**Master's Thesis**

# First evidence of Cas9 expression during CRISPR-Cas containing megaphage infection

**Meeri Niemi**

**University of Jyväskylä**

Department of Biological and Environmental Science

08 April 2024

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Mathematics and Science
Department of Biological and Environmental Science
Master's Degree Programme in Cell and Molecular Biology

Niemi Meeri

First evidence of Cas9 expression during CRISPR-Cas containing megaphage infection

MSci Thesis
42 p., 3 appendices (11 p.)
Supervisors: Assistant Professor Elina Laanto and Professor Lotta-Riina Sundberg
Reviewers: PhD Anniina Runtuvuori-Salmela and PhD Matti Jalasvuori
April 2024

Keywords: flavobacteria, host-phage interaction, host range, megaphage, phage morphology, viral life cycle

Bacteriophages, or phages, are viruses that infect bacteria. Recently, many megaphages with large (over 500 kilo base pairs) genomes have been discovered. Metagenomic analyses show that several phages, including some megaphages, encode a CRISPR-Cas system, an adaptive immune system commonly associated with prokaryotes. These systems consist of CRISPR repeat-spacer arrays and Cas proteins that enable the system to target foreign genetic material in a sequence-specific manner. Some phage-encoded CRISPR-Cas systems have been shown to target host bacterium genome, while others are thought to target competing phages, but empirical evidence is scarce. This research seeks to elucidate the functionality of CRISPR-Cas systems in megaphages, which often exhibit incomplete systems with unclear purpose. In this study, five megaphages (genome size over 600 kilo base pairs) isolated from Lake Jyväsjärvi were studied. Phage host range was determined with titration experiments, and the phages were found to primarily infect species of *Flavobacterium*. The life cycle of phage Elf16 was determined by measuring host optical density, calculating plaque forming units during infection, and by performing an adsorption to host test. The Elf16 life cycle was found to be lytic, with lysis occurring gradually and starting at roughly three hours post-infection. Thin section samples of the phage infection were imaged using transmission electron microscopy, and the phage particles were shown to be large and have a contractile tail. Images also depict the Elf16 life cycle, which matches the life cycle demonstrated by measuring the optical density and plaque forming units. Elf16 genome was analysed, but no match was found between the CRISPR spacers of the phage and the genome of its isolation host. It is possible that the purpose of the phage CRIPSR-system is to attack host genome, but it is not a requisite for a successful infection. To assess the functionality of the Elf16 CRISPR-Cas system, Cas9 gene expression was measured using qPCR. The findings suggested significant Cas9 expression during early stages of infection, therefore hinting at a potentially functional system.

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta
Bio- ja ympäristötieteiden laitos
Solu- ja molekyylibiologian maisteriohjelma

| | |
|---|---|
| Niemi Meeri | Cas9 geenin ilmentyminen CRISPR-Cas järjestelemän omaavan megafaagin infektion aikana |
| Pro gradu tutkielma: | 42 s., 3 liitettä (11 s.) |
| Työn ohjaajat: | Apulaisprofessori Elina Laanto ja Professori Lotta-Riina Sundberg |
| Tarkastajat: | FT Anniina Runtuvuori-Salmela ja FT Matti Jalasvuori |
| Huhtikuu 2024 | |

Hakusanat: faagimorfologia, flavobakteeri, isäntä-faagi vuorovaikutus, isäntäkirjo, megafaagi, viruksen elinkierto

Bakteriofagit eli faagit ovat viruksia, jotka infektoivat bakteereja. Viime aikoina on löydetty useita megafaageja, joilla on suuri (yli 500 kiloemäsparia) genomi. Metagenomianalyyseissä on löydetty faageja, myös joitakin megafaageja, joilla on CRISPR-Cas järjestelmä, adaptiivinen immuunijärjestelmä, joka on yleinen esitumallisilla. CRISPR-järjestelmä koostuu CRISPR-sekvenssistä ja Cas-geeneistä, joiden avulla bakteeri voi tunnistaa ja tuhota vierasta geneettistä materiaalia sekvenssin tarkkuudella. Tämä tutkimus pyrkii selvittämään CRISPR-järjestelmän toimivuutta megafaageissa. Niiden CRISPR-järjestelmä on usein vajavainen, ja sen toimintaa ei juurikaan tunneta. Joidenkin faagien CRISPR-järjestelmien on osoitettu tuhoavan isäntäbakteerien genomia, ja joidenkin arvellaan tuhoavan kilpailevien faagien genomia, mutta empiirinen näyttö on vähäistä. Tässä työssä tutkittiin viittä Jyväsjärvestä eristettyä megafaagia (genomi yli 600 kiloemäsparia). Faagien isäntäkirjo selvitettiin tippatitrauksella. Todettiin, että faagit infektoivat pääasiassa eri *Flavobacterium* lajeja. Faagin Elf16 elinkierto selvitettiin mittaamalla optista tiheyttä, laskemalla plakin muodostavien yksiköiden määrä infektion aikana ja suorittamalla adsorptiokoe. Faagin elinkierto osoitettiin lyyttiseksi. Hitaasti etenevä lyysis alkaa noin kolme tuntia infektion alun jälkeen. Infektion aikana otetuista näytteistä tehtiin ohutleikkeitä, jotka kuvattiin läpäisyelektronimikroskoopilla. Faagipartikkelit olivat suuria, ja ne omasivat kontraktiilin hännän. Kuvat osoittavat myös faagin elinkierron, joka täsmää muissa mittauksissa todettuun kiertoon. Faagin Elf16 genomia analysoitiin, mutta ei havaittu vastaavuutta faagin CRISPR spacereiden ja isäntäbakteerin genomin välillä. On mahdollista, että faagin CRISPR-järjestelmän tarkoitus on leikata isännän genomia, mutta se ei kuitenkaan ole edellytys onnistuneelle infektiolle. Faagin Elf16 CRISPR-järjestelmän toimivuutta testattiin mittaamalla Cas9-geenin ilmentymistä qPCR:n avulla. Cas9:ää ilmentyy infektion alkuvaiheessa, mikä saattaa viitata toiminalliseen CRISPR-Cas järjestelmään.

# TABLE OF CONTENTS


**1 INTRODUCTION ..... 1**

1.1 Phage life cycle ..... 1
1.2 Phages in nature ..... 3
1.3 Huge phages ..... 4
1.4 CRISPR-Cas systems ..... 5
1.4.1 Phage-encoded CRISPR-Cas systems ..... 7
1.5 Research questions and hypotheses ..... 9

**2 MATERIALS AND METHODS ..... 10**

2.1 Materials ..... 10
2.2 Methods ..... 11
2.2.1 Host range testing ..... 11
2.2.2 AMG and CRISPR-Cas analysis of Elf16 genome ..... 12
2.2.3 Determination of optimal conditions for Elf16 infection ..... 13
2.2.4 OD development of strain B330 over an extended time period ..... 14
2.2.5 Elf16 adsorption to isolation host *Flavobacterium* sp. B330 ..... 14
2.2.6 CRISPR-Cas system containing megaphage life cycle ..... 15

**3 RESULTS ..... 17**

3.1 Host specificity of studied phages ..... 17
3.2 Elf16 spacers have no match to isolation host genome ..... 20
3.3 Potential AMGs found in the Elf16 genome ..... 21
3.4 Optimal conditions for Elf16 infection ..... 22
3.5 Elf16 life cycle is lytic ..... 24
3.5.1 Host OD continues to decrease for several hours after Elf16 infection ..... 25
3.6 Elf16 life cycle imaged from adsorption to lysis ..... 26
3.7 Megaphage Elf16 encoded Cas9 is expressed at the early stages of infection ..... 27

**4 DISCUSSION ..... 29**

4.1 Studied phages mostly infect *Flavobacterium* sp. ..... 29
4.2 Elf16 latent phase is long ..... 30
4.3 Correlation between Elf16 genome and capsid size ..... 32
4.4 Cas9 is expressed during early infection ..... 33
4.5 Genome analysis of Elf16 phage-host interactions ..... 34

**5 CONCLUSIONS ..... 34**

**ACKNOWLEDGEMENTS ..... 36**

**REFERENCES ..... 37**

**APPENDIX 1 FULL HOST RANGE RESULTS ..... 45**

**APPENDIX 2 ELF16 SPACER SEQUENCES ..... 52**

**APPENDIX 3 ELF16 POTENTIAL AMGS ..... 53**

# TERMS AND ABBREVIATIONS

## Terms

| | |
|---|---|
| **(Bacterio)phage** | A virus that infects bacteria |
| **Megaphage** | A phage with a genome size of over 500 kbp |
| **Virion** | A complete virus particle |

## Abbreviations

| | |
|---|---|
| **AMG** | auxiliary metabolic genes. |
| **CRISPR-Cas** | clustered regularly interspaced short palindromic repeats and CRISPR associated proteins |
| **crRNA** | CRISPR RNA |
| **Pfu** | plaque forming units |
| **qPCR** | quantitative polymerase chain reaction |
| **TEM** | transmission electron microscopy |

# 1 INTRODUCTION

The existence of bacteriophages or "phages", viruses infecting bacteria, has been known for over a hundred years, but their abundance and diversity is still not fully comprehended. Phages are likely the most abundant biological entities on earth, and there is significant variation in their morphology, genomes, and life cycles. As such, phage classification is based on both genomic and morphological factors (Ackermann 2009). Though there is morphological variation in phages, most known phage virions are tailed. Of these virions, most have a long, noncontractile tail, but phages with a contractile tail or a short stubby tail are also known. Polyhedral, filamentous, or pleomorphic phage virions have also been imaged (Ackermann 2007). Altogether, most known phages have double-stranded DNA (dsDNA) (Brum et al. 2013), although phages with single-stranded DNA (ssDNA) (Lim et al. 2015), single-stranded RNA (ssRNA) (Loeb and Zinder 1961), and double-stranded RNA (dsRNA) (Mertens 2004) are also known. Genome size also varies among phages, from the smallest ssRNA phages with a genome of only 3.8 kilo base pairs (kbp) (Friedman et al. 2009) to megaphages which have genomes exceeding 500 kbp (Devoto et al. 2019). For perspective, the current estimate of average phage genome size is approximately 55 kbp, based on metagenomic analysis (Al-Shayeb et al. 2020). Phages are abundant in many different environments, including soil and aquatic environments. In these environments, they have many roles, such as nutrient cycling (Fuhrman 1999).

## 1.1 Phage life cycle

In addition to structural differences, phage life cycles also vary. Most often phage life cycles are either temperate or lytic. In a temperate life cycle, the phage's genomic information is inserted into the host genome, making the phage a

prophage. As the host bacterium divides, the prophages multiply as their genetic information is transferred to every daughter cell. Often the lysogenetic cycle is only an intermediate stage, and stress to the host can activate the prophage to enter the lytic life cycle (Kohm and Hertel 2021). A lytic life cycle means that after infecting bacteria, phages use the host's replication machinery to make more phages, and then burst out of the host cell, killing it in the process (Young 1992). The phage infection cycle begins with the attachment of the phage to a specific receptor on the surface of its bacterial host. The attachment is very specific, and phage host ranges vary greatly (Dennehy and Abedon 2021a). Phage attachment begins with the initial reversible attachment, which is then followed by irreversible attachment (Nobrega et al. 2018). As mentioned before, most phage virions are tailed. This tail plays a crucial role in the infection process and may even have specialised tail fibers that aid in the attachment of the phage. Once the phage has attached to the bacterial receptors, conformational changes in the tail structure are triggered. These changes can be contraction or reorganization of the tail structure. For example, the baseplate of the phage tail can undergo a contraction, which exerts pressure on the tail sheath. This contraction builds up energy. This energy is then utilized for the injection, in which the genetic material of the phage is inserted into the bacterium. In the injection, part of the tail acts as conduit for the viral genome as it is delivered into the cell. It can also aid in the puncture of the bacterial cell wall (Hardy et al. 2022).

At this point of the infection cycle, the viral genome is in the cytoplasm of the host cell. Because phages lack their own replication machinery, they take control of the host cell's machinery to enable replication. In a lytic life cycle, the phage's genetic material directs the host cell's DNA replication machinery to replicate the viral genome. The replication of phage genome may outcompete the replication of the host cell's own genome (Dennehy and Abedon 2021b). One enzyme with a vital role in phage genome replication is helicase. This enzyme functions to unwind the double-stranded nucleic acid helix, separating it into its individual strands. During the replication phase of a phage infection, this

unwinding allows for the synthesis of new viral DNA strands (Perumal et al. 2010). During viral genome replication, so called early genes are usually transcribed first. These genes can include nucleases that degrade host DNA, proteins that modify the host cell's transcription and translation machinery, or inhibit host cell transcription, translation, or other essential cellular functions. This ensures that the host cell's resources are redirected to support phage replication. As the infection progresses, the expression of late genes is triggered. Late genes encode structural proteins necessary for the assembly of new phage particles (Dennehy and Abedon 2021b).

Towards the end of the infection cycle, the newly synthesized viral components are assembled into virions. These virions then undergo maturation, during which the viral components are organized and packaged into a fully functional form. During this step, the phage genome is packaged into the virions. With the phage replication process complete, the phage induces the host cell to undergo lysis. This is achieved by the production of lytic enzymes, such as endolysin, that degrade the bacterial cell wall. The release of newly formed phage particles into the surrounding environment allows them to infect other susceptible bacterial cells and start the infection cycle anew (Dennehy and Abedon 2021b).

## 1.2 Phages in nature

Phages are extremely abundant in nature and exist in every environment where bacteria occur, having many different roles within their ecosystems. For example, in ocean environments different lytic phages regulate nutrient cycles by releasing organic matter to be consumed by other bacteria when the phages lyse their host bacteria (Fuhrman 1999). Phages also mediate gene transfer, alter host metabolism, and control host populations in different environments (Salmond and Fineran 2015). Some of the environments in which phages are found and have been studied include soil, aquatic environments, and the human body (Batinovic et al. 2019). Of these, phages are especially abundant in aquatic, both

marine and freshwater, environments (Bergh et al. 1989). Aquatic environments are also rich with bacteria, many of which can serve as host for phages.

One such genus of bacteria is *Flavobacterium*, a group of Gram-negative, aerobic rods, which are primarily found in soil and freshwater environments. *Flavobacterium* species have been found all over the world, but they are most abundant in rivers, lakes, and soil in frigid and temperate environments (García-López et al. 1999, Bernardet and Bowman 2006). Many phages infecting *Flavobacterium* are known, utilizing both lytic and temperate life cycles. Known *Flavobacterium* infecting phages are myovirus, siphovirus, and podovirus dsDNA phages (e.g., Borriss et al. 2003, Laanto et al. 2011), as well as the ssDNA phage FLiP (Laanto et al. 2017, Mäntynen et al. 2020). As with bacteria in general, phages have many effects on populations of *Flavobacterium*. The presence of phages causes flavobacteria to diversify faster as they evolve different resistance mechanisms to combat phage infection (Middelboe et al. 2009). As a result of evolving this resistance, the bacterium's metabolism, virulence, and susceptibility to other phages may change (Middelboe et al. 2009, Laanto et al. 2012).

