New *Microviridae* (pro)-phages infecting marine *Alphaproteobacteria*

Von der Fakultät V für Mathematik und Naturwissenschaften der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades und Titels

Doktor der Naturwissenschaften (Dr. rer. nat.)

angenommene Dissertation

von

Falk Zucker

Geboren am 18.10.1985 in Frankfurt (Oder)

Gutachter:	Prof. Dr. Meinhard Simon
Weitere Gutachter:	Assoc. Prof. Dr. Karin Holmfeldt
	PD Dr. Bert Engelen

Tag der Disputation: 12.12.2022

Abstrakt

Bakteriophagen der *Microviridae* Familie stellen eine der Hauptklassen von einzelstrangigen (ss)DNA-Phagen dar. Ihre kultivierten Vertreter sind lytisch und infizieren *Proteobacteria*, *Bacteroidetes* und *Chlamydiae*. In den Genomen von *Bacteroidales*, *Hyphomicrobiales* und *Enterobacteriaceae* konnten Prophagen vorhergesagt werde und diese gruppieren sich innerhalb der Unterfamilien „Alpavirinae“, „Amoyvirinae“ und *Gokushovirinae*. Bereits zuvor wurde der Bakteriophage „Ascunsovirus oldenburgi“ ICBM5 aus der Nordsee isoliert und hat sich als entfernt verwandt mit bekannten *Microviridae* erwiesen. Der neuartige Phage infiziert *Sulfitobacter dubius* SH24-1b und verwendet dabei sowohl eine lytische als auch eine Trägerzustands-Lebensstrategie. Ein kompletter Infektionszyklus wurde mittels eines kombinierten Versuchsaufbaus aus einstufigen Wachstumskurven und Phagen gerichteter fluorescence *in situ* Hybridisierung (FISH) durchgeführt und analysiert. Dadurch konnte der Prozentsatz infizierter Zellen sowie die Anzahl viraler Genome pro Zelle im Verlauf einer Infektion quantifiziert werden. Darüber hinaus konnte gezeigt werden, dass der Phage ICBM5 sein Genom mindestens 50 min nach Beginn der Infektion repliziert und etwa 110 min nach der Infektion eine bakterielle Lyse verursacht. Während der Lyse wurde die „burst size“ auf ungefähr 250 Phagenpartikel berechnet, was ungefähr der Quantifizierungsanalyse entspricht, die durch phageFISH generiert wurde. Darüber hinaus wurden durch die Verwendung von ICBM5-Proteinen Sequenzen 65 neue *Microviridae*-Prophagen und -Episomen in Bakteriengenomen aus zugänglichen Ressourcen gefunden und 47 umweltvirale Genome (EVGs) aus verschiedenen Viromen aufgedeckt. Genom-Clustering basierend auf Proteingehalt und phylogenetischer Analyse zeigte, dass ICBM5 zusammen mit Rhizobium-Phagen, neuen Prophagen, Episomen und EVGs innerhalb zweier neuer phylogenetischer Kladen gruppiert werden kann. Diese bekamen hier vorläufig den Rang einer Unterfamilie zugewiesen und wurden „Tainavirinae“ und „Occultatumvirinae“ genannt. Beide infizieren Mitglieder der Ordnung Rhodobacterales. Occultatumviren infizieren außerdem auch Hyphomicrobiales, einschließlich stickstofffixierender Endosymbionten aus kosmopolitischen Leguminosen. Eine biogeografische Verteilungsanalyse zeigte, dass Tainaviren und Okkultatumviren weltweit in terrestrischen und marinen Umgebungen verbreitet sind. Der hier isolierte neue Phage wirft Licht auf neue und vielfältige Zweige des *Microviridae*-Baums, was darauf hindeutet, dass ein Großteil der ssDNA-Phagenvielfalt weiterhin im Dunkeln bleibt.

Abstract

Bacteriophages of the family *Microviridae* are one of the major clades of ssDNA phages. Their cultivated members are lytic and infect *Proteobacteria, Bacteroidetes,* and *Chlamydiae.* Prophages have been predicted in genomes from *Bacteroidales, Hyphomicrobiales,* and *Enterobacteriaceae* and cluster within the subfamilies "Alpavirinae", "Amoyvirinae" and *Gokushovirinae*. Previously, the bacteriophage "Ascunsovirus oldenburgi" ICBM5 was isolated from the North Sea and is distantly related to known *Microviridae*. The novel phage infects *Sulfitobacter dubius* SH24-1b and uses both a lytic and a carrier-state life strategy. An entire infection cycle was analysed in replicates using a combinational approach of one-step growth curves and phage targeted direct-geneFISH. Thereby, the percentage of infected cells as well as the number of viral genomes per cell could be quantified over the course of an infection. Furthermore, it was shown that phage ICBM5 replicates its genome at least as early as 50 min after the start of infection and causes bacterial lysis around 110 min post-infection. During lysis, the burst size was calculated to be approximately 250 phage particles, which roughly corresponds to the phage genome copy numbers calculated with phage targeted direct-geneFISH. Additionally, using ICBM5 proteins as a query, we retrieved 65 new *Microviridae* prophages and episomes in bacterial genomes from publicly available resources and uncovered 47 environmental viral genomes (EVGs) from various viromes. Genome clustering based on protein content and phylogenetic analysis showed that ICBM5, together with *Rhizobium* phages, new prophages, episomes, and EVGs cluster within two new phylogenetic clades, here tentatively assigned the rank of subfamily and named "Tainavirinae" and "Occultatumvirinae". They both infect members of the order *Rhodobacterales*. Occultatumviruses also infect *Hyphomicrobiales*, including nitrogen-fixing endosymbionts from cosmopolitan legumes. A biogeographical assessment showed that tainaviruses and occultatumviruses are spread worldwide, in terrestrial and marine environments. The new phage isolated here sheds light onto new and diverse branches of the *Microviridae* tree, suggesting that much of the ssDNA phage diversity remains in the dark.

Acknowledgements

This study was carried out from April 2018 until October 2022 at the University of Oldenburg at the Institute for Chemistry and Biology of the Marine Environment (ICBM). During this time, many people helped me to accomplish this work.

First of all, I am grateful to Prof. Meinhard Simon, who gave me the opportunity to join his group and for the support that allowed me to finish this project. Furthermore, I am very thankful to Dr. Cristina Moraru, from whom I have learnt a lot throughout this project and who often found a way to motivate and support me at this time. Additionally, I would like to thank Prof. Karin Holmfeldt as well as Dr. Bert Engelen for providing themselves as second and third reviewer in this PhD project. I kindly thank the entire staff of AG Simon and AG Cypionka for providing help whenever and wherever it was needed.

Thanks also to all my friends from the lab, you truly made my time in Oldenburg special. It was a pleasure to get to know you and I'm looking forward to seeing you again in the near future. A big thank you to Janina Leinberger and Vera Bischoff for proofreading and helping me improve this thesis.

I would also like to express my deepest gratitude to my family for their support and encouragement during my studies. Especially to my mother and my grandparents who were always there and believed in me.

Mulțumesc – Thank you – Vielen Dank

Abbreviations

AMG	Auxiliary metabolic genes
ASW	Artificial seawater
Bp(s)	Basepair(s)
BC	Baltimore class
CFU	Colony forming units
CTX	Cholera toxin
DMSP	Dimethylsulfoniopropionat
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
dsRNA	Double-stranded RNA
EM	Electronic microscopy
EVGs	Environmental viral genomes
FISH	Fluorescence *in-situ* hybridization
GOV	Global Ocean Virome
HGT	Horizontal gene transfer
HT	Horizontal transmission
ICTV	International Committee on the Taxonomy of Viruses
IDT	International DNA technologies
IMG/VR	Integrated Microbial Genomes / Viral Reference
Kbs	Kilo bases
LPS	Lipopolysaccharide
MB	Marine broth
MGA	Meta Gene Annotator
MOI	Multiplicity of infection
NR	Non-redundant
Nts	Nucleotides
OD	Optical density
ON	Overnight
ONT	Oxford nanopore technologies
ORF	Open reading frame

PBS	Phosphate buffered saline
PC	Protein cluster
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PFU	Plaque-forming units
PSC	Protein supercluster
PSSC	Protein super supercluster
RBS	Ribosomal binding site
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT	Room temperature
SAG	Single amplified genome
SBP	Sure border prophage
sRNA	Small RNA
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
TEM	Transmission electron microscopy
tRNA	Transfer RNA
UBP	Unsure border prophage
VGC	Viral genome cluster
VHG	Viral hallmark gene
vOTUs	viral Operational Taxonomic Units
VT	Vertical transmission

List of figures

List of tables

Table of contents

1. Introduction

1.1 History of bacteriophages

Bacteriophages, or phages, are viruses that infect bacteria. The first time they were described was by the English microbiologist Frederick William Twort. In his article from 1915, he reports the occurrence of translucent spots on white *Micrococcus* colonies and already hypothesized about an "ultra-microscopic virus" (Twort 1915). However, he could not continue his work and therefore, the serendipitous discovery of bacteriophages is nowadays attributed to the French-Canadian scientist Felix d'Herelle. In his first study on the subject, published in 1917, he introduced the name "bacteriophage" (eater of bacteria) and identified their potential to act as transmissible agents of bacterial cell lysis. Additionally, he assumed that phage plaques were formed by single phages and already used those spots as a measure for quantifying viral numbers (D'Hérelle 1917). One of the first ways to utilize phages was as a potential treatment strategy against infectious diseases caused by bacteria. Early experiments were promising and were designed as a cure for dysentery and an avian disease caused by *Salmonella gallinarum* (D'Hérelle 1926). As a consequence, large pharmaceutical companies in France, USA and England even started manufacturing commercial products within the 1930s, but the hopeful endeavour of phage therapy was temporarily stopped during the 1940s due to the rise of antibiotic therapy (reviewed in Letarov 2020). In the first two decades after the discovery of bacteriophages, their viral nature was intensively debated. The opponents of d'Herelle understood the principle of phage-induced lysis as an autocatalytic enzyme activation supposedly as part of a metabolic function and probably built into the genetic apparatus of prokaryotic organisms. During this venture of disproving the virus idea, Bordet and his colleagues were the first to describe the phenomenon of lysogeny (Bordet 1925) (reviewed in Letarov 2020). Finally, an end was found to the ongoing controversy through the introduction of electron microscopy (EM) into the field of virology. The first visualization of bacteriophages was performed by Ernst and Helmut Ruska (Ruska 1940) and for the first time morphological and structural characteristics of prokaryotic viruses could be observed by the human eye. Since then, the electron microscope has become one of the most valuable tools for virologists and the morphology of virions was used in a first classification approach.

1.2 Classification of bacteriophages

The presence of a protective protein coating is a feature shared by all bacteriophages and its structure was first visualized via electron microscopy. The majority of bacterial viruses have the structural element known as capsid in an icosahedral geometry with perfect symmetry (see Fig. 1). However, some exceptions are known and a small number of bacteriophages can have an additional lipid layer surrounding the protein hull or even employ helical or filamentous capsid formation. In all cases, the capsid harbours genomic information in the form of a single nucleic acid molecule. Some viruses also carry additional viral proteins inside the capsid (reviewed in Louten 2016). For example, members of the family *Cistoviridae* package several RNA-dependent RNA polymerases (Xiaoyu Sun et al. 2018). The nucleic acid can be either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), single- or double-stranded, linear or circular. Both virion morphology and nucleic acid type architecture have been used as criteria in early bacteriophage classification systems.

1.2.1 Traditional classification of bacteriophages

This classification was performed by David Bradley in 1967 and his work still contains the most studied phages to this day. He already described six different morphological appearances and determined their corresponding nucleic acid type (Bradley 1967). Three of them possessed a more complex morphology and had a tail-like structure attached to their capsid. Two of them employed small icosahedral capsids and the remaining one appeared in filamentous shape (see Fig. 1). Bradley's morphotypes have laid the foundation for the classification scheme used by the International Committee on Viral Taxonomy (ICTV). At the second ICTV meeting in 1975, all of his six morphological groups were already associated with distinctive phage families (eight families were mentioned in total) (Wildy 1971; Fenner 1976) (reviewed in Sharma et al. 2017). The group of tailed phages possessed linear dsDNA genomes and were differentiated according to their differences in tail structure. Myoviruses (from the family *Myoviridae*) had a rigid and contractile tail of 80 - 485 nm in length, while siphoviruses (from the family *Siphoviridae*) had a long flexible and non-contractile tail of 79 – 539 nm in length. The last in the group, the podoviruses (from the family *Podoviridae*) had a small and non-contractile tail with 3 – 40 nm in length (Ackermann 1998). All of the tailed phages were placed together in the order *Caudovirales* and members of this group belonged to the best-

studied bacteriophages in general (Ackermann 1998). The remaining morphotypes were categorized into three different families. The *Microviridae* are phages with small circular ssDNA genomes and icosahedral capsids of 30 nm in diameter. The *Inoviridae* are filamentous phages with small circular ssDNA genomes and helical capsids of 2 µm in length and 7 nm in width. The smallest morphotype was placed in the *Leviviridae* family, which were phages with ssRNA (+) genomes and icosahedral capsids of 26 nm in diameter (King 2012). However, the majority of phages lacked an ordered and comprehensive taxonomy and even though the genomic characteristics, as well as morphological attributes, are important, they are not sufficient when addressing the delicate challenge of viral classification. Another attempt on organizing viruses into a sorted and structured system was performed by David Baltimore in 1971. He based his approach on variations in nucleic acid type and the differences in viral replication systems (Baltimore 1971). He distinguished six Baltimore Classes (BC) in his original publication, while a seventh class was added a few years later (reviewed in Kuhn 2019). His work has been the gold standard in viral taxonomy for the last 50 years until recently the entire taxonomy was overhauled.

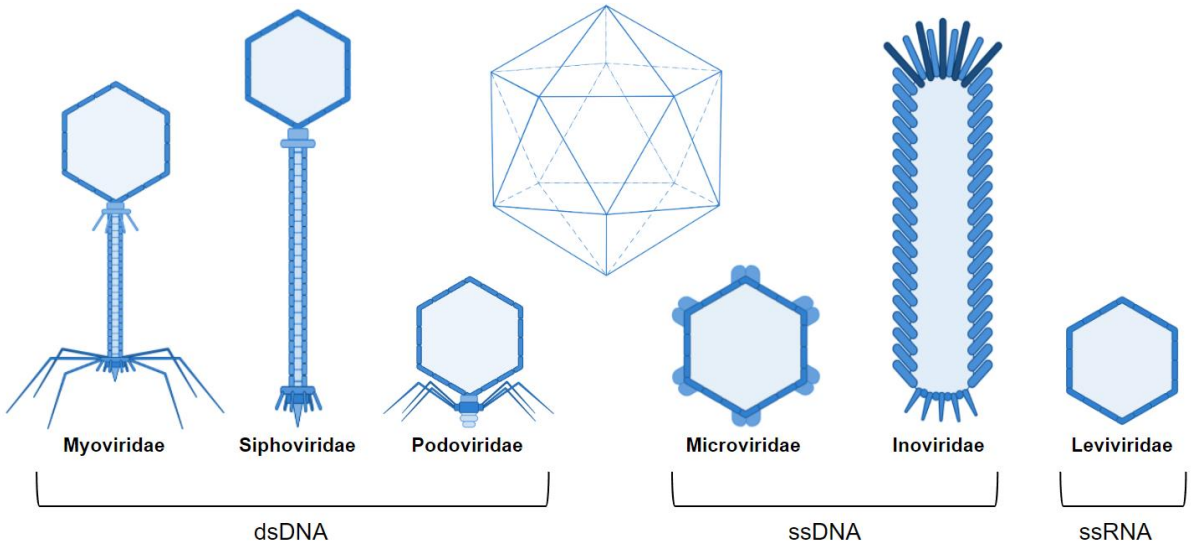

Fig. 1: Six different morphological types of bacteriophages, as presented in Bradley's early classification approach (Bradley 1967). The family description has been added indicating taxonomical association to the different appearances. The centre structure shows a 3D-shaped icosahedron which represents the majority of viral capsid geometry. The number of areas and edges can vary, also some icosahedral capsids can be elongated or stretched, but the basic shape always appears in perfect symmetry. This image was created using BioRender (https://biorender.com/).

1.2.2 Modern classification of bacteriophages

Due to the many advances in genome sequencing, the rise of computational power and the staggering amount of data which comes along with these developments, the methodology for classifying viruses has been tremendously updated (Bin Jang et al. 2019; Gorbalenya et al. 2020). In recent years, major changes were performed in viral classification through the introduction of a megataxonomy (Koonin et al. 2020). This new approach uses viral genome features like shared genes that are common within distinctive groups instead of applying one of the traditional procedures, for example, methods that were described in the previous section. Furthermore, the authors concluded that the traditional BCs do not sufficiently describe the evolution of viruses and therefore lack taxonomic relevance and require an overhaul (Koonin et al. 2020). One of the predicaments for viral taxonomy is the lack of a universally conserved gene. Cellular organisms share a common ancestor and consequently, a set of universal genes has been preserved and can be used for phylogenetic relationship analysis. Viruses do not share a common ancestor and have probably evolved through multiple independent occasions (Krupovic and Koonin 2017; Kazlauskas et al. 2019). However, to circumvent the deficiency of a shared marker gene and still be able to apply phylogenetic methods, comparative genome analysis can be performed using a bipartite network approach. In comparison to phylogenetic methods, which exclusively are based on multiple alignments of homologous genes, the principle behind these networks functions as follows: 1) A clustering of all viral proteins is performed based on homology, for example via BLASTp. 2) Then, viral genome networks are created considering their shared protein clusters. Afterwards, the protein clusters shared by groups of viruses can be identified, which marks the so-called core proteins. These can then be used to construct true phylogenetic trees based on multiple alignments, which allows for the reconstruction of viral evolution within that particular virus group. These viral network analyses have been used to reflect levels of relationship between phages and have led to the definition of viral hallmark genes (VHGs) (Iranzo et al. 2016). These VHGs are shared by many diverse groups of viruses and can be viewed as distinguishing characteristics of the “virus state” (Koonin et al. 2006). Furthermore, their occurrence in viral genomes has been used to define six viral realms: *Adnaviria, Duplodnaviria, Monodnaviria, Riboviria, Ribozyviria* and *Varidnaviria* (Koonin et al. 2020; Koonin et al. 2022). Viruses of bacteria are found in all of these six realms, however, most of them are organized in the class *Caudoviricetes* within the realm *Duplodnaviria*. For each realm,

the “virosphere” is split into 15 hierarchical ranks, ranging from kingdom to species which closely aligns with the Linnaean taxonomic system (Gorbalenya et al. 2020). To classify lower rank levels, for example, genus and species of dsDNA-tailed viruses, intergenomic nucleic acid identities are used and the ICTV recommends a threshold of 70% and 95% for genus and species level, respectively. For the remaining virus groups, the distinguishing characteristics are less clear at the moment of writing. The criteria to organize intermediary ranks like order or family are also yet to be determined but could be based on whole viral proteomes focusing on shared orthologous proteins (Turner et al. 2021). Finally, the highest hierarchical taxon should be classified on the presence of specific VHGs detected via protein clustering and annotation methods (Moraru 2021b). However, even though there are still many “white spots on the map of the virosphere” the overall organization of viral classification is becoming more transparent (Koonin et al. 2020). Apart from the above-mentioned principles to categorize bacteriophages, additional important distinguishing characteristics are the differences regarding viral life cycles.

1.3 Life cycle of bacteriophages

The interaction between phage and host is mediated through specific receptors like lipopolysaccharides, flagella, teichoic acid or proteins (Bertozzi Silva et al. 2016). These individual binding targets are located on the surface of the outer membrane of the host bacterium to which the phage attaches. The binding specificity between the phage and the bacterial receptor defines the host range and partially explains why certain phages can only infect certain bacteria (reviewed in Batinovic et al. 2019). However, other factors require consideration when analysing the first step of infection, the attachment. Growth conditions, virulence and the bacterial membrane type (Gram-positive and Gram-negative) can influence the phage adsorption procedure (Bertozzi Silva et al. 2016) (reviewed in Sharma et al. 2017). Once the phage has successfully attached to the corresponding host receptor, the second step of infection takes place and the phage injects its genomic material into the host organism (see Fig. 2). This procedure is quite phage-specific since tailed phages utilize their tail to pierce through the bacterial membrane(s), while non-tailed phages have to apply different strategies to transfer their nucleic acid into the host organism (Rakhuba et al. 2010). For example, members of the family *Microviridae* utilize their pilot protein to form a tube-like structure, which most certainly functions to deliver the phage DNA across the host’s periplasmic space

into the cytoplasm (Sun et al. 2014). After the genetic material has been successfully inserted, the progression of the infection cycle ranges from complete lysis of the bacterial host cell, and the release of new viral particles, to introducing new viral genetic material into the bacterial host genome. The way a bacteriophage interacts with its host is defined by its lifestyle. There are many different lifestyles. The best investigated are the lytic and lysogenic infection cycles (see Fig. 2). Nevertheless, some alternative strategies like the chronic cycle, pseudolysogeny or the carrier state life cycle are known. Most of our knowledge about viral reproductive strategies comes from the study of cultivated phage-host pairs. Considering that the vast majority of environmental phage-host pairs resist cultivation, it is likely that additional reproductive strategies are to be yet discovered. Recent findings even suggested that phage infection strategies could be considered a continuum instead of distinctive categories, ranging from efficient productive infections to persistent non-productive infections (see Fig. 3) (reviewed in Chevallereau et al. 2021).

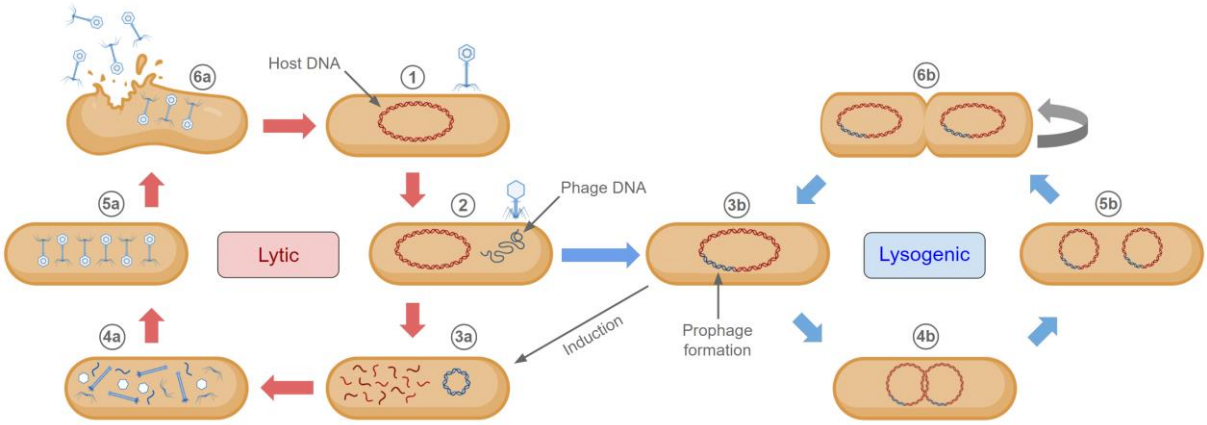

Fig. 2: Shows the differences between lytic and lysogenic life cycles. 1) Phage attaches to the outer membrane of the host. 2) Phage injects its genomic material into the cytoplasm by penetrating the bacterial cell wall. 3a) Phage DNA circularizes while host DNA is degraded 4a) New phage DNA and proteins are produced and 5a) assembled into mature virions. 6a) Bacterial cell lysis induced by the virus leads to the release of virions capable of infecting new hosts. 3b) Phage DNA integrates into the bacterial chromosome becoming a prophage. 4b) Lysogenic bacterium replicates its chromosome normally. 5b) Bacterial duplication proceeds with two identical chromosomes harbouring the prophage 6b) leading to two new lysogenic cells. Upon environmental trigger, the prophage excises from the chromosome and enters the lytic cycle. The figure has been adapted and adjusted from (Batinovic et al. 2019).

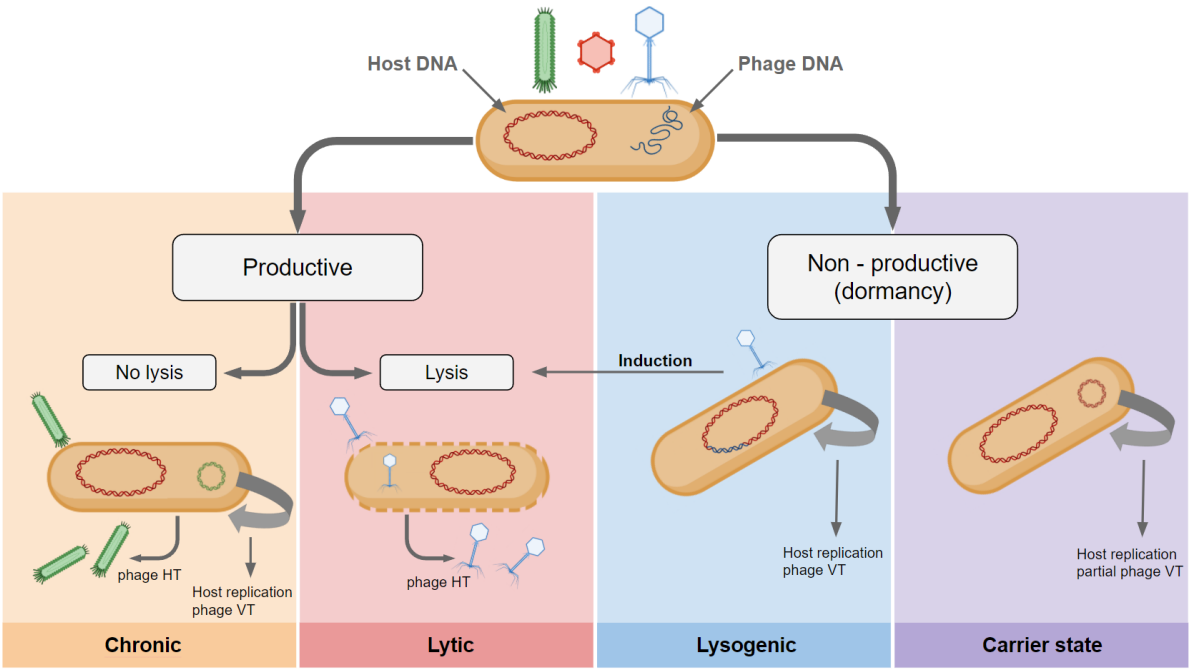

Fig. 3: Overview of different phage infection cycles. After entering the host cell via penetration, phages can enter either a productive replication cycle which leads to the release of new viral particles without host lysis (chronic) or by lysing the host cell (lytic). Non-productive strategies may be via integration of the viral genome into the host chromosome (lysogenic). Alternatively, the phage persists as an extrachromosomal plasmid and gets only partially replicated during cell division (carrier state). VT - vertical transmission, without lysis, HT - horizontal transmission, with lysis. The figure has been adapted and adjusted from (Chevallereau et al. 2021).

1.3.1 Lytic life cycle

During the lytic phase of infection, different stages can be distinguished (see Fig. 2) (1) attachment, 2) genome insertion, 3) genome replication, 4) protein synthesis, 5) capsid assembly and 6) lysis). In the early stages, once the phage is attached to the outer membrane and inserted its genomic material into the host cytoplasm, the host metabolism is reorganized towards the production of new phage particles (Weinbauer 2004). Once a certain number of possible phage genomes has been replicated, the next stage of the lytic cycle begins and the assembled genomes are packed into capsids. Finally, during the last step of the lytic cycle, the host cell lyses, thereby releasing new phages into the environment (see Fig. 2). Achieving phage-induced lysis can be mediated by different pathways, depending on the phage and the host. For larger dsDNA phages of Gram-negative bacteria, two pathways are known, both involving several different proteins, like holins, endolysins and spanins (Young 2013). In short, holins form micron-scale openings within the inner membrane, endolysins degrade the peptidoglycan within the periplasmic space via hydrolytic reaction and spanins disrupt the outer membrane, allowing the release of newly produced virions (Young 2014). In smaller

ssDNA and ssRNA phages, an alternative approach is known where single lysis proteins are used to inhibit the bacterial cell wall synthesis similar to antibiotics like penicillin (Bernhardt et al. 2002).

