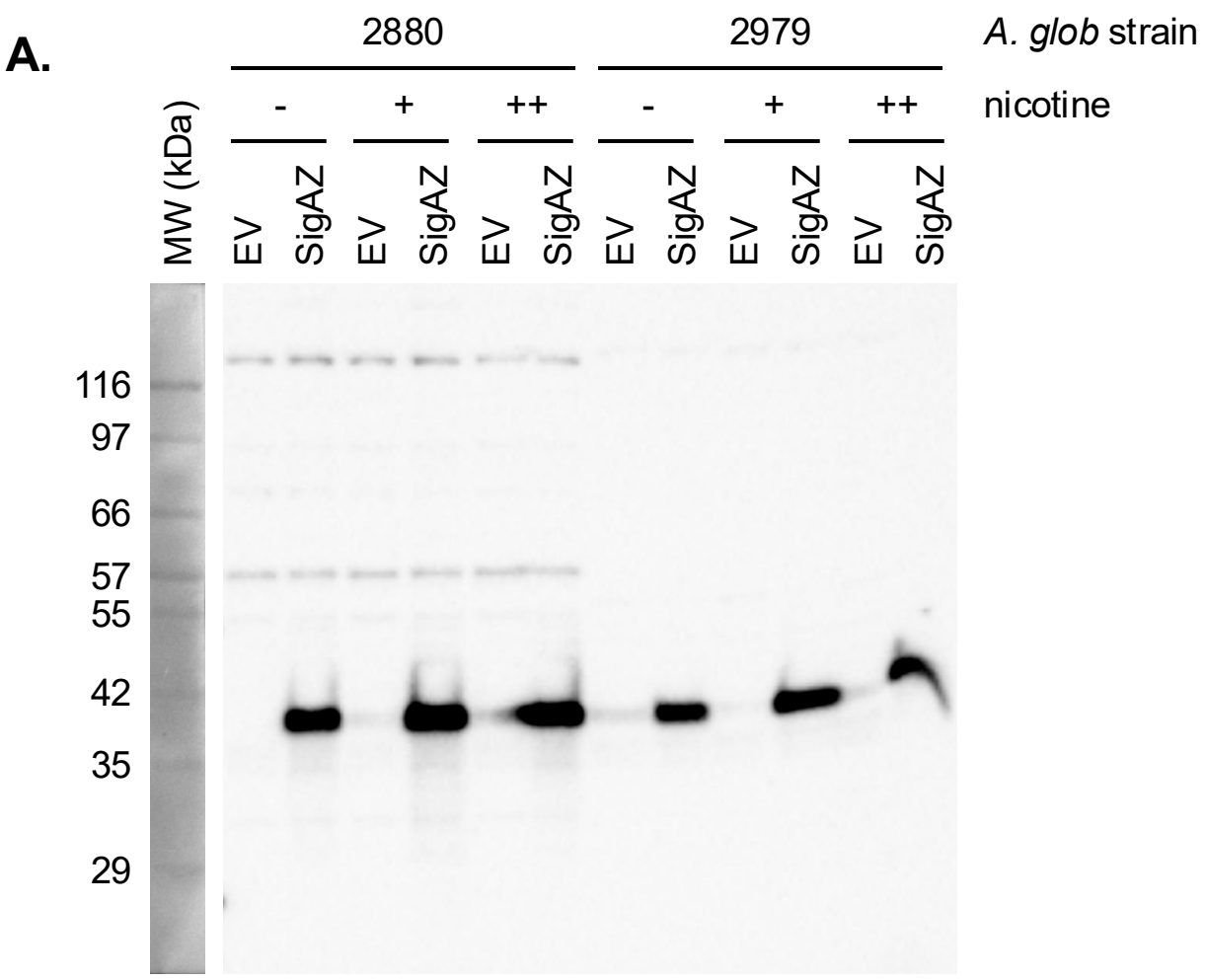
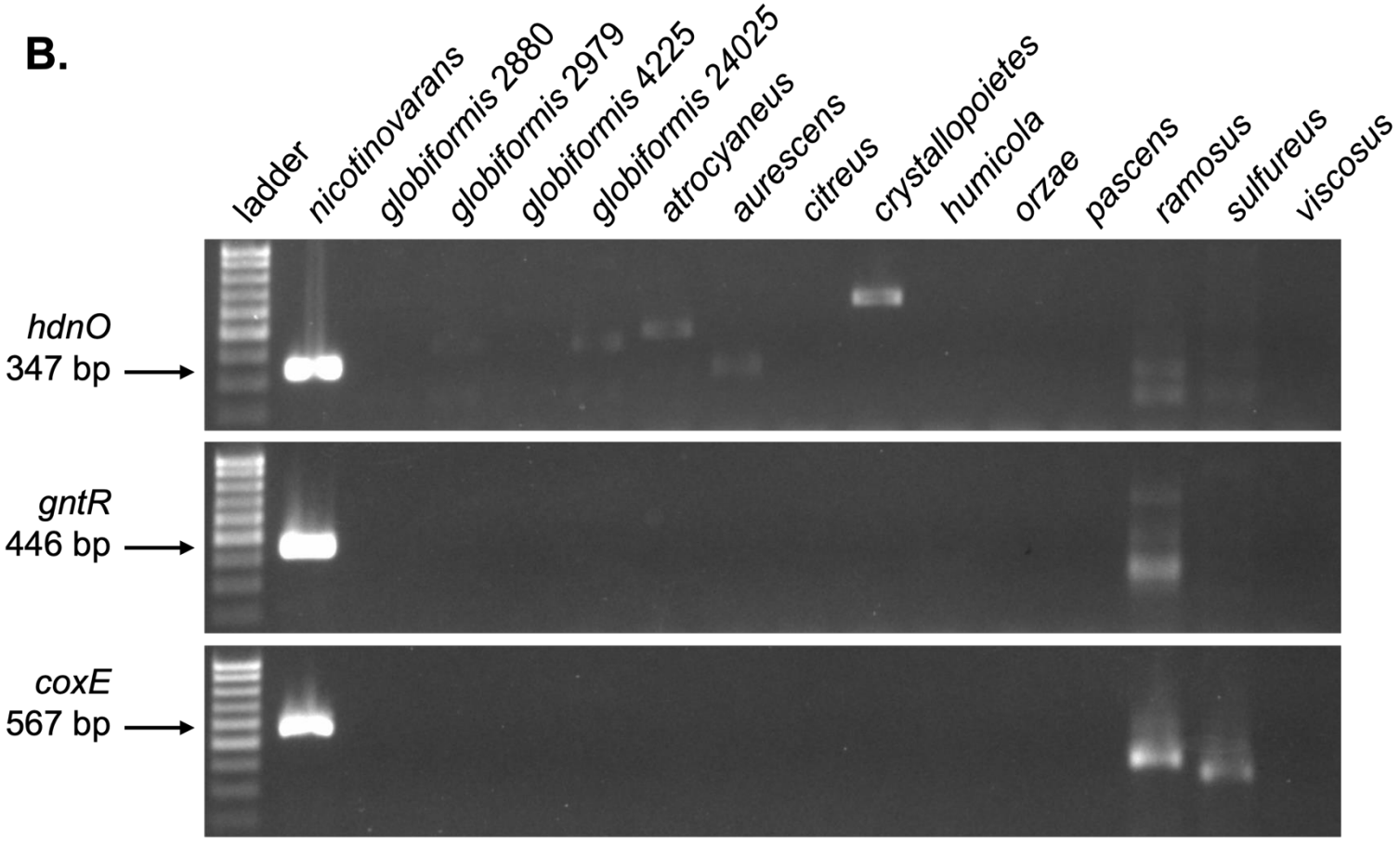


Figure 3. Roles of SigAZ and ThbSP in lysis and lysogeny. A. SigAZ drives lytic replication. Phage dilutions from 10e-1 to 10e-6 spotted on *A glob* 2880 harbouring pART3-*sigAZ* (Crewmate homolog) or -EV as a control. Representative of n=3 biological replicates. **B. ThbSP gives immunity against Crewmate and ObiToo but not SI escape mutants.** Phage dilutions from 10e-2 to 10e-7 spotted on *A glob* 2880 harbouring pART3-*thbSP* (Crewmate homolog) or -EV as a control. Top agar contains 0.05 % nicotine. n=1 biological replicate (not able to be reproduced).


A.
2880
2979
A. glob strain
-
+
++
-
+
++
nicotine
MW (kDa)
EV
SigAZ
116
97
66
57
55
42
35
29


B.
ladder
nicotinovarans
globiformis 2880
globiformis 2979
globiformis 4225
globiformis 24025
atrocyaneus
aurescens
citreus
crystallopoietes
humicola
orzae
pascens
ramosus
sulfureus
viscosus
hdnO
347 bp
gntR
446 bp
coxE
567 bp

C.