## 1.3 Huge phages

As stated previously, the variation among phages extends to their genome size. With the discovery of several phages with unusually large genomes, the terms jumbophage (for phages with a genome size of over 200 kbp) and megaphage (for phages with a genome size of over 500 kbp) have been used to separate phages with large genomes. The term huge phage has been used to refer to both groups simultaneously (Al-Shayeb et al. 2020). Several hundred huge phage genomes have been found in metagenomic analyses, with the largest phage genome recorded being 735 kbp, although the phage has not been isolated and only its genome has been studied. Relatively few megaphage genomes have been found compared to jumbophage genomes, of which there are over 200 isolates (Harding et al. 2023). Many huge phages have been grouped together into clades,

indicating that the large genome size is a relatively stable trait. However, based on the metagenomic data available, huge phages are still a diverse group, as they've been detected in various environments and ecosystems, infecting a multitude of hosts (Al-Shayeb et al. 2020).

Many jumbophages have been found to encode auxiliary metabolic genes (AMG) (Breitbart et al. 2007). These genes can improve the fitness of the phage by modulating the metabolism of the host to better serve phage reproduction. AMGs can influence the diverting of host resources or modify cellular processes. For example, AMGs can be related to both phosphate (Martiny et al. 2009) and nitrogen metabolism (Sullivan et al. 2010) or have roles relating to photosynthesis (Frank et al. 2013) and other metabolic processes (Willenbücher et al. 2022). AMGs can also provide the host with new genetic traits. For example, they could give advantage to the phage by influencing host nutrient utilization. AMGs are one factor through which phages influence the communities surrounding them (Willenbücher et al. 2022). For example, novel megaphage Mar_Mega_1 genome includes AMGs encoding putative dihydrofolate reductase, phosphoesterase and peptidase enzymes, which also have homologues in other megaphages (Michniewski et al. 2021). Current research indicates that AMGs, specifically ones related to carbohydrate and amino acid uptake, are prevalent in marine phage systems and viral metagenomes, suggesting their importance in phage-host interactions (Warwick-Dugdale et al. 2019).

## 1.4 CRISPR-Cas systems

To protect themselves from phage infection, bacteria have developed a multitude of defence mechanisms. One such mechanism is the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas (CRISPR-associated proteins) system, an adaptive immune system of bacteria. The system consists of CRISPR repeat-spacer arrays that can be transcribed into CRISPR RNA (crRNA), trans-activating CRISPR RNA, and different Cas genes. The CRISPR-Cas defence process can be divided into three stages: (1) adaptation or spacer integration, (2)

processing of the primary transcript of the CRISPR locus and maturation of the crRNA, and (3) genome interference (Makarova and Koonin 2015).

Cas genes code Cas proteins which allow the system to acquire new spacers and target invading genetic material. Cas1 and Cas2, which are present in most known CRISPR-Cas systems, are active during the spacer integration stage. Other Cas proteins function during the later stages (Makarova and Koonin 2015). When bacteria are invaded by foreign genetic material, it can be cut into fragments by Cas proteins, and these pieces can be added into the CRISPR array as spacers by the Cas1-Cas2 integrase (Makarova et al. 2015). If the bacterium comes into contact with the same genetic material again, crRNA recognizes and pairs with it, guiding Cas proteins to cleave the targeted sequences. Thus, the CRISPR-Cas system allows the bacteria to gain adaptive immunity and protect itself against infection (Makarova et al. 2015).

CRISPR systems are divided into two classes, each further divided into three types, based on the architectures of their nuclease effector modules which function in crRNA processing and interference with foreign genetic material (Figure 1). Though the effector modules vary, Cas1 and Cas2 are present in the majority of known functioning systems (Makarova et al. 2020). Class 1 systems have nuclease effector modules which contain multiple subunits. This class consists of type I, type III, and type IV systems. Class 2 nuclease effector modules are single large proteins. Class 2 consists of type II, type V, and type VI systems (Makarova and Koonin 2015).

One of the best-known Cas proteins is Cas9. It is an RNA-guided DNA endonuclease enzyme that causes both strands of a double-stranded DNA to break. Cas9 unwinds foreign DNA and checks it for regions that are complementary to the sequence of the Cas9 guide RNA. Upon contact with a complementary region, Cas9 cleaves the foreign DNA. Because it is Cas9 that cuts DNA, it is crucial for the functional interference of type II CRISPR-systems (Redman et al. 2016).

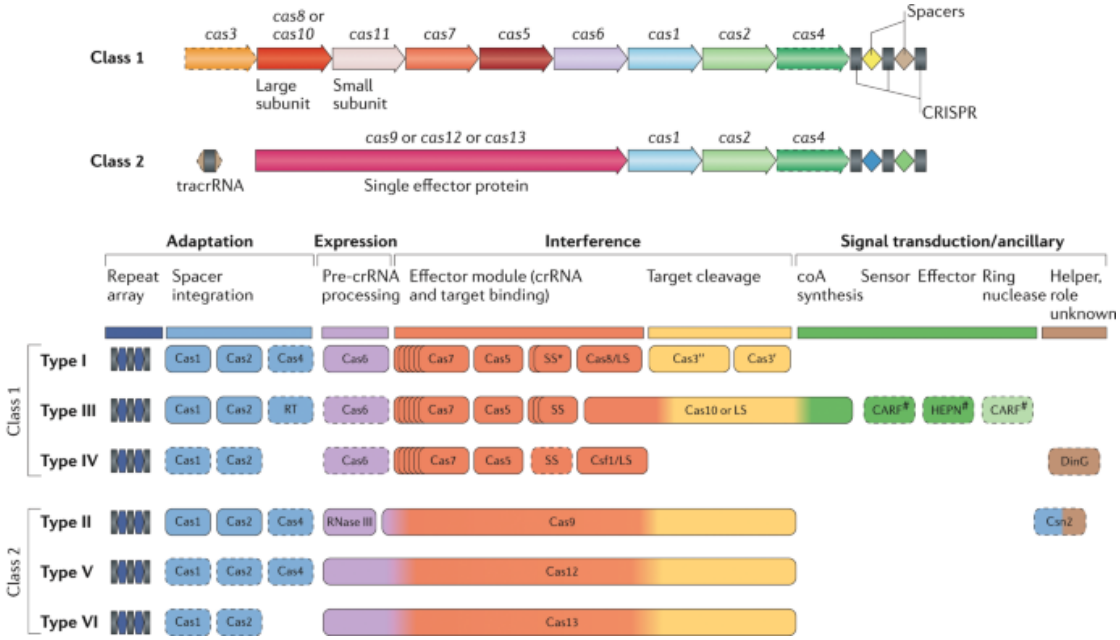



Figure 1. CRISPR–Cas system types and their modular organization. (Original image from Makarova et al. 2020, Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants, Nature Reviews Microbiology, volume 18, page 69, Springer Nature. Permission to use acquired through Copyright Clearance Center.)

### 1.4.1 Phage-encoded CRISPR-Cas systems

As bacteria and phages have evolved side by side, some phages have also acquired the ability to encode CRISPR-Cas systems (Seed et al. 2013, Al-Shayeb 2020). In fact, members of all six types of CRISPR systems have been found to be encoded by phages, with over 6000 known CRISPR-encoding phages identified in metagenomic analysis. However, CRISPR-Cas systems are rare in phages when compared to prokaryotes (Al-Shayeb et al. 2022). 42% of bacteria and 85% archaea genomes encode CRISPR-Cas systems (Makarova et al. 2020), while only 0.4% of phage genomes do (Al-Shayeb et al. 2022). Still, CRISPR-Cas systems can be found diversely in different phage subtypes. Some huge phages with CRISPR-Cas systems have been found in metagenomic analyses, but most CRISPR-Cas encoding phages have genome sizes close to an average phage. CRISPR-Cas systems in phage genomes have many unique properties. For example, most of the phage-encoded CRISPR-Cas systems are incomplete. Notably, less than 10% of phages coding CRISPR-Cas systems also encode the machinery required for

spacer acquisition (Al-Shayeb et al. 2022), meaning the Cas1-Cas2 integrase (Makarova et al. 2015).

Furthermore, of phage genomes with CRISPR arrays, only 6% include Cas effectors encoded nearby, meaning they only carry the repeats and spacers. It is possible that phages lacking these could produce their own guide RNAs but use the Cas effectors of their hosts. This would require that the host bacterium carries CRISPR systems that are genetically compatible with the crRNAs expressed by the phage. Similarly, ~1% of CRISPR-encoding phages encode Cas1-Cas2 integrase but no other Cas enzymes. Some phage-encoded Cas1 proteins are fused to another protein, e.g., reverse transcriptase, which could allow for the acquisition of new spacers similarly to a Cas1-Cas2 integrase. Most phage-encoded CRISPR-Cas systems target DNA, with only some RNA targeting systems identified. Phage-encoded CRISPR systems can also have modified type III or VI systems. For example, some phage-encoded type III systems are associated with CRISPR arrays that can target phage tail proteins, transposases, or other vital RNA transcripts (Al-Shayeb et al. 2022).

The role of CRISPR-Cas system in phages can be related to different goals, such as evading the bacterial host immune response. For example, in a study conducted by Seed et al. (2013) spacers in the CRISPR-Cas system encoded by phage ICP1 show 100% identity to sequences within a phage inhibitory chromosomal island of the host bacterium *Vibrio cholerae*. They show that to successfully infect the bacterium and enter a lytic cycle, sequence identity between phage-encoded CRISPR-spacers and bacterial chromosomal island is required. When such identity is not present, the phage is able to acquire new spacers into the leader end of its CRISPR array to ensure phage replication. This indicates that ICP1 has evolved to target parts of the host genome detrimental to phage infection and has a fully functioning adaptive immune evasion system (Seed et al. 2013).

The concept of phages targeting other phages with their CRISPR- systems to gain advantage in a super infection has been presented based on metagenomic

evidence (Figure 2). CRISPR-Cas systems encoded by phages have been found to have spacer sequences matching to other phage genomes, supporting the possibility of this tactic (Al-Shayeb et al. 2020). However, no experimental studies have been conducted on the subject, as these phages have not been isolated. However, phage-encoded CasΦ, a functional single-protein CRISPR-system, has been expressed and found functional in a laboratory setting (Pausch et al. 2020).

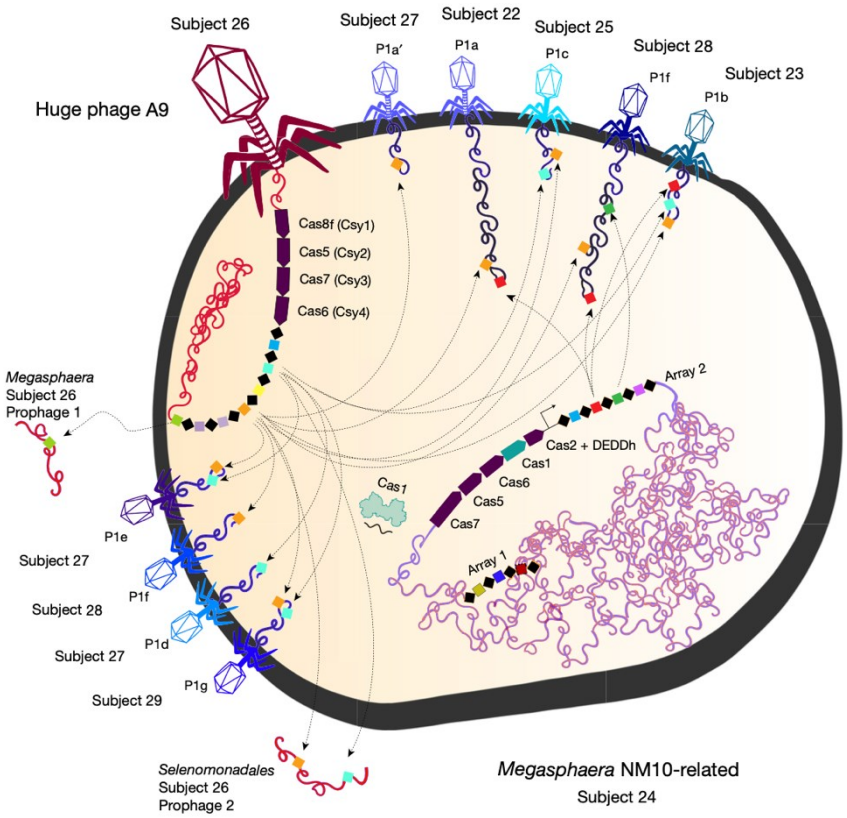



Figure 2. Cell diagram of bacterium-phage and phage-phage interactions involving CRISPR targeting. Arrows indicate CRISPR-Cas targeting. (Original image from Al-Shayeb et al. 2020).

## 1.5 Research questions and hypotheses

In this work, five putative CRISPR-Cas encoding megaphages were studied. For phages of this kind, all published data so far is based on metagenomic analysis, and no studies have been done with isolated phages, as no isolates have been available thus far. Therefore, performing experiments using actual isolates is an important step to gain better understanding of these phages and their host-phage interactions. The first aim of the study was to determine the way the phages

interact with their hosts. Because the phages have such large genomes, it was hypothesised that they would be able to infect a multitude of hosts (Nazir et al. 2021). However, we also hypothesized that their reproduction would likely be limited because the size of the host cell limits the number of new phage particles that can be produced. The second aim was to determine the morphology of the megaphage. Based on genomic information and genome size we hypothesised that the phage particles would be tailed and relatively large. The final aim was to determine the purpose and functionality of the CRISPR-system of the phage. We hypothesized that the phage could target the host bacterium genome to downregulate host gene expression or destroy host genome, which would help the phage reproduce in the host cell (Al-Shayeb et al. 2020). The phage genome, especially its CRISPR-array, was also analysed to better understand phage-host interactions.

# 2 MATERIALS AND METHODS

## 2.1 Materials

Five phages were used for the experiment: JoK79, JoK80, JoK81, JoK82, and Elf16. Each was isolated in 2021 from Lake Jyväsjärvi, and their genomes were sequenced, assembled, and annotated with Pharokka (Bouras et al. 2023) prior to thesis work (Mäkelä et al. unpublished). Each phage has a putative CRISPR-system identified with CRISPRCas finder (Couvin et al. 2018) and a genome of over 600 kbp. JoK80 encodes Cas3, Cas6 and Cas7, and its CRISPR-system most closely resembles a class 1 system. The other four phages encode Cas9, and their CRISPR-systems most closely resemble a class 2 type II system. All of the phages lack both Cas1 and Cas2. Elf16 infects B330 *Flavobacterium* sp. strain B330 and the other four phages infect *Flavobacterium* sp. strain B114 (Mäkelä et al. unpublished). B114 has been previously isolated from River Kevojoki (Laanto et al. 2011) and B330 from Lake Jyväsjärvi (Laanto et al. 2017). Both have been stored

in 20% glycerol at -80 °C. Medium used for the experiments was Shieh medium (Song et al. 1988).