1.3.2 Lysogenic life cycle

When a bacteriophage enters the lysogenic state, it shifts into a dormant condition where its genome is either incorporated into the host chromosome or stays as an extrachromosomal plasmid within the cytoplasm (Casjens and Hendrix 2015; Ravin 2015). The molecular mechanisms of the lysogenic life strategy have been studied in detail for a few *Escherichia coli* phages, for example, Lambda and Mu. These temperate phages completely integrate into the bacterial chromosome via site-specific recombination or random transposition, respectively (reviewed in Howard-Varona et al. 2017). The stably integrated viral genome, known as prophage, is then replicated as part of the host chromosome. As a result, prophages are vertically transmitted to progeny cells, which means transmission without lysis (reviewed in Shivam et al. 2022). During the state of lysogeny, the expression of lytic phage genes is suppressed, for example by the transcription factor CI in phage Lambda (Bednarz et al. 2014). On certain occasions, the prophage can be induced and the repression of lytic genes is no longer suppressed. This leads to the excision of the phage genome and the re-entering of the lytic cycle. The transition between infection cycle strategies, sometimes also referred to as switch, is mediated via different mechanisms for example through phage regulatory genes or via external environmental signals (reviewed in Shivam et al. 2022). Another feature of this dormant life cycle is the phenomenon of lysogenic conversion. Due to the integration of a viral genome into the bacterial chromosome, the genetic inventory of the bacterial host gets extended and some of the foreign genes can lead to an alteration of the bacterial phenotype (Brüssow et al. 2004). This alteration can allow the host to colonize new environmental niches or even introduce attributes of virulence as it has been the case for the filamentous vibriophage CTX Phi, which encodes for the cholera toxin and integrates into the host chromosome of *Vibrio cholera* (Waldor and Mekalanos 1996). Additionally, lysogenic conversion may even protect the bacterial host from infection by other prokaryotic viruses (Touchon et al. 2017).

1.3.3 Chronic life cycle

The chronic infection cycle was discovered first in the early 1960s together with filamentous ssDNA phages from the family *Inoviridae* (Loeb 1960). During a chronic infection, the host organism is not lysed by the virus, but the phage uses its cellular machinery to replicate the viral genome and produce new viral particles (see Fig. 3). The newly assembled virions are then exported continuously out of the cell, usually by utilizing a complex extrusion apparatus (Feng et al. 1999). For a long time, only a few chronic phages have been described. But due to a recent study (Roux et al. 2019), more than 10.000 putative inovirus-like sequences were discovered and the number of known inoviruses expanded by two orders of magnitude. Furthermore, it was shown that members of this particular group display a high level of diversity and its classification is currently being updated. Most likely, the whole inovirus group will be upgraded to a higher taxonomic level. They are capable of infecting hosts across the domains Archaea and Bacteria and can be found in diverse ecosystems, ranging from aquatic to terrestrial and host-associated, indicating the abundant occurrence of a chronic infection style (Roux et al. 2019; reviewed in Chevallereau et al. 2021).

1.3.4 Carrier state life cycle or pseudolysogeny

Apart from the common canonical life cycles, some phages apply an alternative approach of persistence which has been described as a ‘carrier state’ or ‘pseudolysogeny’. One of the first mentioning’s of the carrier state goes back to the year 1953 when André Michel Lwoff describes the phenomena of “carrier strains” (Lwoff 1953). He described them as a mixture of bacteriophages and bacteria which are “in a more or less stable equilibrium”. Furthermore, he observed that the majority of bacteria in this mixture are resistant and only a small fraction is sensitive to extrinsic phage infection therefore both bacterial, as well as phage populations, are sustained. Once the genome has been injected into the host organism, the virus does not enter any of the above-mentioned life cycle strategies but remains as a non-replicative extrachromosomal unit, which is asymmetrically transferred to one of the daughter cells during bacterial division (see Fig. 3). It has been assumed, that the phenomenon of a carrier state lifestyle occurs when host populations encounter nutrient-limited conditions. In doing so, a coexistence of bacterium and virus would be ensured for an elongated period or until more favourable conditions are met. In this case, the phage switches back to a replicative form and enters either the lytic or the lysogenic cycle (reviewed in Chevallereau et al. 2021).

A recent study (Mäntynen et al. 2021) goes even further in distinguishing between the different alternative, non-productive states. They proceeded with Lwoff's definition of the carrier state and described that the presence of strains prone to viral infections within resistant bacterial populations ensures the perpetual distribution of new phage particles. Therefore they defined the term carrier state life cycle as a population-level-based phenomenon.

1.4 Bacteriophages and their environmental impact

Prokaryotes are responsible for producing approximately half of the oxygen on this planet and are additionally involved in global biogeochemical cycles by producing key components that influence major processes like carbon and nitrogen cycles (Whitman et al. 1998; Madsen 2005; Falkowski et al. 2008). Bacteriophages have long been neglected in terms of their impact on environmental systems, but due to the methodological advancements during the last few decades, our understanding of their roles has dramatically changed. It is now generally accepted that phages play key roles in the biology and ecology of microbes, which themselves impact environments on a large scale (Fuhrman 1999; Clokie et al. 2011). Bacteriophages are ubiquitous and occur in every ecological niche in aquatic and terrestrial environments. Additionally, they can be associated with all types of life harbouring bacterial communities, including animals, plants, protists and fungi. As the world's oceans cover up to 70 % of the Earth's surface, they are the largest ecosystem on the planet. In marine systems, prokaryotic communities constitute up to 90 % of the oceanic biomass and are dominant in terms of abundance, diversity and metabolic function (Suttle 2007). Naturally, the viral portion of these groups is equally or even more variable and abundant and therefore of tremendous significance to biogeochemical processes (Batinovic et al. 2019). The sheer abundance of viral particles in marine ecosystems was revealed by quantitative analysis of seawater samples within the late 1980s (Bergh et al. 1989; Børsheim et al. 1990; Proctor and Fuhrman 1990). Since then, it is well known that bacteriophages are the most abundant biological entities on this planet and outnumber their prokaryotic hosts by an order of magnitude (Hendrix et al. 1999; Mushegian 2020). In marine ecosystems, viruses can reach abundances of up to 100 million viral particles per millilitre of seawater (Wommack and Colwell 2000). To have a better estimation of this number, assuming all marine viruses would be lined up one after the other,

this string of viruses would span a distance (5.4 x 10^{22} m) equivalent to 12 billion roundtrips from earth to mars (Breitbart et al. 2018).

1.4.1 Influence of phages on bacterial mortality and environmental processes

Marine viruses infect about 10^{23} cells per second in the oceans and are responsible for the release of 10 billion tons of organic carbon globally per day (Suttle 2007; Breitbart 2012). The majority of marine viruses are phages and they are responsible for up to 50 % of bacterial mortality via viral lysis and therefore capable of shaping community compositions on a large scale (Noble and Fuhrman 2000). The remaining 50% of microbial mortality is attributed to grazing by protozoa. These unicellular eukaryotes are then consumed by zooplankton, fish and finally larger organisms, and therefore involved in moving carbon to higher trophic levels (Breitbart et al. 2018). In contrast, by inducing bacterial lysis, phages are responsible for releasing a great amount of cellular content into the environment. These components are dissolved organic matter (DOM) and particulate organic matter (POM) which are both released into the environment through cell lysis. The released resources, especially the DOM, can be used by heterotrophic microorganisms for metabolic processes and are therefore kept away from being transferred into higher trophic levels. This environmental phenomenon of nutrient cycling is termed the “viral shunt” (see Fig. 4) (Wilhelm and Suttle 1999; Gasol and Kirchman 2018). However, there is recent evidence that bacterial lysis induced by phages can also lead to sticky aggregates that sink to deeper ocean layers. Thereby, organic carbon would also be efficiently shuttled from the surface to the deep ocean where it then is further separated in a process known as the “biological pump”. These assumptions require further evaluation of marine viruses and their role in marine food webs to eventually identify the conditions necessary in which viruses participate more as a shunt or a shuttle (Breitbart et al. 2018).

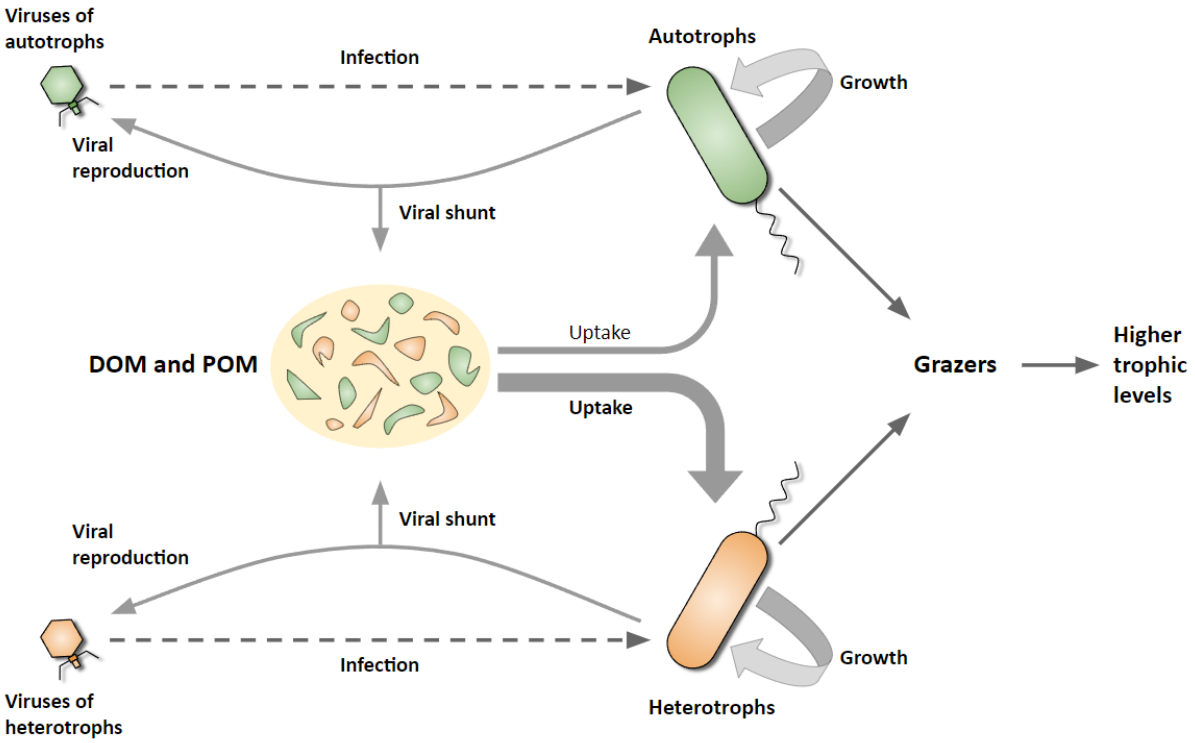

Fig. 4: Schematic overview of the viral shunt. Viral infections can cause bacterial lysis and therefore the release of DOM and POM. This organic matter is mainly utilized by heterotrophic bacteria (thick grey arrow linking POM/DOM with heterotrophs), and to a lesser extent by autotrophs. Consequently, some of the cellular material is not available for higher trophic levels. The figure has been adapted from (Weitz and Wilhelm 2012).

1.4.2 Host genomic enhancements mediated by bacteriophages

Apart from their impact on bacterial mortality, phages can alternate evolutionary trajectories of microbial communities by transferring an estimated amount of 10^{29} genes per day worldwide (Gregory et al. 2019). As already described in the lysogenic life cycle chapter, lysogenic conversion is one way the process of horizontal gene transfer (HGT) between prokaryotic organisms takes place. Another important mechanism is transduction, where a phage mistakenly packages some host DNA in its capsid and carries it over to the next host (Paul 1999). Besides transporting genes between hosts via the aforementioned principles, many bacteriophages also carry genes which are involved in host metabolic functions and can enhance the host metabolism. These genes have been termed auxiliary metabolic genes (AMGs) and are engaged in cellular metabolism, as defined by the Kyoto Encyclopaedia of Genes and Genomes (KEGG) metabolic pathway database. (Hurwitz et al. 2015). AMGs were discovered first in cyanophages (phages that infect *Cyanobacteria*). It was found that proteins translated from phage-encoded genes were part of the photosynthetic apparatus and could support the maintenance of its activity during infection (Sullivan et al. 2005; Lindell et al.

2005). Therefore, these genes confer a fitness advantage to cyanophages (Bragg and Chisholm 2008). From this time on, several hundred genes have been detected in different phages and many of the new AMGs are involved in a variety of different biogeochemical cycles and not only in photosynthesis (Roux et al. 2016b; Huang et al. 2021; Jian et al. 2021). For example, a recent review (Wang et al. 2022) gave an overview of the viral contribution to marine nitrogen cycling by expressing AMGs and thereby regulating host nitrogen pathways like nitrification, denitrification, anammox or nitrogen membrane transport.

RNA-based features like transfer RNAs (tRNAs) or small RNAs (sRNAs) can be encoded on the phage genome and aid in the interaction between the virus and the host. Transfer RNAs are involved in protein synthesis by transporting amino acids and are additionally responsible for recognizing specific codons on the mRNA during translation. In most cases, phages infect hosts with a similar codon usage and tRNA profile, because the phage protein translation machinery would be less efficient otherwise. Nonetheless, it has been observed that some phages harbour their own tRNAs, so they can circumvent a deficiency of a particular amino acid transporter and ensure successful protein translation during infection (Bailly-Bechet et al. 2007). This strategy has been observed to broaden the host range as it has been the case for a variety of different cyanophages (Enav et al. 2012). The second mentioned example, the sRNAs, play an important part in the regulation of gene expression at the post-transcriptional level and are key components of elemental cellular processes (Gottesman and Storz 2011). Furthermore, sRNAs have also been detected in bacteriophages, where they are involved in regulating genes important for the early infection stages, but also in regulating late structural genes (Blasdel et al. 2017). Additionally, sRNAs also regulate genes responsible for maintaining lysogeny and are capable of mediating the transition between phage life cycles (Bloch et al. 2021). Some prophage-encoded sRNAs have been observed to control host metabolism as well as cell division (Balasubramanian et al. 2016). However, investigating these molecular components is a delicate task and more research needs to be done to uncover the true inventory of phage encoded RNA-based regulatory elements. A set of methods on how to reveal unknown features in general but also on how to improve our understanding of phage diversity is presented in the next sections.

1.4.3 How to investigate bacteriophages from environmental samples

The known cultivated bacteriophages reflect only a small portion of bacterial viruses in the environment, their diversity is just being explored and it can be assumed that a large fraction of phages remains to be uncovered. To investigate viruses from the environment, the most common methods can be divided into two different categories. First, culture-dependent techniques, which use phage isolation methods in pure cultures and second, culture-independent techniques, which are based on tools from microbial ecology and work directly on environmental samples (Townsend et al. 2021). Both approaches bring important insights to our understanding but have their individual question-solving approaches in terms of elucidating the unknown. Culture-independent techniques enable insights into larger ecological processes for example by quantifying virus numbers from different ecosystems via flow cytometry or epifluorescence microscopy (Brussaard 2004; Wen et al. 2004). Additionally, culture-independent techniques are capable of producing an enormous amount of metagenomic information through massive sequencing of environmental samples (also known as metagenomics). In comparison, the culture-dependent techniques allow for a more detailed insight into specific phage-host interactions. For example, one-step growth curve experiments can be conducted to determine the different stages during infection as well as quantify the number of released virions (also known as burst size) (Ellis and Delbrück 1939; Middelboe et al. 2010). Furthermore, once a phage has been cultivated, the genomic information can be obtained by using short and long-read sequencing techniques. When using the latter, the entire genome can be sequenced at once while additionally conserving information like methylation patterns and thereby providing insights into epigenetic processes (Rand et al. 2017; Simpson et al. 2017).

Several steps are involved in the successful isolation of an environmental virus and one way of doing that can be via phage enrichment. There, phage particles must be separated from bacterial cells within the environmental sample. Subsequently, the obtained phages are usually combined with a liquid host culture during early exponential growth and the release of new host-specific phages is monitored by checking for cell lysis events. This is done either by plaque assays or by turbidity measurements. In case a lysis event occurs, the newly released phage particles are further purified by successive plaque assays or by dilution-to-extinction assays, depending on the growth characteristics of the host strain (Buchholz et al. 2021). The ability to work on cultivated phages is highly advantageous, especially when a more detailed

analysis of the phage-host interaction is the main objective. However, successfully isolating a virus from environmental systems is an intricate procedure and many obstacles can be a challenge. For example, the inability of cultivating most marine bacteria leads to an inability of cultivating their respective phages. Some techniques can circumvent this problem and preserve the link between the phage and its host without necessarily cultivating the host strain.

One technique to preserve the link between phage and host is fluorescence *in situ* hybridization (FISH) for example phage targeted direct-geneFISH, as it can be directly applied to environmental samples. Using phage targeted direct-geneFISH for this task is done by hybridizing intracellular phages with fluorescently labelled polynucleotides targeting specific regions on the phage genome (Allers et al. 2013; Barrero-Canosa et al. 2017). At the same time, bacterial cells are also hybridized using different fluorescently labelled polynucleotides aiming for the 16S ribosomal RNA (rRNA) gene on the bacterial chromosome. This approach allows for precise taxonomical identification of the host organism. Finally, both hybridized fractions are further analysed via microscopy and if the phage signal overlaps with the 16S signal, this particular cell is considered to be infected by the phage of interest (Barrero-Canosa and Moraru 2021). An alternative approach to establishing the connection between phage and host identity from environmental samples is by sequencing genomic information directly from individual cells. Genomes generated by this method are termed single amplified genomes (SAGs), as the DNA is extracted only from one single cell and then amplified via an enzymatic reaction to produce the required amount necessary for sequencing (Martinez-Hernandez et al. 2017). Doing it this way allows for a complete recovery of the entire genomic information that was present inside the cell at the moment of DNA extraction. Therefore, if a cell was infected by a bacteriophage, it will be sequenced together with the host genome and allow for an insight into the connection of phage and host (Martinez-Hernandez et al. 2017). Additionally, direct sequencing of the infected cell enables a proximate estimation of the stage of infection (reviewed in Bischoff et al. 2021). Extensive sequencing of environmental samples is of enormous advantage for gaining knowledge of the ecological virus diversity and is most commonly achieved via metagenomics.

A metagenome provides a snapshot of all collected DNA molecules inside a selected sample and allows for a deeper understanding of genomic information while bypassing the necessity for isolation and laboratory cultivation. The term viral metagenome, also called

virome, is used when the fraction of phage particles are separated from prokaryotes and other microorganisms before DNA extraction is performed. This procedure ensures that only the viral DNA is sequenced. Sequencing itself is usually performed using high-throughput sequencing technologies, for example, Illumina. This procedure generates small fragments of viral DNA (also termed reads) which, after sequencing, can be assembled into complete or partial phage genomes. However, due to the enormous amount of sequence diversity, only a few phage genomes can be assembled completely (reviewed in Bischoff et al. 2021). Various additional efforts have been made to gain additional knowledge about virus diversity in environmental systems. In terms of metagenomic analyses, the largest datasets were generated from three major scientific expeditions that were conducted over several years throughout the world's oceans. Those expeditions are known as Malaspina, Tara Oceans and Tara Oceans Polar Circle (Karsenti et al. 2011; Duarte 2015; Roux et al. 2016b; Gregory et al. 2019). As a result, two enormous datasets have been created named Global Ocean Virome (GOV) and GOV2 containing dsDNA viruses. The first dataset contained 1,380,834 viral contigs in total which were further grouped, first into 15,222 viral populations and then into 867 viral clusters. By analysing the 867 viral clusters and by comparing them to known viruses, 60 % did not affiliate with any of the cultivated phages. Additionally, when the paper was published, only three of the 867 viral clusters have shown an affiliation with known marine phage groups (Roux et al. 2016a). The GOV2 comprises all 145 samples from all three expeditions. With 195,728 viral populations, this dataset is more than ten times bigger than the one analysed in GOV1. The contigs from GOV2 were at least 10 kb in size and about 90% of them could not be assigned to any known viral family (Gregory et al. 2019). It can be assumed that a big fraction of the unassigned viruses are indeed bacteriophages. In addition to the mentioned endeavours to extend our knowledge of the earth's virosphere, the Integrated Microbial Genomes / Viral Reference (IMG/VR) database provides access to the largest collection of viral sequences obtained from metagenomes (Paez-Espino et al. 2016; Paez-Espino et al. 2017; Roux et al. 2021). In comparison to the above-mentioned datasets GOV and GOV2, the IMG/VR database also contains ssDNA and RNA viruses. Within the latest version (IMG/VR v4), the database contains more than 5 million high-confidence viruses and almost 3 million high-confidence viral Operational Taxonomic Units (vOTUs) (https://img.jgi.doe.gov/cgi-bin/vr/main.cgi). Especially the ssDNA viruses are of particular interest to this work, as it has

been found that this group of phages seem to be more abundant than previously anticipated (Labonté and Suttle 2013b).

1.5 ssDNA phages and the family *Microviridae*

The majority of cultivated bacteriophages have a tail-like structure and were described in the order *Caudovirales* (or today in the class *Caudoviricetes)* (Ackermann 1998; Koonin et al. 2020). For a long time, it was assumed that viruses from this group are environmentally dominant, but recent quantitative electron microscopy analyses and metagenomic studies demonstrated that non-tailed phages make up between 51-92% of global surface oceans (Brum et al. 2013; Ignacio-Espinoza and Fuhrman 2018). A fraction of these non-tailed phages is comprised of ssDNA phages. Furthermore, most of the viromic studies were oriented towards dsDNA phages and consequently, only little is known about the diversity of ssDNA phages in different ecological systems. Depending on the sample preparation and viral particle purification, the differences in filter pore size can select against the smallest or largest phage groups (Székely and Breitbart 2016). Additionally, the nucleic acid of ssDNA phages cannot be sequenced by conventional techniques. Prior to sequencing, the genome of ssDNA phages has to be converted to dsDNA and achieving this can be done by multiple methods. One approach of how this can be achieved is by using an adaptase-linker amplification technique. Thereby, it was shown that ssDNA phages represent not more than 5% of aquatic DNA phage communities. Within these communities, phages of the family *Microviridae* are the most occurring ssDNA viruses within a sample and this particular phage group is of special interest to this work (Roux et al. 2016c).

1.5.1 History of *Microviridae*

The first member of the *Microviridae* was discovered early in the history of bacteriophages and was isolated from a Parisian sewage during the 1920s. Nevertheless, it took a few more decades until the ssDNA nature of this phage was truly understood (Sertic and Boulgakov 1935; Sinsheimer 1959). The small virus infecting *Escherichia coli* was termed ΦX174 and represents the best-studied virus of the family *Microviridae* which was recently placed in the realm *Monodnaviria*. Due to its small genome and its well-described host, ΦX174 belongs to the best-studied bacteriophages in general. For example, ΦX174 has been

thoroughly used to investigate the fundamental mechanisms of DNA replication and capsid assembly (reviewed in Cherwa and Fane 2011). Furthermore, ΦX174 was the subject of the first genome sequencing project performed by Frederick Sanger in the 1970s and a total genome size of 5,386 kb was identified (Sanger et al. 1977a; Sanger et al. 1977b). Since its discovery around 100 years ago, several additional variants have been described and members of this group are very common and diverse. They are found in most ecosystems and are known to infect very different bacterial hosts (Doore and Fane 2016).

1.5.2 Classification and diversity of *Microviridae*

Microviridae in general have a small icosahedral capsid with an approximate size of 30 nm in diameter (see Fig. 1). Their genomes range between 4.4 - 6.3 kb in size and are of circular ssDNA nature (Doore and Fane 2016). Due to the recent taxonomic overhaul, the classification of *Microviridae* has been comprehensively updated (Koonin et al. 2020). Members of this phage family are now placed within the realm of *Monodnaviria*, in the kingdom *Sangervirae*, the phylum *Phixviricota*, the class *Malgrandaviricetes* and the order *Petivirales*. Furthermore, due to inconsistencies in the classification of different ssDNA virus groups, it is currently a topic of discussion to elevate presently recognized families of ssDNA viruses (like the *Microviridae*) to higher taxonomic levels for example order or even class (Koonin et al. 2020). However, at the moment of writing the ICTV has listed two accepted subfamilies, the *Bullavirinae* and the *Gokushovirinae.* Both have been classified according to their genome composition and differences in capsid structure. The *Bullavirinae* (former *Microvirinae*) infect bacteria from the order *Enterobacterales*, while gokushoviruses infect obligate parasitic bacteria, such as *Chlamydia, Spiroplasma* and *Bdellovibrio* (Chipman et al. 1998; Brentlinger et al. 2002; Everson et al. 2002). An additional distinguishing feature between the two subfamilies is the mushroom-like protrusions on the major capsid protein (MCP) which are formed by *Gokushovirinae* (Chipman et al. 1998). Apart from the already recognized subfamilies, several new potential subfamilies were identified in different studies, for example: “Alpavirinae”, “Aravirinae”, “Pichovirinae” and “Stokavirinae” (Krupovic and Forterre 2011; Roux et al. 2012; Quaiser et al. 2015). In addition, new ssDNA phages have been recently cultivated and they infect the marine bacteria *Citromicrobium bathyomarinum* RCC1878, a *Sphingomonadaceae*, and *Ruegeria pomeroyi* DSS-3, a member of the family *Rhodobacteraceae* (Zheng et al. 2018; Zhan and Chen 2019). The phage infecting the

Citromicrobium was suggested to belong to a new subfamily – the "Amoyvirinae", while the two novel *Ruegeria* phages vB_RpoMi-Mini and vB_RpoMi-V15 are considered as unclassified *Microviridae* (Zhan and Chen 2019). Besides clearly identified *Microviridae* phages, other ssDNA viruses with similar characteristics have been isolated and described in the past. For example, (Holmfeldt et al. 2012) mentions the cultivation of eight icosahedral ssDNA phages, isolated from heterotrophic marine *Bacteroidetes* as well as from bacterioplankton. Five of them were placed in a size category similar to *Microviridae* viruses and all of them infect the *Cellulophaga* genus. Furthermore, (Bartlau et al. 2021) proposed a new family of ssDNA phages which was tentatively named "Obscuriviridae". This new family harbours four of the already mentioned *Cellulophaga* phages isolated by (Holmfeldt et al. 2012) and one additional phage isolated within the (Bartlau et al. 2021) study.

1.5.3 Lifestyle of *Microviridae*

Members of the family *Microviridae* employ lytic infection cycles, thereby producing and releasing new viral particles into the environment. Most of the accumulated knowledge concerning this lifestyle strategy has been derived from ΦX174, where the first step of infection is mediated through the recognition of LPS from the outer cell wall (Ilag et al. 1994; reviewed in Cherwa and Fane 2011). Once the phage got attached to the cellular surface of its host, it injects its genomic material across the periplasmic space into the cytoplasm. This is achieved by utilizing the phage pilot protein H (see Fig. 5 for a complete genome overview of phage ΦX174) to form an oligomeric tube-like structure (Sun et al. 2014). After the phage has entered the cytoplasmic space, genome replication starts by converting the ssDNA into a dsDNA molecule, the replicative form (also known as RF I). Then, protein A initiates the replication by rolling circle and thereby producing new ssDNA template molecules (see Fig. 6). Simultaneously, protein synthesis generates components necessary for capsid assembly and DNA packaging. Each newly replicated ssDNA molecule is stored inside the procapsid and the last step of infection starts (reviewed in Cherwa and Fane 2011). In comparison to larger dsDNA phages, *Microviridae* do not encode for a two-component lysis mechanism which is usually composed of an endolysin and a holin. Alternatively, they translate the small protein E which acts by inhibiting the peptidoglycan biosynthesis of the host organism (Bernhardt et al. 2000). This function does not lyse the cell immediately, but as bacterial cell division

continues, the cell wall becomes very sensitive to osmotic pressure due to its impaired structural stability and finally bursts (reviewed in Cherwa and Fane 2011).