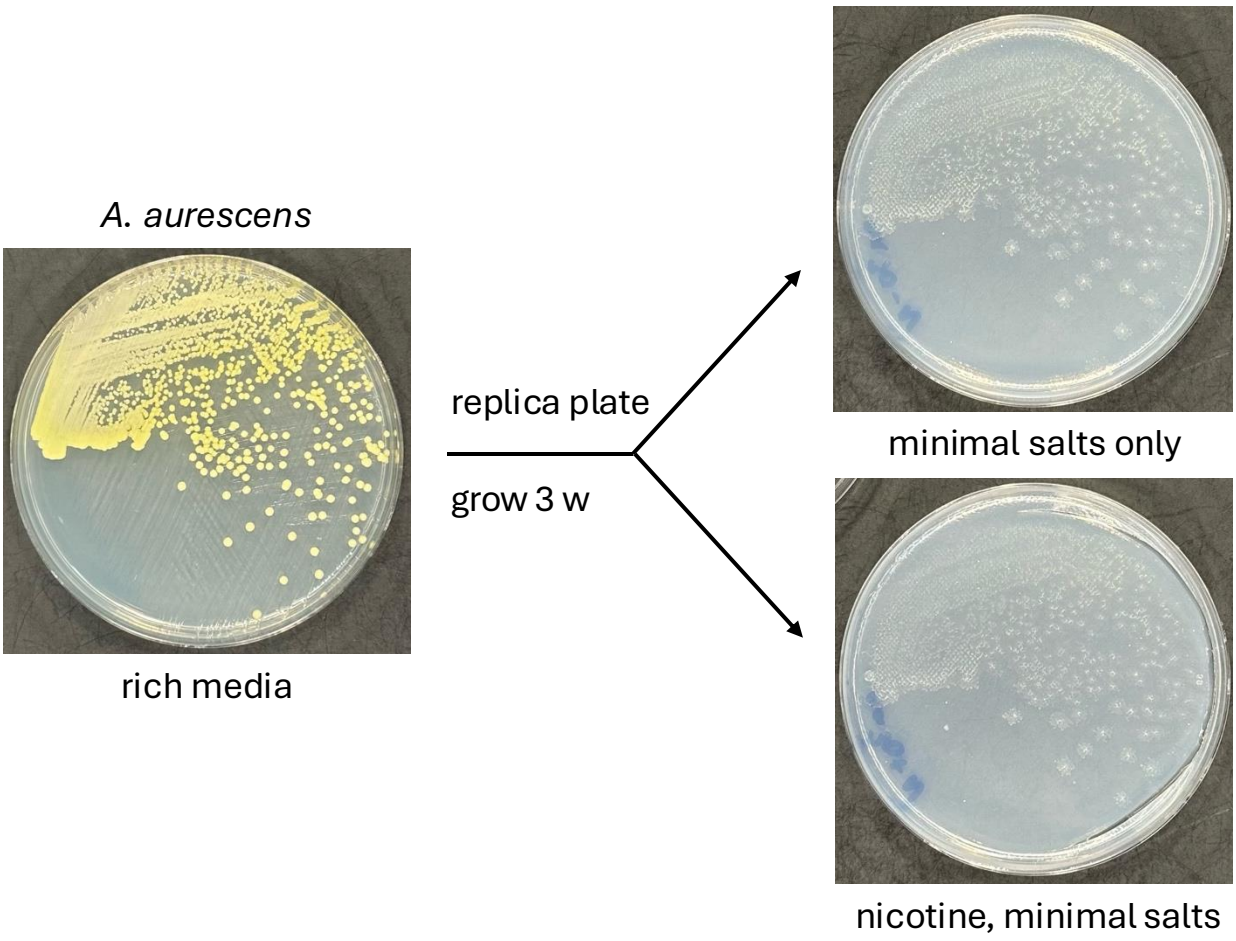


Figure 4. A two-pronged screen reveals *Arthrobacter* strains do not catabolize nicotine. A. pART3 is not substantially or dose dependantly induced in *A. glob.* Western blot on exponentially growing cultures of *A. glob* harbouring pART3-*sigAZ* (or -EV) induced with nicotine for five hours at 0.05 % (++), 0.01 % (+), or uninduced (-). Probed with anti-SigAZ antibody. n=1 biological replicate. **B.** Most *Arthrobacter* strains do not harbour pAO1. 1.2 % agarose gels with the products from mixed colony PCR on 15 strains with primers amplifying fragments of three pAO1 genes (*hdnO*, *gntR*, and *coxE*). The expected product sizes on each gel are indicated by arrows. The ladder has a fragment every 100 bp from 200-1000. Representative of n=2 biological replicates. **C.** Methodology and results of a screen for nicotine catabolism in *Arthrobacter* strains. M9 minimal salts (Sigma-Aldrich), 0.05 % nicotine. Result representative of the 14 strains tested. n=1 biological replicate.

In the most studied case, genes responsible for nicotine catabolism by *Arthrobacter* are encoded exogenously on the megaplasmid pAO1, whose sequence is available in GenBank. Hence, I designed three pairs of primers to amplify fragments of the *gntR*, *coxE*, and *hdnO* genes present on pAO1 and screened our *A. glob* strains for the plasmid by mixed colony PCR. (HdnO is the enzyme responsible for the first step in nicotine catabolism[30].)

I also screened other *Arthrobacter* species from our collection, hoping that if any were positive for pAO1 while *A. glob* 2880 was not, they might be able to act as a donor for its conjugation or transformation into 2880.

Of the 15 strains tested only *A. nicotinovarans*, a known harbourer of pAO1 that we recently procured, efficiently amplified products of the expected size by PCR (Figure 4B).

As some off target amplification did occur, I validated the nicotine catabolic activity of our strains by attempting to culture them on a medium with nicotine as the sole carbon source. Growth on such a medium has been previously reported possible by *A. nicotinovarans*[41]. Figure 4C outlines the method for this and shows a representative result: none of the 14 tested strains produced relevant growth on nicotine minimal media (*A. nicotinovarans* was not tested).

So, not only is nicotine catabolism not possible by our *A. glob* strains, these data also indicate it is not possible by *Arthrobacter* at large, rendering pART3 an ineffective option for inducible expression across the genus. The presence of pAO1 however, and thus nicotine catabolic activity, might vary between strains of the same species which could explain a previous report of induced expression from pART3 in *A. glob*[32].

3.3 SigAZ is cleaved *in vivo* during lysogeny

Next, I probed for the presence of SigAZ in lysogenic cells by western blot with my anti-SigAZ antibody. As revealed by Figure 5A., not only is SigAZ expressed in the Crewmate lysogen, but it only appears in a shorter, cleaved form. Based on a structural homology analysis with Foldseek[42], SigAZ contains only essential sigma factor domains and so cleavage should result in its inactivation as a transcriptional regulator[25]. This supports the idea of SigAZ as a driver of lysis that must be inactivated to maintain lysogeny.

As I cultured AZ lysogens to sample for western blotting, I noticed that many did not grow synchronously with the others or with *A. glob* 2880. Thinking that this might be due to some level of lytic phage replication occurring, I performed a spot test on the supernatants of lysogen cultures and indeed saw released phage (Figure 5B.). The levels varied drastically between lysogens, reaching 1e8 pfu/mL for Amyev and Pixelle but no more than 100 pfu/mL for ObiToo and Crewmate, and correlated with the