## 2.2 Methods

### 2.2.1 Host range testing

Host range testing was conducted with all five phages. Selected bacteria were plated from stock that had been stored in 20% glycerol at -80°C (Table 1). Liquid cultures were made from bacterial colonies. Bacteria were then plated with 3 ml of Shieh medium containing 0.4% low melt agarose. 100 µl of liquid bacterial culture was used for all other bacteria but B185, B245, B067, for which 300 µl of culture was used. Phages, undiluted and diluted to $10^{-1}$, were pipetted on the plates as 10 µl drops. Results were read after two days of incubating at room temperature (RT) and categorized as + (plaques/clear lysis areas) or – (no visible signs of infection). Bacteria/phage pairs with + results were chosen for the second round of testing. In the second round, triplicate plates were made from each bacterium, and the phage was pipetted on top as 10 µl drops as a dilution series of $10^{0}$ – $10^{-6}$. After two days of incubating at RT, plaque forming units per ml (pfu/ml), i.e. the titer, was determined.

TABLE 1. Bacteria chosen for host range testing.

| | | |
|---|---|---|
| *Flavobacterium* sp. | B28 | Laanto et al. 2011 |
| *Flavobacterium* sp. | B80 | 〃 |
| *Flavobacterium* sp. | B105 | 〃 |
| *Flavobacterium* sp. | B121 | 〃 |
| *Flavobacterium* sp. | B127 | 〃 |
| *Flavobacterium* sp. | B169 | 〃 |
| *Flavobacterium* sp. | B171 | 〃 |
| *Flavobacterium* sp. | B174 | 〃 |
| *Flavobacterium* sp. | B176 | 〃 |
| *Flavobacterium* sp. | B178 | 〃 |
| *Flavobacterium* sp. | B180 | 〃 |
| *Flavobacterium* sp. | B202 | unpub |
| *Flavobacterium* sp. | B205 | 〃 |
| *Flavobacterium* sp. | B206 | 〃 |

| | | |
|---|---|---|
| *Flavobacterium* sp. | B207 | Laanto et al. 2011 |
| *Flavobacterium* sp. | B208 | unpub |
| *Flavobacterium* sp. | B209 | Laanto et al. 2011 |
| *Flavobacterium* sp. | B221 | unpub |
| *Flavobacterium* sp. | B222 | Laanto et al. 2011 |
| *Flavobacterium* sp. | B223 | 〃 |
| *Flavobacterium* sp. | B224 | 〃 |
| *Flavobacterium johnsoniae* | UW101 | McBride et al. 2009 |
| *Flavobacterium psychrophilum* | MH1 | Stenholm et al. 2008 |
| *Flavobacterium psychrophilum* | 950106-1/1 | Castillo et al. 2012 |
| *Xylophilus* sp. | B14 | Laanto and Oksanen 2023 |
| *Caulobacter* sp. | B15 | 〃 |
| *Polaromonas* sp. | B16 | 〃 |
| *Pseudomonas* sp. | B20 | 〃 |
| *Herbaspirillum* sp. | B21 | 〃 |
| *Sphingomonas* sp. | B54 | 〃 |
| *Serratia* sp. | B71 | unpub |
| *Curvibacter* | B82 | 〃 |
| *Pseudomonas* sp. | B116 | 〃 |
| *Janthinobacterium* sp. | B126 | 〃 |
| *Aeromonas* sp. | B135 | Almeida et al. 2019 |
| *Exiguobacterium* sp. | B157 | unpub |
| *Aeromonas* sp. | B158 | 〃 |
| *Unidibacterium* sp. | B181 | 〃 |
| *Janthinobacterium* sp. | B193 | 〃 |

### 2.2.2 AMG and CRISPR-Cas analysis of Elf16 genome

Elf16 genome had been previously assembled and annotated (Mäkelä et al. unpublished). After the annotation, out of the 1156 coding sequences, 1070 were left without function and designated as 'hypothetical proteins'. To identify potential AMGs, Elf16 coding sequences were annotated using Phyre2 (Kelley et al. 2015). Phyre2 hits with high confidence (>90%), coverage (>50%), and functions linking them to bacterial metabolisim, were further analysed using BlastP (Altschul 1990).

Elf16 CRISPR spacers were identified using CRISPRCasFinder (Couvin et al. 2018) and analysed using BLAST (Altschul 1990). Elf16 spacers were mapped to isolation host B330 genome (unpublished data) to identify possible matches with MUSCLE 5.1 (Edgar 2022) in Geneious prime (2022.2.2,

https://www.geneious.com). To compare the CRISPR repeat sequences of Elf16 and flavobacteria, two strains of *Flavobacterium columnare* were used. *Flavobacterium columnare* was chosen due to a well characterised CRISPR system (Hoikkala et al. 2021). *Flavobacterium columnare* strain B185 (NZ_CP010992) (type II-C repeat sequence) and strain B245 (NZ_CP071008) (type II-c repeat sequence and type VI-B repeat sequence) were aligned with Elf16 repeat sequence to identify possible matches with MUSCLE 5.1 (Edgar 2022) in Geneious prime (2022.2.2). In addition, Cas9 of both strains were similarly aligned to Elf16 Cas9.

### 2.2.3 Determination of optimal conditions for Elf16 infection

The rest of the experiments were conducted with phage Elf16 only. To find the optimal circumstances for measuring Elf16 life cycle, the development of optical density (OD) of a *Flavobacterium* sp. B330 culture infected with Elf16 was measured in several different conditions. The conditions tested were temperature, culture volume, shaking, and the growth phase (based on OD) at which the culture was infected. OD was always measured at 595 nm (MultiSkan FC, SkanIt 4.1, Thermo Scientific). The protocol for screening Elf16 growth parameters with OD measuring was as follows: a *Flavobacterium* sp. B330 culture grown overnight at RT was diluted ~1:10 to OD ~0.1, and diluted cultures were grown in 150 rpm. Multiplicity of infection (MOI) used was 10. After the experiment, a follow-up sample was taken from an overnight infected culture. The samples were pipetted as 10-fold diluted series of 10 µl drops on plates containing 100 µl of the host bacteria in soft medium (Shieh medium containing 0.5% low melt agarose) to determine pfu/ml.

To test the effect of bacterial growth phase (determined based on OD of bacterial culture) at infection, OD of the control culture was measured hourly, and other cultures were infected at the desired OD points, after which their OD was measured hourly. Cultures were grown in RT, and their volume was 15 ml. In the remaining experiments, bacterial cultures were always infected at OD 0.4.

To test the effect of temperature on phage infection, an infected culture and a control (volume 15 ml) were each grown in 3 temperatures: RT (approx. 25 °C), 20 °C and 22 °C. Temperatures were chosen due to strain B330 growing best at approx. 20 °C (K. Mäkelä, personal communication, 2023). After infection, OD was measured hourly. Cultures in 22 °C were moved to RT and off shaker at 3h and 15min post-infection. In the remaining experiments, cultures were always grown in 20 °C.

To test the effect of culture volume and low shaking (here, low shake means 50 rpm) on phage infection, two 15 ml and two 30 ml cultures were infected. Two cultures (one of each volume) were put on low shake immediately after infection, and the remaining two were put on low shake 2h post-infection. After infection, OD was measured hourly. At 5h post-infection, shaking was turned off and cultures were left overnight at 20 °C.

### 2.2.4 OD development of strain B330 over an extended time period

During the testing of optimal conditions, it was noted that the OD of infected strain B330 culture continued to decrease overnight. To measure the effect of phage infection on strain B330 growth over a longer time period, a 15 ml culture of B330 was infected with Elf16 (MOI10). After 50 minutes of infection, 10 replicates (200 µl each) of both the infected culture and the control were divided onto a Bioscreen plate. Phage lysate in Shieh medium was used as a negative control. OD was measured every 15 minutes with Bioscreen C (FP-1100-C, Bioscreen Oy Growth Curves Ab Ltd).

### 2.2.5 Elf16 adsorption to isolation host *Flavobacterium* sp. B330

To test Elf16 adsorption to host cells, three replicate *Flavobacterium* sp. B330 cultures were diluted to an OD of ~0.2, and infected with Elf16 (titer $3x10^5$ pfu/ml). From the infected cultures, 50 µl samples were collected at 1, 5, 10, 20, 30, and 60 minutes, and added to pre-chilled Eppendorf tubes containing 950 µl of medium. These samples were plated with the host bacteria in soft medium

(Shieh medium containing 0.5% low melt agarose). Experiment setup was modified from one described by Kropinski (2009).

### 2.2.6 CRISPR-Cas system containing megaphage life cycle

To determine Elf16 life cycle, several simultaneous experiments were conducted with infected *Flavobacterium* sp. B330 cultures (Figure 3). An overnight B330 culture was diluted 1:10, divided to seven 15ml cultures and grown to ~0.4 OD. Four replicate B330 cultures were infected with Elf16 MOI10 and grown in 20 °C at 150 rpm for 4 hours, after which they were taken off the shaker. Three replicate B330 control cultures were grown similarly, but Shieh medium was added instead of phage lysate.

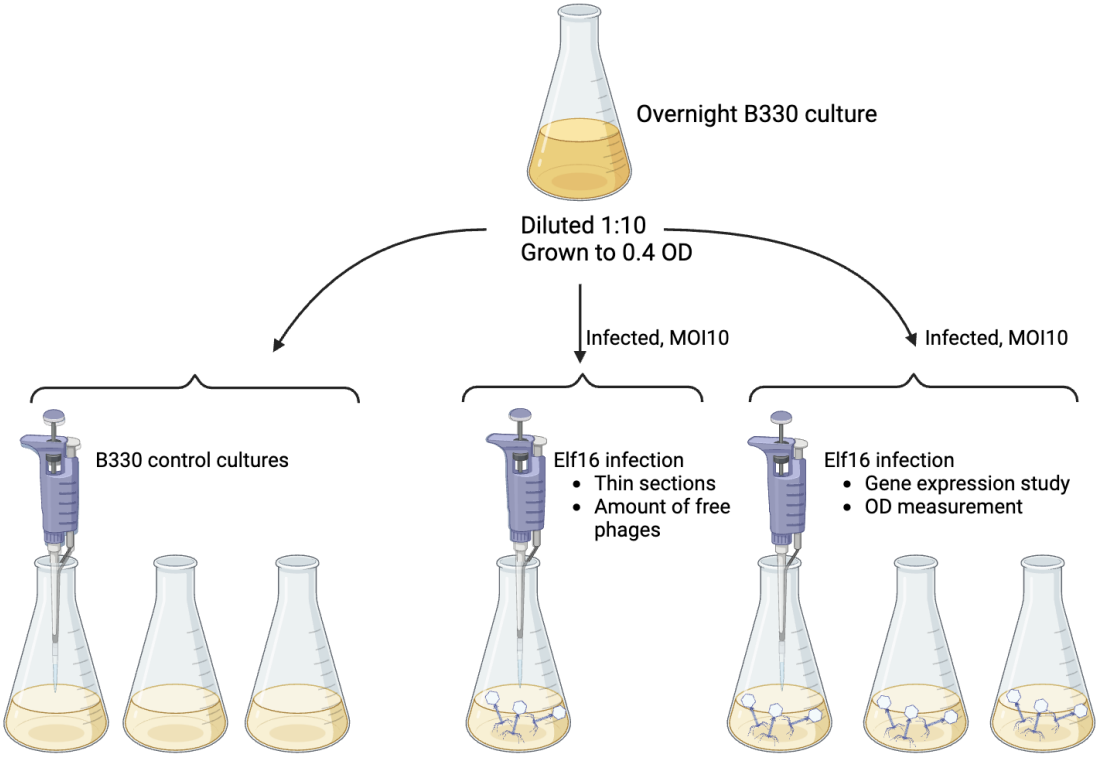



Figure 3. Overview of Elf16 life cycle experiment setup and sampling (Created with BioRender.com)

#### 2.2.6.1 Life cycle determination

To determine the *Flavobacterium* sp. B330 growth curve during Elf16 infection, OD was measured hourly from three of the infected cultures, and the three control cultures. Due to the limitation of culture volume, an additional culture was used to determine the number of free phages. This was calculated by syringe

filtering (Filtropur S 0.2 µm, Sarstedt) and titrating samples taken at 0 minutes, 30 minutes, 1 hour, 2 hours, 4 hours, and 22 hours.

#### 2.2.6.2 Thin section imaging

In order to examine the morphology of Elf16, as well as its life cycle, thin sections were made from an Elf16 infected *Flavobacterium* sp. B330 culture, and then imaged. Samples were collected at 5 minutes, 30 minutes, 1 hour, 2 hours, and 4 hours with the following protocol: 1 ml of liquid culture was collected in a 1.5 ml tube. Cells were pelleted (5 min, 8000 rpm) and supernatant removed. 1 ml of fresh 2.5 % glutaraldehyde in 50 mM potassium phosphate (pH 7.2) was added on top and incubated for 2 hours at room temperature. Sample was centrifuged (1 min, 8000 rpm) and the supernatant was removed. Cells were resuspended in 1.5 ml of 50 mM potassium phosphate (pH 7.2). Samples were sent to the University of Oulu for thin sectioning, in which the samples were post-fixed with osmium tetroxide, dehydrated and embedded in plastic resin (Epon). Ultrathin sections (70-80 nm) were post-stained with uranyl acetate and lead citrate. After receiving the thin sections, imaging was done with transmission electron microscopy (TEM) using JEOL JEM-1400 (JEOL Ltd), operated at 80 kV. Images were recorded with a bottom mounted Quemesa CCD camera with a 4008×2664-pixel resolution. From the TEM images, Elf16 size was measured using ImageJ (version 2.14.0/1.54f, Schneider et al. 2012).

#### 2.2.6.3 Gene expression

To determine the expression of Cas9 and helicase during Elf16 infection, three replicates of Elf16 infected *Flavobacterium* sp. B330 cultures were used. Samples were taken at 0, 5, 15, and 30 minutes, and 1, 2, 3, and 4 hours, with 3 replicates for each time point. From the samples, RNA was extracted using RNeasy® Protect Bacteria Mini Kit (Qiagen) according to kit instructions. Extracted RNA was treated using DNase I, RNase-free according to kit instructions (Thermo Scientific). RNA concentrations were measured using NanoDrop One (ThermoFisher Scientific). RNA was converted to complementary DNA (cDNA)

using the high-capacity cDNA reverse transcription kit (Applied Biosystems) without RNase inhibitor. Primers were designed for Elf16 Cas9 and helicase, and B330 16S (Table 2). First, the annotation of phage genes was confirmed using HHpred (Gabler et al. 2020), and hits with a high probability (>95%) were considered confirmed. Geneious Prime (2022.2.2) primer designing tool was used for the primer design. GoTaq® qPCR Master Mix (Promega) was used for the quantitative polymerase chain reaction (qPCR) according to kit instructions. The reference gene used for qPCR data normalization was B330 16S. qPCR was performed using Bio-Rad CFX96 Touch Real-Time PCR system. Relative quantification (RQ) values were determined using the $2^{-\Delta\Delta Ct}$ method (Microsoft Excel (version 16.77.1, Microsoft Corporation, 2018)) and significance was tested using one-way ANOVA (GraphPad Prism (version 10)).