Fig. 5: Shows the genomic map of the ssDNA phage ΦX174 including all of its known protein coding sequences. Gene A) encodes for the replication initiation protein, gene A* is of unknown function, gene B) encodes for the internal scaffolding protein, gene C) is required for ssDNA synthesis and inhibits dsDNA synthesis during replication, gene D) encodes for the external scaffolding protein and gene E) encodes for the lysis protein. Gene F) encodes for the major capsid protein, gene G) encodes for the major spike protein, gene H encodes for the pilot protein while gene K is of unknown function. Figure has been adapted and adjusted from (Cherwa and Fane 2011). The image was created with SnapGene (http://snapgene.com)

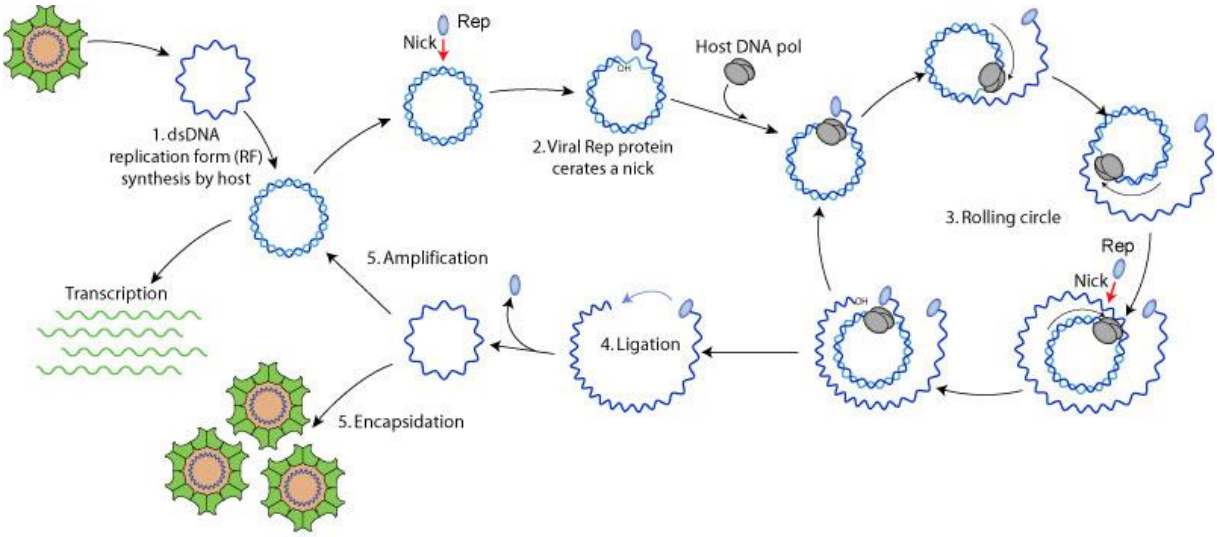

Fig. 6: Shows the different steps of replication for *Microviridae* phages. After the DNA has entered the cytoplasm, 1) the replicative form is generated by host enzymes. 2) Then, the viral replication protein creates a nick and 3) replication via rolling circle takes place. Thereby, a new ssDNA molecule is created which is 4) ligated by the viral replication protein. This new nucleic acid molecule can then either serve as 5) a template to create new replicative forms or 5) to get packaged into assembled capsids. The image was taken from ViralZone, SIB Swiss Institute of Bioinformatics (https://viralzone.expasy.org/1941).

Nevertheless, for a very long time, it was well established that *Microviridae* employ only a lytic infection strategy and are incapable of lysogeny (Fane et al. 2006). This has been assumed until prophages were predicted *in silico* in the genomes of *Bacteroidetes* (Krupovic and Forterre 2011). Additional studies predicted more *Microviridae*-like prophages in other *Bacteroidetes* (Roux et al. 2012; Quaiser et al. 2015), in a *Caenibius tardaugens* strain, an alphaproteobacterium (Zheng et al. 2018) and in *Enterobacteriaceae* (Kirchberger and Ochman 2020). In the last study, it was detected that gokushovirus prophages are capable of forming viable virus particles (Kirchberger and Ochman 2020). *Microviridae* lack of an integrase gene. Instead, they integrate their genome into the host chromosome by using the host chromosome dimer resolution system (Krupovic and Forterre 2011; Kirchberger and Ochman 2020).

1.5.4 *Microviridae* and their distribution in different environments

Within recent years, hundreds of complete *Microviridae* genomes have been sequenced and assembled from diverse environments. Microviral genomes have been detected in marine environments (Székely and Breitbart 2016; Labonté and Suttle 2013a), in freshwater habitats (Roux et al. 2012; López-Bueno et al. 2009), in human gut and faeces (Roux

et al. 2012), in stromatolites (Desnues et al. 2008), in dragonflies (Rosario et al. 2012), in sewage as well as in sediments (Hopkins et al. 2014; Quaiser et al. 2015). These newly discovered environmental microviruses help to improve our understanding of the diversity and distribution of this particular phage family. However, our comprehension regarding *Microviridae* diversity in different environmental systems is still relatively limited; thus, more work will be required to elucidate the unknown.

1.6 Members of the Roseobacter group and their respective phages

The *Roseobacter* group – which are heterotrophic *Alphaproteobacteria* representing marine *Rhodobacteraceae* – are key players in biogeochemical cycles and comprise up to 30% of bacterial communities in pelagic environments (Simon et al. 2017). Members of this bacterial group have been found in a large variety of habitats, ranging from coastal regions to deep-sea sediments and from polar ice to tropical regions. Furthermore, they are capable of utilizing a plethora of organic compounds and can apply a multitude of different metabolic processes, for example, sulphur oxidation, aerobic anoxygenic photosynthesis, and oxidation of carbon monoxide or the methylation of Dimethylsulfoniopropionat (DMSP) (Brinkhoff et al. 2008). Additionally, bacteria of this clade are in commensal relationships with marine phytoplankton, invertebrates and vertebrates. They have also been found to be free-living or particle-associated (reviewed in Buchan et al. 2005).

Bacteriophages infecting the *Roseobacter* group are known as roseophages and in comparison to their well-investigated hosts, only little is known about them. The majority of known roseophages belong to the realm of *Duplodnaviria* and have a podoviral or siphoviral morphology (reviewed in Bischoff et al. 2021). The first described phage infecting a marine *Rhodobacteraceae* was SIO1 which also was the first sequenced marine bacteriophage (Rohwer et al. 2000). Twelve years later, highly similar phages (>96% nucleotide identity) have been isolated from the same geographic region, demonstrating that marine viruses can persist as separate populations over long periods (Angly et al. 2009) (reviewed in Bischoff et al. 2019). In recent years, more roseophages have been identified and isolated. For example, the isolation of two strictly lytic bacteriophages, vB_LenP_ICBM1 and vB_LenP_ICBM2, lead to the delineation of a new *Caudoviricetes* family – the Zobellviridae, in which several roseophages grouped into the *Cobavirinae* subfamily. By phylogenetic analysis, it was shown that this

particular phage group comprises 15 different phages and they infect members of the genera *Lentibacter, Sulfitobacter and Celeribacter* (Bischoff et al. 2019). In addition, a recent study (Forcone et al. 2021) identified 173 putative prophages in 79 genomes of *Rhodobacteraceae*. These findings indicate that prophages are an important factor in terms of genomic diversity within the *Roseobacter* group and contribute to the ecology of this widely distributed bacterial group. However, all of the above-mentioned roseophages belong to the realm Duplodnaviria and so far only two roseophages have been isolated which can be placed in the realm *Monodnaviria.* These two phages infect the *Ruegeria* genus and with only 4.2 kb and four predicted open reading frames (ORFs), vB_RpoMi-Mini and vB_RpoMi-V15 show the smallest genome size within known ssDNA phages (Zhan and Chen 2019).

1.7 Preliminary results and foundation for this work

The preliminary findings described in this section are not part of this work but are necessary for the entire story, as they describe the discovery and the description of phage ICBM5. A graphical summary of the preliminary findings is shown in figure 8. The isolation of Phage ICBM5 was performed by Dr. Cristina Moraru, Vera Bischoff and other students that have worked with Dr. Moraru during that time. The marine virus was isolated from surface seawater collected from the coastal region of the North Sea (53°42'09.8"N 7°41'58.9"E) in June 2015. It was found that the virus infects a marine *Rhodobacteraceae, Sulfitobacter dubius* SH24-1b, which was also previously isolated from the North Sea during an algae bloom (Hahnke et al. 2013). Additionally, phage ICBM5 formed turbid clearing spots during plaque assays, which usually indicates the presence of both a lysogenic lifestyle of the phage and a resistant bacterial strain. This indication for an alternative lifestyle has been further investigated by collecting surviving bacterial cells from the turbid spots and plating them to obtain single colonies. These colonies were screened by polymerase chain reaction (PCR) for the presence of phage ICBM5 in their genome. The bacterial genotypes harbouring ICBM5 were challenged with the phage in a spot assay which did not lead to the formation of clearing zones. This observation indicates that the purified genotypes were indeed resistant to phage ICBM5. To investigate the putative integration into the host chromosome, PacBio long-read technology was used with a 7 kb size threshold for library preparation. However, ICBM5-specific sequences were not found by this approach and therefore no evidence for a lysogenic lifestyle could be obtained. A host range assay has shown, that apart from its original host

phage ICBM5 also infects *Sulfitobacter dubius* DSM 16472. Furthermore, the morphological structure of ICBM5 was investigated via transmission electron microscopy (TEM) on samples stained with uranyl acetate. This has revealed an icosahedral capsid structure without tail formation (see Fig. 7A). The diameter of the capsid structure was measured to be 28.68 +/- 1.95 nm (100 phages were measured, three measurements for each phage). Genome sequencing of phage ICBM5 was done via Illumina sequencing technology and followed by an *in silico* analysis. The reads were assembled in Tadpole and open reading frames (ORFs) were predicted with MetaGeneAnnotator (MGA) (Noguchi et al. 2008). After assembly, direct terminal repeats were detected at the end of the contig, indicating that the contig can be circularized and that the genome is complete. The entire phage genome has a size of 5,581 nucleotides (nts) in length and six genes, of which two are encoding for hypothetical proteins (see Fig. 7B). The other four genes were identified using Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) against a non-redundant database (http://ncbi.nlm.nih.gov/). This way, the pilot protein, the major capsid protein (MCP), the lysis protein as well as the replication initiation protein (REP) were identified (Zucker et al. 2022).

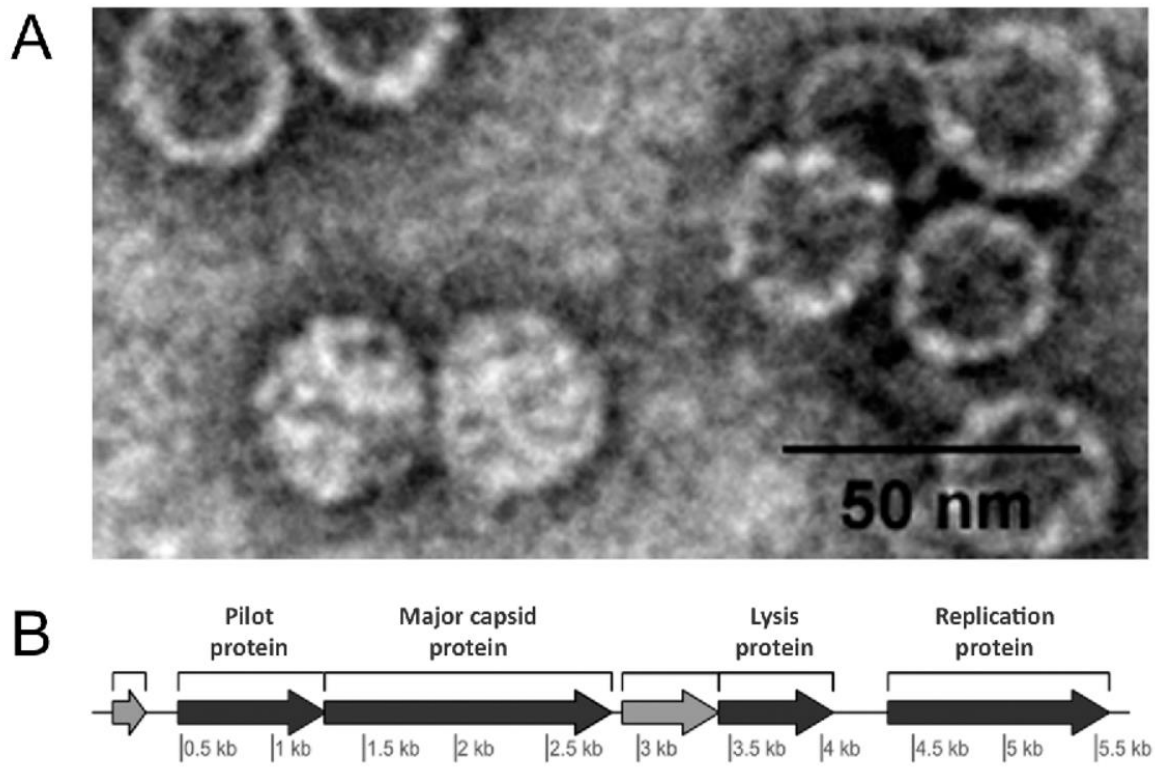

Fig. 7: A) Morphology of phage ICBM5 determined by TEM of uranyl acetate stained virions. B) Genome map of ICBM5. In dark grey - identified proteins, with labels on top of each gene. In light grey - hypothetical proteins.

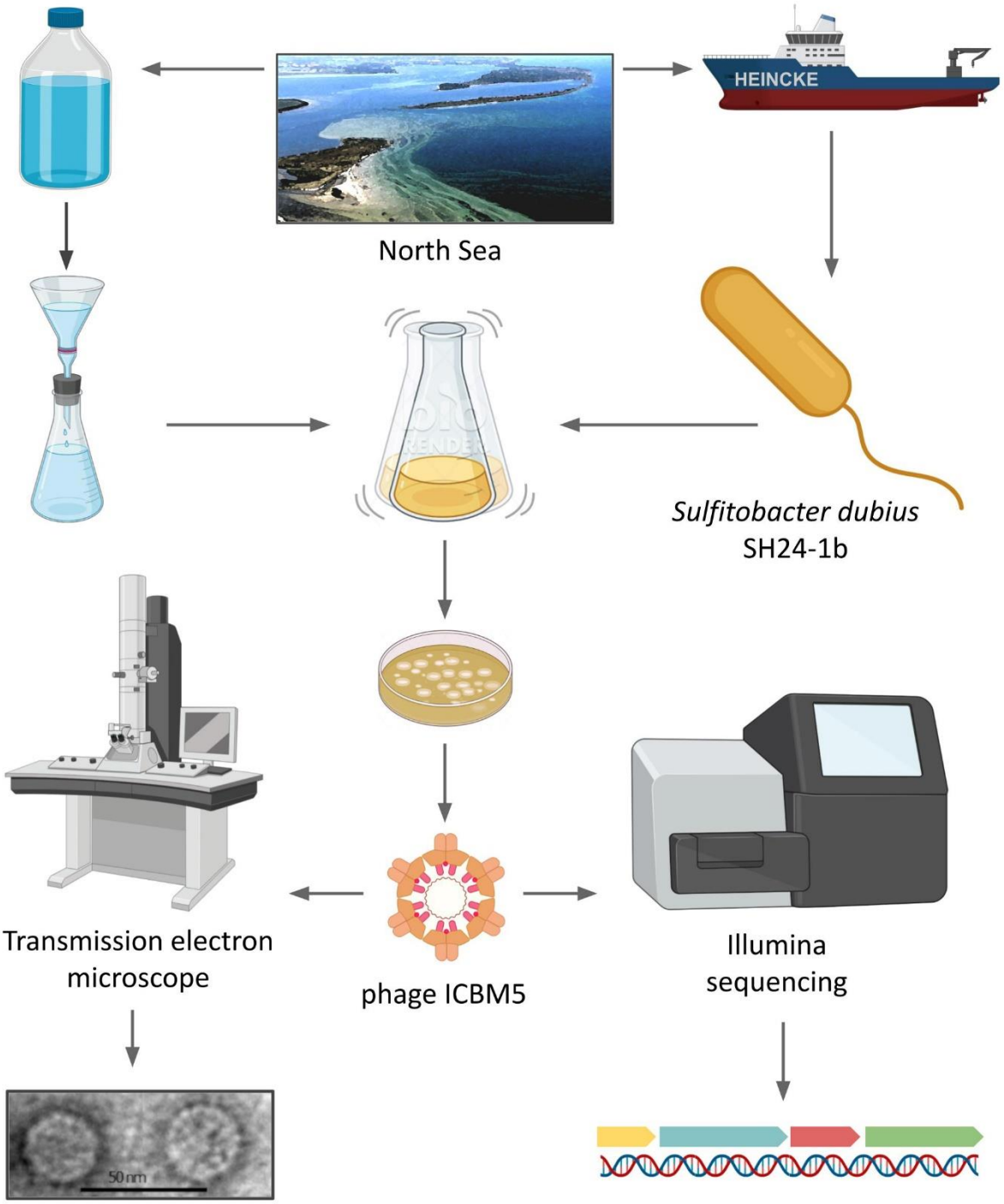

Fig. 8: Overview of the workflow summarizing the isolation of phage ICBM5 from the North Sea as well as the additional preliminary findings. The image was created using BioRender (https://biorender.com/).

1.8 Aims of this study

This study aimed to discover and characterize new ssDNA phages – especially from the family *Microviridae* – which infect marine *Alphaproteobacteria*. The model organism used in this study was a bacteriophage that was recently isolated from the North Sea and tentatively named ICBM5. Phage ICBM5 employed a lytic infection cycle strategy and additionally demonstrated indications for a temperate infection cycle. Initial genome analysis of phage ICBM5 showed that the virus is distantly related to known *Microviridae* and further Bioinformatics prediction showed the presence of *Micriviridae*-like prophages in a handfull of bacterial genomes. Therefore, the main objectives of this thesis were:

1) to investigate the life cycle of phage ICBM5;

2) to search for related prophages in bacterial genomes;

3) to place phage ICBM5 and the related prophages into a taxonomic and phylogenetic context of other ssDNA phages; and

4) to further characterize phage ICBM5 regarding its genomic and regulatory features.

2. Material and methods

2.1 Growth media

Tab. 1: Overview of ingredients for the applied media solutions, mentioned quantities are used for 1.000 ml of filtered water.

Marine broth medium (MB)

5 g	Bacto peptone
1 g	Yeast extract
0.1 g	Fe(III) citrate
19.45 g	NaCl
12.6 g	$MgCl_2x6H_22O$
3.24 g	Na_2SO_4
2.38 g	$CaCl_2x2H_2O$
0.55 g	KCl
0.16 g	$NaHCO_3$
0.01 g	$Na_2HPO_42H_2O$
0.08 g	KBr
22 mg	H_3BO_3
34 mg	$SrCl_2$
4 mg	Na-silicate
2.4 mg	NaF
1.6 mg	$(NH_4)NO_3$

Artificial seawater base medium (ASW)

24.32 g	NaCl
10 g	$MgCl_2x6H_22O$
1.5 g	$CaCl_2x6H_2O$
0.66 g	KCl
4 g	Na_2SO_4
2.38 g	HEPES
1 ml	KBr (0.84 M)
1 ml	H_3BO_3 (0.4 M)
1 ml	$SrCl_2$ (0.15 M)
1 ml	NH_4Cl (0.4 M)
1 ml	KH_2PO_4 (0.04 M)
1 ml	NaF (0.07 M)

Marine broth (MB) was used both for the liquid cultures as well as for the plaque and spot assays. After autoclavation, the media was completed by adding 1 ml/l of a multi-vitamin solution (Balch et al. 1979). Artificial seawater (ASW) base medium was used for phage purification, plaque assays and one-step infection experiments. MB agar plates were prepared by using an MB medium which was prepared with MB difco (Thermo Fisher Scientific, USA).

2.2 Growth analysis and quantification of *Sulfitobacter dubius* SH24-1b

To determine the time points of the different growth phases as well as the generation time of *Sulfitobacter dubius* SH24-1b, growth curve experiments were conducted and combined with colony-forming unit (CFU) assays. For technical replicates, three Erlenmeyer flasks were filled with 80 ml of MB medium and inoculated with an overnight (ON) culture aiming for a final optical density (OD) at 600 nm of 0.01. The OD was measured every 2h over a total period of 24h. Additionally, 1 ml samples were taken in sterile reaction tubes from each flask and stored on ice. A dilution series (10^0 - 10^{-6}) was prepared using the collected samples and sterile MB medium. 100 µl of the diluted samples were evenly distributed on MB agar plates using an alcohol-flamed Drigalski spatula. Afterwards, the plates were incubated for at least 24h at 20°C and when colonies have formed they were counted on plates with at least 30 but not more than 300. By this approach, the number of bacteria could be determined for each sampling point and therefore associated with the different stages of growth. Finally, growth curves and CFU numbers have been used to calculate the mean generation time (see Appendix A, Tab. A1 and Tab. A2).

$$\mathrm{CFU/ml} = \frac{number\ of\ single\ colonies}{dilution\ x\ start\ volume}$$

2.3 Phage plaque assays using the double-agar layer method

To obtain single plaques for quantification, serial dilutions were prepared from phage fractions by combining them with MB medium or ASW base medium. Therefore, 100 µl of diluted phage solution was mixed with approximately 250 µl exponentially growing host culture (final OD in 3 ml = 0.023) and then incubated for 15 min on ice to allow for adsorption. Sequentially, the mixture was pipetted on MB agar plates and combined with 3 ml MB soft agar (0.6 % low melting point Biozym Plaque GeneticPure agarose, Biozym, kept warm at 37°C). The solution was carefully mixed by swirling the plates on the table for a few seconds. After the top layer has solidified, the plates were incubated overnight at 20°C. On the next day, phage plaques were detected, defined as clearing zones and counted for quantification. By this approach, the concentration of phage stock or phage lysate could be determined.

$$\mathrm{PFU/ml} = \frac{number\ of\ single\ plaques}{dilution\ x\ start\ volume}$$

2.4 Preparation of phage lysate and phage ICBM5 stocks

To generate a higher volume of phage lysate, plaque-forming unit (PFU) assays were conducted with 60 plates, using a dilution of an ICBM5 phage solution resulting in confluent lysis on bacterial lawn. Approximately 24 h after the PFU assay or when the plaques have formed, 5 ml of SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.4) was poured on each plate. Afterwards, the plates were incubated at 4°C for 6h. Then, the liquid was collected in sterile 50 ml centrifugation tubes. To remove excess cells and debris, the solution was centrifuged for 15 min, 4.000 g and 4°C and sequentially filtered through a 0.2 µm syringe filter. The final phage lysate was stored at 4°C or until further use. For long-term storage, two types of glycerol stocks were prepared: 1) a stock of free phage particles consisting of 1 part phage fraction and 1 part MB with 50% glycerol and 2) a stock of infected host cells consisting of 1 part infected cells (375 µl phage fraction combined with 375 µl host culture), 15 min on ice to allow for adsorption and then finished with 1 part MB media with 50% glycerol.

2.5 ICBM5 purification via CSCl gradient ultracentrifugation

To concentrate phages, polyethylene glycol (PEG) precipitation was conducted first. Therefore, phage lysate from 2.4 (Preparation of phage lysate and phage ICBM5 stocks) was mixed with PEG (final concentration 10%) and NaCl (final concentration 0.6 mM). The mixture was incubated for 2h at 4°C and then centrifuged for 2h, 7,197 x g and 4°C. Sequentially, the supernatant was discarded, the pellet resuspended in SM buffer (100 mM NaCl, 8 mM MgSO4, 50 mM Tris-HCl pH 7.4) and then incubated for 30 min at 4°C.

Further purification of phages was performed by caesium chloride gradient ultracentrifugation. Therefore, a density gradient was set up in UltraClear™ centrifuge tubes (Beckman Coulter, USA) using caesium chloride solutions with different densities (from bottom up): 1.5 ml of 1.65 g/ml CsCl, 2 ml of 1.5 g/ml CsCl, 2 ml of 1.4 g/ml CsCl and 1 ml of 1.2 g/ml CsCl. On top of the gradient, the PEG concentrated phage fraction was carefully added. The ultracentrifugation was running for 4h, 25,000 rpm at 20°C (Beckman SW 41 Ti, USA). The visible band representing the phage fraction was collected (~ 500 µl) using a syringe and needle by piercing through the wall of the ultracentrifuge tube. Removal of caesium chloride residues was done via dialysis using Slide-A-Lyzer® G2 Dialysis Cassettes 10K MWCO (Thermo Fisher Scientific, USA) in an ASW base buffer solution for 21 h. The ASW buffer was

exchanged after 3 h and again after 18 h. Finally, the purified phage solution was tested for lysis by spot assay. Spotting dilutions (10^{0}-10^{-11}) of the concentrated phage solution were pipetted on double agar layer plates containing exponentially growing *Sulfitobacter dubius* SH24-1b followed by incubation at 20°C.

2.6 Phage genome extraction using phenol-chloroform

The purified phage solution from step 2.5 (ICBM5 purification via CSCl gradient ultracentrifugation) was mixed with the same amount of phenol:chloroform:isoamyl solution (25:24:1) (Carl Roth, Germany). The mixture was gently inverted and then centrifuged for 15 min, 12,000 x g and 4°C. The aqueous phase was carefully collected and mixed with an equal amount of ice-cold absolute ethanol (Th. Geyer, Germany). The DNA was precipitated at -80°C for 30 min and subsequently pelleted by centrifugation for 20 min, 12,000 x g and 4°C. Finally, the DNA pellet was dissolved in nuclease-free water (Thermo Fisher Scientific), purified with the NucAway spin column kit (Thermo Fisher Scientific, USA) and quantified via the Nanodrop 2000 spectrophotometer.

2.7 Enzymatic digestions for testing the ssDNA nature of the ICBM5 phage

To determine the genomic architecture, extracted phage DNA from 2.6 (Phage genome extraction using phenol-chloroform) was exposed to four different enzymes: S1 nuclease (Thermo Fisher Scientific, USA), TURBO DNAse (Thermo Fisher Scientific, USA), Exonuclease VII (New England Biolabs, USA), and Hind III (New England Biolabs, USA). S1 and Exonuclease VII strictly target single-stranded DNA, while TURBO DNase digests both, single and double-stranded DNA. Hind III targets only dsDNA. For each enzyme, a 50 µl reaction was set up, by adding 1 µl enzyme, 1 µg of extracted phage DNA, corresponding reaction buffers, and water. Each of the reaction tubes were incubated for 30 min at 37°C, which was followed by 10 min at 95°C, to deactivate the enzyme activity. For visualization of the final digestion products, 2 µl of digested DNA were mixed with 5 µl loading buffer (BlueJuice Gel Loading Buffer, Thermo Fisher Scientific, USA) and loaded on a 0.9% agarose gel. The gel was run for 30 min at 80 V and pre-stained with SYBR gold (Thermo Fisher Scientific, USA). Finally, the gel was analyzed with FAS Digi Gel Documentation System (NIPPON Genetics Europe, Germany) and evaluated using BioDocAnalyze software (Biometra GmbH, Germany).

2.8 Sequencing of the host strain lysogenic for ICBM5 phage via Nanopore

Extracted DNA from the resistant host strain was checked for purity via Nanodrop 2000 spectrophotometer before library preparation was done. To make sure that the pores are not clocked by impurities during the run, the absorbance ratio of 260/280 needed to be between 1.8 and 2.0, and an absorbance ratio of 260/230 between 2.0 and 2.2. Once purity had been validated, library preparation was performed in accordance with the manufacturer's instructions (Oxford Nanopore Technologies, SQK-RBK004, UK). The sequencing was then performed on a MinION Flow Cell (R9.4.1), controlled by the MinION software.

After sequencing, the quality control was performed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (fastqc -t 16 $input"01_reads.fastq" -o $Step2_out --nano). Adapters were removed using PoreChop (https://github.com/rrwick/Porechop) (porechop -i $input"01_reads.fastq" -t 16 -v 2 -o $output"03_reads_trimmed.fastq" > $output"03_porechop.log"). Assembly was performed using a set of tools from the Pomoxis suite (https://github.com/nanoporetech/pomoxis). Minimap2 (minimap2 -x ava-ont -t $threads $output"01_reads.fastq" $output"01_reads.fastq" | gzip -1 > $mapping"08_mapping.paf.gz") and miniasm (miniasm -f $output"01_reads.fastq" $mapping"08_mapping.paf.gz" > $mapping"08_miniasm_reads.gfa") were used to map the .fastq files onto each other to find overlaps. Afterwards, the .gfa file was converted to .fasta (awk '$1 ~/S/ {print ">"$2"\n"$3}' $mapping"08_miniasm_reads.gfa" > $assembly"09_miniasm_reads.fasta"). Assembly was done by minimap2 (minimap2 $assembly"09_miniasm_reads.fasta" $output"01_reads.fastq" > $assembly"09_minimap_reads.paf") and polishing was done by racon (racon -t $threads $output"01_reads.fastq" $assembly"09_minimap_reads.paf" $assembly"09_miniasm_reads.fasta" > $polished"10_racon.fasta"). A last quality control was performed by Quast (quast.py -t $threads -o $Step10_QC $polished"10_racon.fasta" $assembly"09_miniasm_reads.fasta") (Gurevich et al. 2013). Further analysis of the final fasta files, as for example genome alignments, was done in Geneious v 9.1.5 (http://www.geneious.com (Kearse et al. 2012)).