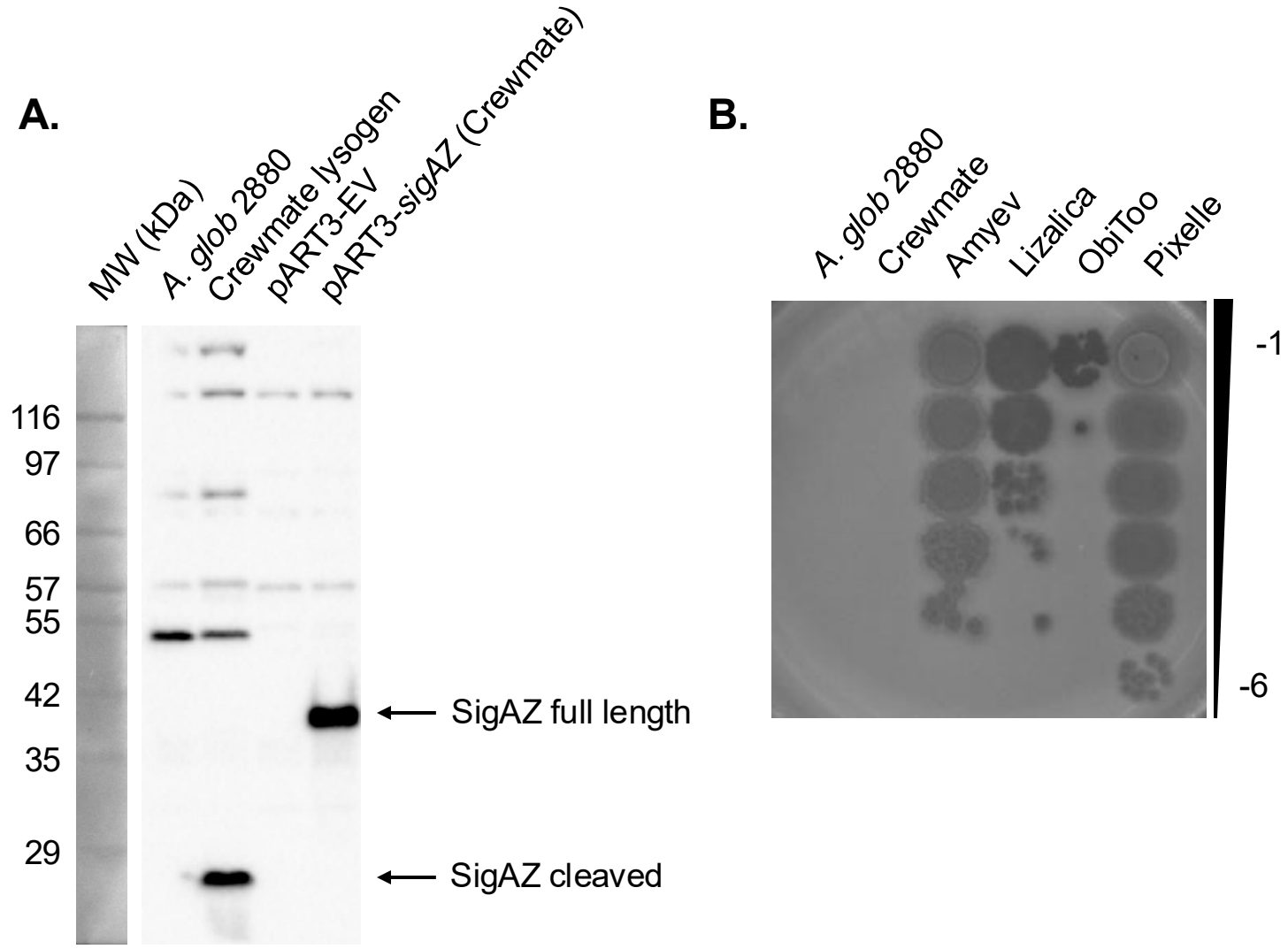


Figure 5. Two characteristics of AZ lysogens. A. SigAZ is cleaved during lysogeny in Crewmate. Western blot on a culture of Crewmate lysogen probing with anti-SigAZ antibody. Cultures of WT *A. glob* 2880, the native host of Crewmate, and 2880 transformed with either pART3-EV or -*sigAZ* serve as controls. Representative of n=2 biological replicates. **B.** AZ lysogens release varied amounts of phage. Spot assay on the filtered supernatants of exponentially growing lysogen culturesl. 10e-1 to 10e-6 plated on *A. glob* 2880. Representative of n=3 biological replicates.

differences I observed in growth rate. Overall, this demonstrates that spontaneous lysogeny exit is common among AZ phages.

3.4 A mechanism for AZ phage lysogeny maintenance and superinfection immunity

Based upon the information presented thus far, I propose a mechanism where the phage encoded sigma factor, SigAZ, co-opts the host RNA polymerase to direct transcription for lytic replication. To form and maintain lysogens, AZ phages produce ThbSP which binds to an 8 bp palindromic sequence in the *sigAZ* promotor and proteolytically cleaves SigAZ, preventing it from inducing its own transcription and thus attenuating the lytic signal. Prophage produced ThbSP also binds to the palindromic sequence in secondary infecting AZ phages, granting superinfection immunity. However, mutations in the palindromic sequence that prevent ThbSP binding allow for SI escape (Figure 6.).

Demonstrating ThbSP activity *in vitro* has remained elusive due to its poor solubility and probable dependence on Zn^{2+} cofactors that are readily stripped by several of the chemical agents commonly used in purification[43,44]. As any evidence linking ThbSP directly to SigAZ cleavage is missing, I indicated this part of my proposed mechanism with a question mark in Figure 6..

3.5 Structural modeling permits a more precise prediction of the ThbSP binding sequence

I used AlphaFold3 to create a structural model of two copies of ThbSP from Amyev, four Zn^{2+} ions, as well as the first fifty base pairs of the promotor upstream of *sigAZ* in Crewmate (Figure 7A).

AlphaFold3 predicts with high confidence that, much like the solved structure of human SPRTN[27], ThbSP will dimerize and bind to DNA, with each subunit sequestering one Zn^{2+} ion in its C-terminal domain, and another in the N-terminal domain at the HEXXH active site. However, unlike SPRTN, the N- and C-termini of each subunit straddle the DNA, connected by a linker which passes through the major groove. ThbSP also introduces a noticeable bend to DNA through its binding, through it's difficult to know whether this would be biologically relevant or if is just an artifact of the computer model.

In the prediction, ThbSP is bound directly to the 8 bp palindromic sequence defined by the SI escape mutants. I identified two fully conserved residues, Q156 and T158 in the linker, which appear to mediate sequence specific interaction by hydrogen bonding with nucleobases at 3 of 4 positions in one half of the palindrome per subunit (Figure 7A,B.).

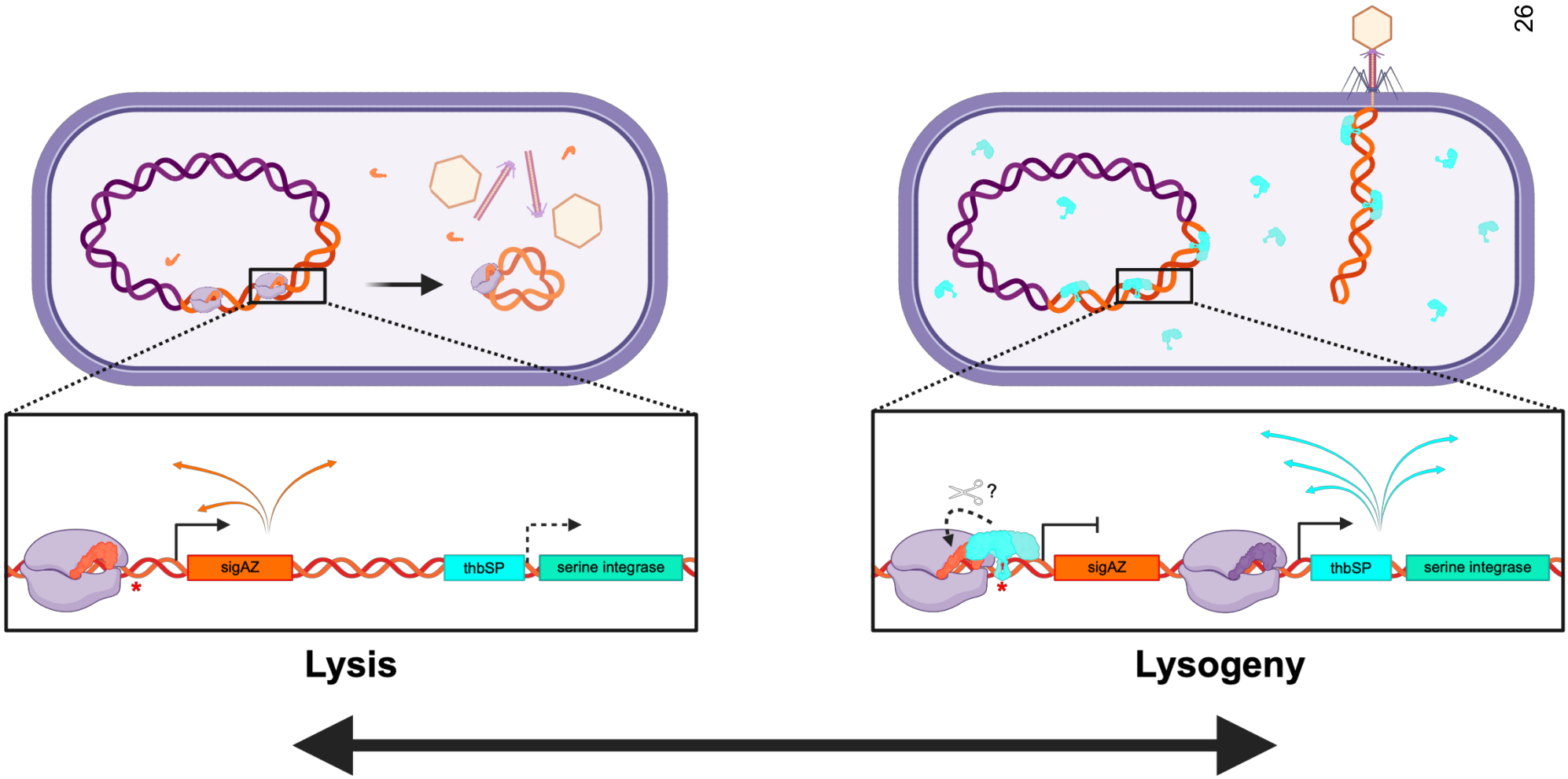


Figure 6. The AZ lysis-lysogeny equilibrium. Simplified maps of gene transcription resulting in lysis or lysogeny and superinfection immunity in AZ phages (not to scale). RNA polymerase is shown in purple with either a host sigma factor (purple) or SigAZ (orange). The red asterisks indicate the place where SI escape mutations prevent ThbSP binding. Created with BioRender.