TABLE 2. Primers used for qPCR. For the primer design, genes were annotated using HHpred (Gabler et al. 2020) and primers designed using Geneious prime (2022.2.2) primer designing tool.

| | Target gene | Sequence (5′ to 3′) |
|---|---|---|
| B330 | 16S (forward) | AACTTGCGTTCGTACTCCCC |
| | 16S (reverse) | GTAGTCCACGCCGTAAACGA |
| Elf16 | Cas9 (forward) | CCCTGCATCAATTACGAAAGGTC |
| | Cas9 (reverse) | GTTCAGTTCACGAGCCATTTCT |
| Elf16 | Helicase (forward) | TAGACAAGAACAGGATAGTGCGC |
| | Helicase (reverse) | ACCACTTCCCGTTGGTAAATTCA |

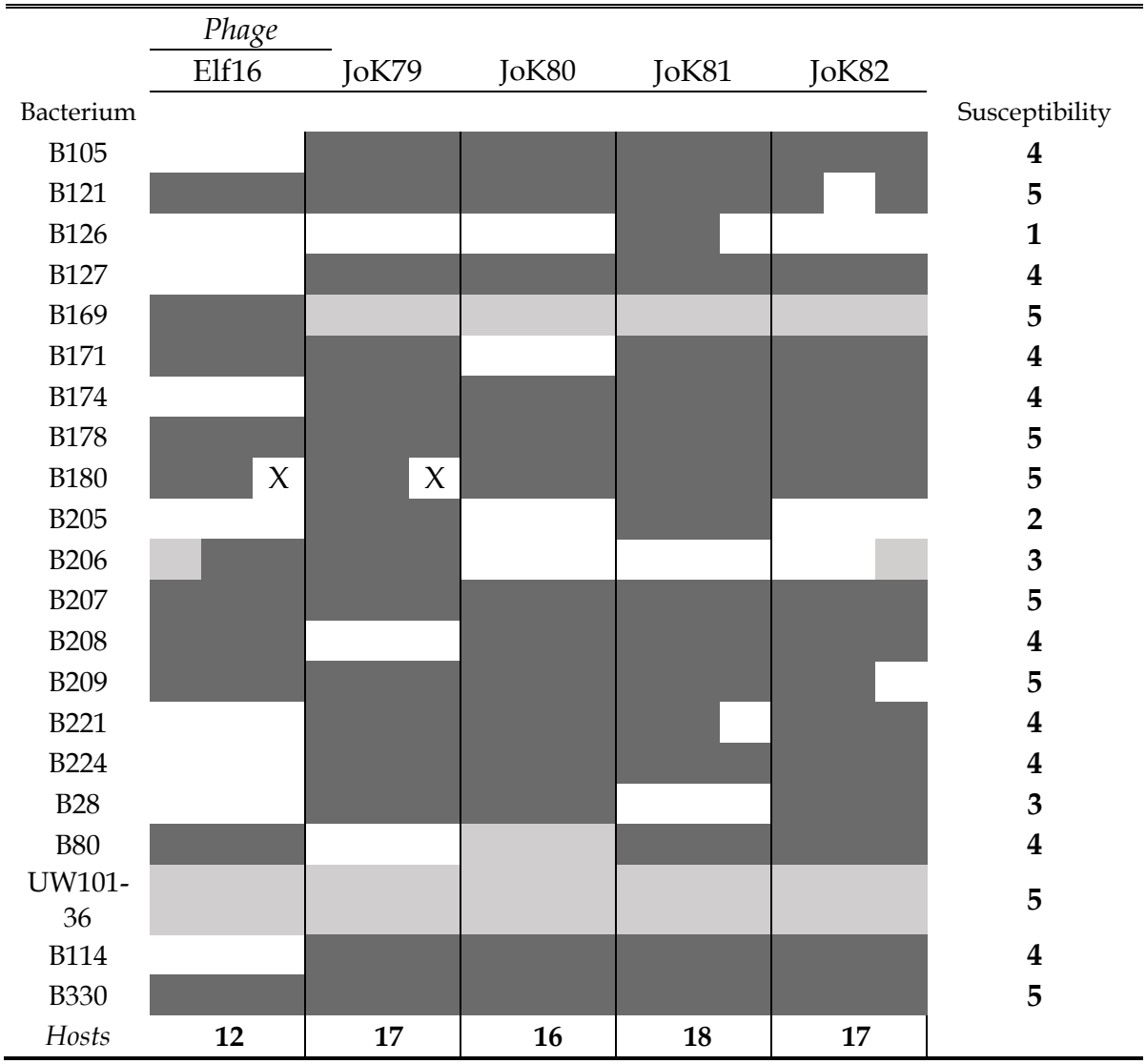
# 3 RESULTS

## 3.1 Host specificity of studied phages

In the host range test, phages Elf16, Jok79, Jok80 and Jok82 only infected species of *Flavobacterium*. JoK81 mostly infected species of *Flavobacterium*, and in addition to one species of *Janthinobacterium*. Of the bacterial strains tested (Table 1), 21 had no visible interaction with the phage. In total, 12 hosts were identified for Elf16, 17 for JoK79, 16 for JoK80, 18 for JoK81, and 17 for JoK82 (Table 3, see

Appendix 1 for expanded). Many of the plates showed visible signs of infection only on the edges of the phage drop (Figure 4).

TABLE 3. Results of the host range test. These results have been summarised from the $10^0$ and $10^{-1}$ dilutions, with each phage having three replicate plates per bacteria. Dark grey cells indicate infection, light grey cells indicate inhibition, and white cells indicate no infection. "Susceptibility" means number of tested phages that infect the bacterium. "Hosts" means number of tested bacteria that the phage can infect. "X" indicates failed replication plates from which results could not be recorded.



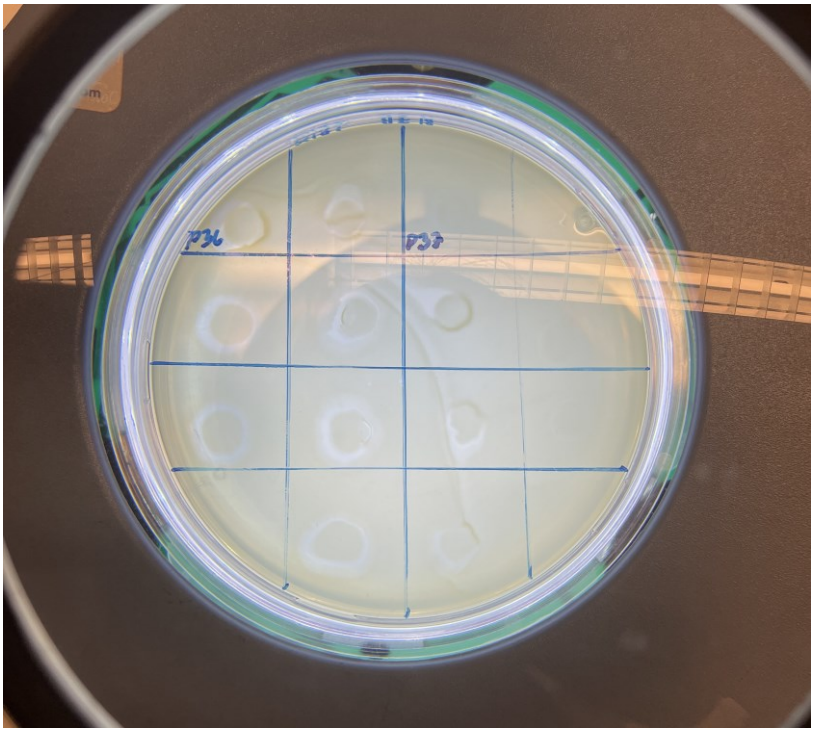



Figure 4. Example of a host range testing plate with only partially clear areas. Phages JoK79 (P36) and JoK80 (P37) infecting *Flavobacterium* sp. strain B127.

## 3.2 Elf16 spacers have no match to isolation host genome

The CRISPR-Cas area of the Elf16 genome was analysed. 27 spacers were identified in the Elf16 genome (Figure 5). Of those, 25 spacers were identified with CRISPRCasFinder (Couvin et al. 2018) and two spacers (11 and 12) were identified manually (see spacer sequences in Appendix 2). Elf16 repeat sequence had a length of 36 nucleotides. No match was found between Elf16 spacers and the bacterial host B330 genome. Elf16 repeat sequence had an 83.3% identity with B185 repeat sequence (type II-C), compared to a 30.6% identity with B245 repeat sequence (type II-c) and 38.9% identity with B245 repeat sequence (type VI-B). Elf16 Cas9 and B245 Cas9 nucleotide sequences had a 69.6% nucleotide identity, and Elf16 Cas9 and B185 Cas9 nucleotide sequences had a 65.7% nucleotide identity.

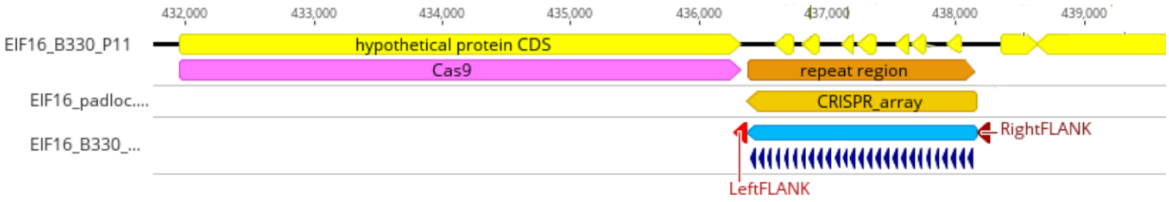



Figure 5. CRISPR-Cas system of Elf16 (Geneious prime 2022.2.2). Elf16 encodes Cas9, but no other Cas genes were detected. Dark blue sections show the 27 spacers found in the repeat spacer array.

## 3.3 Potential AMGs found in the Elf16 genome

In total, 20 potential AMGs were found in the Elf16 genome based on the Phyre2 annotations. BlastP hits for potential AMGs gave mostly identities of <50%, with only five hits with an identity higher than 60%. The highest identity hit had an identity of 67,38%. Of the 20 potential AMGs, nine had homologues in other phages, with an e-value of < 1e-5 (Table 4, see Appendix 3 for expanded).

TABLE 4. List of potential Elf16 AMGs. "Homologues in other phages" refers to number of those homologues with an e-value of <1e-5.

| Locus | Phyre annotation | Closest BlastP hit | BlastP E-value | Homologues in other phages |
|---|---|---|---|---|
| 292 | dna processing chain a (dpra) | DUF2493 domain-containing protein [*Elusimicrobiota bacterium*] | 6E-30 | 1 |
| 311 | atp-dependent clp protease proteolytic subunit | ATP-dependent Clp protease proteolytic subunit [*Candidatus Muirbacterium halophilum*] | 2E-80 | |
| 470 | grea protein | transcription elongation factor GreA [*Paludibacteraceae bacterium*] | 1E-43 | |
| 477 | leucine--trna ligase | class I tRNA ligase family protein [*Pseudomonas* sp. Irchel 3H3] | 1E-136 | |
| 487 | phosphate starvation-inducible protein | PhoH family protein [*Actinomycetota bacterium*] | 3E-107 | 3 |

| | | | | |
|---|---|---|---|---|
| 522 | ribonucleotide reductase r1 protein | ribonucleotide reductase [*Tenacibaculum* phage PTm1] | 1E-162 | 3 |
| 557 | adp-sugar pyrophosphatase | NUDIX domain-containing protein [*Verrucomicrobiota bacterium*] | 1E-73 | 1 |
| 562 | atp-dependent clp protease proteolytic subunit | ATP-dependent Clp protease proteolytic subunit [*Flavobacteriaceae bacterium*] | 4E-86 | 1 |
| 577 | thioredoxin | MAG: thioredoxin [*Sulfurovum* sp.] | 4E-25 | |
| 584 | glutamate--trna ligase | glutamate--tRNA ligase [*Flavobacterium branchiophilum*] | 1e-163 | |
| 701 | guanosine 5'-monophosphate reductase | IMP dehydrogenase [*Candidatus Dojkabacteria bacterium*] | 4E-101 | 20 |
| 873 | dihydrofolate reductase | type 3 dihydrofolate reductase [*Candidatus Thioglobus* sp.] | 6E-32 | |
| 922 | ribose-phosphate pyrophosphokinase 1 | glycosyltransferase [*Candidatus Elulimicrobium humile*] | 9E-154 | 1 |
| 924 | mazg-like nucleoside triphosphate pyrophosphohydrolase | nucleoside triphosphate pyrophosphohydrolase family protein [*Bacteroidota bacterium*] | 1E-53 | |
| 934 | heat shock protein 70 | molecular chaperone DnaK [*Candidatus Muirbacterium halophilum*] | 0.0 | |
| 940 | lon protease | AAA family ATPase [*Candidatus Muirbacterium halophilum*] | 5E-137 | 12 |
| 966 | nicotinamide phosphoribosyltransferase | nicotinate phosphoribosyltransferase [*bacterium*] | 0.0 | |
| 967 | ribose-phosphate pyrophosphokinase 1 | ribose-phosphate diphosphokinase [*Spirochaetota bacterium*] | 5E-105 | |
| 992 | thioredoxin reductase | thioredoxin-disulfide reductase [*Formosa algae*] | 3E-139 | |

## 3.4 Optimal conditions for Elf16 infection

When the optimal conditions for studying Elf16 in liquid were tested, the bacterial growth phase at infection affected the success of the lysis. If the

*Flavobacterium* sp. B330 culture was infected too late (past OD 0.4), its OD continued to increase, and no lysis occurred. If the culture was infected earlier (OD 0.2), lysis occurred, but it was not as clear (Figure 6a). Based on this, OD 0.4 was chosen as the optimal infection point. Cultures grown in 22 °C had the largest difference in OD between control and infected cultures (Figure 6b). B330 grew to a higher OD and lysis was clearer in 15 ml cultures compared to 30 ml cultures. Lower shaking immediately after infection did not seem to have an effect on bacterial growth or lysis (Figure 6c). The phage titer measured from overnight infected cultures was typically approx. $1x10^{10}$ (n=6).