2.9 One-step growth experiment with phage ICBM5 and fixation of collected samples

To identify the different stages during an infection cycle, one-step growth curve experiments were conducted. First, it was important to ensure stable and consistent growth for the host strain of interest, *Sulfitobacter dubius* SH24-1b. This was achieved by growing the host culture in MB media at 20°C and 100 rpm for three consecutive generations. Then, when the culture reached an early exponential growth in the last generation (an OD of approximately 0.3), the purified ICBM5 phage from step 2.5 (ICBM5 purification via CSCl gradient ultracentrifugation) was added using a multiplicity of infection (MOI) of 6.5. The MOI was calculated using the CFU numbers from step 2.2 (Growth analysis and quantification of *Sulfitobacter dubius* SH24-1b) and the PFU assay of the purified phage stock, one day before the one-step infection experiment. At an OD of 0.5, the host culture concentration of *Sulfitobacter dubius* SH24-1b was approximately $1*10^8$ CFU ml^{-1} and the phage stock concentration after purification was about $1.2*10^{11}$ CFU ml^{-1}. By multiplying the desired MOI with the host culture concentration, the final concentration of phage stock could be calculated ($6.5*10^8$ PFU ml^{-1} in $1*10^8$ CFU ml^{-1} gives a final MOI of 6.5). To dilute the phage stock, sterile ASW base medium was used. Additionally, a negative control culture was prepared where ASW base medium was added instead of the phage stock. After the start of the infection, phage adsorption took place for 20 min at 20°C. Free phages were removed afterwards via centrifugation for 10 min at 6.000 x g and the pellets were resuspended in 1.5 ml MB medium (the initial volume was 2 ml, 1 ml host culture and 1 ml diluted phage stock). The resuspended infected culture was added to an Erlenmeyer flask with a final volume of 75 ml and incubated at 20°C and 100 rpm for 3.5 h. During this time frame, samples were collected every 30 min for phage quantification via PFU assays, these numbers were used to construct the final one-step growth curve. Therefore, plaque assays were used to quantify both the free and total (free and cell-bound) phages. The free fraction was obtained by filtering the infected culture through a 0.22 µm syringe filter. Additionally, samples were also collected every 15 min and fixed with paraformaldehyde (PFA) (Electron Microscopy Sciences, USA) to be further analysed in fluorescence *in situ* hybridization (FISH) experiments. Therefore, 1.5 ml of samples (infected culture as well as negative control) was mixed with 500 µl 16% paraformaldehyde each (f.c. 4%) and incubated for 1h at RT. The mixture was centrifuged for 10 min at 5.000 x g and 4°C, the supernatant was discarded and the pellet was resuspended in 1 ml 1 x phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4) (Invitrogen). The

solution was centrifuged again for 10 min at 5.000 x g and 4°C, the supernatant was discarded afterwards and the pellet was carefully resuspended in 100 µl of 1x PBS solution. Then, 100 µl of absolute ethanol (Th. Geyer, Germany) was added and the fixed culture was stored at -20°C until further use. The PFU-based burst size was determined from the one-step growth infection cycle experiment:

$$\text{Burst size} = \frac{\text{average free phages after lysis} - \text{average free phages before lysis}}{\text{average free phages average at T1}}$$

2.10 Construction and labelling of ICBM5 genome probes

As preparation for Fluorescence *in situ* hybridization (FISH) experiments, eight specific dsDNA polynucleotides (see table S1) were designed *in silico* to target the ICBM5 genome. For a successful hybridization, it was important for all polynucleotides in the mixture to have a similar melting temperature (not more than 5°C of variation). Additionally, to have specific target detection, it was important to ensure that the probe mixture does not contain polynucleotides potentially binding to other, non-target nucleic acid sequences. GeneProber addresses the mentioned objectives and was therefore used for constructing the probes (gene-prober.icbm.de) (Moraru 2021a).

The dsDNA polynucleotides were then synthesized by Integrated DNA Technologies (IDT, USA) and delivered as dried pellets. These were resuspended in 5 µl labelling buffer (5 mM Tris-HCl and 1 mM EDTA), to a final concentration of 100 ng/µl. 2 µl of each probe was collected in a 0.2 ml thin-walled PCR tube (Thermo Fisher Scientific, USA) and filled up to 20 µl with labelling buffer. This was followed by denaturing the probes at 95°C for 5 minutes and placing them on ice. Immediately afterwards, 30 µl of Alexa Flour 594 labelling reagent (Thermo Fisher Scientific, USA) was added and then incubated at 80°C for 30 minutes. The probes were purified using the NucAway Spin column kit (Thermo Fisher Scientific, USA). The concentration and labelling efficiency was measured using a Nanodrop 2000 spectrophotometer. Finally, base-to-dye ratios were calculated according to the instructions given with the ULYSIS Nucleic Acid Labeling Kit (see Appendix B, Thermo Fisher Scientific, USA).

2.11 ICBM5-targeted phage targeted direct-geneFISH or phage targeted direct-geneFISH

To detect intracellular phages as well as to visualize the different stages of an infection cycle a modified version of the phage targeted direct-geneFISH protocol was used (Barrero-Canosa et al. 2017; Barrero-Canosa and Moraru 2021). First, the fixed cells from step 2.8 (One-step growth experiment with phage ICBM5 and fixation of collected samples) were immobilized on a solid support. This was done by spotting 20 µl of fixed cell suspension (infected culture as well as control culture) on SuperFrost Plus glass slides (Electron Microscopy Sciences, USA). On these, silicone isolators (Grace Bio-Lab, USA) were placed to create wells. The spotted cells were then dried at 37°C and washed afterwards for 1 min in 0.22 µm filtered, deionized and autoclaved water (MQ water). An additional rinsing step was done for 30 seconds in absolute ethanol (Th. Geyer, Germany). Next, permeabilization and RNA removal was performed. Hence, samples were treated with a solution containing 0.5 mg/ml lysozyme (Sigma Aldrich, USA), RNase Cocktail (500 Uml^{-1} RNase A, 20.000 Uml^{-1} RNase T1) (Thermo Fisher, USA), 0.05 M EDTA pH 8.0, and 0.1 M Tris-HCl pH 8.0. Afterwards, the slides were incubated for 30 min at 37°C while covering the silicone isolators with a coverslip to prevent evaporation. This was followed by washing the glass slides twice for 5 min in 1 x PBS, 1 min in MQ water, 30 sec in absolute ethanol, and finally, by air drying. Then, denaturation and hybridization were performed. Under a fume hood, a closable glass container was prepared by placing sheets of tissues on the bottom that were soaked in 45% formamide. On these sheets, pipette racks were placed to serve as support for the glass slides. Next, a hybridization buffer was needed containing 45% formamide, 5 x SSC (750 mM NaCl, 0.075 mM sodium citrate), 20% dextran sulfate, 0.1% SDS, 20 mM EDTA, 0.25 mg/ml sheared salmon sperm DNA, 0.25 mg/ml yeast RNA and 1 % blocking reagent for nucleic acids (Roche, Switzerland). To the buffer, the ICBM5 phage genome probes from step 2.9 (Construction and labelling of ICBM5 genome probes) were added aiming for a final concentration of 30 pg/µl for each polynucleotide probe. The final probe-buffer solution was carefully mixed and evenly distributed on the prepared glass slides. The glass container was closed and the samples were denatured for 40 min at 85°C, followed by 2h of hybridization at 46°C. When hybridization was complete, the samples were cleaned in different washing buffer solutions. First, quick rinsing with washing buffer I (2 x SSC, 0.1 % SDS) at room temperature (RT) followed by soaking for 30 min in pre-warmed washing buffer II (0.1 x SSC, 0.1 % SDS) at 48°C. A final washing step was done for 15 min in 1 X PBS, 1 min in MQ water, and then air-dried. A counterstaining

solution was prepared by dissolving DAPI in SlowFade Gold (Thermo Fisher Scientific, USA) aiming for a final concentration of 5 ng ml^{-1}. The samples were then counterstained by using 3 µl of the prepared solution and then covered with a #1.5 high-precision coverslip (Marienfeld, Germany). During and after sample preparation, DAPI-treated slides were protected from sunlight as good as possible and always covered if not used. Therefore, visualization of prepared slides via microscopy was done as quickly as possible, otherwise, samples were stored in a dark container at 4°C.

2.12 Microscopy on samples generated by phage targeted direct-geneFISH

To finally visualize the samples that were generated by phage targeted direct-geneFISH an Axio Imager .72m fluorescent microscope (Carl Zeiss, Germany) was used. For image generation and processing, the associated software AxioVision (version 4.8.2.0) (Carl Zeiss, MicroImaging GmbH, 2006-2010, Germany) was applied. On the microscope, the highest resolution was used in conjunction with immersion oil (Carl Zeiss, Germany). Then, for each field of view, a set of images was generated where bacterial cells were evenly distributed and not too clustered, this allowed for a more precise quantification during analysis. Additionally, different exposure times for the two fluorescent channels, DAPI and Alexa594 were used. For DAPI, the following filter set was used: 365 excitation, 445/50 emission, and 395 Beam Splitter. For Alexa594, the following filter set was used: 562/40 excitation, 624/40 emission, and 593 Beam Splitter. The exposure times were 40, 80, and 150 ms for DAPI, and 80, 200, 600, 1200, 3000, and 5000 ms for Alexa594. To save the images for further analysis, each exposure time was marked individually and saved as .ZVI and as .TIFF file, respectively.

2.13 Image analysis for phage quantification

Images generated from phage targeted direct-geneFISH were analyzed using CellProfiler v. 3.1.9 (McQuin et al. 2018). By using this software for analysis, the fraction of infected cells could be quantified as well as the number of phage genomes. To do so, Cell Profiler was set up to define a semi-automatic workflow for image analysis. First, for every microscopic field of view, images from the two different fluorescent channels (DAPI for cell counterstaining and Alexa594 for phage detection) were imported. The respective images were previously acquired with exposure times allowing for the detection of single, weak phage

signals but also avoiding the overexposure of strong phage signals. Second, the contours of bacterial cells were automatically identified from the DAPI channel and additionally manually curated. Cell clumps were avoided, due to the difficulties in identifying cell borders. In parallel, the contours of 100 small, dot-like phage signals were manually identified in the first sampling time point after infection. These were considered to approximate a single ICBM5 genome. Third, for each corresponding Alexa594 image a mean background level (grey value/pixel) was measured in cell-free image areas and further subtracted from the entire Alexa 594 image to reduce background noises. Fourth, the mean Alexa 594 intensity (grey value/pixel) and area were determined for each cell or phage contour and exported. Fifth, an additional background correction was performed by averaging the Alexa 594 grey values of the cells in the negative control images (samples without an infection) and subtracting these values from the average intensity of the infected cells. Finally, the total Alexa 594 intensity (grey value) was calculated for each cell, by multiplying the average corrected intensity value by the cell area. The amount of per-cell phage genome copies was then calculated by dividing the per-cell total Alexa 594 intensity by the total intensity of a single phage. For each time point, a total of 100 phage-positive cells were quantified. The fraction of infected cells was calculated by manually counting from the images the cells with visible Alexa 594 signals, and then reporting this number to the total cell number, as defined from the DAPI images. For each time point, a total of 500 cells were investigated.

2.14 Detection and curation of ICBM5-like regions in bacterial genomes

To search for similar variants of phage ICBM5, protein sequences (see Appendix C) were used to query the non-redundant (NR) database from NCBI, using DELTA-BLAST, with two iterations. When a protein hit was detected, the GeneBank file was downloaded and imported into Geneious v 9.1.5 (http://www.geneious.com (Kearse et al. 2012). This procedure was done over a period of approximately two years and was repeated four times to consider updates of genome databases and find the newest entries possible. Detected entries were further identified as being part of a viral or bacterial genome based on their organism name and their taxonomy. Bacterial strains having hits with at least two different ICBM5 phage proteins were considered to potentially harbor ICBM5-like prophages and were selected for further analysis.

To identify the prophage borders, a selection of close relatives (same species) for each bacterial strain was used to search in the GenBank sequence database and aiming for very similar bacterial genomes both with and without the prophage. The selected genomes were aligned using MAUVE (Darling et al. 2004) and in doing so, prophage regions could be precisely defined. From here on, we refer to these as “sure border prophages” (SBPs). The remaining sequences have been designated as “unsure border prophages” (UBPs). For them, a larger region in adjacency of the phage-like genes was selected and additionally refined. In the first step, open reading frames (ORFs) of these UBPs were predicted using MetaGeneAnnotator (MGA) (Noguchi et al. 2008) and clustered together with proteins of ICBM5, other publicly available ssDNA phages and also of the sure border prophages. Next, an all against all BLASTp (e-value < 1e−5 and bitscore > 50) was performed, followed by protein clustering using the mcl program and the parameters “-I 2 --abc”. To further identify the UBPs, the defined protein clusters were processed via the following steps: 1) the UBPs were considered as phage genes if they were annotated as major capsid protein (MCP), replication initiation protein (REP), lysis protein, pilot protein or if they grouped in protein clusters with proteins from the SBPs or the reference ssDNA phages. 2) When genes translating for hypothetical proteins were located between previously determined phage genes, they were kept and labelled as unknown phage genes. 3) When neither of the above-mentioned cases was true and the genes located at the periphery of a region encoding phage genes, they were regarded as bacterial origin and UBPs were truncated accordingly. Additionally, genes from the contig that occurred in the reverse direction (3’ to 5’ instead of 5’ to 3’) were removed as well as those genes at the border that were in a unique cluster or unidentifiable. If a ribosomal binding site could be determined for the first gene, this has been used as the starting point for the prophage region. Otherwise, the start of the prophage region was determined at the start codon of the first gene.

2.15 Clustering *Microviridae* proteins and genomes

To classify phage ICBM5, a dataset was assembled by combining reference *Microviridae* genomes from publicly available sequence databases (NCBI), the newly detected ICBM5-like regions in bacterial genomes, ICBM5-related EVGs as well as the ICBM5 phage genome. The reference was defined by choosing those microvirus genomes, be it from phage isolates or EVGs that have been previously assigned to different *Microviridae* subfamilies.

Additionally, phage genomes from the *Obscuriviridae* family were included in the analysis (Bartlau et al. 2021).

Hierarchical clustering from the above-mentioned dataset of ssDNA phages was performed with VirClust, using a clustering approach based on their protein super-supercluster content (Moraru 2021b). Therefore, open reading frames (ORFs) were predicted, translated into proteins and then subjected to three clustering steps. First, the proteins were grouped into protein clusters (PCs) using the following parameters: BLASTp-based similarity (evalue > 0.00001, bitscore >= 30 and coverage > 70) and clustering based on log evalues. Second, PCs from the previous step were grouped into protein super-clusters (PSCs), applying the following parameters: Hiden Markow-Model (HMM) based similarity (probability >= 90, coverage >= 60, no threshold on alignment length) and clustering based on log evalues. Third, PSCs from the previous step were grouped into protein super-superclusters (PSSCs), using the following parameters: HMM-based similarity (probability >= 90, no threshold on coverage and alignment length). From here on we refer to PSSCs as “protein clusters”. The next step was to cluster the phage genomes hierarchically applying a clustering distance of 0.9.

The created protein clusters were annotated by VirClust comparing the individual protein sequences to various databases. The NR database from NCBI was queried using BLASTp 2.6.0+ and the InterPro database v 66.0 (Finn et al. 2017) was queried using the InterProScan 5.27-66.0 tool (Jones et al. 2014). The prokaryotic Viruses Orthologous Groups (pVOGs) database (Grazziotin et al. 2017), the Virus Orthologous Group (VOGDB) database (Kiening et al. 2019), and the Prokaryotic virus Remote Homologous Groups (PHROGS) database were searched using hhsearch (Steinegger et al. 2019). Finally, the efam database (Zayed et al. 2021) was searched using hmmscan (Eddy 2011). In the last step, annotations for each protein were manually curated and compared with annotations of the other proteins in the same protein cluster. In doing so, the annotation of each protein cluster represented a consensus of the annotations of all proteins in the same cluster.

2.16 Prediction of additional genomic and transcriptomic features of phage ICBM5

To search for further genomic as well as transcriptomic features, a set of bioinformatical prediction tools was used on the genome of phage ICBM5. RegRNA 2.0 was used on the viral sequences which already incorporates the prediction of the most common functional RNA motifs as well as an additional open reading frame prediction (Chang et al.

2013). To predict additional possible Rho-independent terminator sequences, ARNold was also used (Naville et al. 2011). ORFs and ribosomal bindings sites (RBSs) were also predicted using MGA (Noguchi et al. 2008). ORFs found on the reverse sequence have been further queried in Virclust for additional protein function identification. To further investigate for sRNAs on the ICBM5 genome (see Appendix D) a length of 101 nts was defined as a sliding window which moves over the entire genome sequence in incrementing steps of 50 nts, thereby creating 112 putative sRNA candidates. These input sequences were then further processed using presRAT in default settings to predict their likelihood of being sRNAs (Kumar et al. 2020). In case a putative sRNA with a high score (seven and higher) was found, the respective entry was further analysed with RNAfold to perform secondary structure prediction (Gruber et al. 2008). Afterwards, CopraRNA was used to search for possible mRNA targets (Wright et al. 2014). To operate correctly, CopraRNA required at least four NCBI reference sequences as additional input which should be closely related to the host strain of interest (see Tab. 2).

Tab. 2: Overview of additional input genomes for CopraRNA to aid in mRNA target prediction.

Accession No.	**Name**	**Identity to SH24-1b**
CP020694	*Sulfitobacter* sp. D7	99.7 %
CP045372	*Sulfitobacter* sp. THAF37	97.7 %
CP018076	*Sulfitobacter alexandrii* strain AM1-D1	97.0 %
CP022415	*Sulfitobacter pseudonitzschiae* strain SMR1	96.6 %

3. Results Objective 1 - Life cycle investigation of phage ICBM5

Graphical abstract:

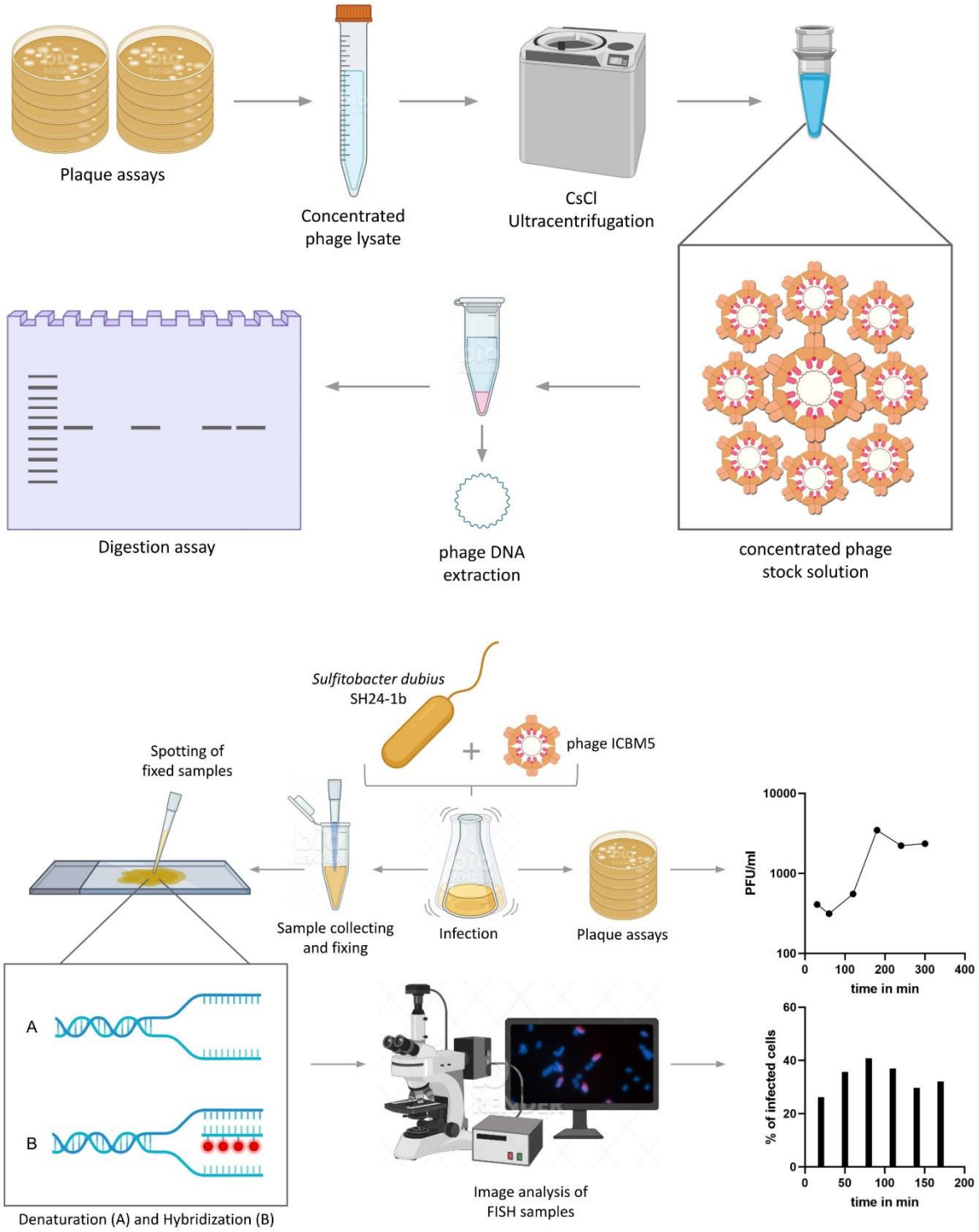

Fig. 9: Schematic overview of the workflow showing the main experiments performed within chapter I, the life cycle investigation of phage ICBM5. The image was made with BioRender (https://biorender.com/).

3.1 Growth behaviour of SH24-1b in laboratory settings

For a better understanding of the phage–host interaction and for following experiments, it was important to investigate the physiological characteristics of the host strain *Sulfitobacter dubius* SH24-1b. Therefore, growth experiments were conducted in MB media over a period of 24h. Samples were taken every 2 h and the OD at 600 nm was used to identify the different growth stages (see Fig. 10). In the given experiment conditions, exponential growth can be observed until about 12 h after the experiment has started. Exponential growth ended around 14 h when the culture transitioned into the stationary phase. In addition to the spectrophotometric analysis, bacterial cell numbers were quantified via CFU assays and compared with the growth curve (see Fig. 10). The time points for the different stages of growth corresponded to each other when comparing both findings. Furthermore, it could be determined that an OD_{600} of 0.5 corresponded to ~$1*10^8$ CFU ml^{-1}. Cell counts were also used to calculate the generation time. *Sulfitobacter dubius* SH24 has a mean generation time of 2h when grown in MB full medium at 20°C and 100 rpm.

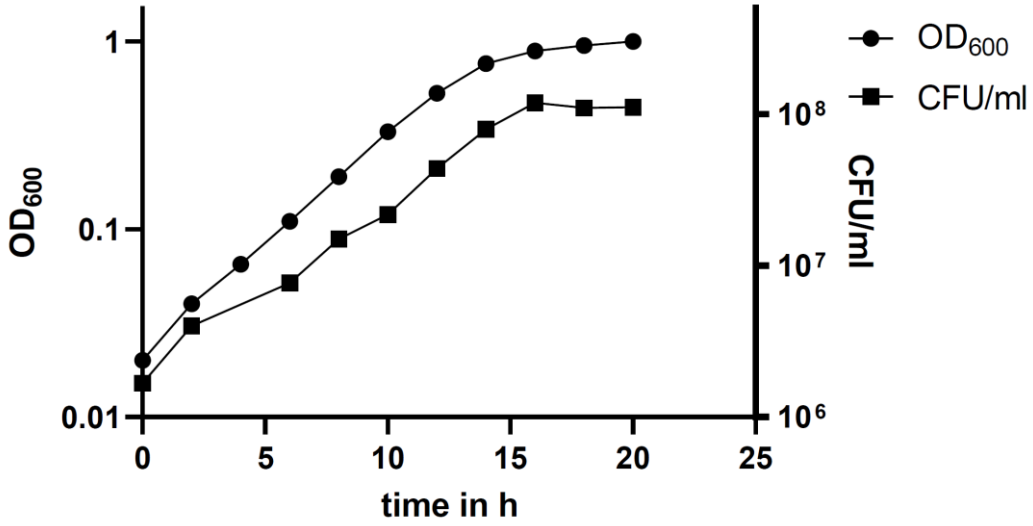

Fig. 10: CFU assay of *Sulfitobacter dubius* SH24-1b, dots representing the optical density (OD) values over time, while the squares show the development of CFUs/ml over time.

3.2 Purification of phage ICBM5 via caesium chloride ultracentrifugation

Several experiments performed in this study were dependent on a purified and concentrated phage stock solution. To acquire such a stock solution, purification by caesium chloride ultracentrifugation was conducted. Phages collected from several double agar-layer

plates showing confluent ICBM5 plaques were purified by centrifugation in the CsCl gradient. Here, the phage band was visibly distinct from the cell and cell-debris band (see Fig. 11), which facilitated its extraction. Using this methodology, phage stocks of ~ $1.2*10^{11}$ PFU ml^{-1} (average value taken from three replicates) were obtained. These phage stocks were used in further experiments.

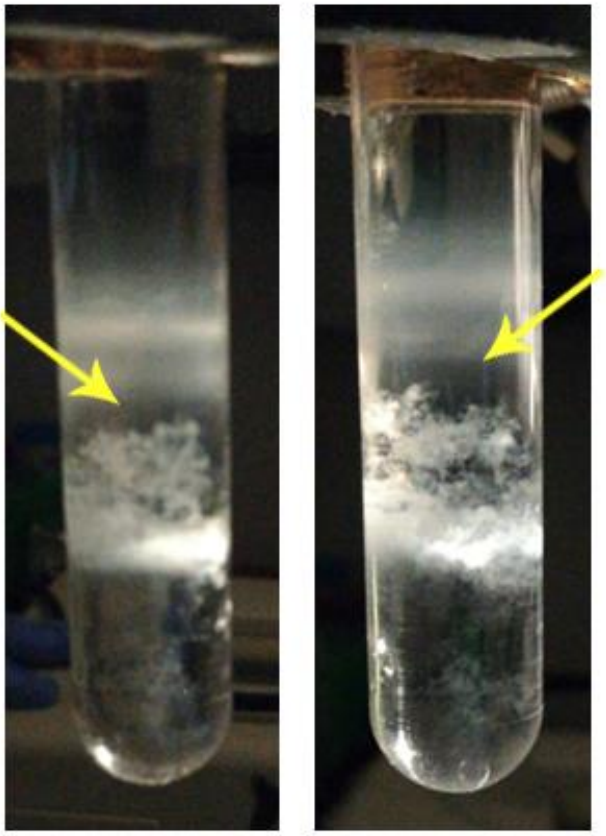

Fig. 11: Different bands after ultracentrifugation, phages were extracted below the second band, as indicated by the yellow arrows.

3.3 Investigating the ICBM5 genome type via enzymatic digestion

Genome analysis from the preliminary results indicated that phage ICBM5 might be of ssDNA nature. To confirm the previous findings, an enzymatic digestion assay was performed on phage DNA extracted using a phenol-chloroform procedure. Exonuclease VII and S1 nuclease were used to strictly target ssDNA, TURBO DNase digests ssDNA as well as dsDNA. Hind III targets exclusively dsDNA. Control samples were created by using the buffer solution without the enzyme and results were visualized by gel electrophoresis. The DNA of phage ICBM5 was completely digested by TURBO DNase, Exonuclease VII and S1 nuclease and no

digestion took place when left untreated or treated with Hind III or different buffer solutions (see Fig. 12). This indicates that the nucleic acid structure of phage ICBM5 is of ssDNA nature.