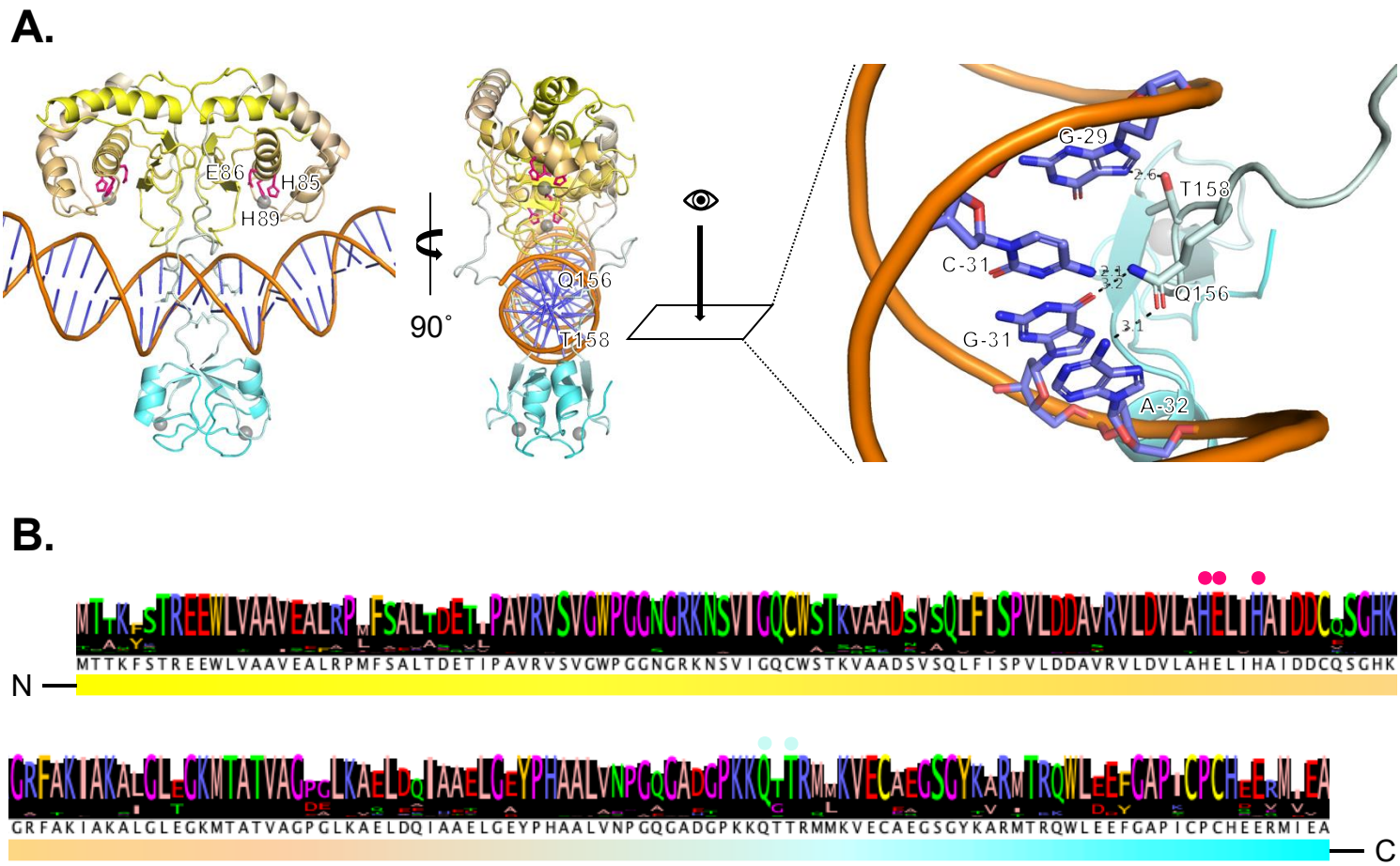


Figure 7. AGNCGNCT is a probable ThbSP binding site. A. AlphaFold prediction of Amyev ThbSP dimer bound to the Crewmate *sigAZ* promotor. Conserved residues of the active site and those predicted to form sequence-specific interactions with DNA are shown as sticks in pink and pale cyan respectively; Zn^{2+} ions are shown as spheres. In the final pane, one subunit of ThbSP along with non-interacting nucleobases were removed for clarity. Dashed lines denote measurements between atoms in Ångströms. **B.** ThbSP consensus sequence among 68 AZ phages. Fully conserved residues of the active site and those predicted to form sequence-specific interactions with DNA are marked with dots using to the same colour scheme as above**. C.** *sigAZ* promotor from ObiToo and Crewmate marked by shaded blue dots in positions where sequence-specific interactions with ThbSP are shown in A. Unshaded dots mark symmetrical positions interacting with the second ThbSP subunit. Red asterisks indicate the positions where SI escape mutations occurred; the position with two was affected in two independent mutants. Every tenth base is underlined.

Across both subunits, these interactions include nucleobases at 6 of 8 positions in the palindrome. When compared, 4 of 5 SI escape mutations map to the all the exact same positions considering the symmetry of the palindrome (Figure 7C.).

Not only does this further implicate ThbSP as responsible for SI but it enables a more precise prediction of ThbSP binding sequence as the quasipalindrome AGNCGNCT, where the unspecified base N is permitted at the two positions that are unaffected by SI escape mutations and do not form sequence specific interactions with ThbSP in the structural model.

3.6 The promotor upstream of *sigAZ* demonstrates a long history of natural evolution driven by opposing selective pressures for lysis or lysogeny

Through multi-sequence alignment and phylogenetic analyses, I examined and compared the promotors upstream of both *thbSP* and *sigAZ* across AZ phages. The *thbSP* promotor (P_*thbSP*) has a fully conserved -35 motif (TTACAC) which is similar to consensus -35 site TTGACA used by the housekeeping sigma factor in GC-rich bacterium *Thermus aquaticus*[45]. It also has a conserved set of GAGA repeats around the -10 site which resemble a Shine-Dalgarno sequence (Figure 8A.)[46]. Given these conserved elements, I suspect ThbSP expression is strong and driven primarily by a host sigma factor during lysogeny.

In comparison, the region upstream of *sigAZ* (P_*sigAZ*) is poorly conserved and lacks identifiable proximal promotor elements at the -10 and -35 sites. The sequence alignment is filled with far more long stretches of low nucleotide occupancy where insertions have occurred in only a few phages. And even in regions of the alignment with complete occupancy, no single nucleotide is completely conserved like in P_*thbSP* (Figure 8A.).

That said, P_*sigAZ* still has regions where conservation is enriched, up to 97% for some bases. In two of these regions, sequences resembling the ThbSP binding site appear, where in one case there is a common base substitution at one of the critical positions, and in the other it is mutated by an insertion in many phages.

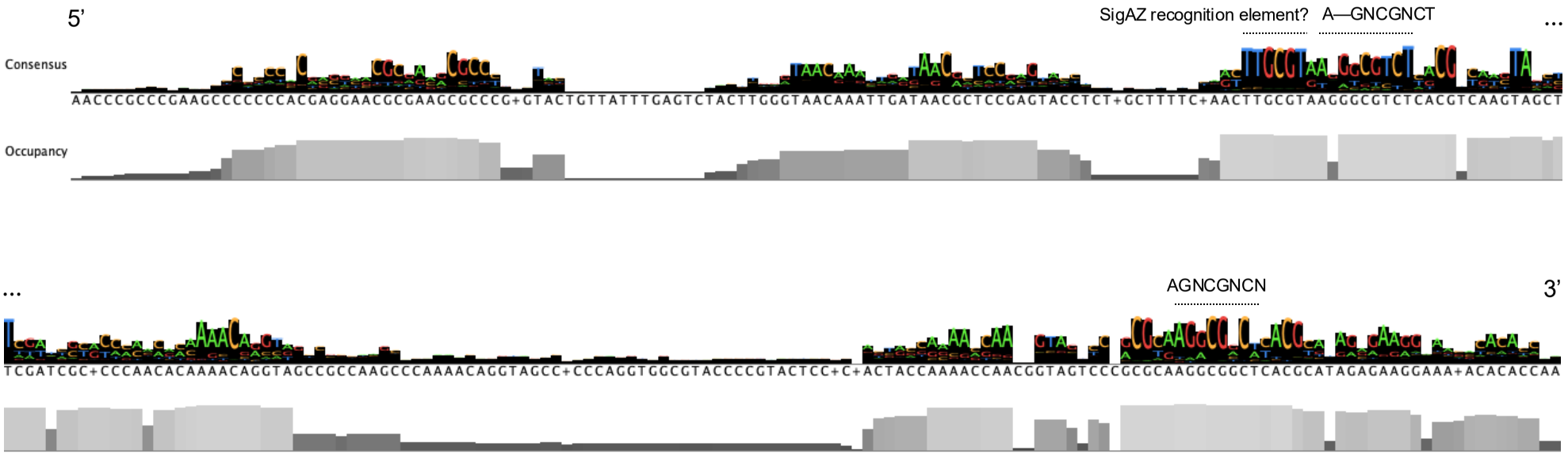
Immediately upstream of the ThbSP binding site affected by that insertion is the well conserved 6 bp sequence TTGCGT. Although it appears around -100 in P_*sigAZ*, I noticed this resembles common -35 motifs in the form of its sequence (TTNNNN), perhaps implicating it as part of a distal promotor for an alternative sigma factor like SigAZ.