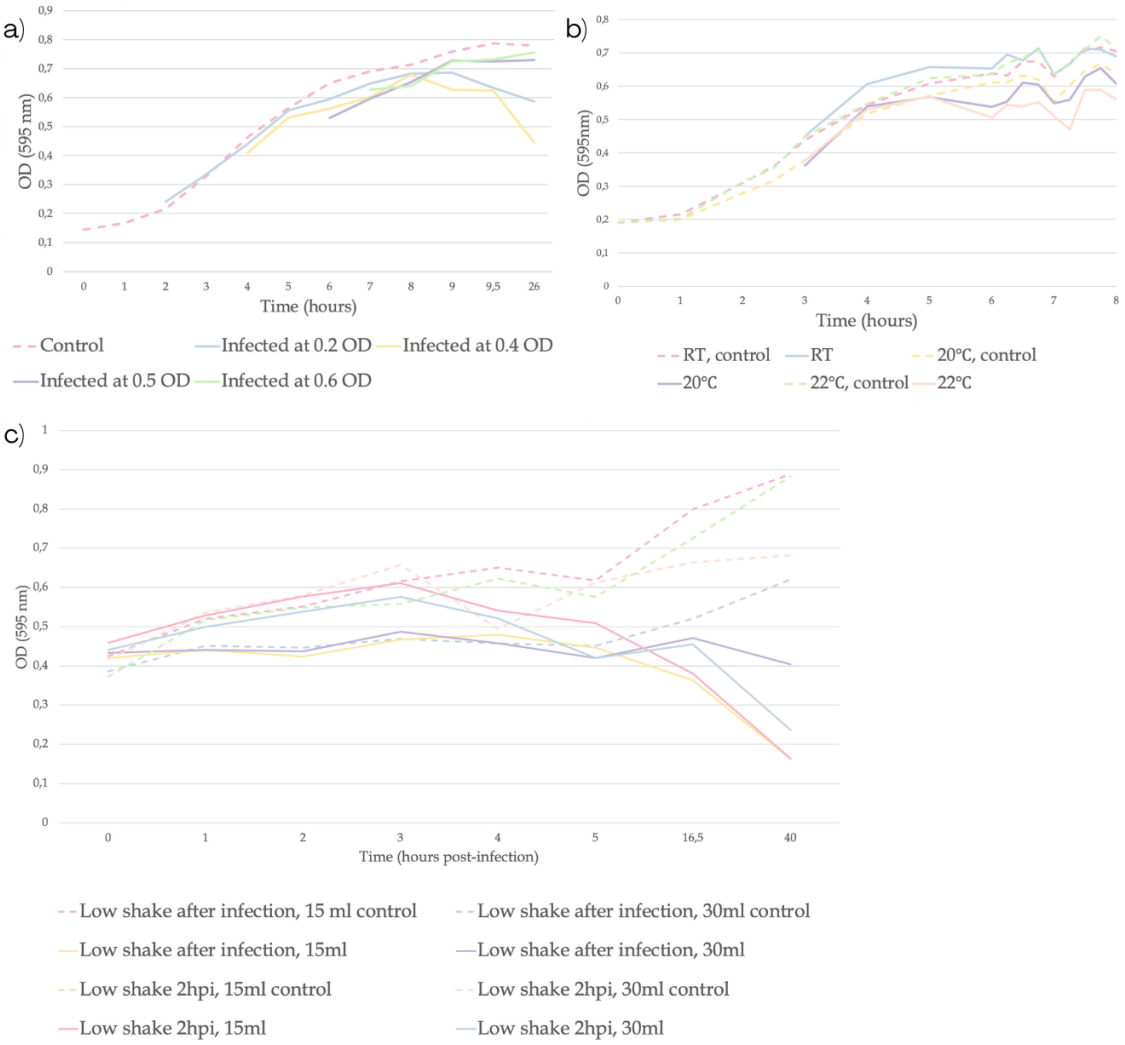



Figure 6. The effect of (a) growth phase (determined based on OD) at infection, (b) temperature, (c) and culture volume and shaking on the development of OD (measured at 595 nm) of *Flavobacterium* sp. B330 culture as Elf16 infects cells of

strain B330. Here, low shake means 50 rpm. "Low shake after infection" means that cultures were moved to low shake immediately after infection, and "low shake 2hpi" means that cultures were moved to low shake 2 hours post-infection. Note that the x-axes are not evenly scaled.

## 3.5 Elf16 life cycle is lytic

Elf16 life cycle was shown to be lytic. In the adsorption test, as Elf16 infected *Flavobacterium* sp. B330, the phage titer decreased until 10 minutes post-infection, after which it continued to rise until 60 minutes post-infection (Figure 7). However, based on the OD measurements, the lysis started approximately three hours after infection. Lysis also happened very gradually. The largest drop in bacterial density happened between three- and four-hours post-infection. The number of free phages decreased until two hours post infection and increased between two- and four-hours post-infection (Figure 8).

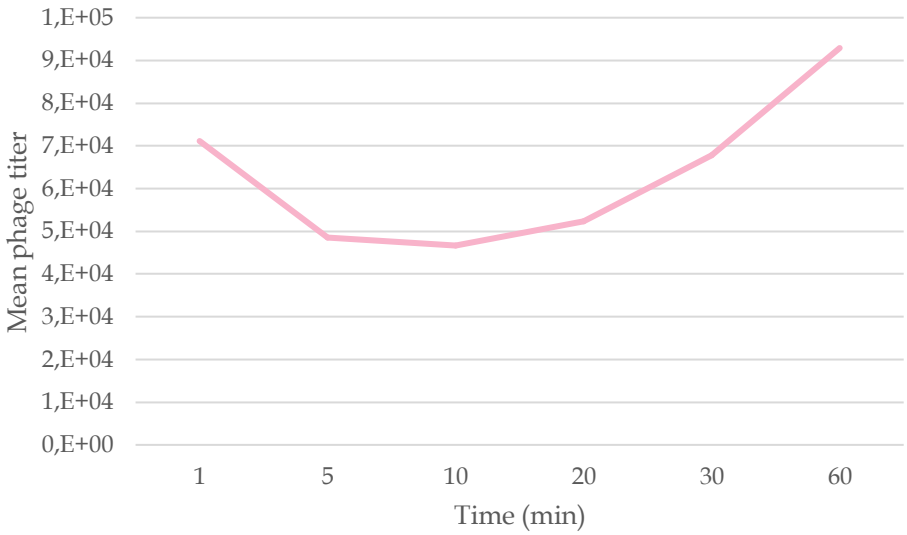



Figure 7. Adsorption to host based on phage titer development as Elf16 infects cells of *Flavobacterium* sp. strain B330. Titers were measured from three replicate cultures.

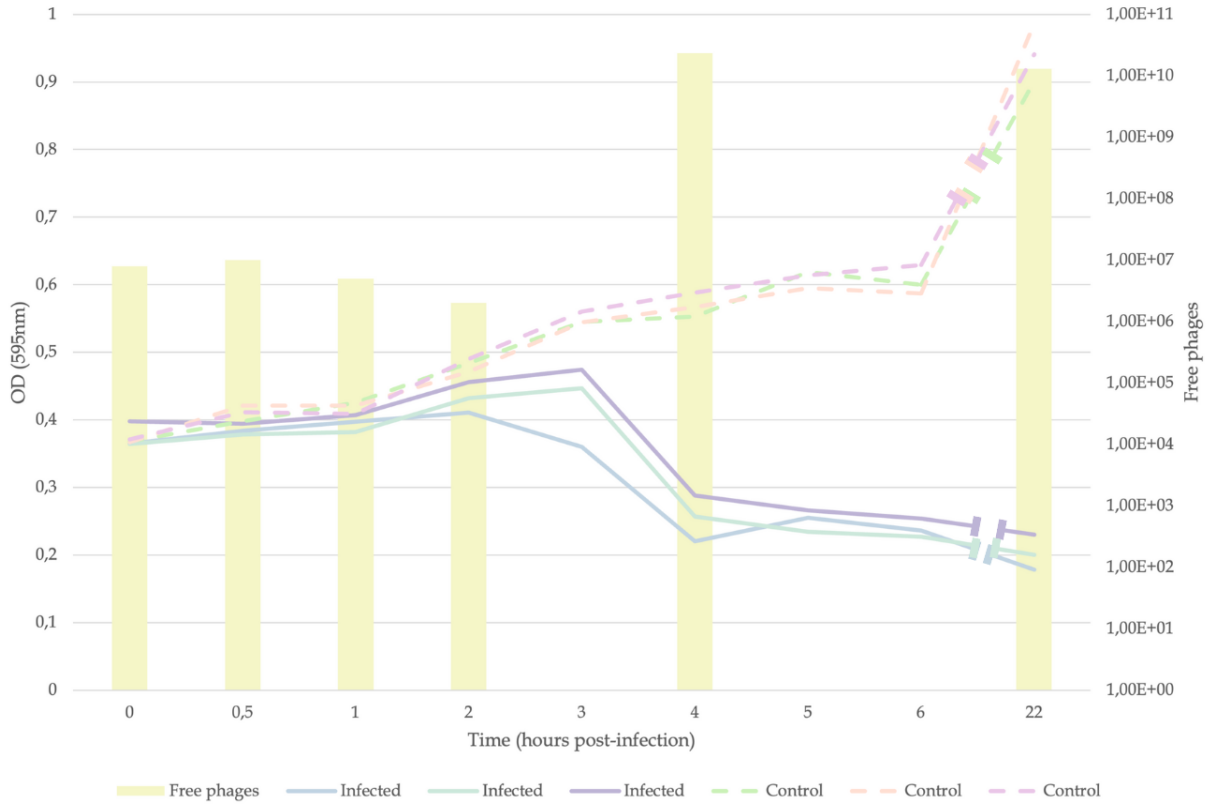



Figure 8. The development of OD (measured at 595 nm) of *Flavobacterium* sp. B330 culture as Elf16 infects cells of strain B330. Infected cultures are replicates and control cultures are only B330. Number of free phages is shown as bars on the secondary Y-axis. Free phages were measured from only one culture.

### 3.5.1 Host OD continues to decrease for several hours after Elf16 infection

During the life cycle determination, it was noted that the OD continued decrease overnight. To determine the effect of phage infection over a longer time period, Bioscreen was used to measure the OD for 27 hours post-infection every 15 minutes. The OD of the infected cultures continued to clearly decrease for approximately 16 hours post-infection. During those 16 hours, the OD of the control cultures continues to increase (Figure 9).

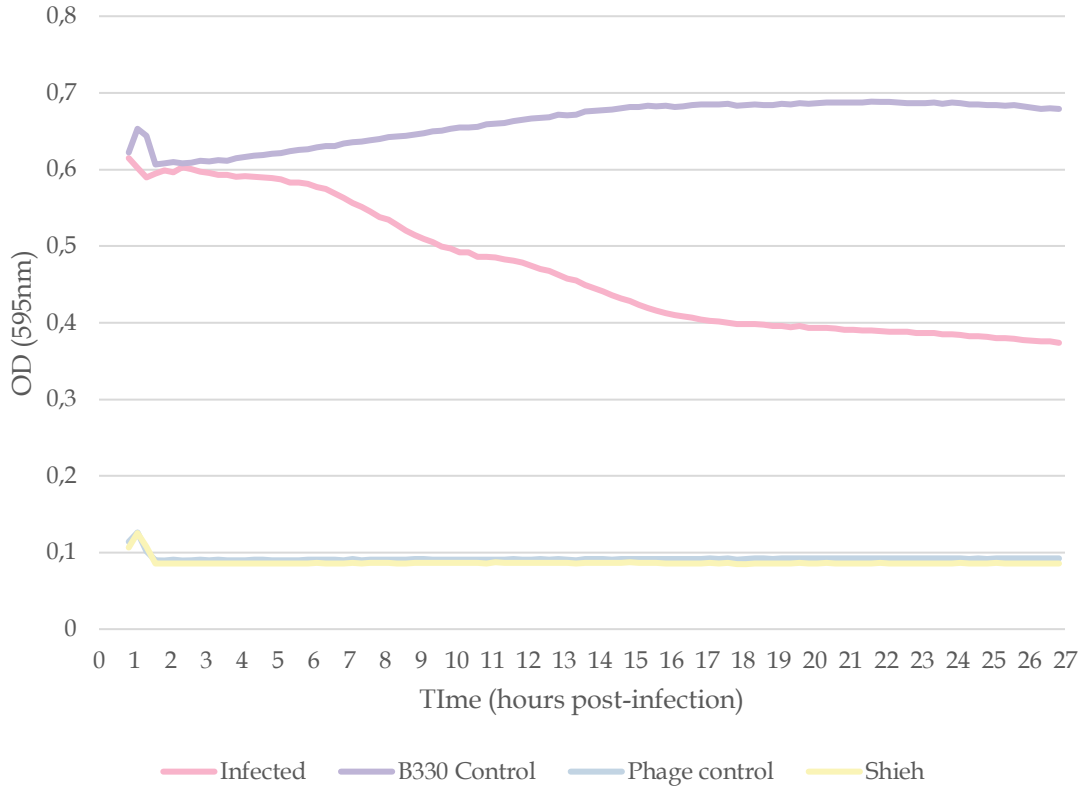



Figure 9. The development of OD (measured at 595 nm) of *Flavobacterium* sp. B330 culture as Elf16 infects cells of strain B330. B330 control is uninfected B330 culture, phage control is only phages in Shieh medium, and "Shieh" is a control of only Shieh medium. 10 replicates of each sample type were used.

## 3.6 Elf16 life cycle imaged from adsorption to lysis

Different parts of Elf16 life cycle were seen from TEM images taken from thin sections. At 5 minutes post-infection, phage particles had attached to the bacteria (Figure 10a). At 30 and 60 minutes the phage capsids on the cells were empty, but no new particles were visible inside bacteria (Figure 10b and c). However, at 120 minutes new phage particles could be detected inside hosts (Figure 10d). At 240 minutes, new phages were visible both outside and inside hosts, and most of the bacteria had lysed (Figure 10e and f). From the images, the capsid size of Elf16 was determined to be approximately 130 nm (vertex-to-vertex) (n=6), while the tail size varied between 80-155 nm (n=6), with unattached phage particles showcasing notably longer tails (non-contracted) compared to those phages that were attached (contracted).

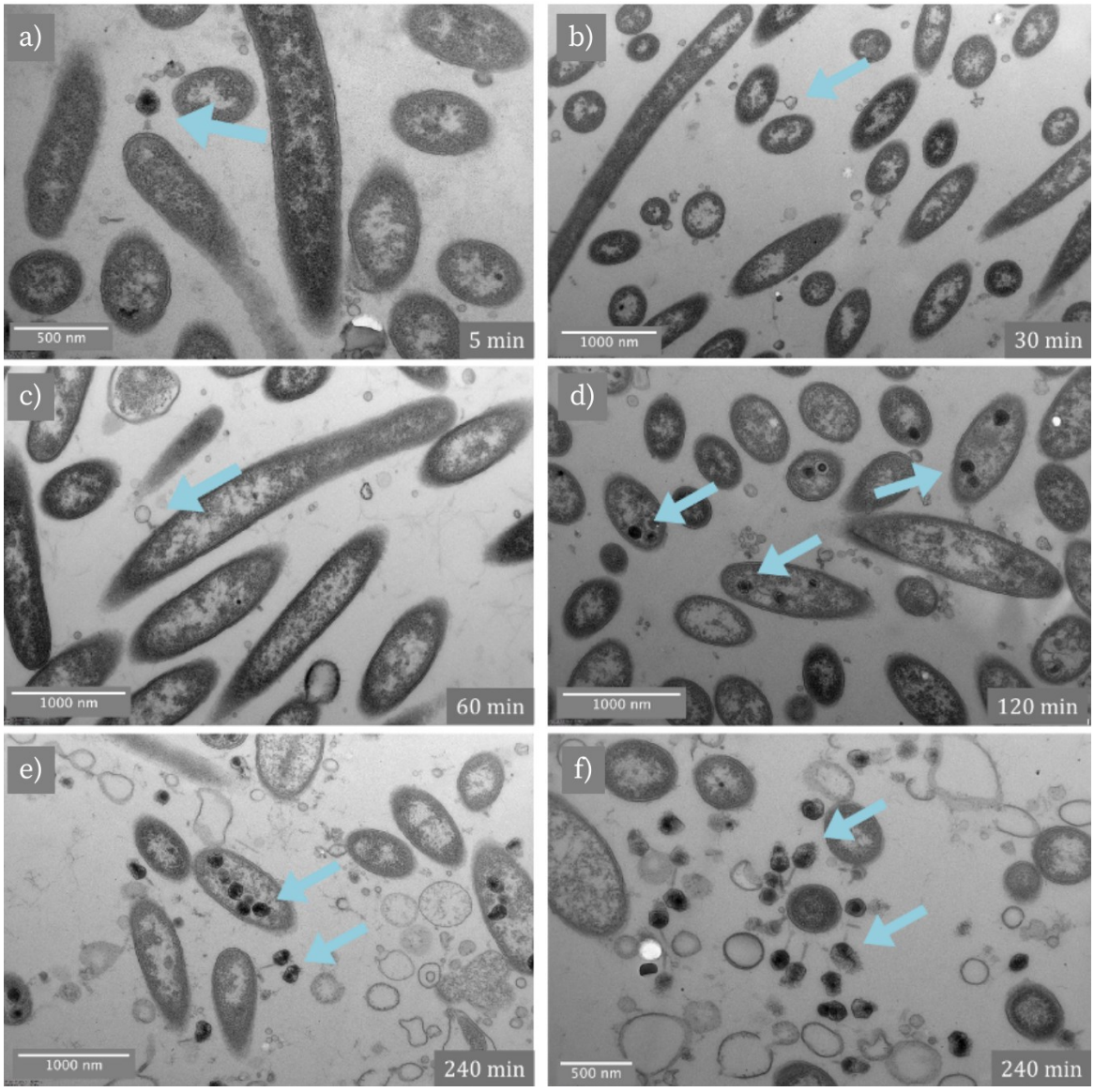



Figure 10. TEM images of thin slices made during Elf16 infection of *Flavobacterium* sp. B330 cells. Blue arrows indicate phage particles. Time stamps indicate minutes since beginning of phage infection.