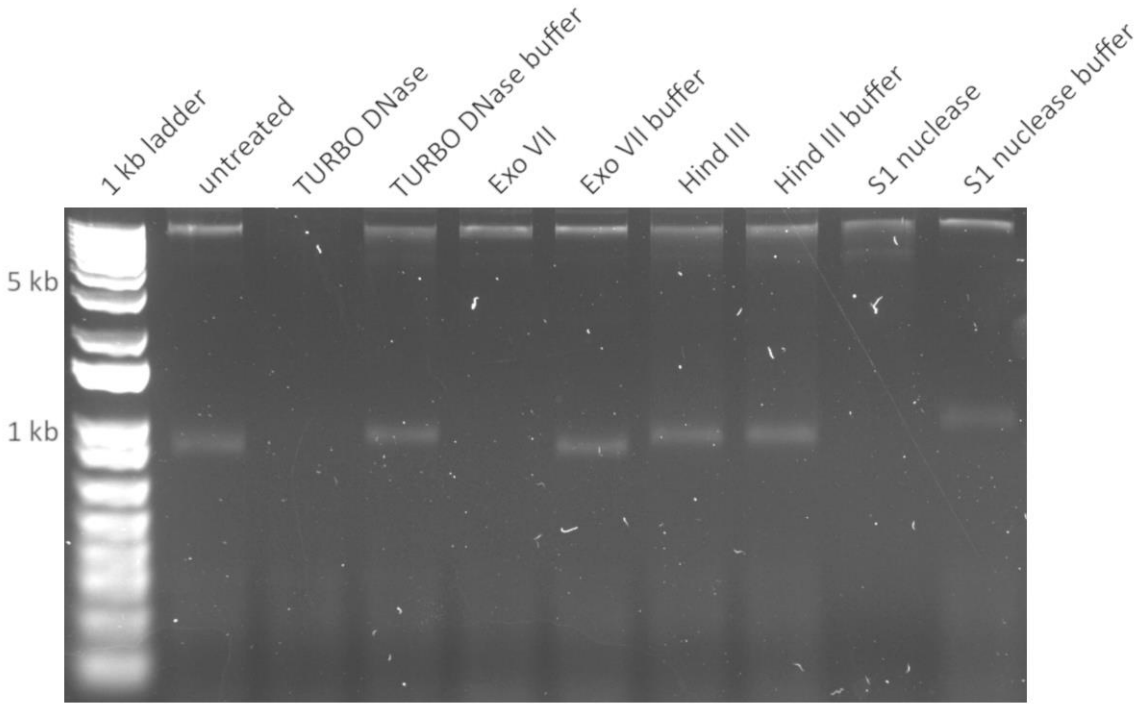

Fig. 12: Agarose gel shows enzymatic digestion of ssDNA phage ICBM5. The DNA was digested by TURBO DNase, Exo VII, and S1 nuclease, but was not affected by treating it via restriction enzyme Hind III, which only targets dsDNA, or the exclusion of nucleases (usage of only buffer). The 1kb plus ladder was used to track the DNA migration. However, it was not used to infer the size of the ICBM5 genome, because the ladder comprises from linear dsDNA molecules, in contrast to the ICBM5 genome, which comprises of a circular, ssDNA molecule.

3.4 Sequencing the putative lysogen *Sulfitobacter dubius* SH24-1b c2 via ONT sequencing

To determine the location (integrated into the host chromosome, or as an extrachromosomal element) of the ICBM5 genome in the ICBM5-carrier *Sulfitobacter dubius* SH24-1b strain, long-read sequencing technology was conducted via Oxford nanopore technologies (ONT). If the ICBM5 genome would have been integrated into the host chromosome, this would have been captured by long reads containing parts of the host genome surrounding the phage genome. Such a read would be evidence for an incorporated phage genome and therefore for a temperate lifestyle. The MinIon device has worked for 24h and produced 73.224 reads ranging from 27 – 141.894 nucleotides (nts) in length. In total, 451 Mbs were sequenced, so the whole genome read coverage was at around 111 fold (the bacterial chromosome plus plasmids have a combined size of 4.05 Mbs). After assembly and annotation, the host chromosome of *Sulfitobacter dubius* SH24-1b was reconstructed and

compared against the original sequence using Mauve whole genome alignment (Darling et al. 2010). The overall pairwise identity was 98.5% and the deviation in similarity can most probably be explained by the known high error rate of nanopore technology. Phage ICBM5 was also detected through ONT sequencing and was found to be present as an independent circular contig, but not to be integrated within the host chromosomes or plasmids. These findings showed that the phage ICBM5 can persist within its host *Sulfitobacter dubius* SH24-1b, but not as an integrated part within the host genome.

3.5 Design and synthesis polynucleotides for phage targeted direct-geneFISH

To design specific polynucleotide probes targeting the phage ICBM5 genome, the geneProber software was used (Moraru 2021a). In direct-geneFISH, multiple polynucleotides are used to detect a single target, to increase sensitivity. Therefore, for ICBM5, 8 polynucleotides were designed. Each had a length of 300 bps, and covered 2.400 bps /5.581 bps, representing ~43% of the ICBM5 genome (see Fig. 13). The polynucleotides were of a dsDNA nature. Even though phage ICBM5 has an ssDNA genome when encapsidated, it becomes dsDNA when the genome is replicated. Therefore, the dsDNA probe will detect the genome more efficiently, which leads to a stronger signal for later quantification analysis. Furthermore, the base-to-dye ratio was calculated to be 8.3 bases:dye (for the complete calculation, see Appendix B). An overview of all probes as well as their melting profile is shown in table 3 and figure 14.

Tab. 3: Overview of designed polynucleotides showing names, sequences, start and end position as well as GC content.

Names	Sequences	Start	End	Length	GC%
Poly_11	aggatgcgcgacgataaaattagctcttgcaaaccccgtccaaattcgactaaggtca atttagggaatggttcccggcgcgggtttatcccgtggtgctaacaaaggaaaaacaca tggctgataaagcaacttccccgactaccgtctctaaaaaagatcgtgatgattttgagc ttcatgttcggctcttgagccgggcaggttacaaaactgccgaagctcggacaattgca tggctgcgcggccctgctgggctgcaagaaatgcttgcgggcgtcaaggtcgcaagct aacata	11	310	300	50
Poly_1891	tcgaaggcgtcggcgttgaatacaatgccgttccagtgactacccccggttggaacgat acagagaatgaccgtgctgctgatcctcactatttctcggaccacgccgcaggcattcg ccttgcgattgcctcggagaatggtcgtccagcggtctacgctgttggcgagggtcaaa atgccggcggcatttcccttactgacttctacaacgccgaactgatggacagtcttgtgc gccagatgcggcagattgtcgatgataaccctgagtatggcgaagaaatggtcacacg ctggg	1891	2190	300	55
Poly_2229	attttgcaccaaagccagcaaatgttcggcaatcaattccgccgtgccatggatggtgc aaacctcgatgttgctcaatccgacatgatgcaaagccttgaattcactgtccccgtgcc tccaacggaattgggcggggtcgtgattacctttgcatctgtgaagcctgacgaaacac taggctctcagccgcacccgttcctgtcgacaacatgggaggcgactaactacgtctcc	2229	2528	300	54

	gacgagttggcacgcgaccctgaggccgtgaccatgcgtcacctgaacagtgacgttc ctatg				
Poly_2538	gacactcgtgctctctacatcggtcacaacggtctgaaaaaagcctacatctcatatgg ctttaaccggcatctcgaccctactacggtcgaagcgaaaactgccatctggcagcttg aggtgccaatgtccgtaacgcctgaaagcgttatctatcccgaggacttggaccattatc cgtttgcggatcagcttgcagaggtctgcacctatcaggttagttcgaccgctacagtgc ggacgcctatggtcttcggtccgacaccagttgaggaactggcccagattgagacaga caac	2538	2837	300	52
Poly_2868	ggttatgccgcccctcaataaaccttaaaaaactggaaagatgaaatgcaagttaatg atgcaaaacgctccgttcttctcgtcggccttaatggcggcgaagtttctgtaatctccgc agatggtgagattatcgcgaccgaaggtgtaaccgctggtcggcataaatgctcctctt gggtgccatttatgtccaatgagggcgatgaattgagcttctccggcgatgtggtcccaa tggtgccaaatggcggtcgtgtccggcctatggcctacggccccggtcaatttgaaagc ggt	2868	3167	300	51
Poly_3876	ctatagacctcgtgcatggttatgattggggatgggacagtgcccatattgagattgctt tgtggcgtcaagttccaaagcgtcaaattggcaaaactggctctaattatcccccgtcac cttgggaacgtgaacagcgtttcaaggagttgcttcccgctgtttatgcggcgcaacatc ggccatagacttatggcaatctatgcaaactaacaagcgggcggcctctatgcgagag ccgcccgttttgcatccctgcgggccagtcggagactggccagcggcaactatattccg ggc	3876	4175	300	53
Poly_4400	tttgcctgtcgtaagtgtaatgagtgcatcaccgctaggaaaaatggttgggttgctcgt gcgatggctgaaaaggccgtcactgctgaaactttcagcgtaaccctaacttataatga tgctactcaggaaagccgcgatgcggcaaagacgttcgagtatcgacacgtcaaaaac tggattaagaaccttaggcgtcaaatagagtacaccaccggccagactggccttctccg ctaccttgtagcgggtgagcgcggttccgataagggccgctgccactggcatgtaatctt gttc	4400	4699	300	51
Poly_4700	Tgtaacgccgatattttaacgctaggaaaaatgacgcactggccctctggcaaacttgc tgccgatagggccgaaattataactacaggaaaaaggaaaaaacgtattaactggtct ctctggccgtatggcttcgtcagctttcaagagcctgaccagtggggcatggagtacgc cttagcgtatgccctcaaagatcagtttaacattgtctccgctgccggtacggcgcggga ggctcacgtttcccgcacttctgcgggtatgttccgcatgtccaaaaaaccccccaatcgg tttc	4700	4999	300	51

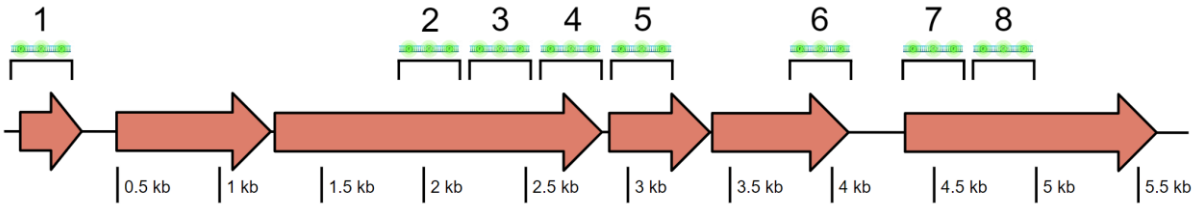

Fig. 13: Overview of the eight designed polynucleotides spanning over the entire genome of phage ICBM5. The shown probe regions are each 300 base pairs (bps) in length and cover almost half of the genome.

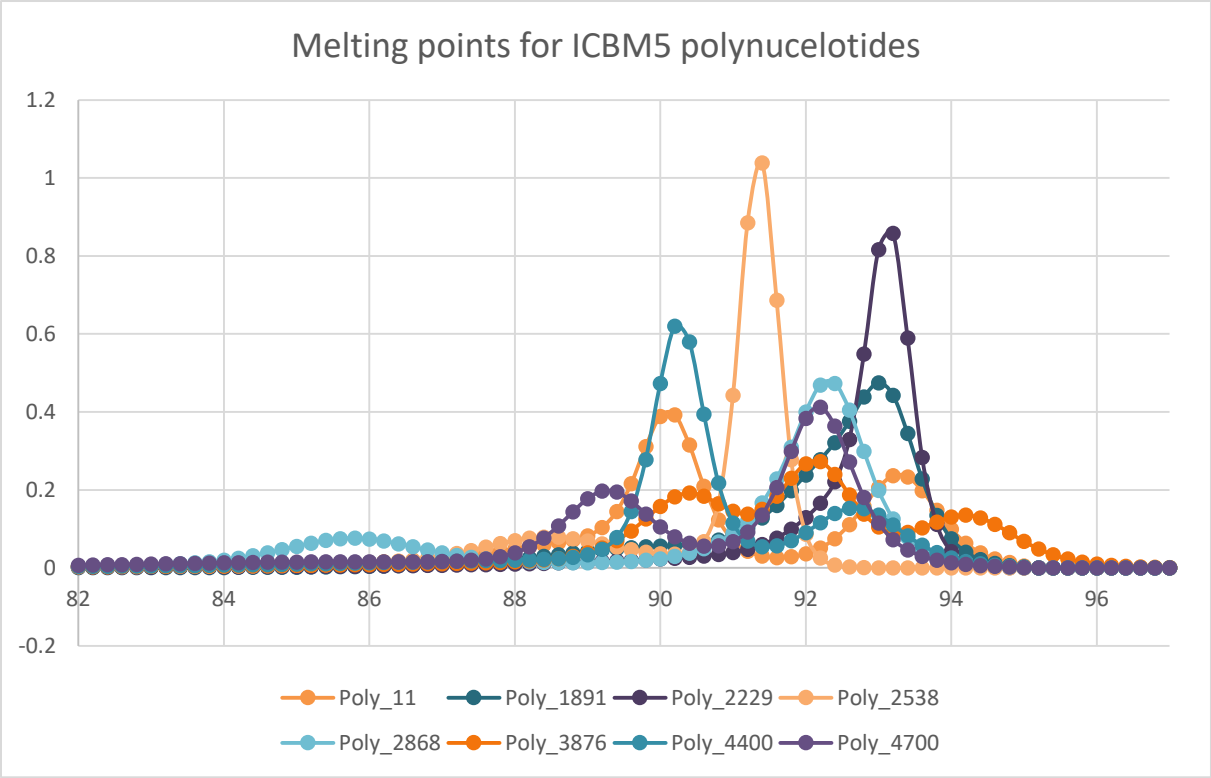

Fig. 14: Melting profiles of the eight polynucleotides used for phage targeted direct-geneFISH on phage ICBM5, x-axis shows the temperature in °C and on the y-axis the -dFoldedFraction/dTemperature is shown (d stands for derivative and the folded fraction stands for dsDNA). The predicted melting points of the polynucleotides had a temperature difference of a maximum 5 °C, as required by efficient hybridization of polynucleotides mixtures.

3.6 One-step growth curves as preparation for the infection cycle investigation

To analyse the infection cycle of phage ICBM5 on its host *Sulfitobacter dubius* SH24-1b, a set of one-step growth experiments was performed. Each of these experiments was performed at an MOI of 0.1, as suggested by Middelboe et al. 2010, to avoid multiple infections per cell. Furthermore, a filtration step was implemented to also quantify the free phage fraction. The first experiments were conducted with sampling points every 60 minutes to narrow down the first lysis event. Later, 30-minute intervals have been used for sampling and quantification. The plotted curve (see Fig. 15A) shows an increase in PFU counts starting 50 min after the start of the infection. Already, these findings indicate that bacterial cell lysis takes place between 50 min and 180 min post-infection which defined the time frame for upcoming experiments.

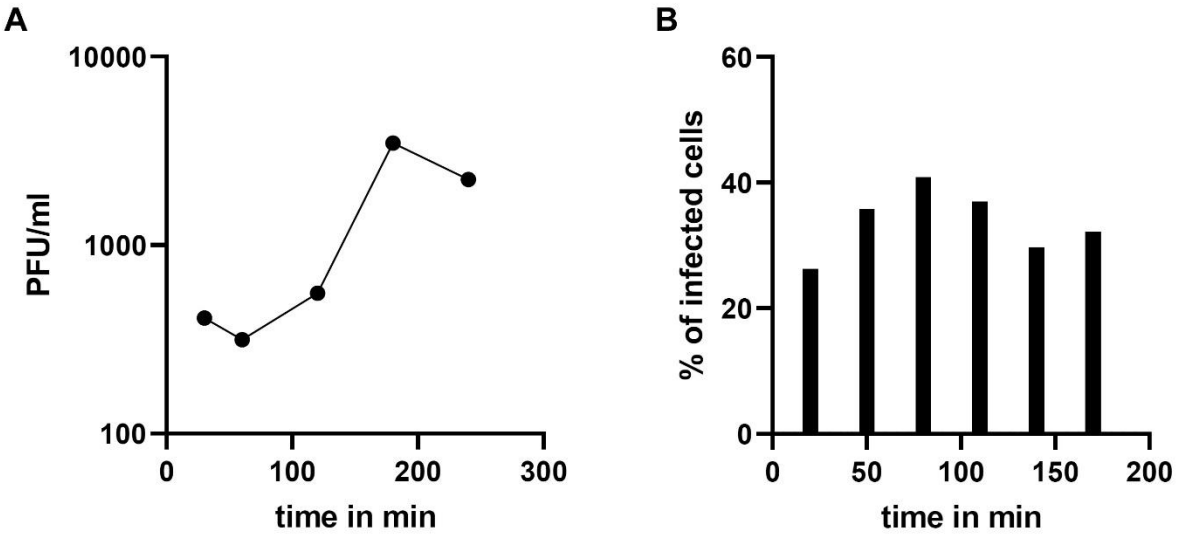

Fig. 15: A) One-step growth curve of phage ICBM5 infecting *Sulfitobacter dubius* SH24-1b. The curve represents the free phage fraction. The x-axis describes the time for the course of the infection, while the y-axis shows the PFUs/ml. Between 60 min and 180 min, an increase in PFU/ml can be observed which indicates a release of new phage particles and therefore a probable lysis event. B) Percentage of infected cells over time represented as bar diagram.

3.7 Determining parameters for phage targeted direct-geneFISH

To determine the optimal conditions for phage targeted direct-geneFISH, a series of different experiments were conducted. An MOI that resulted in a proportion of infected cells which can be detected with phage targeted direct-geneFISH needed to be identified. Therefore, phage targeted direct-geneFISH was performed using three different MOIs (1, 10 and 100) on an infected culture during early exponential growth. An MOI of 1 has shown very little to no proportion of infected cells while an MOI between 10 and 100 has shown acceptable proportions of infected cells (see Fig. 16) and were therefore used for future experiments. Additionally, the signal intensity over the course of infection was quantified at an MOI of 10. The first samples were taken 20 min after incubation for adsorption. Sequentially, samples were taken every 30 min until 170 min post-infection and images were visualized afterwards via fluorescence microscopy. These images have been used to detect the total amount of cells per image as well as the number of infected cells by counting them manually. In doing so, an estimate of the percentage of infected cells per time point could be determined and the development of infected cells over the course of infection could be plotted (see Fig. 15B). The percentage of infected cells increased until 80 min post-infection. From there, a drop occurred and the percentage of infected cells started to decrease indicating the release of new phages.

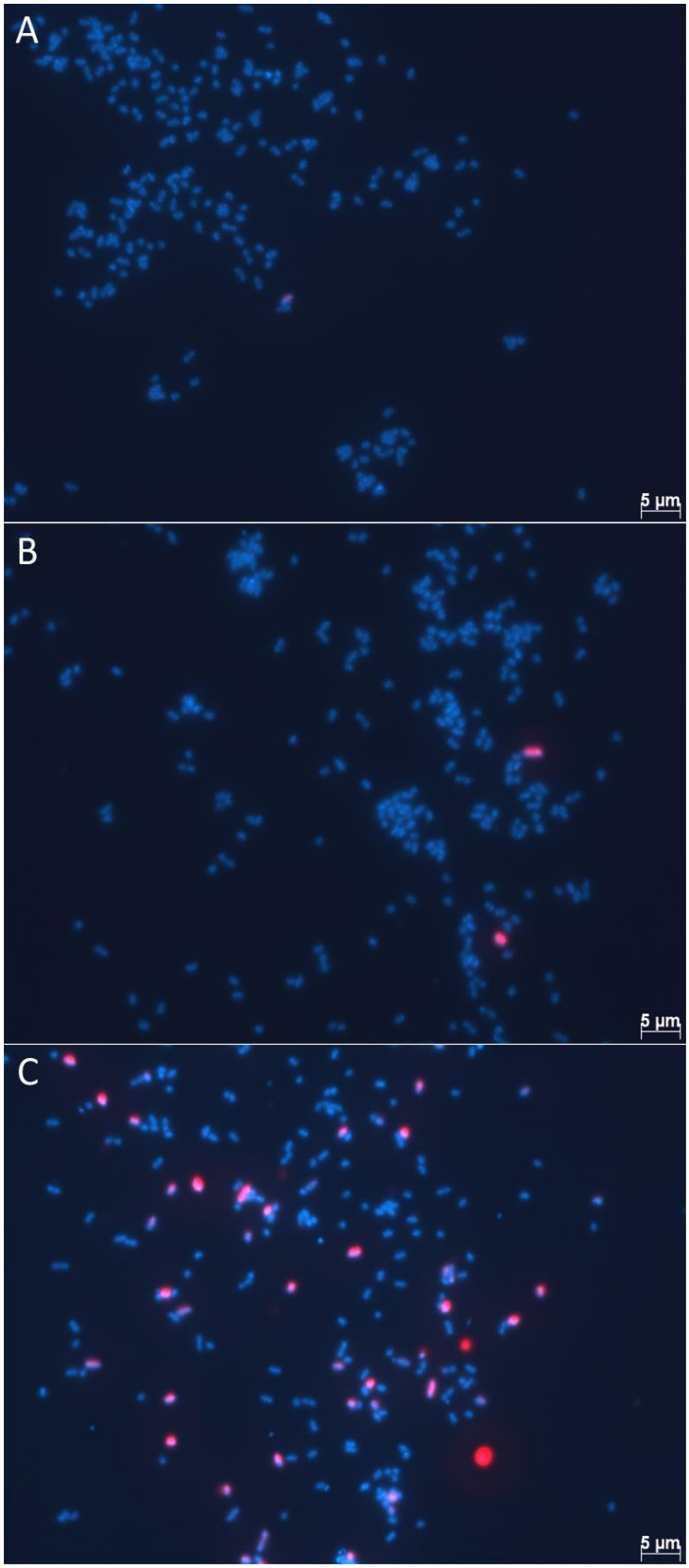

Fig. 16: Visualization of an infected culture at early exponential growth using different MOIs, A) shows an MOI of 1, B) shows an MOI of 10 and C) shows an MOI of 100. Blue dots represent the DAPI stained bacteria, while the red dots represent the phage signal visualized by ICBM5 genome-targeted probes.

3.8 Investigating the life cycle of phage ICBM5 via one-step growth curves and phage targeted direct-geneFISH

The findings from the previous experiments pointed in the direction that bacterial lysis occurred in between 80 and 140 min after the start of infection (see Fig. 15B). Additionally, a suitable range of MOIs has been determined to generate quantifiable results for further analysis. To get a better understanding of the life cycle of phage ICBM5, an experiment was conducted to follow an entire infection cycle by performing one-step growth curves and phage targeted direct-geneFISH simultaneously. Additionally, the whole experiment was replicated on another day. The one-step growth curve experiments were performed to identify the different stages of infection by quantifying PFU assays. For every time point taken for PFU assays, the free phage fraction as well as the total phage fraction was obtained and quantified. This was done by collecting two samples per time point and filtering one of them through a 0.22 µm syringe filter before performing plaque assays. Phage targeted direct-geneFISH was applied to visualize the different stages of infection while saving the images for later quantification of intracellular phage genomes. For collecting samples during a complete wave of infection, phages and bacterial cells were brought together in a small volume at an MOI of 6.5 and incubated for 20 min at 20°C. From there on, samples were collected and fixed for phage targeted direct-geneFISH every 15 min and collected and processed for PFU assays every 30 min. In total, 11 time points were acquired within 185 minutes. The samples that were taken for phage targeted direct-geneFISH have been used to determine the percentage of infected cells as well as the numbers of ICBM5 genomes per infected cell (see Fig. 17). A gradual increase of ICBM5 genomes was observed from 50 min post-infection (p.i.) which was growing until 110 min p.i. as seen in the total phage fraction (see Fig. 17). Furthermore, 50 min after the start of infection, the median of phage genomes per-cell was 6.5x and 2x higher than at 35 min p.i. for experiment 1 and experiment 2, respectively. Additionally, 110 min after the infection started, it was 61x and 81x higher than at 35 min p.i. for experiment 1 and experiment 2, respectively. This shows a start of genome replication at least as early as 50 min p.i. Moreover, bacterial cell lysis events could be visually observed between 110 min and 140 min p.i. (see Fig. 18) in both experiments. Congruently, the free phage particles increased in numbers starting at ~110 min p.i. This corresponded with the decline of the cell population, with a high amount of ICBM5 genomes, and the emergence of a cell population with a low amount of ICBM5 genomes, as indicated by the progressive drop in the median number of

ICBM5 genomes per cell (see Fig. 19). However, the population with a small number of ICBM5 genome copies maintained stable from 155 min p.i. until the end of the experiment. This indicates that the bacterial cell lysis took place starting at 110 min post-infection, leading to a release of new viral particles infecting new cells and forming a second wave of infection. Taken together, these findings demonstrated that phage ICBM5 is capable of undergoing a complete lytic cycle on *Sulfitobacter dubius* SH24-1b within approximately 110 min.

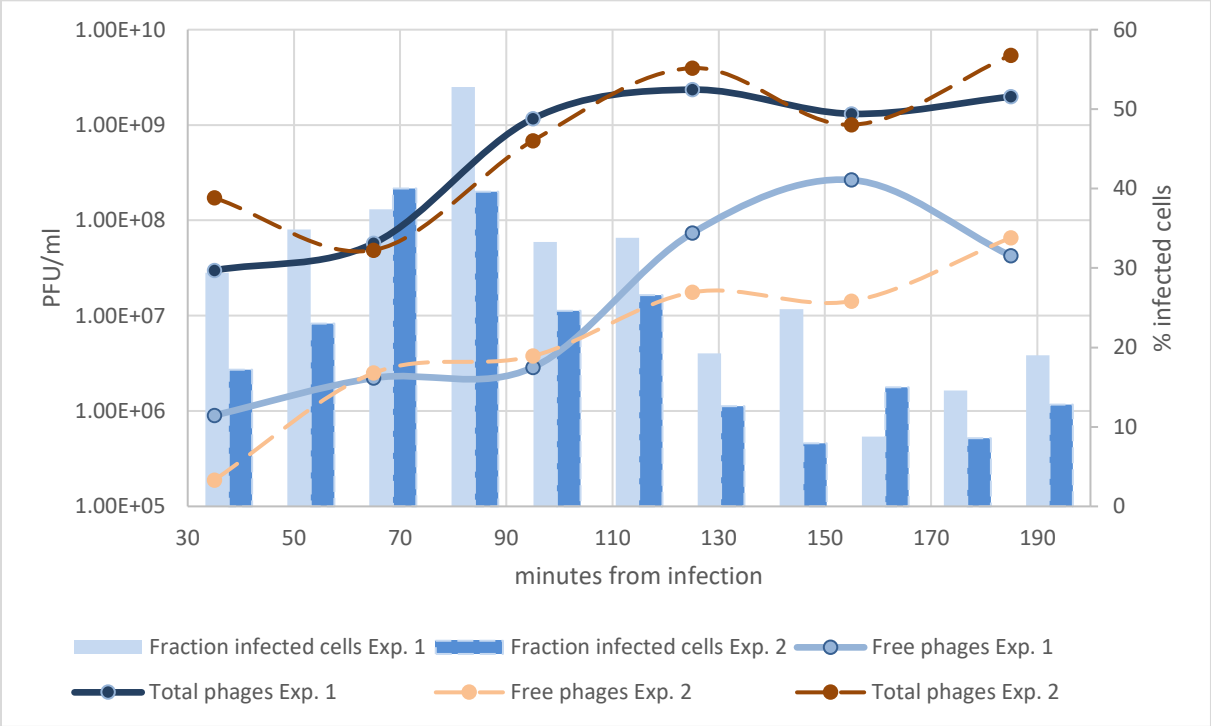

Fig. 17: The fraction of ICBM5 infected cells (bars) and the abundance of free (extracellular) and total (intra- and extracellular) phages, given in PFU/ml (lines). The fraction of infected cells was calculated from the proportion of cells showing a phage signal, after subtraction of the false-positive signals detected in the negative control cultures.