This is also the site affected by the one SI escape mutation that did not appear directly within a ThbSP binding site (Figure 2C. The mutant substituted TTG<u>C</u>GT→TTG<u>T</u>GT). If the site indeed represents a SigAZ recognition element, the mutation may have offered

A.

P_*thbSP*

P_*sigAZ*



B.

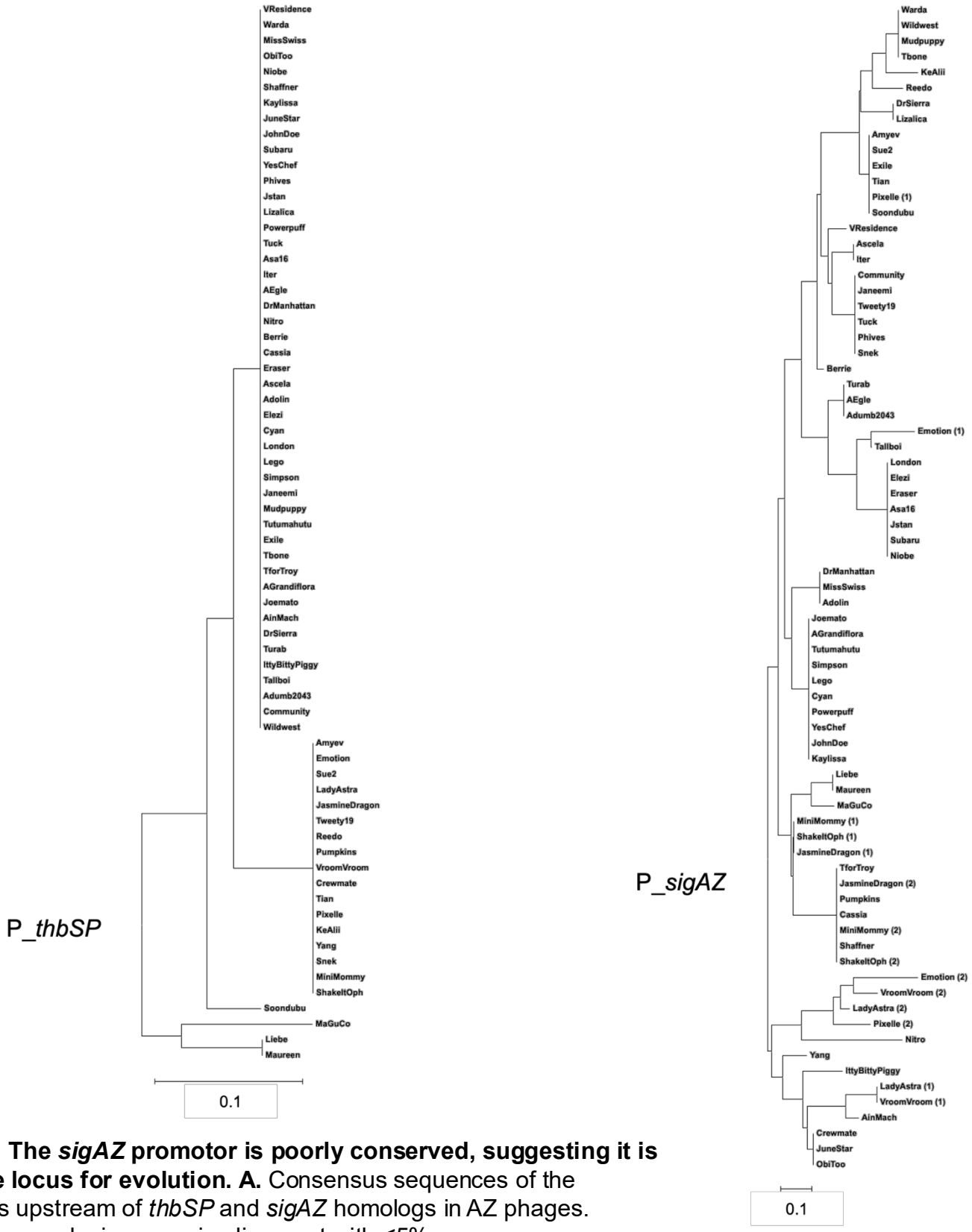


Figure 8. The *sigAZ* promotor is poorly conserved, suggesting it is an active locus for evolution. A. Consensus sequences of the promotors upstream of *thbSP* and *sigAZ* homologs in AZ phages. Sequences producing gaps in alignment with ≤5% occupancy were removed for clarity; the final consensus represents 52/68 P_*thbSP* sequences and 65/91 P_*sigAZ* sequences. Notable features are indicated above each consensus. **B.** Phylogenetic trees of the promotors upstream of *thbSP* and *sigAZ* homologs in AZ phages prepared in MEGA by neighbour joining. P_*sigAZ* sequences from the AZ5 subcluster were removed from the tree for clarity and comparability. In phages where a tandem duplication of *sigAZ* arose, the promotor upstream of the first (1) and second (2) homolog are indicated.

Figure 9. The *sigAZ* promotor is replete with evolutionary relics of the ThbSP binding site. The promotor upstream of all 91 *sigAZ* homologs in the AZ cluster. Predicted ThbSP binding sites are shown in dark blue. Near hits to the ThbSP binding site, allowing one nucleotide subsitution, are shown in light blue. In phages where a tandem duplication of *sigAZ* arose, the promotor upstream of the first (1) and second (2) homolog are indicated. Relevant individual phages form a separate block at the top.

SI escape by increasing SigAZ affinity, strengthening the promotor so SigAZ expression overpowered ThbSP repression and induced lysis.

This potential SigAZ recognition element also appears upstream of the proximal promotor of *thbSP* for a subset of phages, where SigAZ may upregulate ThbSP expression as an autoregulation mechanism to prevent premature exit from lysogeny.

I theorized that the general lack of P_*sigAZ* conservation stemmed from a history of evolutionary selection for all forms of SI escape mutations at this critical locus, where natural AZ phages who share mostly the same host range, must often compete with other AZ prophages over the same host. This evolution is clear when comparing the phylogeny of P_*sigAZ* to P_*thbSP*—across P_*sigAZ* there is evidence of far more evolutionary divergence (branching), and the evolutionary distance between diverged promotors is far greater (Figure 8B.).

Of course, the most common SI escape mutations we observed in the lab with Crewmate and ObiToo were base substitutions at the ThbSP binding site, and this seemed as though it may be broadly true based on a similar substitution appearing in the P_*sigAZ* consensus. To gain more insight, I decided to look at exact sequence matches to the ThbSP binding site (AGNCGNCT), as well as near matches, permitting one base substitution at the specified positions, across all *sigAZ* promotors (Figure 9.)

This revealed that almost every *sigAZ* promotor has at least one near match to the ThbSP binding site, excluding only promotors upstream of the second *sigAZ* homolog in a handful of AZ4 phages where a duplication likely arose as an alternative adaptation for SI escape. In only about half of the phages is the exact ThbSP site intact, and in those phages, it is more common to have multiple sites as opposed to just one, with one group of phages having four sites! I suspect that sets of multiple sites are evolutionarily stable because they allow for more autoregulation of ThbSP through autocleavage, allowing SI escape, as it has been shown in the human homolog, SPRTN, that autocleavage occurs in trans[47].

It is also clear from this that the most common position to have a ThbSP site is proximal to the *sigAZ* translation start site, around -30. This position, being downstream of the predicted SigAZ binding element near -100, likely offers stronger transcriptional repression of *sigAZ* than distal ThbSP sites upstream of -100. This is supported by the lysogen stability data in Figure 5B., where Amyev and Pixelle, with distal ThbSP sites, demonstrate far more spontaneous lysogeny exit than Crewmate and ObiToo, which have proximal sites.

Chapter 4: Discussion

4.1 ThbSP is an intrinsically adaptive lysogeny maintenance system suited for phage-phage competition and host immune evasion

In the Darwinian model of evolution, the entity that can best compete over resources in the natural environment is favoured by selection. For phages, and all viruses, that resource is their host. However, because no resource is limitless, true evolutionary longevity is awarded to the entity that can best compete against the others without depleting that resource for itself[48].

Take the example of a purely lytic phage in an isolated population of host—soon enough the host will be dead, and with nothing left to infect, that phage's branch of the evolutionary tree will have reached a terminus—extinction! So lysogeny has arisen as a resource conservation strategy, enabling evolutionary longevity in phages.