## 3.7 Megaphage Elf16 encoded Cas9 is expressed at the early stages of infection

The expression of phage-encoded Cas9 during different points of infection can be used as an indication of the functionality of the phage's CRISPR-system. The expression level of Cas9 showed significant upregulation at 30 minutes post-infection, with relatively little expression at other time points (Figure 11). The expression level of helicase was at its highest at 120 minutes post-infection, with

some expression at 30 and 180 minutes, and no expression at 15 minutes (Figure 12).

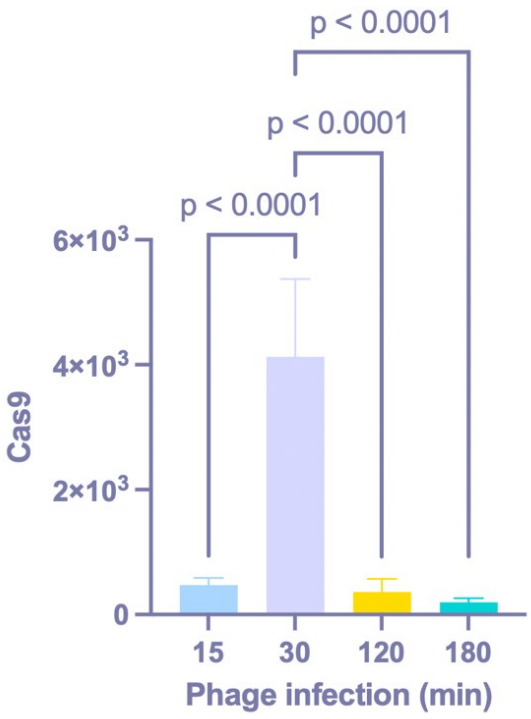



Figure 11. Phage-encoded Cas9 expression. Y-axis showcases the RQ value of Cas9 expression at each time point, compared to the calibrator sample (time point zero). Values were calculated as averages from three biological replicates, and two technical replicates. The p-values shown were calculated using one way ANOVA.

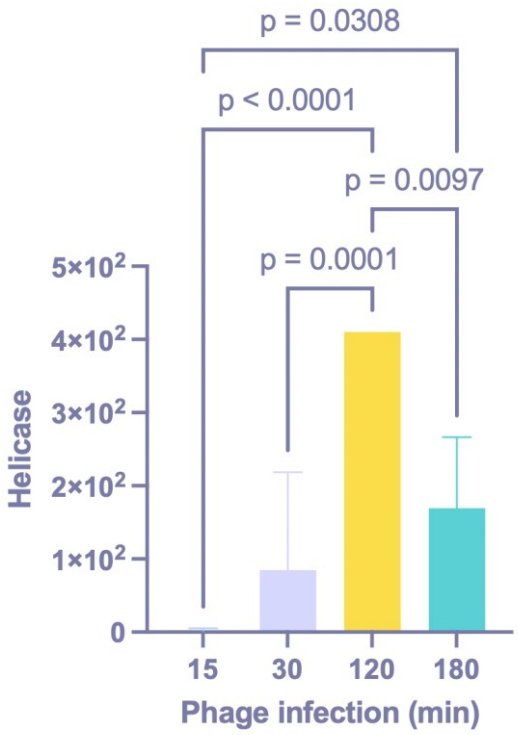



Figure 12. Phage-encoded helicase expression. Y-axis showcases the RQ value of helicase expression at each time point, compared to the calibrator sample (time point zero). Values were calculated as averages from three biological replicates, and two technical replicates. The p-values shown were calculated using one way ANOVA.

# 4 DISCUSSION

In this work, five putative CRISPR-Cas encoding megaphages were studied. Previous knowledge of these types of phages has been based only on metagenomic analysis, so there is no information regarding their biology. Therefore, this study aimed to use the first megaphage isolates to determine the host range, morphology, life cycle, and the functionality of the CRISPR-system of a CRISPR-Cas encoding megaphage.

## 4.1 Studied phages mostly infect *Flavobacterium* sp.

It was hypothesized that phages with large genome sizes could have wide host ranges. However, in this study the phages infected several strains of

*Flavobacterium* sp., but no other aquatic bacteria. The only exception to this was JoK81, which infected one strain of *Janthinobacteriumı* sp. in addition to strains of *Flavobacterium* sp. This indicates that these phages have a multitude of hosts within the *Flavobacterium* genus but might be specific to flavobacteria. Although phage specificity and host range vary greatly, from phages infecting a specific strain of bacteria to ones capable of infecting bacteria across genera, it is not uncommon for phages to be limited to a single genus (Koskella and Meaden 2013). This is due to the restrictions of bacterial surface receptors that phage receptor binding proteins or tail fibre proteins can bind to (Dowah and Clokie 2018).

It has been suggested that phages with a larger number of tRNAs might have a wide host range (Nazir et al. 2021). A larger genome might potentially accommodate more tRNA genes, which could mean that huge phages have wide host ranges. However, a larger genome does not necessarily correlate to more tRNAs, which would explain the specificity of the phages studied here. Additionally, it should be noted that many of the plates in the second round of host range testing showed visible signs of infection only on the edges of the phage drop area, instead of having unambiguous plaques. This complicated the categorising of the results and could have potentially led to mistakes in the results.

## 4.2 Elf16 latent phase is long

Lysis of *Flavobacterium* sp. B330 during Elf16 infection happened very gradually, with host OD dropping slowly over several hours. The period between infection starting and lysis beginning was also relatively long. For some comparison, phage T4 begins lysis after ~20 minutes of infecting *Escherichia coli* in optimal conditions (Couse 1968, Hadas et al. 1997), while phage λ lysis begins ~50 minutes after induction (Wang 2006). However, it is worth noting that there is much variation within phage lysis times, and many factors contribute to lysis. It is possible that the conditions tested here were suboptimal for Elf16 or otherwise

contributed to the long latent period in its life cycle. Furthermore, the OD measurements conducted showcased notable variation in Elf16 lysis time, suggesting that a change in conditions may easily affect the latent period of this phage. It is also possible that the large genome or particle size might affect the replication period, as it might take longer to assemble larger phage particles, as well as replicate and pack a longer genome.

Although all of the OD measurements suggest lysis beginning at ~3 hours, the adsorption to host test showed the phage titer already beginning to rise at 10 minutes after induction. In an adsorption test, the phage titer first decreases as phages attach to bacteria, and then rises as new phage particles emerge through lysis (Kropinski 2009). Based on OD measurements and the thin section images, this should not happen at 10 minutes, but much later. Since the adsorption test was performed only once, compared to the several different OD measurements, it is more likely that the results of the OD measuring are correct. For example, errors in sample handling, or variation in the physiological state of the bacterial culture could explain the differing results. It is also possible that the initial dip in phage titer is due to the reversible attachment of phages to host bacteria, and the subsequent rise in titer is due to the detachment of these phages, instead of new phage particles being released through lysis. Reversible attachment could also explain why the OD of an infected B330 drops gradually and seemingly at different paces, instead of all at once.

The thin section EM images also illustrate the Elf16 life cycle, supporting the OD measurement data. At 5 minutes, phages have attached to bacteria, and the dark phage capsid still contains the genome. At 30 and 60 minutes, the particles are empty, meaning the phage genome has been injected. However, no phage particles are yet visible inside the cells, indicating that the first steps of phage replication are in process. The large particle size might contribute to how long it takes for new particles to be visible inside cells. At 120 minutes, it is possible to observe new particles being assembled inside host cells, but no lysis has occurred, as the host bacteria are still intact. At 240 minutes, most bacteria

have lysed, and new phage particles have been released from the cells, although some particles are visible inside cells. Based on the lysis time of ~3 hours, these new particles are likely virions that have already started the infection cycle anew. The time points imaged match with the results of the OD based life cycle measurement. However, it should be noted that due to the sparse time points imaged, the specific time of genome injection or cell lysis cannot be determined.

## 4.3 Correlation between Elf16 genome and capsid size

The phage morphology shown in the EM images is very typical to most known phages. The phage particle is tailed, and the length of the tail varies based on whether the phage is attached to a host. This indicates that the tail is contractile. The variation in tail length observed with a contractile tail is due to the extension and contraction of the sheath during the infection process (Leiman and Shneider 2012). The tail appears longer when it is extended for attachment, and it contracts after the injection of the phage genome. The contracted tail can be seen in images of timepoints 30 and 60 minutes. During this injection, the sheath contracts to approximately half of its original length (Leiman and Shneider 2012), which in the case of Elf16 means a contraction from ~155 nm to ~80 nm.

It was hypothesized that Elf16 would have a large capsid to accommodate its large genome. At 130 nm, it is in line with the capsid size of jumbo phage φKZ, which has a capsid diameter of 120 nm and a tail length of 180 nm (Mesyanzhinov et al. 2002). However, φKZ has a genome of 280 kbp (Mesyanzhinov et al. 2002), which is much smaller than Elf16. It is also worth noting that phage T4 has a capsid of 120 × 86 nm, but a genome of only ~171 kbp (Rao et al. 2023). When it comes to phages with similar size capsids but vastly different genome lengths, it is possible that there are differences with how efficiently and tightly the genome is packaged into the capsid. For huge phages, their genome packaging machinery would possibly need to be highly efficient to package the long genome into the capsid during infection. The genome would also need to be packaged more tightly to accommodate for the genome length. It might be that this could aid

genome injection since a higher pressure within the capsid would drive the genome out.

Another megaphage with a genome size similar to Elf16 has been imaged with cryogenic electron microscopy (cryo-EM), and its capsid was determined to be 160 nm (unpublished data), which is notably larger than Elf16. It is possible that genome size does correlate with capsid size, but only to a certain point, or with several other factors. There can also be inaccuracy related to measuring capsid size from thin sections. It is worth noting that many phages have much smaller particles. For example, phage φX174 particles are only 25 nm in diameter (Yazaki 1981), and phage λ particles are approximately 60 nm in diameter (Bayer and Bocharov 1973).

## 4.4 Cas9 is expressed during early infection

Because Cas9 is so crucial for the CRISPR-Cas system as a whole (Redman et al. 2016), its expression can be used as an indicator of CRISPR-Cas system functionality. Here, both phage-encoded Cas9 and helicase expression were detected during phage infection. Helicase is needed during the genome replication phase of phage infection (Perumal et al. 2010). Cas9 was expressed at an earlier time point than helicase. This would mean it is expressed before replication, during the early stage on infection. One possible function of a phage-encoded CRISPR-Cas systems is evading host immune response (Al-Shayeb et al. 2020). Bacterial innate immune response begins rapidly after phage genome injection (Hampton et al. 2020). Therefore, for the purpose of evading this response, Cas9 would likely need to be expressed in relatively large amounts and early in the infection cycle. Therefore, the results presented here indicate that the Elf16 CRISPR-Cas system could possibly be functional. However, Cas9 expression alone is not enough to confirm this, and further research on the topic is required.

## 4.5 Genome analysis of Elf16 phage-host interactions

It was hypothesized that the phage could target the host bacterium genome to downregulate host gene expression or destroy host genome. This would require a match between phage CRISPR-array spacers and the host genome (Seed et al. 2013, Al-Shayeb et al. 2022). In the case of Elf16 and its host B330, there is no match. However, it is possible that Elf16 spacers match the genome of some of its other hosts. Further genome analysis with other known Elf16 hosts could reveal more about this possibility. Furthermore, it is possible that a CRISPR-Cas containing phage could target other phages with its spacers (Al-Shayeb et al. 2020). Therefore, it is possible that the function of the Elf16 CRISPR-Cas system is to compete with other phages, instead of attacking the host genome. Notably, there being no match between Elf16 spacers and B330 genome confirms that the phage having a matching spacer is not a requisite for successful Elf16 infection, since Elf16 can infect B330 without such a match.

Based on the functions of Phyre-annotated proteins, Elf16 genome has 20 potential AMGs. Many of the potential AMGs had close BlastP hits to proteins found in bacteria, suggesting that the gene in question could influence bacterial metabolism. However, this genome analysis alone is not enough to define genes as being AMGs. Therefore, these results are only preliminary, providing possible potential for future work, but not enough to make any conclusions based on this research alone.

# 5 CONCLUSIONS

Recently, many phages (viruses infecting bacteria) with unusually large genomes have been discovered in metagenomic analyses. However, no isolate studies have been conducted on these huge phages, until this thesis. Some phages encode a CRISPR-Cas system, which is an immune system typically found in prokaryotes. In this study, isolates of CRISPR-Cas containing megaphages were studied. All five phages infected mostly *Flavobacterium* sp.

species, with only one of them infecting one other bacterial species. Phage Elf16 life cycle was determined to be lytic, with a long latent phase preceding lysis. Elf16 particles were imaged, and they were shown to be large, with a contractile tail. Phage encoded Cas9 expression was shown to occur at the early phase of infection. This indicates that the phage could have a functional CRISPR-Cas system. However, genome analysis shows no match between Elf16 spacers and *Flavobacterium* sp. B330 strain genome. Since Elf16 is able to infect B330, a spacer match is not a requisite for a successful infection. Still, it is possible that Elf16 spacers target either other hosts, or other competing phages.

## ACKNOWLEDGEMENTS

I would like to thank the Emil Aaltonen Foundation for the funding of this project. Special thanks to my supervisors Elina Laanto and Lotta-Riina Sundberg for their advice and guidance during this project, and to Daniel de Oliveira Patricio for his invaluable help during the qPCR experiments.

Lohtaja April 8, 2024
Meeri Niemi

## REFERENCES


Ackermann H.-W. 2007. 5500 Phages examined in the electron microscope. *Archives of Virology* 152: 227–243.