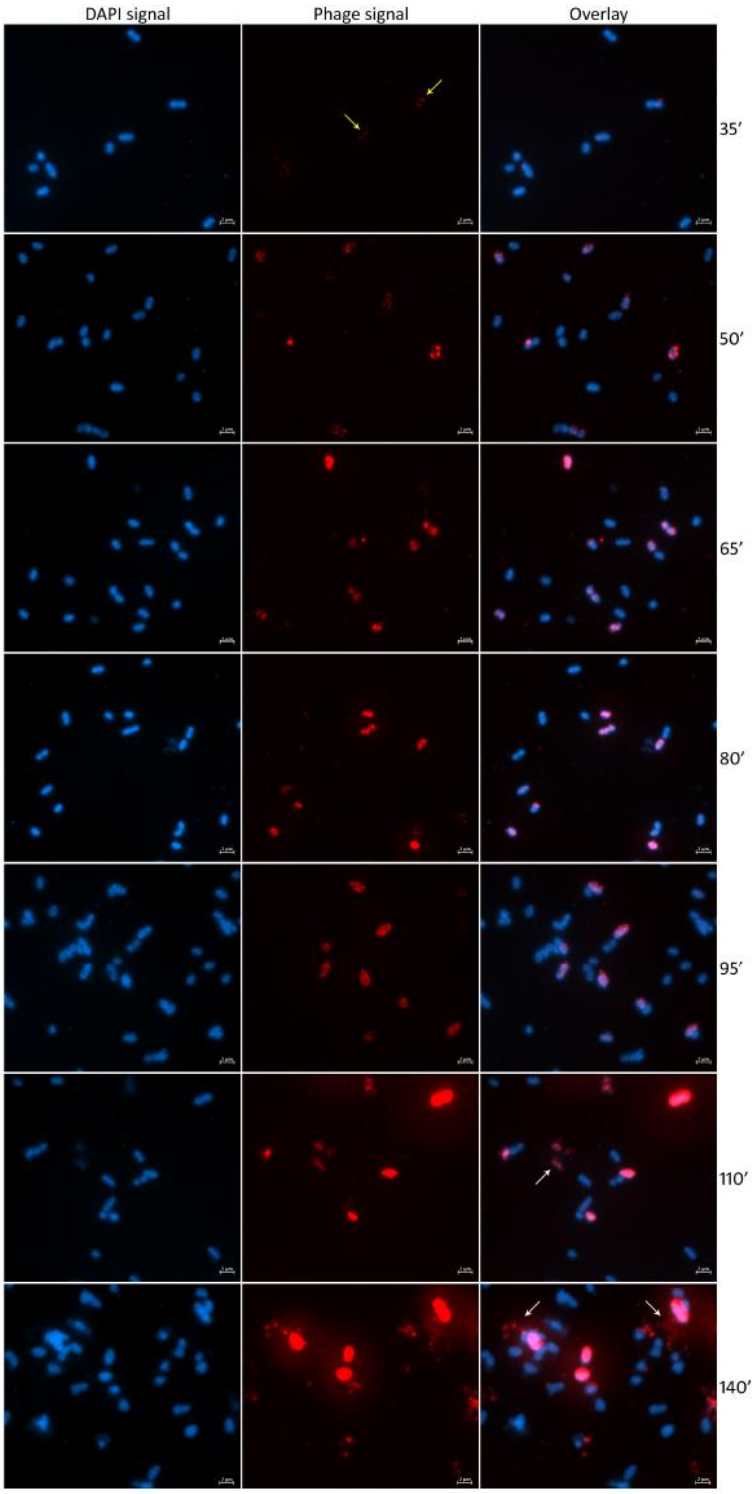

Fig. 18: Visualization of the infection of *Sulfitobacter dubius* SH24-1b by phage ICBM5 using phage targeted direct-geneFISH. First column: cells as visualized by DAPI staining. Second column: intracellular phages, as visualized by ICBM5 genome-targeted probes. Third column: overlay of DAPI and phage signal. The phage signal increases progressively through the infection. The very small dots in the first time point (yellow arrows) represent early infections, with just a few copies per cell. Larger, cell-wide signals in the following time points represent phages in the replication and maturation phase. And finally, at 110' and 140', cell lysis events can be noticed (white arrows). For a higher resolution image, open file externally.

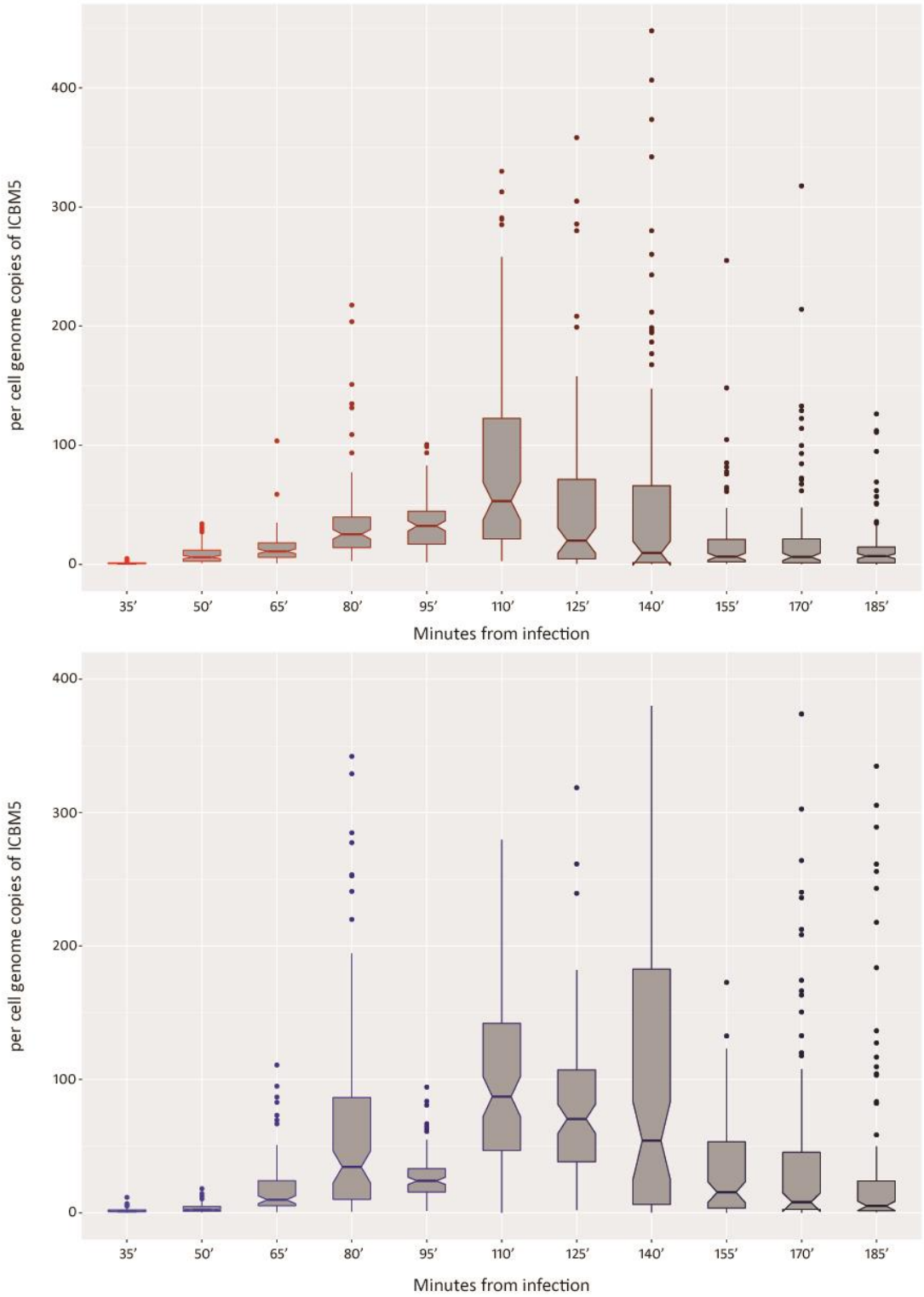

Fig. 19: The variation of the per-cell ICBM5 genome copies through the infection, as calculated from measuring the phage signal intensities using phage targeted direct-geneFISH. The box plot borders represent the 1st and the 3rd quartile, and the middle line represents the 2nd quartile. The whiskers extend from the 1st or 3rd quartiles with 1.5 * interquartile range (distance between first and third quartiles). The data beyond the whiskers are plotted individually. The plot was generated using the ggplot2 R package (Wickham 2016).The different shades of red for the box plots are indicating the progress of the infection time, from bright red at the beginning of the experiment, to dark red at the end of the experiment. The second plot shows the replicated experiment in blue colours.

3.9 Phage ICBM5 shows the potential for a carrier life infection strategy

To further investigate the potential for an alternative lifestyle of phage ICBM5, data generated by phage targeted direct-geneFISH were additionally consulted and searched for hints towards this direction. It was of particular interest, that the fraction of infected cells reached a maximum point at 80 min post-infection. Afterwards, the percentage of infected cells progressively declined until the end of the experiment. These findings from phage targeted direct-geneFISH - combined with the preliminary results that ICBM5 is forming turbid clearing spots during PFU assays and that positive cultures do not form clearing zones - have led to the assumption that a resistant *Sulfitobacter dubius* SH24-1b sub strain has emerged. In addition, long-read sequencing of the ICBM5 positive strain is also signalling the potential for an alternative infection strategy. To analyse a temperate lifestyle, phage targeted direct-geneFISH was also performed on the resistant culture and it could be shown that the ICBM5 genome was present in about 4.5% of host cells (see Fig. 20). However, the phage targeted direct-geneFISH signal varied in size and shape in between the cells and in some infected cells, small dot-like signals were observed (see Fig. 21). This is usually an indication for a low number of ICBM5 genome copies as seen during early stages of infection. Some other cells have shown larger and sometimes diffuse signals. This would indicate a higher number of ICBM5 genome copies as seen in later stages of infection. However, no cell lysis events were recognized during these experiments. Taken together with the findings from the preliminary studies, the nanopore sequencing, as well as the different phage targeted direct-geneFISH results, phage ICBM5 does not seem to undergo a lysogenic lifestyle as integrated prophage in the tested conditions. Nevertheless, phage ICBM5 can live and replicate within a sub-population of sensitive *Sulfitobacter dubius* SH24-1b cells coexisting with a dominant ICBM5-resistant sub strain. These findings taken together indicate that ICBM5 has a carrier state infection strategy, in addition to its lytic infection strategy.

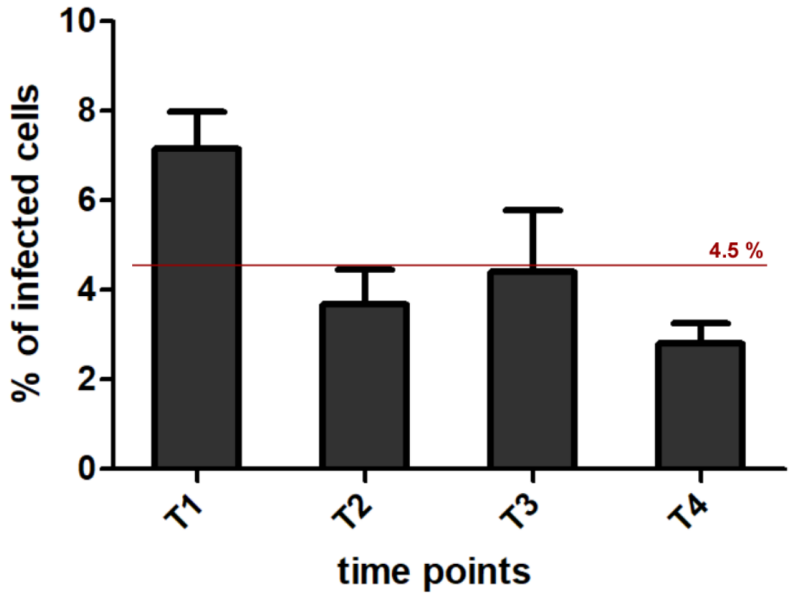

Fig. 20: Percentage of infected cells over time represented as bar diagram. T1 represents an OD of ~ 0.3, early exponential growth. From there on samples were collected every 60 min. For each time point, ten images have been taken and the total cell count as well as the percentage of infected cells have been determined (cell numbers ranged from 29 to 173 per image). The mean value for the percentage of infected cells is 4.5 % for all time points, as indicated by the red line.

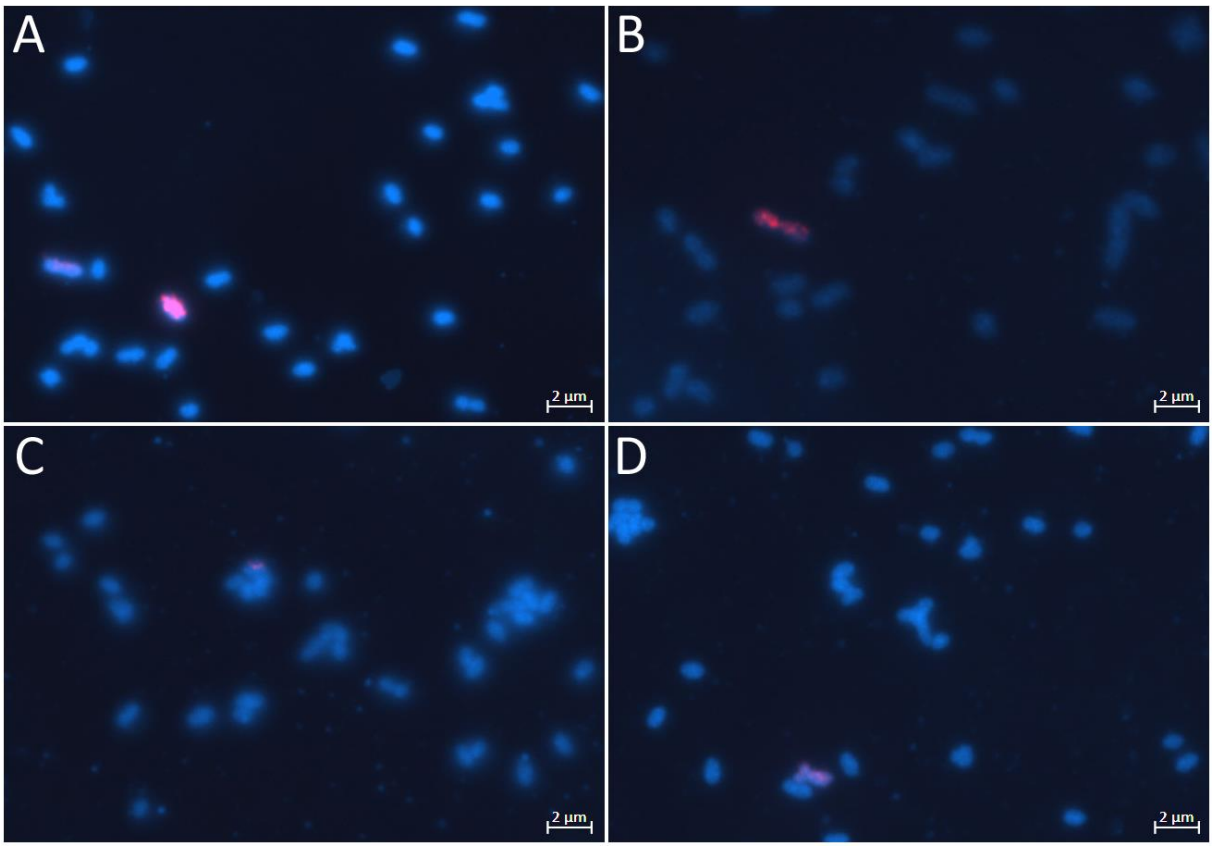

Fig. 21: Visualization of infected cells using phage targeted direct-geneFISH showing different stages in accordance to the time points in figure 20. Image A) shows the infected culture during early exponential growth at T1, B) shows the infected culture 60 min later at T2, C) shows the infected culture 120 min later at T3 and D) shows the infected culture 180 min later at T4. Only very few cells are infected and the phage signal varies in size and intensity.

4. Results Objectives 2 and 3 - ICBM5-like prophages and phylogenetic diversity

Graphical abstract:

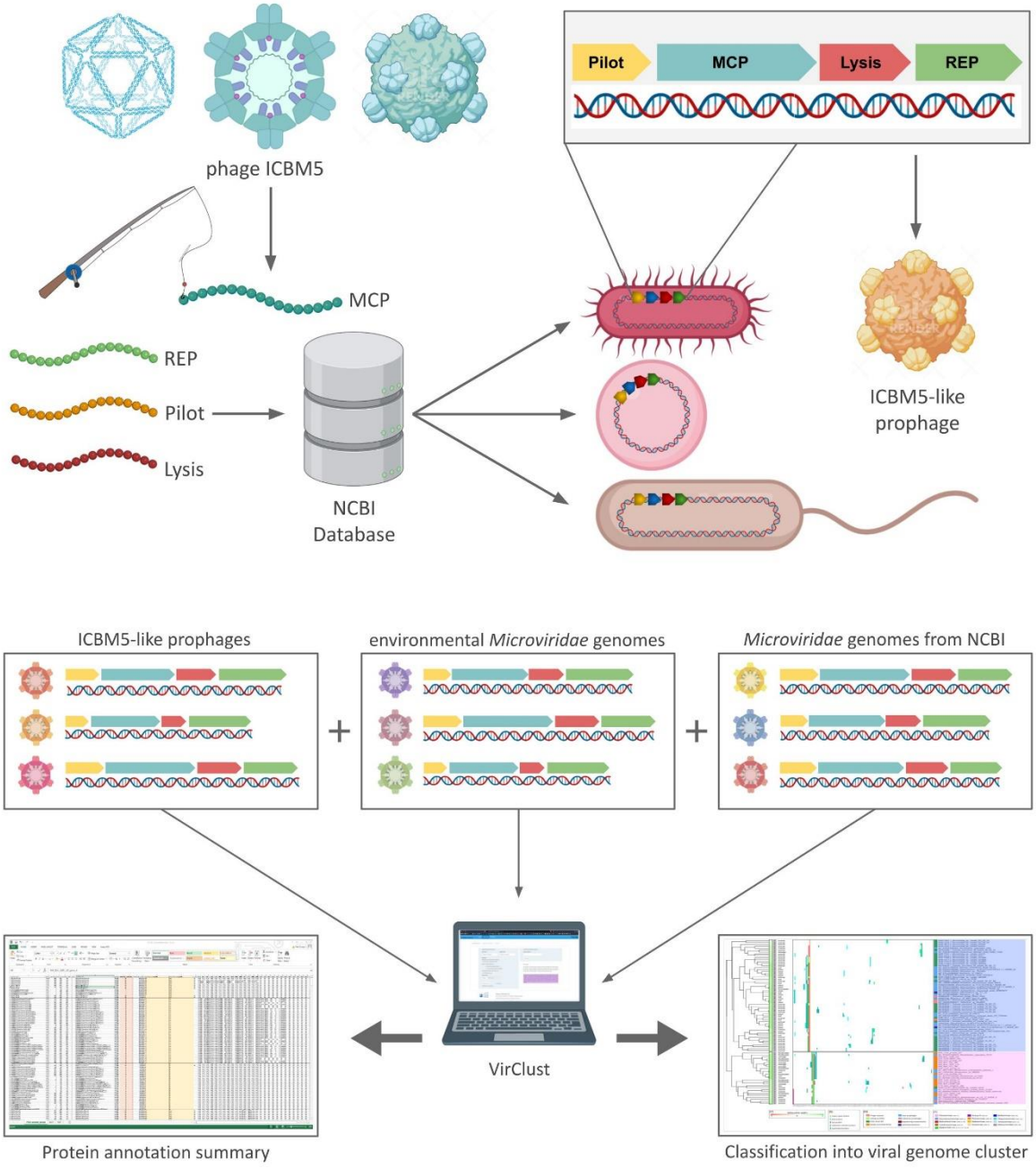

Fig. 22: Schematic overview of prophage mining for ICBM5-like sequences. ICBM5 protein sequences were used as bait against the NCBI online database and several hits were found showing similar entries to be integrated within bacterial chromosomes from different phyla. These prophage sequences have been used for further classification analysis performed with VirClust. The image made with BioRender (https://biorender.com/).

4.1 Using ICBM5 protein sequences as bait revealed similar phages

To get a better understanding of the distribution of ICBM5-related phages in bacterial hosts and environmental samples, as well as their phylogenetic classification, a search for similar genomes was conducted in two publicly available data sources: bacterial genomic data and viral metagenomes. First, protein sequences of phage ICBM5 were used to identify potential prophages within prokaryotic genomes (see Fig. 23). By applying this procedure, 72 ICBM5-like genomes could be retrieved and they mainly occurred in bacterial chromosomes or plasmids. Out of these sequences, seven have been previously described (Krupovic and Forterre 2011; Quaiser et al. 2015; Zheng et al. 2018), while the remaining 65 *Microviridae*-like sequences were completely new to this study (see Fig. 23). Furthermore, for 39 it was possible to determine the borders by performing whole genome alignments against related bacterial strains that were free from these regions. For the other sequences, the borders have been narrowed down by only keeping proteins with a clear phage origin or being surrounded by plain phage genes (see materials and methods). Additionally, the length of the newly found regions demonstrated variation between 3.3 and 6.6 kilo bases (kbs) for prophages with clearly identified border regions and between 3.5 and 8.2 kbs for regions where borders could not be predicted. When investigating the different host taxa, it became clear that the newly detected prophages are associated with a variety of different hosts. The majority of them are infecting members of the class *Alphaproteobacteria* (53.5%), followed by *Bacteroidia* (29.5%) and *Gammaproteobacteria* (5.6%). The remaining viruses were infecting members of the classes *Bacilli*, *Clostridia*, *Erysipelotrichia*, *Negativicutes, Cyanophycea* as well as *Flavobacteriia* (see Fig. 24). Three of the ICBM5-like prophages were found to be localized on bacterial plasmids, ranging in size from 0.1 Mb to 1.6 Mb. Seven prophage sequences consisted of small contigs with a similar size as the ICBM5 genome, probably representing episomes from a carrier state life strategy (see Fig. 23). The remaining majority of 62 ICBM5-like genomic regions were found to be integrated into bacterial chromosomes, displaying the potential for a lysogenic lifestyle.

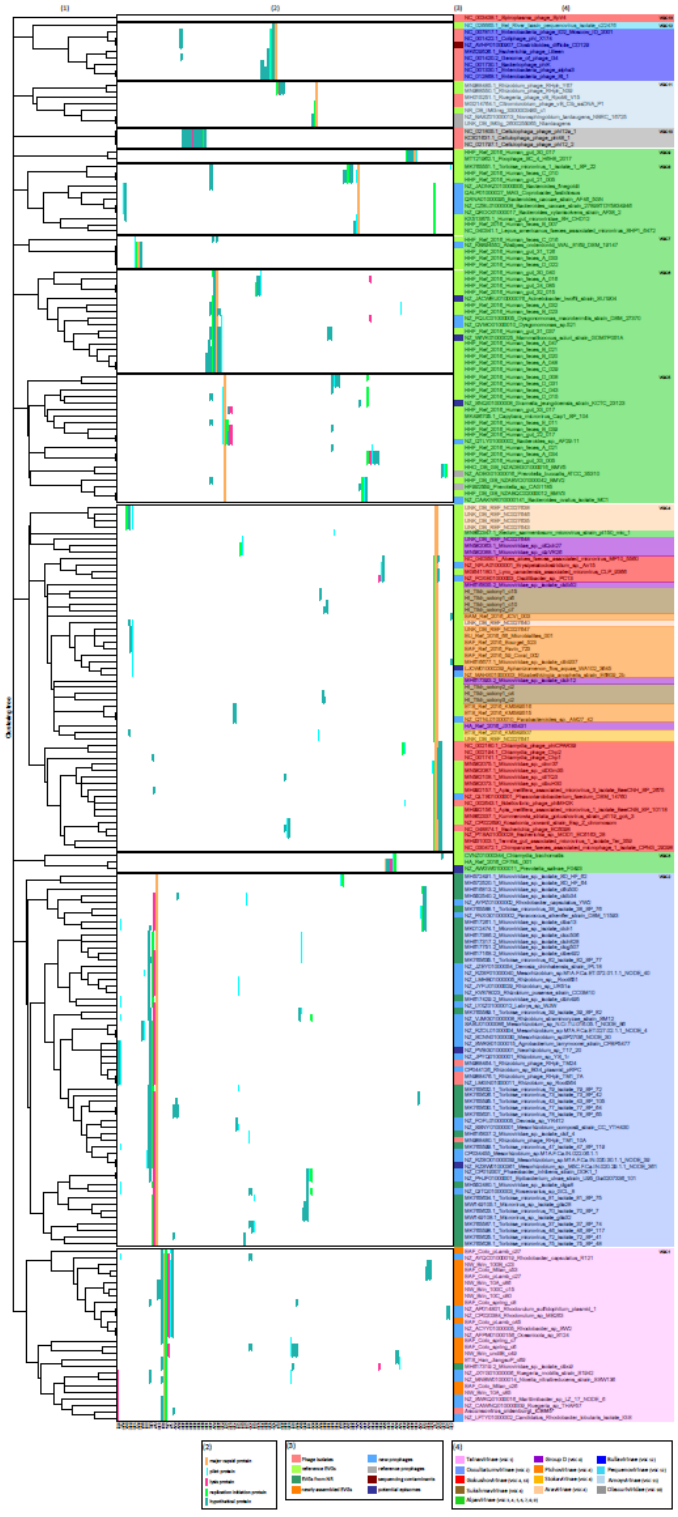

Fig. 23: Hierarchical clustering of the ssDNA (pro)-phage genomes, based on their protein super-supercluster content. The annotations show the following: 1) Hierarchical clustering tree; 2) Distribution of the protein clusters (PSSCs) in each viral genome. Protein clusters not shared with any other phage genomes in this dataset are not shown. The colour encodes different protein annotations; 3) Phage genome category. EVGs = environmental viral genomes; 4) the label given to each phage genome, consisting of accession numbers and names of the phage isolate, or the environmental contig or of the bacterial host in which a (pro)-phage was predicted. Figure was produced in collaboration (Zucker et al. 2022). For a higher resolution image, open file externally.

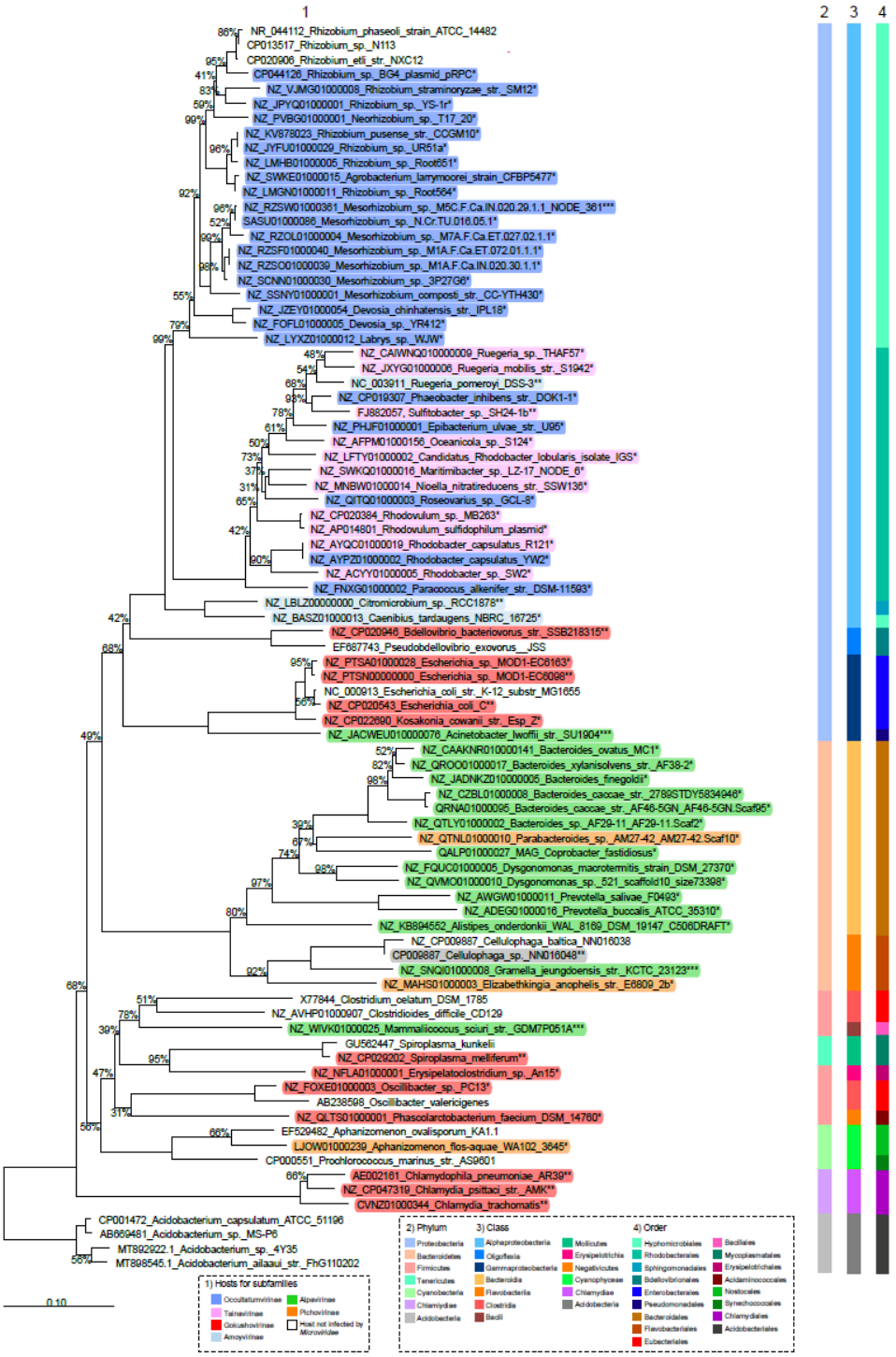

Fig. 24: Neighbour-joining tree based on the 16S rRNA gene sequence similarity showing the position of *Sulfitobacter dubius* SH24-1b and other bacterial hosts for *Microviridae*-like (pro)-phages. The 16S rRNA gene was not detected for four bacterial strains in Figure 23 and therefore not included in this tree. Bootstrap values are derived from 1000 replicates. GeneBank accession numbers are given as prefixes, followed by species and strain names. The bar represents 10 substitutions per nucleotide position. The stars encode the following: * hosts of predicted prophages, integrated into chromosomes or plasmids; ** hosts of isolated phages; *** hosts of predicted episomes, represented by short contigs. Figure was produced in collaboration (Zucker et al. 2022). For a higher resolution image, open file externally.