The equilibrium reached between lysis and lysogeny varies greatly between phages, influenced by the properties and growth characteristics of their host and in kind by the host's natural environment. To an extent, previously characterized lysogeny maintenance systems have been shown to be responsive to changes in the natural environment that affect their hosts, even the presence of other phages[13,49,50,51].

But such responsivity is insufficient to allow for viable competition between phages and prophages when they maintain lysogeny via the same system, like the AZs. For this, a lysogeny maintenance system must be proactive, allowing a given phage to simultaneously reap the benefits of lysogeny, while not being susceptible to superinfection immunity.

I posit that AZ phages have achieved this through selection for an intrinsically adaptable DNA site at which ThbSP mediates lysogeny maintenance and SI: a quasipalindromic sequence. It has been shown that quasipalindromic sequences as short as 7 nt can form stable minihairpins[52] which stall the polymerase during DNA synthesis and substantially increase the rate of base misalignment mutations[53]. This enables the AZ phages to produce diverse progeny at a greater proportion than is achieved by random mutation alone, ensuring that every time lytic replication occurs some progeny have a site that permits strong ThbSP binding and are more lysogenic, while others do not and are primed for lysis and SI escape.

4.1.2 AZ sequences are biased to appear with mutations in the ThbSP binding site

From this theory follows an explanation as to why we observe no intact ThbSP binding sites in the *sigAZ* promotor sequences of Lizalica and Warda, two phages for which we have isolated lysogens (Figure 9.). Leading up to genome sequencing, phages are amplified by taking advantage of their own natural lytic replication, then harvested and passed through a filter which physically prevents passage of any lysogenic host.

Through this process we artificially select for lytic phages, enriching their prevalence in our sequence data beyond what it likely is in nature. This is a critical limitation of single timepoint/single sample sequences—they only represent an infinitesimal slice of an infinite evolutionary timeline.

All to say, I suspect if we sequenced the Lizalica and Warda lysogens, we would find a ThbSP binding site perfectly intact.

4.1.3 The ThbSP binding sequence is well suited to adapt to anti-phage restriction by the host

Evolutionary selection for the ThbSP binding site may have also been influenced by its adaptability in response to restriction enzymes, a host anti-phage defense that targets and cleaves palindromic sequences in the phage genome to prevent phage replication. The inclusion of unspecified nucleotides in the binding site might ensure that phage progeny are not universally susceptible to the certain restriction systems that target strict palindromes.

A search through the currently unfinished genome sequences of *A. glob* with DefenceFinder[54] revealed 7 restriction related genes within three operons among the few anti-phage defenses present in 2979, giving some weight to this theory. There are however, known restriction systems that include unspecified nucleotides in their recognition sites so it is certainly possible that the host could develop counteradaptations that would allow it to efficiently target AZ phages at the ThbSP binding site.

4.2 ThbSP likely stabilizes and maintains lysogeny by acting at additional sites outside the *sigAZ* promotor

Although P_*sigAZ* is the critical locus of ThbSP mediated lysogeny maintenance, where it would directly repress *sigAZ* expression, I searched AZ phage sequences for other ThbSP binding sites and found a pattern in those that commonly appear. They exist in promotors or within genes essential for recircularization of the phage genome (integrase, recombination directionality factor[17]) and phage genome replication (primase/helicase, DNA polymerase) upon lysogeny exit. Figure 10. shows the sites in Crewmate and Pixelle as an example.

The appearance of sites within genes has led me to hypothesize that SigAZ may also cleave another RNA polymerase subunit than SigAZ to halt transcription elongation at these loci.

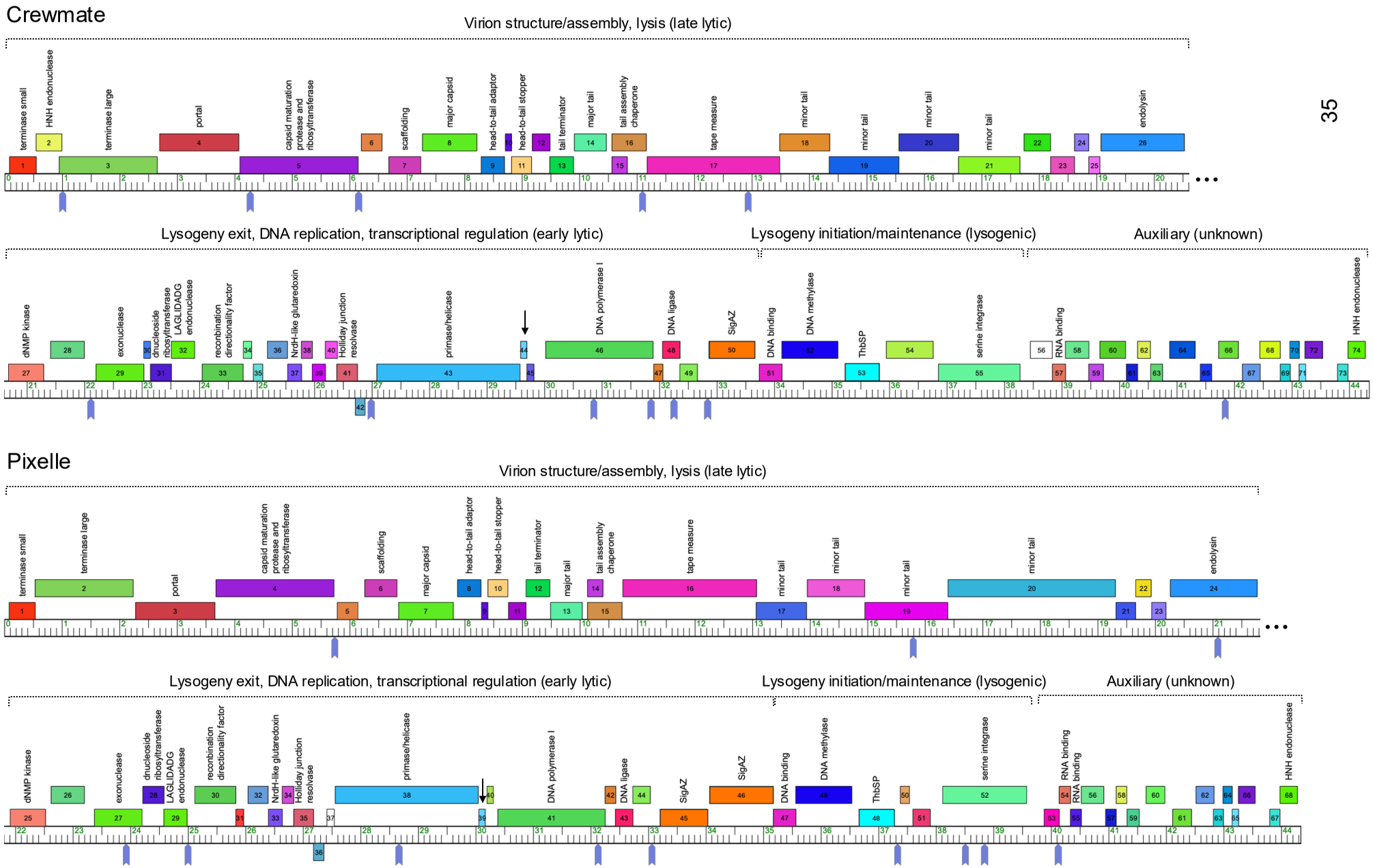


Figure 10. ThbSP likely represses lysogeny exit and stabilizes lysogens by acting at other sites in the prophage genome. Manually annotated genome maps of Crewmate and Pixelle from Phamerator. Groups of genes in each are labelled in one of four categories (early lytic, late lytic, lysogenic, and unknown) based on their presumed role in infection. Predicted ThbSP binding sites are indicated with blue flags. Black arrows denote homologs of the small open reading frame discussed in chapter 4.3. Scale bars represent nucleotide position in kbp.