Ackermann H.-W. 2009. Phage classification and characterization. *Methods in Molecular Biology (Clifton, N.J.)* 501: 127–140.

Almeida G.M.F., Laanto E., Ashrafi R. & Sundberg L.-R. 2019. Bacteriophage Adherence to Mucus Mediates Preventive Protection against Pathogenic Bacteria. mBio 10: 10.1128/mbio.01984-19.

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

Al-Shayeb B., Skopintsev P., Soczek K.M., Stahl E.C., Li Z., Groover E., Smock D., Eggers A.R., Pausch P., Cress B.F., Huang C.J., Staskawicz B., Savage D.F., Jacobsen S.E., Banfield J.F. & Doudna J.A. 2022. Diverse virus-encoded CRISPR-Cas systems include streamlined genome editors. *Cell* 185: 4574-4586.e16.

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Batinovic S., Wassef F., Knowler S.A., Rice D.T.F., Stanton C.R., Rose J., Tucci J., Nittami T., Vinh A., Drummond G.R., Sobey C.G., Chan H.T., Seviour R.J., Petrovski S. & Franks A.E. 2019. Bacteriophages in Natural and Artificial Environments. *Pathogens* 8: 100.

Bayer M.E. & Bocharov A.F. 1973. The capsid structure of bacteriophage lambda. *Virology* 54: 465–475.

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

Bernardet J.-F. & Bowman J.P. 2006. The Genus Flavobacterium. In: Dworkin M., Falkow S., Rosenberg E., Schleifer K.-H. & Stackebrandt E. (eds.), *The Prokaryotes: Volume 7: Proteobacteria: Delta, Epsilon Subclass*, Springer, New York, NY, pp. 481–531.

Borriss M., Helmke E., Hanschke R. & Schweder T. 2003. Isolation and characterization of marine psychrophilic phage-host systems from Arctic sea ice. *Extremophiles* 7: 377–384.

Bouras G., Nepal R., Houtak G., Psaltis A.J., Wormald P.-J. & Vreugde S. 2023. Pharokka: a fast scalable bacteriophage annotation tool. *Bioinformatics* 39: btac776.

Breitbart M., Thompson L., Suttle C. & Sullivan M. 2007. Exploring the Vast Diversity of Marine Viruses. *Oceanography* 20: 135–139.

Brum J.R., Schenck R.O. & Sullivan M.B. 2013. Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. *The ISME Journal* 7: 1738–1751.

Castillo D., Higuera G., Villa M., Middelboe M., Dalsgaard I., Madsen L. & Espejo R.T. 2012. Diversity of Flavobacterium psychrophilum and the potential use of its phages for protection against bacterial cold water disease in salmonids. Journal of Fish Diseases 35: 193–201.

Couse N.L. 1968. Control of Lysis of T4-infected Escherichia coli | Journal of Virology.

Couvin D., Bernheim A., Toffano-Nioche C., Touchon M., Michalik J., Néron B., Rocha E.P.C., Vergnaud G., Gautheret D. & Pourcel C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research* 46: W246–W251.

Dennehy J.J. & Abedon S.T. 2021a. Adsorption: Phage Acquisition of Bacteria. In: Harper D.R., Abedon S.T., Burrowes B.H. & McConville M.L. (eds.), *Bacteriophages: Biology, Technology, Therapy*, Springer International Publishing, Cham, pp. 93–117.

Dennehy J.J. & Abedon S.T. 2021b. Phage Infection and Lysis. In: Harper D.R., Abedon S.T., Burrowes B.H. & McConville M.L. (eds.), *Bacteriophages: Biology, Technology, Therapy*, Springer International Publishing, Cham, pp. 341–383.

Devoto A.E., Santini J.M., Olm M.R., Anantharaman K., Munk P., Tung J., Archie E.A., Turnbaugh P.J., Seed K.D., Blekhman R., Aarestrup F.M., Thomas B.C. & Banfield J.F. 2019. Megaphages infect Prevotella and variants are widespread in gut microbiomes. *Nature Microbiology* 4: 693–700.

Dowah A.S.A. & Clokie M.R.J. 2018. Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophysical Reviews* 10: 535–542.

Edgar R.C. 2022. High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. : 2021.06.20.449169.

Frank J.A., Lorimer D., Youle M., Witte P., Craig T., Abendroth J., Rohwer F., Edwards R.A., Segall A.M. & Burgin A.B. Jr. 2013. Structure and function of a cyanophage-encoded peptide deformylase. *The ISME Journal* 7: 1150–1160.

Friedman S.D., Genthner F.J., Gentry J., Sobsey M.D. & Vinjé J. 2009. Gene Mapping and Phylogenetic Analysis of the Complete Genome from 30 Single-Stranded RNA Male-Specific Coliphages (Family Leviviridae). *Journal of Virology* 83: 11233–11243.

Fuhrman J.A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399: 541–548.

Gabler F., Nam S.-Z., Till S., Mirdita M., Steinegger M., Söding J., Lupas A.N. & Alva V. 2020. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics* 72: e108.

García-López M.-L., Santos J.-Á. & Otero A. 1999. FLAVOBACTERIUM. In: Robinson R.K. (ed.), *Encyclopedia of Food Microbiology*, Elsevier, Oxford, pp. 820–826.

Hadas H., Einav M., Fishov I. & Zaritsky A. 1997. Bacteriophage T4 Development Depends on the Physiology of its Host Escherichia Coli. *Microbiology* 143: 179–185.

Hampton H.G., Watson B.N.J. & Fineran P.C. 2020. The arms race between bacteria and their phage foes. *Nature* 577: 327–336.

Harding K.R., Kyte N. & Fineran P.C. 2023. Jumbo phages. Current Biology 33: R750–R751.

Hardy J.M., Dunstan R.A., Lithgow T. & Coulibaly F. 2022. Tall tails: cryo-electron microscopy of phage tail DNA ejection conduits. *Biochemical Society Transactions* 50: 459-22W.

Hoikkala V., Ravantti J., Díez-Villaseñor C., Tiirola M., Conrad R.A., McBride M.J., Moineau S. & Sundberg L.-R. 2021. Cooperation between Different CRISPR-Cas Types Enables Adaptation in an RNA-Targeting System. *mBio* 12.

Kelley L.A., Mezulis S., Yates C.M., Wass M.N. & Sternberg M.J.E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10: 845–858.

Kohm K. & Hertel R. 2021. The life cycle of SPβ and related phages. *Archives of Virology* 166: 2119–2130.

Koskella B. & Meaden S. 2013. Understanding Bacteriophage Specificity in Natural Microbial Communities. *Viruses* 5: 806–823.

Kropinski A.M. 2009. Measurement of the rate of attachment of bacteriophage to cells. *Methods in Molecular Biology (Clifton, N.J.)* 501: 151–155.

Laanto E., Bamford J.K.H., Laakso J. & Sundberg L.-R. 2012. Phage-Driven Loss of Virulence in a Fish Pathogenic Bacterium. *PLOS ONE* 7: e53157.

Laanto E., Mäntynen S., De Colibus L., Marjakangas J., Gillum A., Stuart D.I., Ravantti J.J., Huiskonen J.T. & Sundberg L.-R. 2017. Virus found in a boreal lake

links ssDNA and dsDNA viruses. *Proceedings of the National Academy of Sciences of the United States of America* 114: 8378–8383.

Laanto E. & Oksanen H.M. 2023. Three Phages from a Boreal Lake during Ice Cover Infecting Xylophilus, Caulobacter, and Polaromonas Species. *Viruses (Basel)* 15.

Laanto E., Sundberg L.-R. & Bamford J.K.H. 2011. Phage Specificity of the Freshwater Fish Pathogen Flavobacterium columnare. *Applied and Environmental Microbiology* 77: 7868–7872.

Leiman P.G. & Shneider M.M. 2012. Contractile Tail Machines of Bacteriophages. In: Rossmann M.G. & Rao V.B. (eds.), *Viral Molecular Machines*, Springer US, Boston, MA, pp. 93–114.

Lim E.S., Zhou Y., Zhao G., Bauer I.K., Droit L., Ndao I.M., Warner B.B., Tarr P.I., Wang D. & Holtz L.R. 2015. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nature medicine* 21: 1228–1234.

Loeb T. & Zinder N.D. 1961. A BACTERIOPHAGE CONTAINING RNA. *Proceedings of the National Academy of Sciences of the United States of America* 47: 282–289.

Makarova K.S. & Koonin E.V. 2015. Annotation and Classification of CRISPR-Cas Systems. *Methods in molecular biology (Clifton, N.J.)* 1311: 47–75.

Makarova K.S., Wolf Y.I., Alkhnbashi O.S., Costa F., Shah S.A., Saunders S.J., Barrangou R., Brouns S.J.J., Charpentier E., Haft D.H., Horvath P., Moineau S., Mojica F.J.M., Terns R.M., Terns M.P., White M.F., Yakunin A.F., Garrett R.A., Oost J. van der, Backofen R. & Koonin E.V. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews. Microbiology* 13: 722–736.

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., Moineau S., Mojica F.J.M., Scott D., Shah S.A., Siksnys V., Terns M.P., Venclovas Č., White M.F., Yakunin A.F., Yan W., Zhang F., Garrett R.A., Backofen R., Oost J. van der, Barrangou R. & Koonin E.V. 2020. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology* 18: 67–83.

Martiny A.C., Huang Y. & Li W. 2009. Occurrence of phosphate acquisition genes in Prochlorococcus cells from different ocean regions. *Environmental Microbiology* 11: 1340–1347.

McBride M.J., Xie G., Martens E.C., Lapidus A., Henrissat B., Rhodes R.G., Goltsman E., Wang W., Xu J., Hunnicutt D.W., Staroscik A.M., Hoover T.R., Cheng Y.-Q. & Stein J.L. 2009. Novel features of the polysaccharide-digesting gliding bacterium Flavobacterium johnsoniae as revealed by genome sequence analysis. *Applied and Environmental Microbiology* 75: 6864–6875.

Mertens P. 2004. The dsRNA viruses. *Virus Research* 101: 3–13.

Mesyanzhinov V.V., Robben J., Grymonprez B., Kostyuchenko V.A., Bourkaltseva M.V., Sykilinda N.N., Krylov V.N. & Volckaert G. 2002. The genome of bacteriophage φKZ of Pseudomonas aeruginosa11Edited by M. Gottesman. *Journal of Molecular Biology* 317: 1–19.

Michniewski S., Rihtman B., Cook R., Jones M.A., Wilson W.H., Scanlan D.J. & Millard A. 2021. A new family of "megaphages" abundant in the marine environment. *ISME Communications* 1: 1–4.

Microsoft Corporation. 2018. Microsoft Excel. Retrieved from https://office.microsoft.com/excel

Middelboe M., Holmfeldt K., Riemann L., Nybroe O. & Haaber J. 2009. Bacteriophages drive strain diversification in a marine Flavobacterium: implications for phage resistance and physiological properties. *Environmental Microbiology* 11: 1971–1982.

Mäntynen S., Laanto E., Sundberg L.-R., Poranen M.M., Oksanen H.M. & Report Consortium I. 2020. ICTV Virus Taxonomy Profile: Finnlakeviridae. *The Journal of General Virology* 101: 894–895.

Nazir A., Ali A., Qing H. & Tong Y. 2021. Emerging Aspects of Jumbo Bacteriophages. *Infection and Drug Resistance* 14: 5041–5055.

Nobrega F.L., Vlot M., Jonge P.A. de, Dreesens L.L., Beaumont H.J.E., Lavigne R., Dutilh B.E. & Brouns S.J.J. 2018. Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology* 16: 760–773.

Pausch P., Al-Shayeb B., Bisom-Rapp E., Tsuchida C.A., Li Z., Cress B.F., Knott G.J., Jacobsen S.E., Banfield J.F. & Doudna J.A. 2020. CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science (New York, N.Y.)* 369: 333–337.

Perumal S.K., Raney K.D. & Benkovic S.J. 2010. Analysis of the DNA translocation and unwinding activities of T4 phage helicases. *Methods* 51: 277–288.

Rao V.B., Fokine A., Fang Q. & Shao Q. 2023. Bacteriophage T4 Head: Structure, Assembly, and Genome Packaging. *Viruses* 15: 527.

Redman M., King A., Watson C. & King D. 2016. What is CRISPR/Cas9? *Archives of Disease in Childhood. Education and Practice Edition* 101: 213.

Richter C.A. & Pate J.L. 1988. Temperate Phages and Bacteriocins of the Gliding Bacterium Cytophaga johnsonae. *Microbiology* 134: 253–262.

Salmond G.P.C. & Fineran P.C. 2015. A century of the phage: past, present and future. *Nature Reviews Microbiology* 13: 777–786.

Schneider C.A., Rasband W.S. & Eliceiri K.W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9: 671–675.

Seed K.D., Lazinski D.W., Calderwood S.B. & Camilli A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* 494: 489–491.

Song Y.L., Fryer J.L. & Rohovec J.S. 1988. Comparison of six media for the cultivation of Flexibacter columnaris. *Fish Pathology* 23: 91–94.

Stenholm A.R., Dalsgaard I. & Middelboe M. 2008. Isolation and Characterization of Bacteriophages Infecting the Fish Pathogen Flavobacterium psychrophilum. *Applied and Environmental Microbiology* 74: 4070–4078.

Sullivan M.B., Huang K.H., Ignacio-Espinoza J.C., Berlin A.M., Kelly L., Weigele P.R., DeFrancesco A.S., Kern S.E., Thompson L.R., Young S., Yandava C., Fu R., Krastins B., Chase M., Sarracino D., Osburne M.S., Henn M.R. & Chisholm S.W. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology* 12: 3035–3056.

Wang I.-N. 2006. Lysis Timing and Bacteriophage Fitness. *Genetics* 172: 17–26.

Warwick-Dugdale J., Buchholz H.H., Allen M.J. & Temperton B. 2019. Host-hijacking and planktonic piracy: how phages command the microbial high seas. *Virology Journal* 16: 15.

Willenbücher K., Wibberg D., Huang L., Conrady M., Ramm P., Gätcke J., Busche T., Brandt C., Szewzyk U., Schlüter A., Barrero Canosa J. & Maus I. 2022. Phage Genome Diversity in a Biogas-Producing Microbiome Analyzed by Illumina and Nanopore GridION Sequencing. *Microorganisms* 10: 368.

Yazaki K. 1981. Electron microscopic studies of bacteriophage phi X174 intact and "eclipsing' particles, and the genome by the staining, and shadowing method. *Journal of Virological Methods* 2: 159–167.

Young R. 1992. Bacteriophage lysis: mechanism and regulation. *Microbiological Reviews* 56: 430–481.