4.2 Expanding the dataset and clustering of new phage sequences by VirClust

To further extend the dataset of new potential prophages, another search for ICBM5-related genomes was performed. This search was conducted amongst environmental viral genomes (EVGs) which have been sequenced from environmental samples and can be found in NR and public viromes. While searching the NR NCBI database, 54 additional environmental phage genomes were retrieved in total and 23 of them were already affiliated with known *Microviridae* subfamilies. These viral genomes were generated by using ten different viromes that are associated with humans, animals or plants (see Fig. 23) and their size ranged between 4.2 and 6.6 kbs. Furthermore, the MCP from phage ICBM5 was used to additionally screen 2.944 previously published viral metagenomes. However, these metagenomes were only available as raw reads and had to be newly assembled within this study. From this analysis, a total of 15 circular contigs were retrieved from eight different viromes, obtained from fresh- or reclaimed water or from terrestrial systems (see Fig. 23). Each of the contigs represented a potential full-length viral genome. For a better characterization regarding the taxonomical classification of the retrieved viral sequences, all newly found prophages, episomes as well as EVGs were pooled together with ICBM5 and other reference *Microviridae* genomes and eventually compared in terms of gene content using VirClust. It could be demonstrated that their proteins are indeed related to each other and form clusters with proteins from cultivated and uncultivated reference *Microviridae* genomes. The majority of genomes found within this study have shown the common inventory of *Microviridae* proteins: pilot, MCP, lysis and Rep proteins, with interspersed hypothetical proteins. The hierarchical clustering of the genome sequences based on their protein cluster content resulted in the formation of 13 major viral genome clusters (VGCs) (see Fig. 23). It was observed, that for each VGC a distinctive set of protein clusters has been identified while a few protein clusters are shared in between the different VGCs. Members of the subfamily *Obscuriviridae,* which are also ssDNA phages with icosahedral capsids, did not share protein clusters with the rest of the dataset and formed their own VGC. In addition, to generate a better understanding of the relationship between the viral sequences, an alternative phylogenetic analysis was performed using the MCP and Rep proteins. Findings from this analysis have been coherent with the VirClust analysis, as the 13 genome clusters mostly corresponded to the major phylogenetic clades (see Fig. 23 and 25).

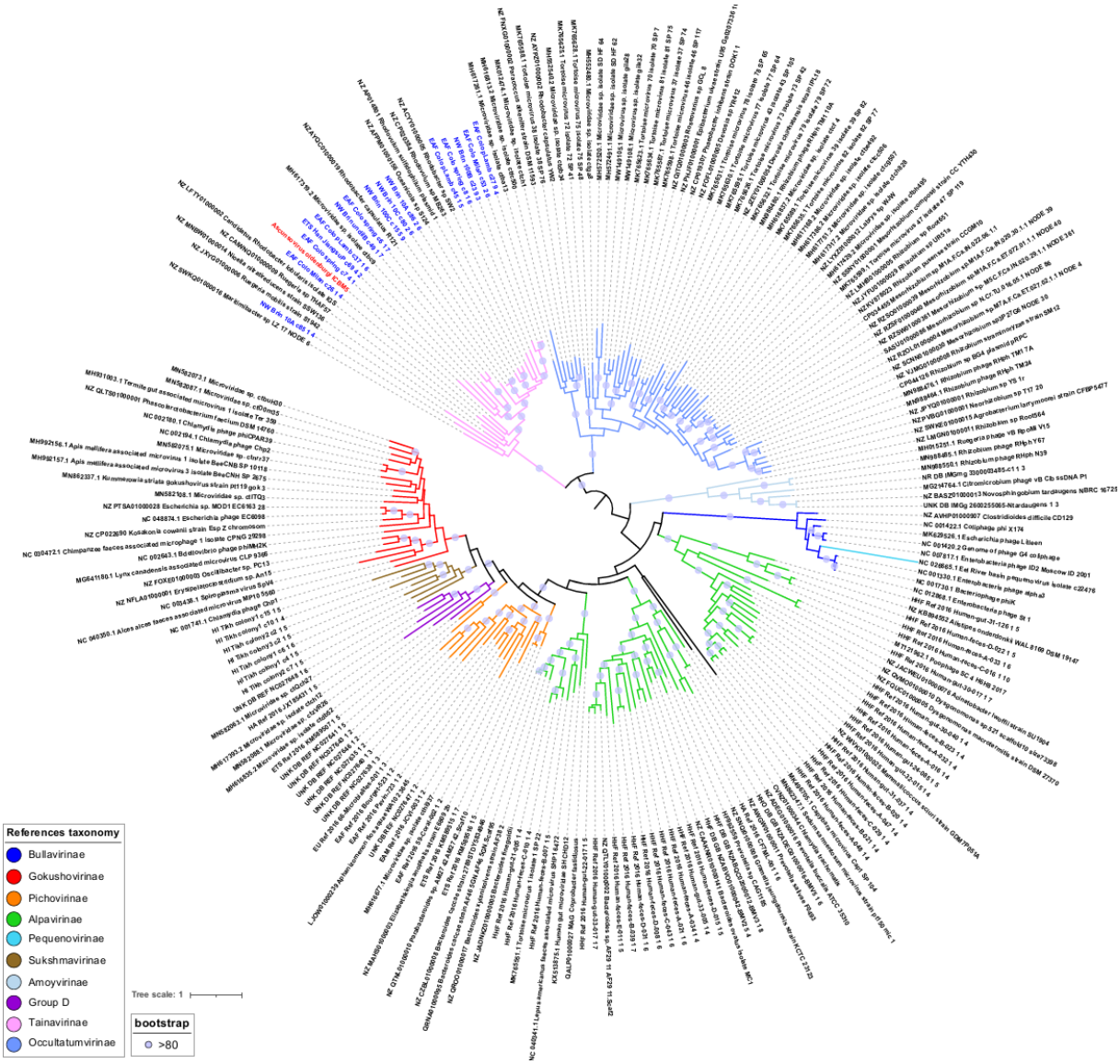

Fig. 25: MCP and Rep based phylogeny. Phylogeny on the concatenated major capsid and replication proteins of the ssDNA (pro)-phage genomes. Branches are coloured according to the taxonomic affiliation into subfamilies. Clear blue circles are placed on internal branches when bootstrap values are over 80. ICBM5 is highlighted in red, while closely related virome contigs are in blue (BLASTp on major capsid proteins, bit-score>300). Figure was produced in collaboration (Zucker et al. 2022). For a higher resolution image, open file externally.

4.3 Two novel subfamilies have been proposed within the *Microviridae*

Both in the hierarchical clustering generated by VirClust as well as in the MCP-Rep phylogenetic tree, phage ICBM5 was found with less than half of the newly predicted prophages, episomes and EVGs in a group which clustered separately from previously defined *Microviridae* subfamilies (see Fig 25). This group was further split into two different phylogenetic clades within the phylogenetic tree and both groups corresponded to VGC1 and VGC2 from the VirClust analysis, respectively (see Fig. 23). The two clades were perfectly

identical in both approaches, the smaller cluster VGC1 contained 27 viral genome sequences, while the bigger group VGC2 is composed of 60 viral genome sequences. Both groups are composed of new prophages as well as of EVGs. The first group VGC1 harboured phage ICBM5, which was also the only cultivated phage in both groups. VGC1 also contained 10 newly identified prophage sequences and all the 15 EVGs newly assembled in this study, coming from eight viromes that were generated in three different previous studies (Colombo et al. 2017; Han et al. 2017; Brinkman et al. 2018). Only one phage sequence within VGC1 (MH617319.2_Microviridae_sp._isolate_ctbc9) has been previously identified as *Microviridae* but was not further classified (Tisza et al. 2020). The larger of the two groups VGC2 is composed of 25 new prophages, 30 EVGs, two episomes and three recently isolated phages that are infecting *Rhizobium* strains (van Cauwenberghe et al. 2021). All of the EVGs that occurred in this group have been previously identified as *Microviridae* and are part of nine different viromes belonging to four recently published studies (Creasy et al. 2018; Orton et al. 2020; Tisza et al. 2020; Collins et al. 2021). Within two of the studies, the respective EVGs were already recognized to represent a group separate from any of the known *Microviridae* subfamilies (Creasy et al. 2018; Orton et al. 2020). However, no further classification of these phages has been performed. Taken into consideration that the two VGCs were generated, undoubtedly separate from each other, and are even more distantly related to known *Microviridae* subfamilies, two new subfamilies have been provisionally proposed: 1) “Tainavirinae” (from the Romanian word “taina” which means secret), representing the clade with phage ICBM5 and 2) “Occultatumvirinae” (from the Latin word “occultatum”, which means hidden), representing the clade with the *Rhizobium* phages. In addition, it was detected that the genomic diversity within the two new subfamilies was very high, with nucleotide-based intergenomic identities ranging between 0.0% and 99.9%. For the majority of phage pairs, an intergenomic identity lower than 40% was shown. Only a few phages showed an intergenomic identity above 95%, which would qualify them to be considered as the same species: five EVGs into two species in the “Tainavirinae” and two *Rhizobium* phages into one species within the “Occultatumvirinae” (see Fig. 26).

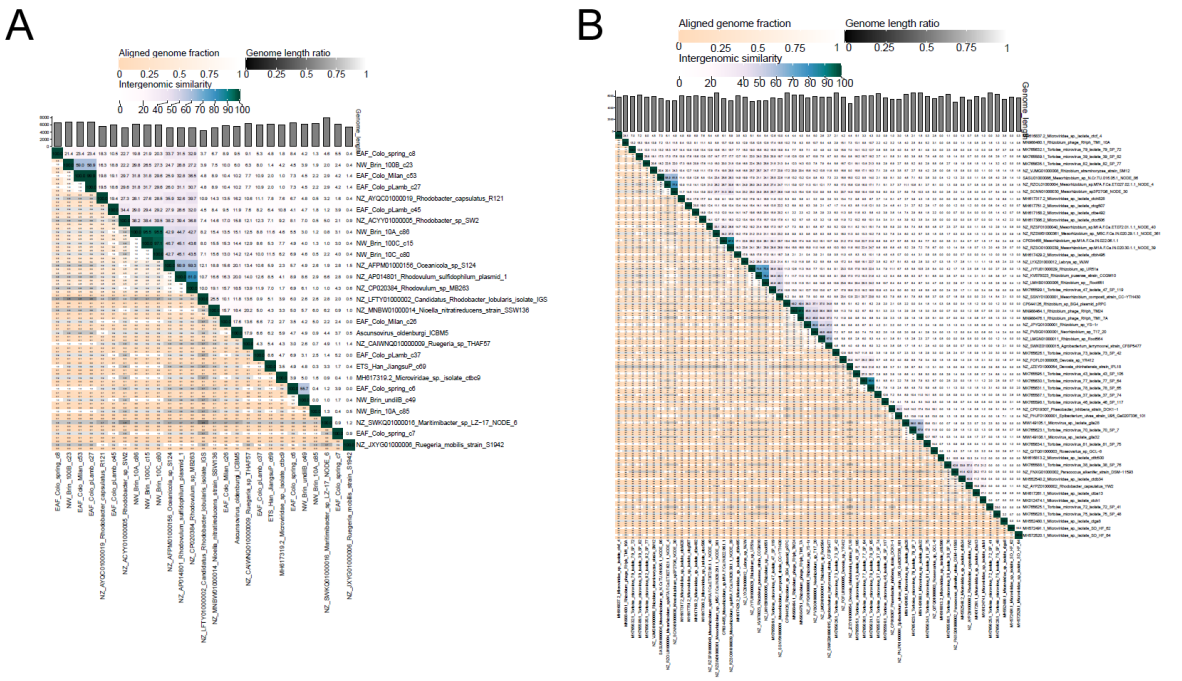

Fig. 26: Overview of the nucleotide-based intergenomic similarities between viruses of the proposed subfamily: A) "Tainavirinae" and B) "Occultatumvirinae" calculated with VIRIDIC (Moraru et al. 2020). For a higher resolution image, open file A externally, or open file B externally.

4.4 Biogeographical distribution and environmental occurrences of ICBM5-like phages

To investigate the biogeographical distribution of the newly classified two subfamilies, as well as to identify their environmental background, the individual isolation origin was retrieved from the respective databases for each of the viral sequences. From there, it was observed, that the two subfamilies comprised phages that are present in different environments and are spread worldwide (see Fig. 27), while only infecting members of the class *Alphaproteobacteria* (see Fig. 24). Furthermore, it was shown that tainavirus prophages, as well as phage ICBM5, have only been found in hosts from the order *Rhodobacterales*. The majority of occultatumvirus prophages, episomes, and cultivated phages infect *Hyphomicrobiales*, while the remaining viruses infect hosts from the related bacterial order *Rhodobacterales.* An overview of the different habitats showed that these viruses and EVGs from the two subfamilies, as well as their respective hosts, were usually found in terrestrial and marine environments, often in association with plants and animals. In addition, occultatumvirus EVGs were found in the lizard *Heloderma suspectum* (Collins et al. 2021), the fishes *Carassius carassius*, *Lutjanus campechanus* and *Pimephales* sp., the snail *Haliotis sp.* (Tisza et al. 2020), the tortoise *Gopherus morafkai* (Orton et al. 2020) and the sea squirt *Ciona*

robusta (Creasy et al. 2018). Moreover, each of the *Hyphomicrobiales* which was infected by an "Occultatumvirinae" member has been isolated from soil or plant root nodules. The *Rhodobacteraceae* infected by occultatumviruses were isolated either from soil or marine algae and seawater. Tainavirus EVGs were detected in viromes from paddy soils (Han et al. 2017), freshwater rivers (Colombo et al. 2017), and wastewater (Brinkman et al. 2018). Finally, the remaining *Rhodobacteraceae* hosts infected by tainaviruses were isolated from terrestrial environments, including ponds, sediments, and soil as well as marine environments, including sediments, water columns, sponges, copepods and dinoflagellates.

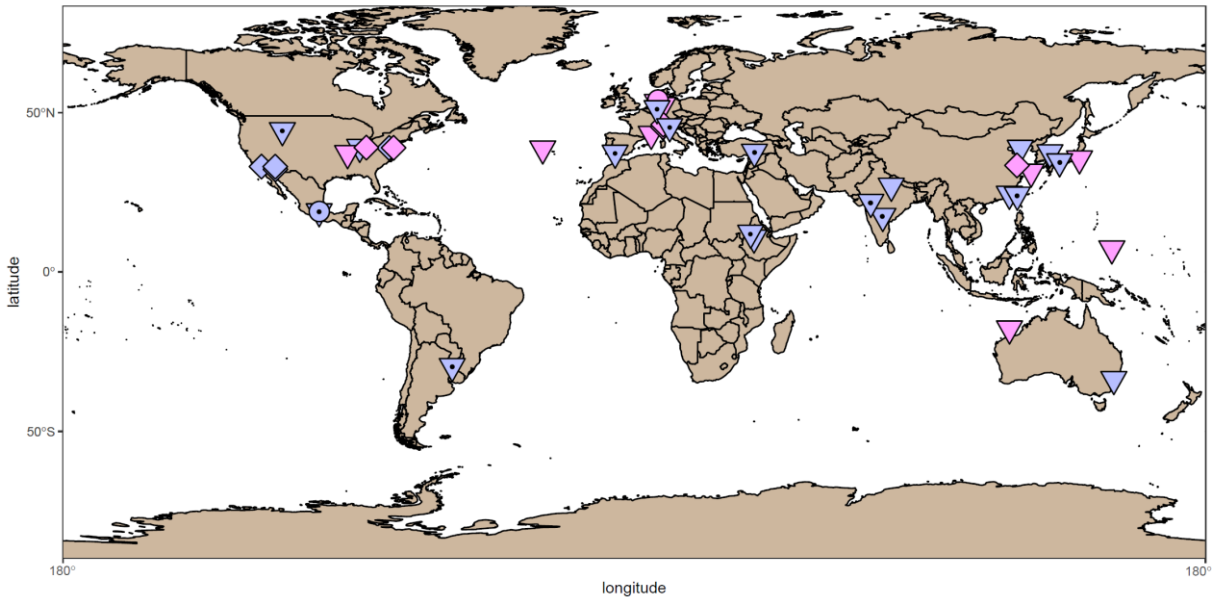

Fig. 27: Biogeographical distribution of tainaviruses (pink) and occultatumviruses (lavender). The points on the map represent the sampling place for phage isolates (circles), for bacterial hosts harbouring prophages and episomes (triangles), and for the viromes from which the EVGs have been assembled (squares). The triangles with a central circle represent locations for *Rhizobium* and *Mesorhizobium* infected by occultatumviruses.

5. Results Objective 4 - Additional genomic features of phage ICBM5

Graphical abstract:

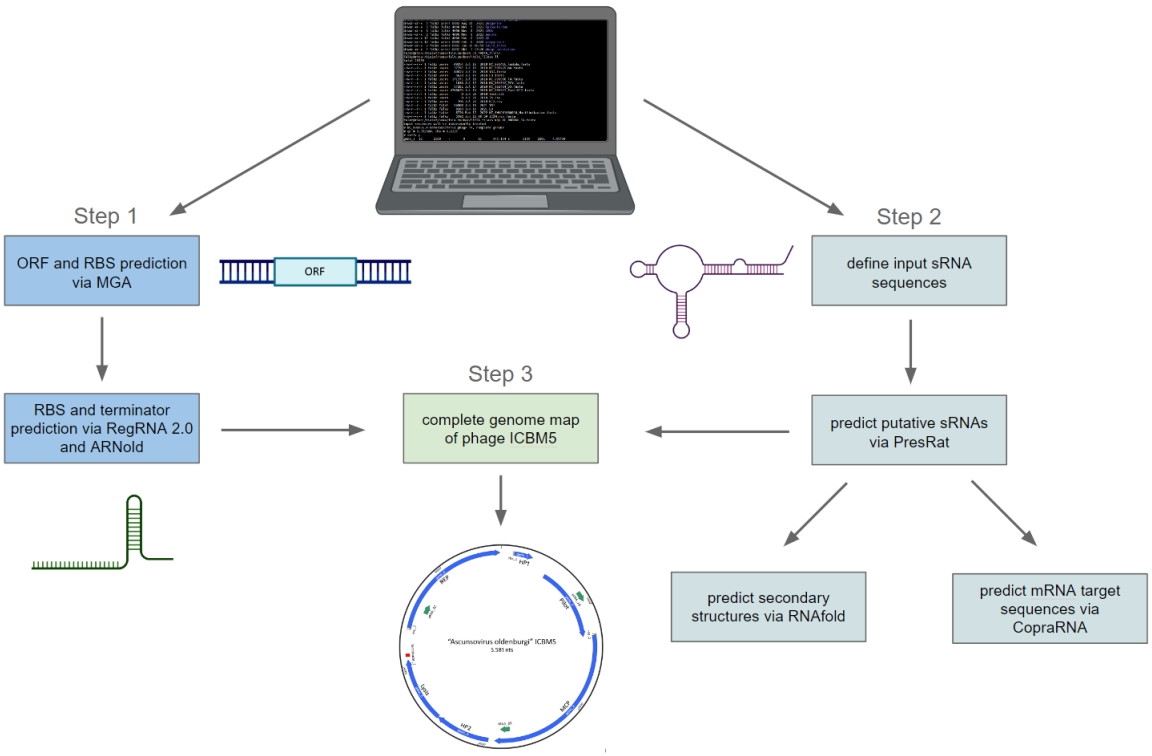

Fig. 28: Schematic overview of the workflow showing the different steps for the prediction of genomic features of phage ICBM5 and for completing its genomic map. The image was made with BioRender (https://biorender.com/).

5.1 Additional genomic and transcriptomic features on the ICBM5 genome

To investigate additional genomic and transcriptomic elements of phage ICBM5, further bioinformatical analysis was performed. To search for additional ORFs that could be encoded on the reverse complemented strand, ORFs have been predicted using MGA incorporated in VirClust (see Tab. 4). By this approach no additional protein translating sequences could be found. Furthermore, for translation to be working, the ribosome needs to be able to interact with a certain region on the mRNA. These regions, also known as ribosomal binding sites (RBSs), have been predicted using MGA implemented in VirClust (see Tab. 5) and also by using RegRNA 2.0 (see Fig. 29). MGA found three different RBSs, the first was located in front of gene_1 which translates for a hypothetical protein. The second RBS was predicted shortly before the MCP gene and the third RBS predicted by MGA was detected in front of the Rep gene (see Fig. 30).

Tab. 4: Predicted ORFs by MGA, showing start and end position as well as strand orientation.

gene ID	start pos.	end pos.	strand
gene_1	127	306	+
gene_2	484	1290	+
gene_3	1284	2855	+
gene_4	2913	3440	+
gene_5	3437	4063	+
gene_6	4364	5572	+

Tab. 5: Predicted ORFs by MGA, showing gene score, as well as location for ribosomal binding sites and their respective score. Three RBS could be identified, before, gene_1, gene_3 and gene_6.

gene ID	gene score	rbs start	rbs end	rbs score
gene_1	11.6717	114	119	2.8478
gene_2	93.0185	-	-	-
gene_3	177.291	1272	1277	-1.93578
gene_4	46.5827	-	-	-
gene_5	55.6991	-	-	-
gene_6	74.3979	4350	4355	0.239135

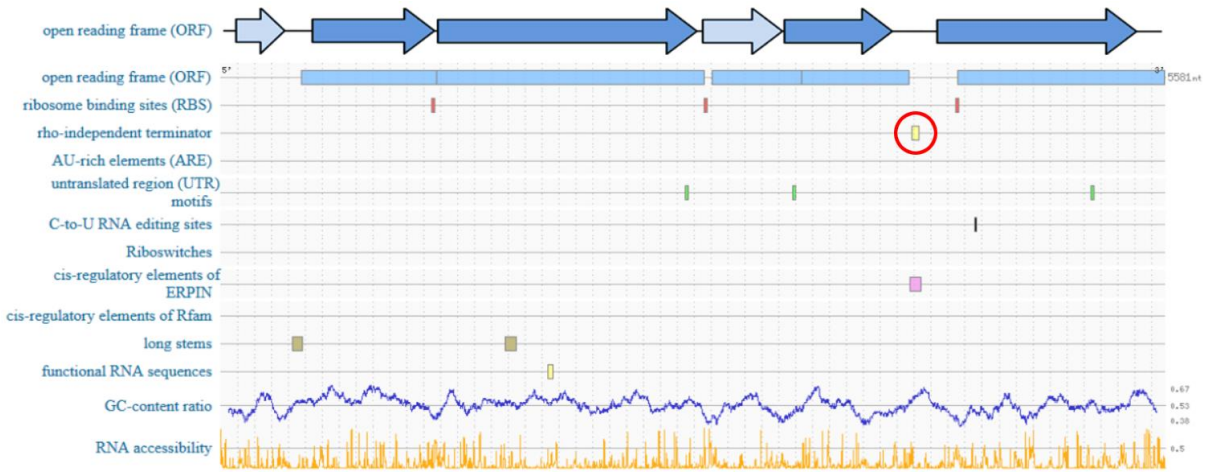

Fig. 29: Prediction of RNA regulatory elements on the phage ICBM5 genome generated by RegRNA 2.0. The first ORF line represents the previous prediction done by MGA and serves as comparison. A putative rho-independent terminator was found at the end of the lysis gene (red circle).

Further analysis regarding RBSs has been done in RegRNA 2.0, which also found RBS regions in front of the MCP and Rep gene. RegRNA 2.0 did not find the first ORF and therefore also not the associated RBS, but predicted an additional RBS in front of gene_4 which encodes for a hypothetical protein. Rho-independent terminators were predicted using RegRNA 2.0 and ARNold (see Fig. 29 and Fig. 31). Both tools predicted possible transcription terminator sequences closely behind the lysis gene, which encodes for the lysis protein. Apart from this

predicted regulatory element which was found by both tools, ARNold predicted five additional transcriptional terminators on the (-) strand and one additional terminator located inside the lysis gene. However, these additional elements on the (-) strand are probably false detections, since no ORFs could be determined on this part of the genome. The terminator inside the lysis gene could work to regulate its activity.