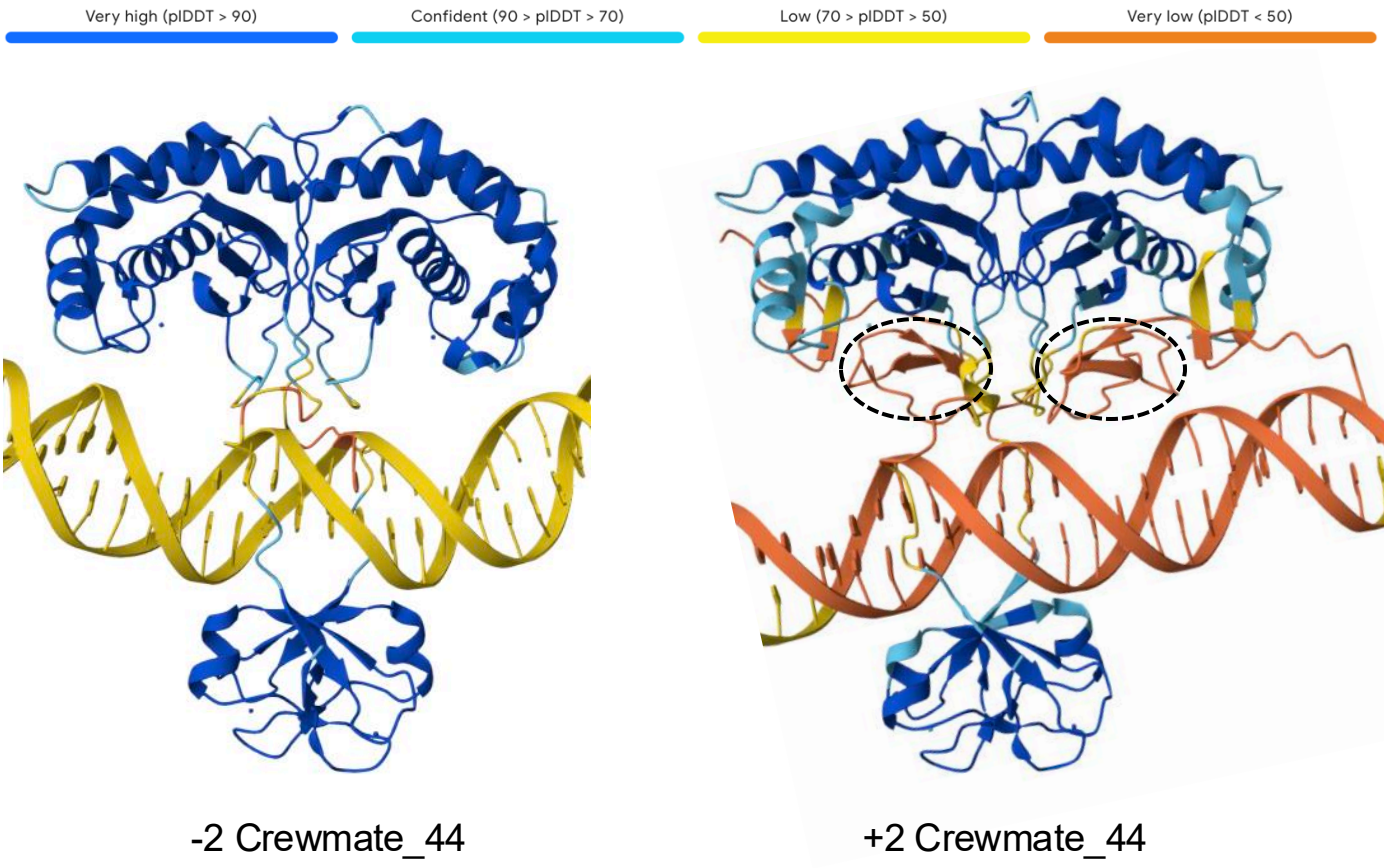


Figure 11. A small protein in the early lytic region may inactivite ThbSP to reinforce commitment to lysis. AlphaFold structures of the first 50 bp of the *sigAZ* promotor from Crewmate, two copies of ThbSP from Amyev, four Zn^{2}+, and with, or without two copies of the small protein from Crewmate (gene 44). Taken directly from the AlphaFold server with colouring by plDDT confidence scores.

Another well conserved site ThbSP binding site exists in the promotor of the gene following capsid maturation protease in the late lytic region of the genome. Although this gene has no known function, it may have a role in signaling and timing lysis

4.3 Commitment to lytic replication is likely reinforced by genes other than *sigAZ*

Thus far, discussion about the lysis-lysogeny equilibrium in AZ phages has remained near exclusively in the context of SigAZ and ThbSP as these are the two genes for which I have in lab data supporting their involvement. However, it would be facile not to acknowledge the roles that other phage genes surely play.

As an example, Silpe et. al. have demonstrated in some *Vibrio* phages that the commitment to DNA damage-independent lysis is linked to genes encoded by small open reading frames (~100 bp). There appear to be many genes like this in AZ phages, and one is conserved downstream of the primase/helicase in all but AZ4 and AZ5 phages, who already have either SigAZ duplications or ThbSP deletions respectively (homologs indicated by arrows in Figure 10.). An AlphaFold prediction shows that the protein produced by this small open reading frame may associate and occlude the ThbSP active site while destabilizing its interaction with DNA, inactivating it and solidifying the commitment to lysis (Figure 11). This could also be the reason for the pattern of ThbSP binding sites that appear in or upstream of the primase/helicase.

4.4 SigAZ and ThbSP outside of the AZ cluster

Homologs of SigAZ and ThbSP are co-conserved upstream of the integrase in EH cluster phages infecting *Microbacterium* where they likely also mediate lysis and lysogeny via the same mechanism as in the AZ1-AZ4 phages.

SigAZ too appears in EB phages infecting *Microbacterium* and *Curtobacterium* and in one FP phage infecting *Arthrobacter*. These phages have a similar genome architecture to the AZ phages, but in all cases the phages are missing at least one critical element for the transmission or maintenance of lysogeny, either an integrase or ThbSP. In this way they resemble the AZ5 phages and have been classified as lytic.

Homologs of ThbSP exist alone in phages of the A, AB, BE, BK, BR, C, CG, CT, DR, FC, GC, and K clusters infecting hosts of the genera *Mycobacterium*, *Curtobacterium*, *Streptomyces*, *Rhodococcus*, *Gordonia*, *Arthrobacter*, and *Microbacterium* where they have probably evolved diverse roles in transcriptional regulation. All this information is available from phagesdb[22] and Phamerator[55].

Chapter 5: Future directions

Like any worthwhile scientific endeavour, I'm left with more questions than answers. To name a few: what are all the genes regulated by SigAZ? Does this include any host genes? Does ThbSP prefer some nucleotides at the unspecified positions over others? What other AZ genes reinforce the commitment to lysis or lysogeny?

To continue my work and answer some of these, the most natural starting point would be to reengineer pART3 (or another plasmid) to permit robust, dose dependant induction of gene expression in *A. glob.* Then, use that plasmid to verify the immunity previously shown by ThbSP against phage infection (Figure 3B.). As follow up, it could be tested whether encoding the small open reading frame discussed in chapter 4.3, or several ThbSP binding sites on the same plasmid as ThbSP nullifies the phenotype on infection by inactivating or titrating away the protease.

This reengineered plasmid would also be important for assessing the effect of other AZ genes on infection and broader use for genetics in *Arthrobacter*. As a specific workaround to repeat the ThbSP immunity phenotype, I've already begun cloning it into pART3 along with its native promotor.

Next, the purification of ThbSP should be optimized to demonstrate its activity against SigAZ *in vitro*. Site directed mutagenesis of ThbSP would be used to produce a catalytic dead form as a control. Additionally, mutants that prevent site specific DNA binding and dimerization would be generated to verify their essentiality for activity. The purified protease would also be used in gel shifts to determine its binding affinity to the quasipalindromic site and mutated forms of the site, as well as assess its preference for different nucleotides at the two unspecified positions in the site.

RNA sequencing could be performed on lysogens, and at timepoints during liquid infection of the host, with or without the overexpression of SigAZ, to gain a deeper understanding about the influence of SigAZ on transcription as well as the transcriptional profile of lysogens. It would also deepen understanding of the host, *A. glob*, in general.

As a prerequisite for RNA seq, genome sequences of the host would need to be finished. At the same time, lysogens could also be sequenced to determine the prophage integration locus and verify if they retain an intact ThbSP binding site.

Chapter 6: Afterword

When faced with the mounting antimicrobial resistance crisis, phages present a promising therapeutic path forward. Administered virions have demonstrated capable of disseminating within the body and clearing even systemic pan-drug resistant bacterial infections with few to no side effects[56]. That is, lytic phage virions. Lysogenic phages are avoided for therapeutic use as they pose risks of transducing further antibiotic resistance or toxin genes to the pathogen[11,57].

Today, phages are ruled out as therapeutic candidates if they so much as encode a prophage transmission system like an integrase. However, it is clear that mutations readily form in phage lysogen maintenance systems which despite the integrase, make them effectively lytic. At a time when simply finding phages that infect clinical isolates is a substantial barrier to therapy, we cannot afford to be ruling out promising candidate phages in this way. With a better understanding of diverse lysogen maintenance systems, we could confidently include more phages in our therapeutic arsenal.

Additionally, a better understanding of lysogen maintenance could allow us to engineer therapeutic phages that escape superinfection immunity, which is relevant as most pathogenic bacteria themselves are lysogens or polylysogens.

So, I put forth this thesis as my small contribution to the human fight against antimicrobial resistance through phages. And as a record of the fascinating evolution in these viruses who will outlast us all on planet earth[58].