# APPENDIX 1 FULL HOST RANGE RESULTS

Table 1 Full results from the second round on host range testing, after the preliminary testing. Each phage was tested with three replicate plates with a dilution series of $10^0$ – $10^{-6}$ titered as 10 µl drops. Results were categorized as C = clear drop area, P = partially clear drop area, + = plaques that could not be counted, number = number of plaques, I = inhibition and - = no visible infection.

| Bacteria | Phage | Dilution | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $10^0$ | $10^{-1}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
| B178 | Elf16 | C | P | - | - | - | - | - |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | JoK79 | + | + | + | + | + | + | + |
| | | P | P | P | P | P | P | P |
| | | P | - | - | - | - | - | - |
| | JoK80 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | JoK81 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK82 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | - |
| B209 | Elf16 | C | C | + | + | + | 23 | + |
| | | C | C | C | + | + | 12 | P |
| | | C | C | + | + | + | 13 | P |
| | JoK79 | + | + | P | P | P | P | P |
| | | + | P | P | P | P | P | P |
| | | + | P | P | P | P | P | P |
| | JoK80 | + | P | P | P | P | P | P |
| | | + | P | P | P | P | P | P |
| | | + | P | P | P | P | P | P |
| | JoK81 | + | P | P | P | P | P | P |
| | | + | P | P | P | P | P | P |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | + | + | P | P | P | P | P |
| | JoK82 | P | P | P | P | P | - | - |
| | | P | P | P | P | P | P | - |
| | | - | - | - | - | - | - | - |
| B80 | Elf16 | C | P | P | P | P | P | P |
| | | C | P | P | P | - | - | - |
| | | C | C | P | P | P | P | - |
| | JoK80 | P | I | I | I | I | I | |
| | | P | I | I | I | I | I | |
| | | P | I | I | I | I | I | |
| | JoK81 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | JoK82 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| B121 | Elf16 | - | - | - | P | - | P | - |
| | | - | - | - | P | - | P | - |
| | | P | P | P | P | P | P | - |
| | JoK79 | - | - | P | - | P | - | - |
| | | - | - | P | - | P | - | - |
| | | - | - | P | - | P | - | P |
| | JoK80 | - | P | P | P | P | P | - |
| | | - | P | P | P | P | P | P |
| | | - | - | - | P | P | P | - |
| | JoK81 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK82 | - | - | - | P | - | P | P |
| | | - | - | - | - | - | - | - |
| | | - | - | P | P | P | P | - |
| B127 | JoK79 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK80 | - | - | P | P | P | P | P |
| | | - | - | P | P | P | P | P |
| | | - | P | P | P | P | P | P |
| | JoK81 | - | - | - | - | - | - | - |
| | | P | P | P | P | P | P | P |
| | | - | C | C | P | P | P | P |
| | JoK82 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | P | P | P | P | P | - | - |
| B105 | JoK79 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK80 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK81 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK82 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| B221 | JoK79 | C | + | + | + | P | P | P |
| | | P | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | JoK80 | C | + | + | + | P | P | P |
| | | C | + | + | + | + | + | P |
| | | C | P | P | P | P | P | P |
| | JoK81 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | - |
| | | - | - | - | - | - | - | - |
| | JoK82 | C | P | P | - | - | - | - |
| | | C | - | - | - | - | - | - |
| | | C | P | P | P | P | P | - |
| B206 | Elf16 | C | I | I | I | I | I | I |
| | | P | P | P | P | P | P | - |
| | | C | P | P | P | P | P | P |
| | JoK79 | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | JoK82 | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | | P | I | I | I | I | I | I |
| B208 | Elf16 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK80 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | JoK81 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | C | P | P | P | P | P | P |
| | JoK82 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | - |
| B171 | Elf16 | P | P | P | P | P | P | - |
| | | P | P | P | P | P | P | - |
| | | P | P | P | P | P | P | P |
| | JoK79 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK81 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK82 | C | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| B207 | Elf16 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | P |
| | JoK79 | P | P | P | P | P | P | P |
| | | C | + | + | + | + | + | P |
| | | C | + | + | + | + | + | + |
| | JoK80 | + | + | + | P | P | P | P |
| | | + | + | P | P | P | P | P |
| | | + | + | P | P | P | P | P |
| | JoK81 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | P |
| | JoK82 | C | 30 | P | P | P | 18 | P |
| | | C | P | P | P | P | 17 | P |
| | | C | P | P | P | P | 19 | - |
| B174 | JoK79 | C | P | P | P | P | P | - |
| | | P | P | P | P | P | P | P |
| | | P | P | + | P | + | P | P |
| | JoK80 | + | + | P | P | P | P | P |
| | | + | + | P | P | P | P | P |
| | | + | + | P | P | P | P | P |
| | JoK81 | P | P | P | P | P | P | - |
| | | P | P | P | P | P | P | - |
| | | P | P | P | P | P | P | - |
| | JoK82 | C | C | P | P | P | P | P |
| | | P | P | P | P | P | P | P |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | P | P | P | P | P | P | P |
| B126 | Elf16 | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | JoK81 | + | + | P | P | P | P | - |
| | | - | - | - | - | - | - | - |
| | | - | + | - | - | - | - | - |
| | JoK82 | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| B205 | JoK79 | - | P | P | P | P | P | P |
| | | P | P | P | P | + | P | P |
| | | P | P | P | P | P | P | P |
| | JoK81 | P | P | P | P | P | P | P |
| | | P | P | P | P | P | P | - |
| | | P | P | P | P | P | P | P |
| B15 | JoK81 | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| B169 | Elf16 | C | C | C | + | + | + | 30 |
| | | C | C | C | C | + | + | 31 |
| | | C | C | C | C | + | + | 33 |
| | JoK79 | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | JoK80 | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | JoK81 | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | JoK82 | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| | | C | I | I | I | I | I | I |
| B158 | JoK79 | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | JoK80 | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| | | - | - | - | - | - | - | - |
| B28 | JoK79 | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | C | P | P | P | P | P | - |
| | JoK80 | C | - | - | - | - | - | - |
| | | C | P | P | P | P | P | P |
| | | C | P | P | P | P | P | - |
| | JoK82 | C | P | P | P | - | - | - |
| | | C | P | P | P | - | - | - |
| | | C | P | P | P | - | - | - |
| B180 | Elf16 | + | + | + | - | - | - | - |
| | | + | + | + | - | - | - | - |
| | JoK79 | P | P | P | P | P | - | - |
| | | C | P | P | - | - | - | - |
| | JoK80 | C | + | + | P | P | P | P |
| | | + | + | + | P | P | P | P |
| | | P | P | P | P | P | P | - |
| | JoK81 | + | + | + | - | - | - | - |
| | | + | + | + | - | - | - | - |
| | | + | + | + | - | - | - | - |
| | JoK82 | C | P | P | + | P | P | P |
| | | C | P | P | + | P | P | P |
| | | C | + | + | + | + | 3 | P |
| UW101-36 | Elf16 | + | I | I | I | I | I | I |
| | | + | I | I | I | I | I | I |
| | | + | I | I | I | I | I | I |
| | JoK79 | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| | JoK80 | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| | JoK81 | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| | JoK82 | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| | | I | I | I | I | I | I | I |
| B224 | JoK79 | C | P | P | P | P | P | - |
| | | C | P | P | P | P | P | - |
| | | C | P | P | P | P | P | - |
| | JoK80 | C | P | P | P | P | P | P |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | C | P | P | P | P | P | P |
| | C | P | P | P | P | P | - |
| JoK81 | C | P | P | P | P | P | P |
| | C | P | P | P | P | P | P |
| | C | P | P | P | P | P | P |
| JoK82 | C | P | P | P | P | - | - |
| | C | P | P | P | P | - | - |
| | C | P | P | P | P | - | - |

# APPENDIX 2 ELF16 SPACER SEQUENCES

Table 1 Elf16 spacer sequences. Spacers were identified CRISPRCasFinder (Couvin et al. 2018). Spacers marked with * have been identified manually.

| | Sequence |
|---|---|
| spacer 1 | AAAATTTTGATTATGGCAATTTTATTAC |
| spacer 2 | TCAAAGAACAAGTACTACATTGTGGGTTC |
| spacer 3 | CAAAGAAACATAGGCTACTGTTCAGAATAA |
| spacer 4 | ATAATATTGTTAAATACTTAACTAAGTAT |
| spacer 5 | TTTGCAAAAAGTGAATACGAAAGAAATAGG |
| spacer 6 | AATTTTCAAATACTTCAATAGCATCTTCTA |
| spacer 7 | CTTTAGTGATAAGGGAAACACGATTGGCTT |
| spacer 8 | ATGCGACTAGTGCAGCAGGTACATCTTACG |
| spacer 9 | ATCATATCTTTTAGAAAGCAGAAGGGCTAA |
| spacer 10 | TCTTTGTTGTGCCATATGCTTAGCATCGTA |
| spacer 11* | ACAAAAGAACTTCAGAGCAGA |
| spacer 12* | CCCAATTTGGCGATGCTGTTGTATGATGG |
| spacer 13 | GTATGATATTTTATCACTCAATACTGT |
| spacer 14 | TCGGACGCATCATAATACCAAAGTTCTTT |
| spacer 15 | TGAATCAACAGTTTACGTGGTTGCAAACAA |
| spacer 16 | TATTAAATGTGATTTTTTTTATGCCAAATG |
| spacer 17 | ACAAGATTGTTATATTAAGATAAATGATTG |
| spacer 18 | ATTGAATTCTAATACTGGTTCTACGATGAA |
| spacer 19 | TGGTCGGAAAGTATATACAACGAGGATTCT |
| spacer 20 | TCAGGACAATTCCATCACAATAGAGATAAA |
| spacer 21 | TTGAATTCAGATAACTTTAACCATCTGAAA |
| spacer 22 | TTGTAGACGTGAGGATACTCTGACTTAAA |
| spacer 23 | ATGCTTAGAGGAGTAACCGATAAACTAAT |
| spacer 24 | TTAATCGATACAATTTTAGGTATTTATACT |
| spacer 25 | GTATAGTAGTAACAGAACATTTCCTGCTGG |
| spacer 26 | AAGAATAGAAATAGATGAAGTCAAAACCAA |
| spacer 27 | ATATGGAATTACTGATGCATTCTCTGTACC |

# APPENDIX 3 ELF16 POTENTIAL AMGS

Table 1 Full results of the Elf16 AMG analysis. AMGs were identified from annotations made with Phyre2 (Kelley et al. 2015). Phyre2 hits with high confidence (>90%), coverage (>50%), and functions linking them to bacterial metabolisim, were then analysed using BlastP (Altschul 1990).

| Locus | Phyre annotation | Closest BlastP hit | BlastP e-value | BlastP identity [%] | BlastP coverage [%] | BlastP accession | Homologues in other phages (e-value <1e-5) |
|---|---|---|---|---|---|---|---|
| 292 | dna processing chain a (dpra) | DUF2493 domain-containing protein [*Elusimicrobiota bacterium*] | 6.00E-30 | 52.68 | 86 | MDR0676078.1 | CAB4143417.1 |
| 311 | atp-dependent clp protease proteolytic subunit | ATP-dependent Clp protease proteolytic subunit [*Candidatus Muirbacterium halophilum*] | 2.00E-80 | 56.42 | 95 | MCK9475552.1 | |
| 470 | grea protein; | transcription elongation factor GreA [*Paludibacteraceae bacterium*] | 1.00E-43 | 51.7 | 94 | MBR1515399.1 | |
| 477 | leucine--trna ligase; | class I tRNA ligase family protein [*Pseudomonas* sp. Irchel 3H3] | 1.00E-136 | 34.79 | 99 | WP_095165231.1 | |
| 487 | phosphate starvation-inducible protein; | PhoH family protein [*Actinomycetota bacterium*] | 3.00E-107 | 67.38 | 93 | NCX04905.1 | CAB4175533.1, CAB4159516.1, CAB4175523.1 |
| 518 | ribonucleoside-diphosphate reductase 1 subunit beta; | ribonucleotide-diphosphate reductase subunit beta [*bacterium*] | 2.00E-149 | 60.4 | 96 | NBP58150.1 | CAG7579770.1, APC44338.1, ATN93506.1 |
| 522 | ribonucleotide reductase r1 protein; | ribonucleotide reductase [*Tenacibaculum* phage PTm1] | 1.00E-162 | 55.17 | 87 | YP_009873733.1 | YP_009873733.1, CAG7579778.1, , BBI90748.1 |
| 557 | adp-sugar pyrophosphatase; | NUDIX domain-containing protein [*Verrucomicrobiota bacterium*] | 1.00E-73 | 55.56 | 98 | NDF16453.1 | CAG7579809.1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 562 | atp-dependent clp protease proteolytic subunit; | ATP-dependent Clp protease proteolytic subunit [*Flavobacteriaceae bacterium*] | 4.00E-86 | 54.07 | 97 | NBU82000.1 | CAB4129906.1 |
| 577 | thioredoxin; | MAG: thioredoxin [*Sulfurovum* sp.] | 4.00E-25 | 44 | 95 | RUM75271.1 | |
| 584 | glutamate--trna ligase; | glutamate--tRNA ligase [*Flavobacterium branchiophilum*] | 1.00E-163 | 51.11 | 98 | WP_097553176.1 | |
| 701 | guanosine 5'-monophosphate reductase; | IMP dehydrogenase [*Candidatus Dojkabacteria bacterium*] | 4.00E-101 | 49.41 | 98 | PCI28490.1 | UVM93634.1, UWD58689.1, UVY03294.1, UVX33254.1, UWG86772.1, YP_010111148.1, DAM35126.1, DAE74178.1, DAW21837.1, UWG87172.1, DAE56667.1, DAR10922.1, DAI39669.1, DAI57668.1, DAQ87643.1, DAG93824.1, DAH01382.1, DAK96519.1, DAK49851.1, DAD66455.1, |
| 873 | dihydrofolate reductase; | type 3 dihydrofolate reductase [*Candidatus Thioglobus* sp.] | 6.00E-32 | 45.8 | 99 | WP_291926873.1 | |
| 922 | ribose-phosphate pyrophosphokinase 1; | glycosyltransferase [*Candidatus Elulimicrobium humile*] | 9.00E-154 | 52.17 | 99 | NBP56719.1 | CAG7581274.1 |
| 924 | mazg-like nucleoside triphosphate pyrophosphohydrolase; | nucleoside triphosphate pyrophosphohydrolase family protein [*Bacteroidota bacterium*] | 1.00E-53 | 57.72 | 97 | MBR9923235.1 | |
| 934 | heat shock protein 70; | molecular chaperone DnaK [*Candidatus Muirbacterium halophilum*] | 0.0 | 66.75 | 98 | MCK9477072.1 | |
| 940 | lon protease; | AAA family ATPase [*Candidatus Muirbacterium halophilum*] | 5.00E-137 | 56.52 | 94 | MCK9477498.1 | CAG7581549.1 , YP_010659354.1, YP_007003156.1, YP_010659445.1, CAB4221149.1, YP_007674104.1, CAB4125928.1, YP_010658547.1, WMM35533.1, CAB4130091.1, YP_010105051.1, DAP73097.1, |

| | | | | | | |
|---|---|---|---|---|---|---|
| 966 | nicotinamide phosphoribosyltransferase; | nicotinate phosphoribosyltransferase [*bacterium]* | 0.0 | 61.13 | 91 | NBO22640.1 |
| 967 | ribose-phosphate pyrophosphokinase 1; | ribose-phosphate diphosphokinase [*Spirochaetota bacterium*] | 5.00E-105 | 50.14 | 97 | TXG80798.1 |
| 992 | thioredoxin reductase; | thioredoxin-disulfide reductase [*Formosa algae*] | 3.00E-139 | 63.43 | 99 | WP_103192558.1 |