Bibliography

1. Forterre, P. The virocell concept and environmental microbiology. *ISME J.* **7**, 233–236 (2013).
2. Fermin, G. Host Range, Host–Virus Interactions, and Virus Transmission. *Viruses* 101–134 (2018) doi:10.1016/B978-0-12-811257-1.00005-X.
3. Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proc. Natl. Acad. Sci.* **96**, 2192–2197 (1999).
4. Keen, E. C. A century of phage research: Bacteriophages and the shaping of modern biology. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **37**, 6–9 (2015).
5. Suttle, C. A. Marine viruses — major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
6. Danovaro, R. *et al.* Marine viruses and global climate change. *FEMS Microbiol. Rev.* **35**, 993–1034 (2011).
7. Georjon, H. & Bernheim, A. The highly diverse antiphage defence systems of bacteria. *Nat. Rev. Microbiol.* **21**, 686–700 (2023).
8. Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nat. Rev. Microbiol.* **3**, 504–510 (2005).
9. Hatfull, G. F. & Hendrix, R. W. Bacteriophages and their Genomes. *Curr. Opin. Virol.* **1**, 298–303 (2011).
10. Jacoby, M. L., Hogg, G. D., Assaad, M. R. & Williamson, K. E. Seasonal trends in lysogeny in an Appalachian oak-hickory forest soil. *Appl. Environ. Microbiol.* **90**, e01408-23 (2023).
11. Wendling, C. C., Refardt, D. & Hall, A. R. Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. *Evol. Int. J. Org. Evol.* **75**, 515–528 (2021).
12. Owen, S. V. *et al.* A window into lysogeny: revealing temperate phage biology with transcriptomics. *Microb. Genomics* **6**, e000330 (2020).
13. Galkin, V. E. *et al.* Cleavage of bacteriophage lambda cI repressor involves the RecA C-terminal domain. *J. Mol. Biol.* **385**, 779–787 (2009).
14. Taslem Mourosi, J. *et al.* Understanding Bacteriophage Tail Fiber Interaction with Host Surface Receptor: The Key "Blueprint" for Reprogramming Phage Host Range. *Int. J. Mol. Sci.* **23**, 12146 (2022).
15. Wang, X., Llopis, P. M. & Rudner, D. Z. Organization and segregation of bacterial chromosomes. *Nat. Rev. Genet.* **14**, 191–203 (2013).
16. Piligrimova, E. G. *et al.* Putative plasmid prophages of Bacillus cereus sensu lato may hold the key to undiscovered phage diversity. *Sci. Rep.* **11**, 7611 (2021).
17. Lewis, J. A. & Hatfull, G. F. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res.* **29**, 2205–2216 (2001).

18. McGinnis, R. J. *et al.* A monomeric mycobacteriophage immunity repressor utilizes two domains to recognize an asymmetric DNA sequence. *Nat. Commun.* **13**, 4105 (2022).
19. Ptashne, M. *A Genetic Switch: Phage Lambda Revisited*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004).
20. Guler, P. *et al.* Arbitrium communication controls phage lysogeny through non-lethal modulation of a host toxin–antitoxin defence system. *Nat. Microbiol.* **9**, 150–160 (2024).
21. Mavrich, T. N. & Hatfull, G. F. Evolution of Superinfection Immunity in Cluster A Mycobacteriophages. *mBio* **10**, (2019).
22. Russell, D. A. & Hatfull, G. F. PhagesDB: the actinobacteriophage database. *Bioinformatics* **33**, 784–786 (2017).
23. Gauthier, C. H. & Hatfull, G. F. PhamClust: a phage genome clustering tool using proteomic equivalence. *mSystems* **8**, e00443-23 (2023).
24. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
25. Paget, M. S. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules* **5**, 1245–1265 (2015).
26. Schwartz, D., Lehmkuhl, B. & Lennon, J. Phage-Encoded Sigma Factors Alter Bacterial Dormancy. *mSphere* **7**, (2022).
27. Li, F., Raczynska, J. E., Chen, Z. & Yu, H. Structural Insight into DNA-Dependent Activation of Human Metalloprotease Spartan. *Cell Rep.* **26**, 3336-3346.e4 (2019).
28. Ghimire, N., Kim, B., Han, S.-R. & Oh, T.-J. Comparative genomics based exploration of xenobiotic degradation patterns in *Glutamicibacter*, *Arthrobacter*, and *Pseudarthrobacter* isolated from diverse ecological habitats. *Heliyon* **10**, e40280 (2024).
29. Ren, L., Shi, Y., Jia, Y. & Yan, Y. Genome Sequence of Arthrobacter sp. YC-RL1, an Aromatic Compound-Degrading Bacterium. *Genome Announc.* **3**, e00749-15 (2015).
30. Brandsch, R. Microbiology and biochemistry of nicotine degradation. *Appl. Microbiol. Biotechnol.* **69**, 493–498 (2006).
31. Sandu, C., Chiribau, C. B. & Brandsch, R. Characterization of HdnoR, the Transcriptional Repressor of the 6-Hydroxy-D-nicotine Oxidase Gene of Arthrobacter nicotinovorans pAO1, and its DNA-binding Activity in Response to L- and D-Nicotine Derivatives *. *J. Biol. Chem.* **278**, 51307–51315 (2003).
32. Sandu, C., Chiribau, C.-B., Sachelaru, P. & Brandsch, R. Plasmids for Nicotine-Dependent and -Independent Gene Expression in Arthrobacter nicotinovorans and Other Arthrobacter Species. *Appl. Environ. Microbiol.* **71**, 8920–8924 (2005).

33. Zhang, H., Li, Y., Chen, X., Sheng, H. & An, L. Optimization of electroporation conditions for *Arthrobacter* with plasmid PART2. *J. Microbiol. Methods* **84**, 114–120 (2011).
34. Untergasser, A. *et al.* Primer3—new capabilities and interfaces. *Nucleic Acids Res.* **40**, e115 (2012).
35. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
36. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
37. Madeira, F. *et al.* The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* **52**, W521–W525 (2024).
38. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
39. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
40. Ganas, P. & Brandsch, R. Uptake of l-nicotine and of 6-hydroxy-l-nicotine by Arthrobacter nicotinovorans and by Escherichia coli is mediated by facilitated diffusion and not by passive diffusion or active transport. *Microbiology* **155**, 1866–1877 (2009).
41. Igloi, G. L. & Brandsch, R. Sequence of the 165-Kilobase Catabolic Plasmid pAO1 from Arthrobacter nicotinovorans and Identification of a pAO1-Dependent Nicotine Uptake System. *J. Bacteriol.* **185**, 1976 (2003).
42. van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
43. Krężel, A. *et al.* Coordination of heavy metals by dithiothreitol, a commonly used thiol group protectant. *J. Inorg. Biochem.* **84**, 77–88 (2001).
44. Kim, Y. J. *et al.* Purifying Properly Folded Cysteine-rich, Zinc Finger Containing Recombinant Proteins for Structural Drug Targeting Studies: the CH1 Domain of p300 as a Case Example. *Bio-Protoc.* **7**, e2537 (2017).
45. Campbell, E. A. *et al.* Structure of the Bacterial RNA Polymerase Promoter Specificity σ Subunit. *Mol. Cell* **9**, 527–539 (2002).
46. Ma, J., Campbell, A. & Karlin, S. Correlations between Shine-Dalgarno Sequences and Gene Features Such as Predicted Expression Levels and Operon Structures. *J. Bacteriol.* **184**, 5733–5745 (2002).
47. Stingele, J. *et al.* Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol. Cell* **64**, 688–703 (2016).
48. Darwin, C. *The Origin of Species*. (Random House, New York, NY).
49. Silpe, J. E. & Bassler, B. L. A Host-Produced Quorum-Sensing Autoinducer Controls a Phage Lysis-Lysogeny Decision. *Cell* **176**, 268-280.e13 (2019).

50. Zamora-Caballero, S. *et al.* Antagonistic interactions between phage and host factors control arbitrium lysis–lysogeny decision. *Nat. Microbiol.* **9**, 161–172 (2024).
51. Silpe, J. E. *et al.* Small protein modules dictate prophage fates during polylysogeny. *Nature* **620**, 625–633 (2023).
52. Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K. & Miura, K. Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. *Nucleic Acids Res.* **20**, 3891–3896 (1992).
53. Klaric, J. A., Perr, E. L. & Lovett, S. T. Identifying Small Molecules That Promote Quasipalindrome-Associated Template-Switch Mutations in Escherichia coli. *G3 GenesGenomesGenetics* **10**, 1809–1815 (2020).
54. Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat. Commun.* **13**, 2561 (2022).
55. Cresawn, S. G. *et al.* Phamerator: a bioinformatic tool for comparative bacteriophage genomics. *BMC Bioinformatics* **12**, 395 (2011).
56. Lin, D. M., Koskella, B. & Lin, H. C. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World J. Gastrointest. Pharmacol. Ther.* **8**, 162–173 (2017).
57. Fidelma Boyd, E. & Waldor, M. K. Alternative Mechanism of Cholera Toxin Acquisition by Vibrio cholerae: Generalized Transduction of CTXΦ by Bacteriophage CP-T1. *Infect. Immun.* **67**, 5898–5905 (1999).
58. Vonnegut, K. Hocus Pocus. in 198–205 (The Berkley Publishing Group, New York, NY, 1990).