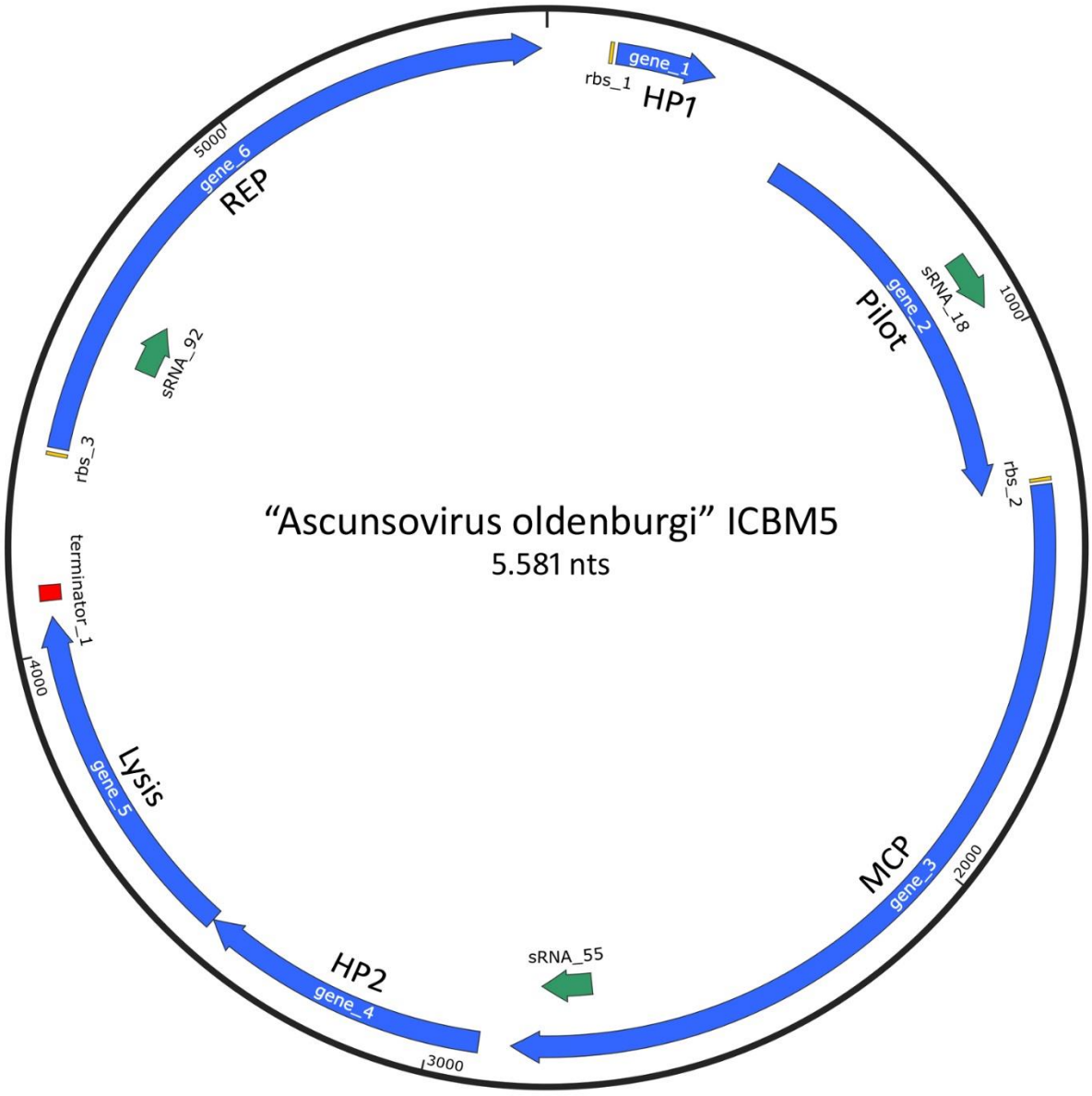

Fig. 30: Overview of the genomic features of phage ICBM5, including the position of the new regulatory elements predicted in this work. ORFs are coloured in blue. Predicted RBSs are coloured in yellow and located before HP1, MCP and REP. The predicted Rho-independent terminator is coloured in red and located behind lysis. The putative sRNAs are coloured in green. The image was created with SnapGene (http://snapgene.com)

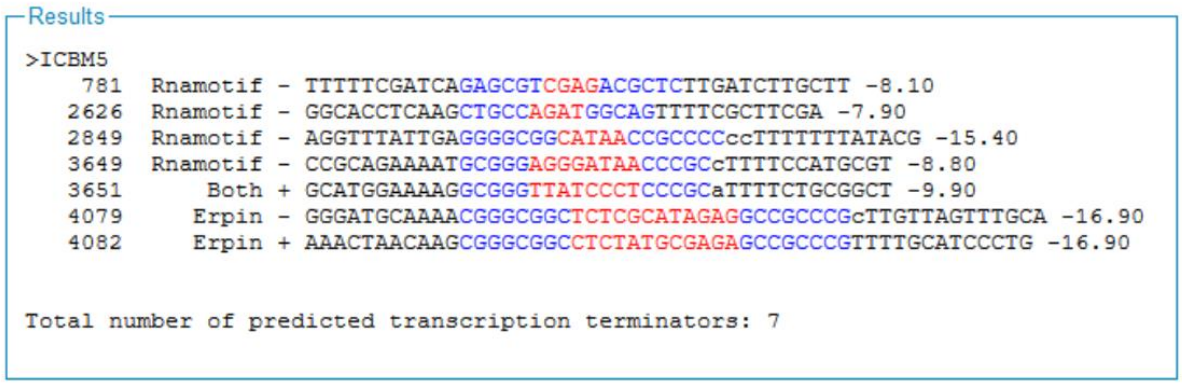

Fig. 31: Overview of the prediction terminator sequences generated by ARNold. The first column shows the position on the genome, the second column indicates the type of regulator, the third column indicates if plus or minus strand, the fourth column shows the sequence and the region of interest is highlighted in different colours. The last column gives a probability score. By this approach, five possible terminators have been found on the negative strand, while two are found on the plus strand.

5.2 Putative RNA based regulatory elements on the phage ICBM5 genome

In addition, it was of particular interest to search for phage-encoded sRNAs on the genome of phage ICBM5. Therefore, a randomised set of small DNA fragments has been selected from the genome, containing 112 sRNA candidates in total, and each of them was processed via PresRAT. From this prediction analysis, 11 candidates were selected according to their probability profiles. These putative sRNA sequences were further investigated for their interaction capacity with potential mRNA targets by using CopraRNA as well as presRAT. From the 11 previously selected sRNA candidates, three showed promising interaction potential with mRNA targets from the queried host genomes. The three final candidates and their presRAT results are shown in Fig. 32, 33, and 34. It was checked, if these genes also exist on *Sulfitobacter dubius* SH24-1b, to make sure the interaction could indeed take place on this particular host (see Tab. 6). The first sRNA target (sRNA_18) that fulfils the described criteria is the *murF* gene, which encodes for a UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase. The second hit (sRNA_55) is an unnamed iron-sulphur cluster assembly accessory protein. The last putative interaction target is the *msrA* gene, which encodes for the protein peptide-methionine (S)-S-oxide reductase.

Tab. 6: Overview of the predicted sRNAs, their location on the phage ICBM5 genome, their associated predicted target host gene as well as their probability score indicated as p-value.

sRNA name	Location	Target host gene	presRAT p-value
sRNA_18	gene_2	*murF*	0
sRNA_55	gene_3	Iron sulphur cluster assembly	0
sRNA_92	gene_6	*msrA*	0

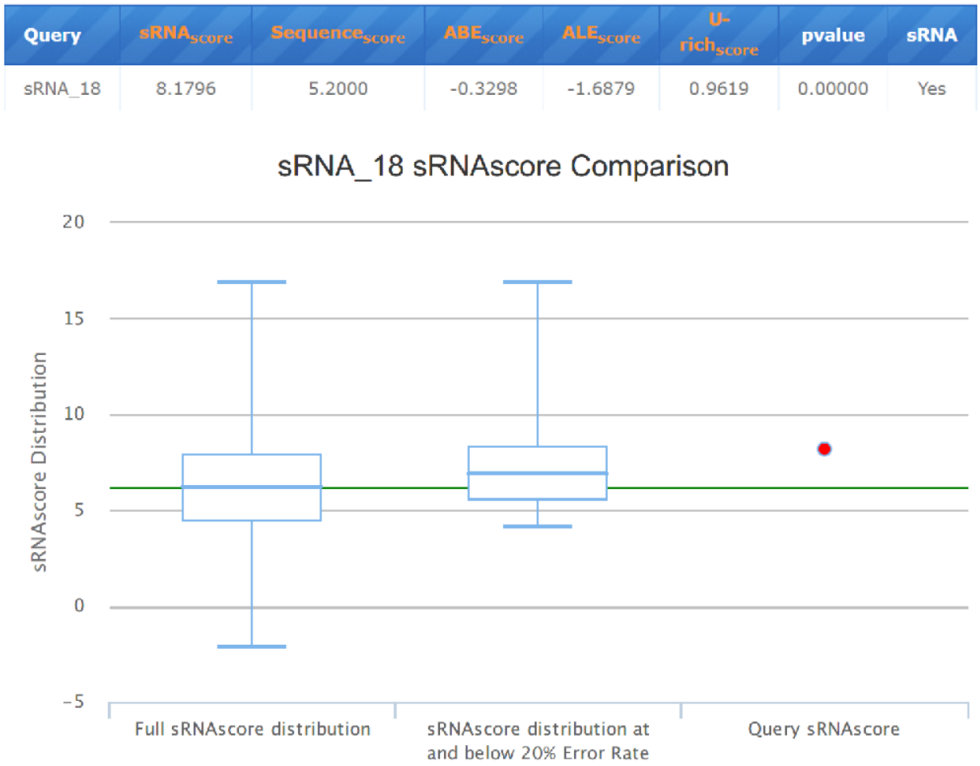

Fig. 32: PresRAT result for sRNA_18, the query sRNA (red dot) scores higher than both mean values generated by known sRNAs indicated by the boxplot graph.

Query	$sRNA_{score}$	$Sequence_{score}$	ABE_{score}	ALE_{score}	$U\text{-}rich_{score}$	pvalue	sRNA
sRNA_55	7.2281	4.5170	-0.3108	-1.6522	0.7481	0.00000	Yes

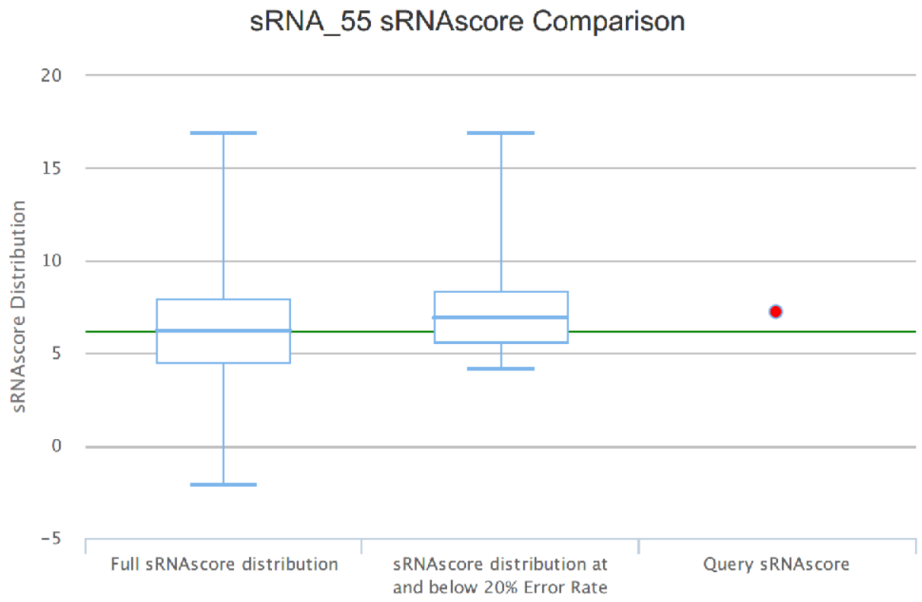

Fig. 33: PresRAT result for sRNA_55, the query sRNA (red dot) scores higher than both mean values generated by known sRNAs indicated by the boxplot graph.

Query	$sRNA_{score}$	$Sequence_{score}$	ABE_{score}	ALE_{score}	$U\text{-}rich_{score}$	pvalue	sRNA
sRNA_92	7.6146	4.0730	-0.2144	-1.8843	1.4429	0.00000	Yes

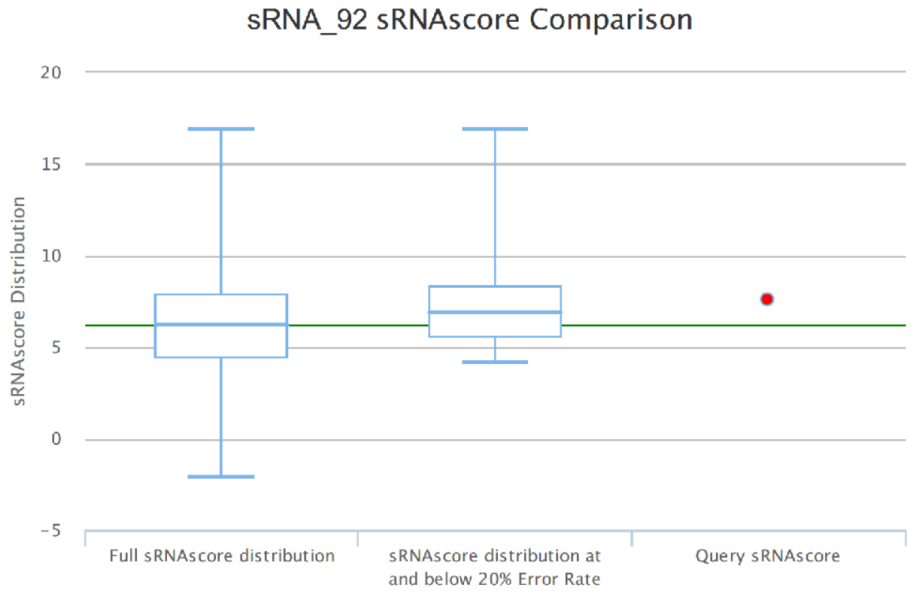

Fig. 34: PresRAT result for sRNA_92, the query sRNA (red dot) scores higher than both mean values generated by known sRNAs indicated by the boxplot graph.

Additionally, another secondary structure prediction was performed (presRAT already performs its own prediction) on all of the three sRNA candidates to check if they would form a stable conformation (see Fig. 35). The first candidate, sRNA_18, is located at the end of the pilot gene. The second putative sRNA (sRNA_55) can be found within the MCP gene while the third possible regulator (sRNA_92) is located inside the replication gene. Interestingly, none of the detected sRNA candidates are located in intergenomic regions (IGRs) or in close proximity to predicted rho-independent terminators, indicating that they would probably be co-transcribed with their associated genes. All three target genes were also found on the genome of *Sulfitobacter dubius* SH24-1b and the potential interaction regions with their respective sRNAs are shown in figure 36.

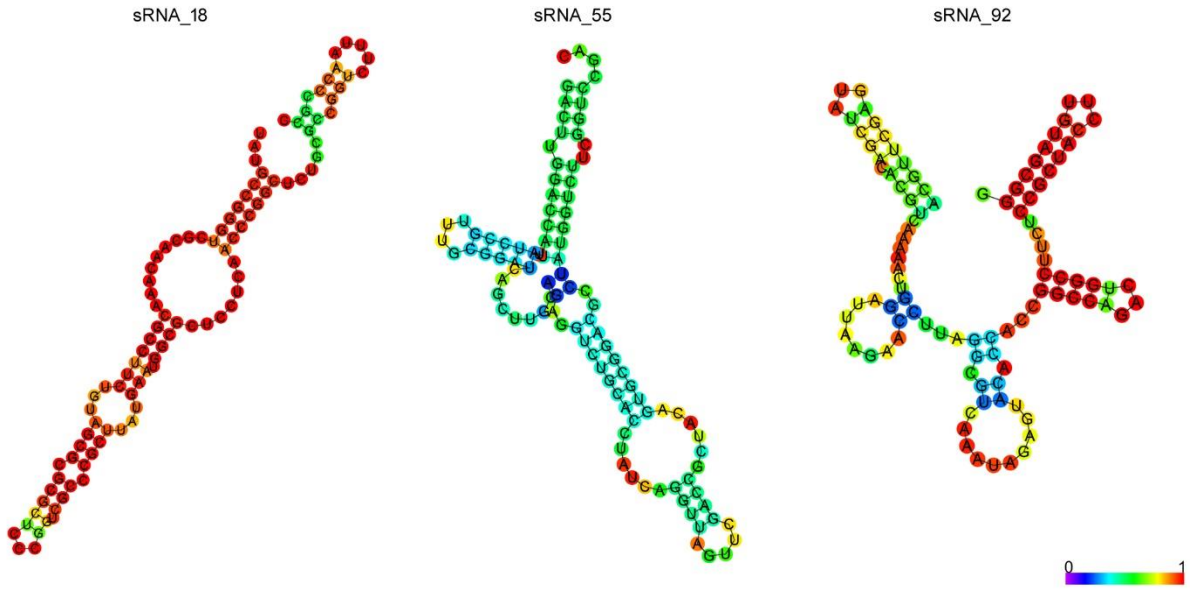

Fig. 35: Secondary structure prediction of three putative sRNA candidates, colours are indicating the base pair probability according to the legend in right bottom corner.

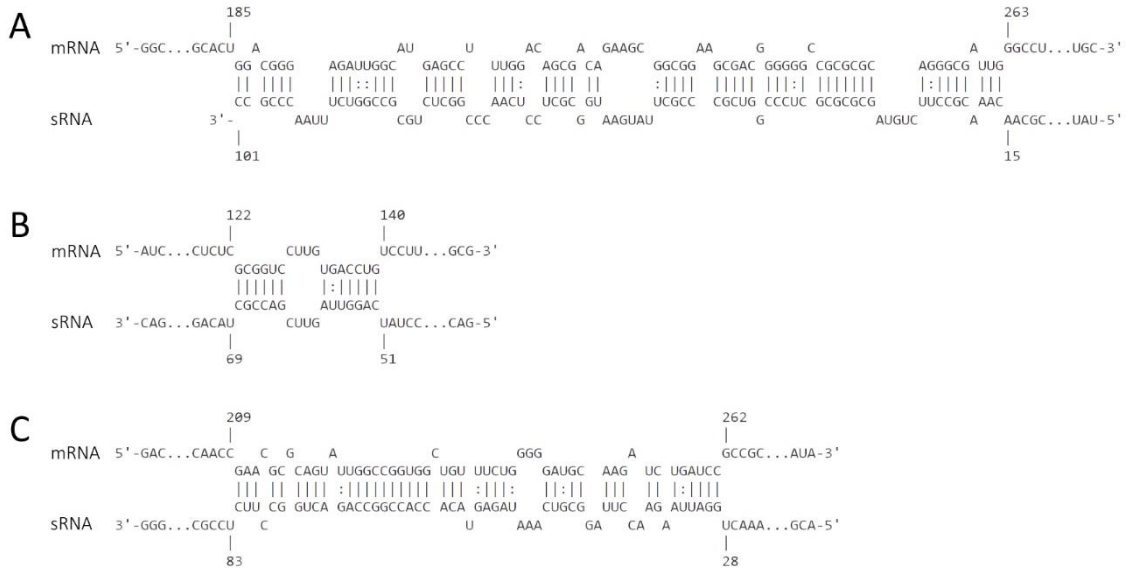

Fig. 36: Regions of the predicted sRNA – mRNA target interaction. A) Shows the putative interaction between sRNA_18 and the *murF* gene. B) Shows the putative interaction between sRNA_55 and the gene which translates for an iron sulphur cluster assembly protein. C) Shows the putative interaction between sRNA_92 and the *msrA* gene.

6. Discussion

6.1 Two new *Microviridae* subfamilies with alternative lifestyle strategies

Accumulating knowledge regarding the diversity of tailless, icosahedral ssDNA phages is still in progress as demonstrated by the constant sequencing of new viral genomes. As a consequence, their classification is regularly revised and it has become fairly clear, that this group of icosahedral ssDNA phages consists of several major clusters. At the moment of writing, these groups are defined as subfamilies – *Bullavirinae, Gokushovirinae,* "Pichovirinae", "Aravirinae", "Alpavirinae", "Stockavirinae", "Pequenovirinae", "Sukshmavirinae", and "Amoyvirinae" –, and all of them are placed under the umbrella of the *Microviridae* family (Creasy et al. 2018; Zheng et al. 2018). As indicated by the *italic* notation, only the first two subfamilies are currently accepted by the ICTV. The remaining groups are proposed subfamilies and therefore written with "quotation marks". Within this study, it could be shown, that the newly isolated phage falls far from other related reference genomes and thereby cannot be classified within existing groups. Furthermore, by using genomic information of this new phage isolate, two new major groups of *Microviridae* – "Tainavirinae" and "Occultatumvirinae" – could be determined. These proposed new subfamilies represent a substantial contribution to the known *Microviridae* diversity, as indicated by the MCP-Rep phylogeny, where the two groups are distant from all of the current subfamilies (see Fig. 25).

The collected data even enlarges our understanding of the *Microviridae* lifestyle. Until now, only a few microvirus-like prophages could be predicted via bioinformatical methods (Krupovic and Forterre 2011; Roux et al. 2012; Quaiser et al. 2015; Zheng et al. 2018; Kirchberger and Ochman 2020). Apart from these findings, not much is known about alternative infection strategies within this particular phage group. In contrast, the other major group of ssDNA phages, the *Inoviridae* family*,* have long been known as being capable to integrate their genome into the host chromosome as part of their chronic life cycle (Mai-Prochnow et al. 2015). Moreover, due to current findings from a large metagenomics study, an enormous diversity of inovirus-like prophages could be revealed which are distributed throughout many bacterial and archaeal taxa (Roux et al. 2019). In this study, by cultivating and sequencing phage ICBM5, the prediction of many more *Microviridae*-like prophages, as well as episomal structures, was possible. So far, the known host clades have been restricted to the *Bacteroidia* class, the *Enterobacteriaceae* family (*Gammaproteobacteria* class*),* and one

Hyphomicrobiales family (*Alphaproteobacteria* class) (Krupovic and Forterre 2011; Roux et al. 2012; Quaiser et al. 2015; Zheng et al. 2018). These groups could now be expanded with *Flavobacteriia*, as well as more *Alphaproteobacteria* and *Gammaproteobacteria.* In addition, several new *Firmicutes* classes were found to be infected by ICBM5-like microviruses, which makes this study the first report for the occurrence of *Microviridae* prophages in Gram-positive bacteria. It is also the first report of *Microviridae* prophages within the *Rhodobacteraceae* family, an environmentally significant clade of the marine ecosystem. A recent study reported the presence of such prophages in three different chromosomes of the *Rhodobacteraceae* family (Forcone et al. 2021), but after further analysis, it could be determined that these genomic elements are identical to ΦX174 and are putatively an unremoved residue from the sequencing process. The new prophages found in *Hyphomicrobiales* and *Rhodobacterales* are not related to the *Amoyvirinae* prophage which was previously identified in *Caenibius tardaugens (Hyphomicrobiales)* (Zheng et al. 2018). Correspondingly, they are also unrelated to the previously isolated *Ruegeria pomeroyi* phages, which according to our analysis, should be part of the *Amoyvirinae* subfamily.

6.2 Two new subfamilies and their occurrence in different environments

The two new subfamilies – “Occultatumvirinae” and “Tainavirinae” – are capable of infecting hosts from the order *Hyphomicrobiales* and *Rhodobacterales*. For the EVGs, the hosts remain unknown for now but due to their phylogenetic similarity to phage isolates and prophages, it can well be assumed that these viruses also infect the same group of prokaryotes. This assumption is additionally supported by the homogeneity of the known bacterial hosts in those groups. Representatives of these two bacterial orders have been found in habitats similar to those in which the EVGs were found, either in diverse aquatic systems (freshwater, marine or sewage), in terrestrial systems or associated with eukaryotes (animals, plants, or microalgae). For instance, *Rhodobacteraceae* has been detected in the gut of a *Ciona* species (Dishaw et al. 2014), indicating that the occultatumvirus EVGs found in *Ciona robusta* could also infect bacteria from the same family. Similarly, *Rhodobacteraceae* and *Hyphomicrobiales* were both found within the gut or on the skin of fish species that were closely related to those from which occultatumviruses were retrieved (DeBofsky et al. 2020; Nielsen et al. 2018; Tarnecki et al. 2022). The same accounts for the tortoise species *Gopherus morafkai,* which harbours both *Hyphomicrobiales* as well as *Rhodobacterales* within their

nasal biome (Weitzman et al. 2018), making members of the two orders very likely the hosts of the occultatumviruses that were found in the faecal viromes from similar tortoises. Additionally, bacteria from the family *Rhodobacteraceae* are distributed among several different ecosystems and are also common in freshwater rivers (Liu et al. 2019), in wastewater (Numberger et al. 2019), in paddy soils (Wang et al. 2018), in marine sediments and the water column (Hahnke et al. 2013). Here, *Rhodovulum* sp. MB263 as well as *Rhodovulum sulfidophilum*, which belong to the same species (Fig. 24) were both identified in terrestrial and in marine systems and their respective tainaviruses seem to be closely related (Fig. 23 and 25). Furthermore, members of the *Rhodobacteraceae* family are also known to have a free-living or associated lifestyle. For example, the host organism *Epibacterium ulvae (Breider et al. 2019)* is known to occur in association with macroalgae. It could well be, that the tainavirus EVGs also infect *Rhodobacteraceae* which live in diverse aquatic systems (freshwater, marine, sewage) but also in terrestrial systems. In regards to the occultatumviruses, they have been shown to infect bacteria from two different bacterial orders, *Rhodobacterales* and *Hyphomicrobiales,* which are closely related. Closer analysis of the MCP-Rep tree revealed that all occultatumviruses which infect *Hyphomicrobiales* form a monophyletic group and therefore seem to share a common ancestor (Fig. 25). Bacteria from the *Hyphomicrobiales* order, especially the *Rhizobium* and *Mesorhizobium* strains, are of special interest. These particular prokaryotes are most commonly found in soil which is in a state of nitrogen starvation. They are known to migrate into the root hairs of legumes, where they transform into nitrogen-fixing endosymbionts (Poole et al. 2018; Clúa et al. 2018). When taking into consideration that both wild and cultivable, economically important legumes are spread worldwide (Sprent et al. 2017), it can well be the case that occultatumviruses are cosmopolitan. Undoubtedly, the variable geographical locations in which occultatumvirus harbouring *Rhizobium, Neorhizobium* and *Mesorhizobium* were found (see Fig. 27) supports this hypothesis. To recapitulate, the two related new subfamilies “Occultatumvirinae” and “Tainavirinae” infect respectively *Rhodobacterales* and *Hyphomicrobiales* or only *Rhodobacterales.* Both groups are related orders within the class of *Alphaproteobacteria* which reinforces our belief that these viral groups are evolutionary linked. One possible way could be that the ancestor of these phages was already infecting the ancestor of this particular branch of *Alphaproteobacteria*. Another possibility could be, that the viral ancestor infected

only *Rhodobacterales* and over time the virus group evolved to also be capable of infecting *Hyphomicrobiales*.

Taking our classification of the two new subfamilies as a reference, some of the remaining *Microviridae* subfamilies that have been proposed by other studies. For example, ”Pichovirinae” and “Alpavirinae”, seem to be polyphyletic when adding additional prophage genomes or EVGs. Furthermore, according to the analysis presented within this study, these two subfamilies infect respectively 2 and 3 different bacterial phyla, which indicates that these subfamilies might be too large and require revision.

6.3 Phage ICBM5 displays the characteristics for a carrier state life strategy

Infection cycle strategies of the best-studied bacteriophages are lytic, lysogenic or chronic cycles, but more unconventional lifestyles such as pseudolysogeny or the carrier state have been observed within different groups of phages and are less investigated so far. However, their definition is not always consistent and both terms have been used interchangeably over the years (reviewed in Mäntynen et al. 2021). Apart from temperate phages, which by definition can apply two different infection cycle strategies, it has been noted that some phages are capable of applying variable lifestyles. For example, phage P22 which infects *Salmonella* Typhimurium displays the following lifestyles during infection: a) a lytic cycle which leads to cell lysis, b) a pseudolysogenic strategy which is characterized by the existence of a P22 episome within the cell cytoplasm. During cell division, this episome is then transmitted only to one of the daughter cells and 3) a lysogenic cycle arises from the cell that inherited the episomal phage (Cenens et al. 2015). Due to the discovery of prophages in other bacteria from the *Rhodobacteraceae* family, it was of particular interest if phage ICBM5 in addition to its lytic lifestyle is also capable of lysogeny. During the investigation of an ICBM5 resistant strain that has been isolated from turbid plaques, no evidence could be found that phage ICBM5 incorporates its genome into the host chromosome of *Sulfitobacter dubius* SH24-1b. Nevertheless, it could be demonstrated that phage ICBM5 undergoes a carrier life strategy and a sub-population of resistant cells carried the ICBM5 genome as an episome which was present in variable numbers within the host cytoplasm. In addition, another numerically dominant sub-population was observed without carrying ICBM5 phages. This alternative sub-population has shown to be resistant to the virus and thereby conferred to the

host strain resistance of ICBM5 which has been observed in spot assays. When taking into consideration that this particular strain has been isolated from a single colony, there are two possible scenarios by which the two sub-populations were generated. The first mechanism could be explained by an emergence of a single cell from the colony that harboured the ICBM5 genome intracellularly. From there, an asymmetrical cell division would have been the outcome meaning that transmission of ICBM5 would have been the case to only one of the daughter cells. The second mechanism could have been established due to an extracellular attachment of ICBM5 phage particles to the cell that initially formed the colony. This scenario would lead to subsequent cell division and therefore a rise of sensitive cells which as a consequence would have become infected by ICBM5. However, the parameters which confer resistance to ICBM5 remain to be unknown for now. In the case of P22 infection, the daughter cells without virus episomes that have resulted from asymmetric cell division of pseudolysogenic cells were temporarily immune to the P22 phage. This resistance was conferred due to immunity factors that have been transmitted cytoplasmically from the mother cell which resulted in an unavoidable dilution by subsequent cell division (Cenens et al. 2015). According to the large portion of non-infected cells in the ICBM5-resistant strain, it is rather unlikely that a comparable mechanism is responsible for conferring resistance. More experiments are required to better characterize the ICBM5 carrier state life strategy and to elucidate the host resistance mechanism.

Furthermore, by findings generated from this study, it could be demonstrated that the carrier state life cycle does not seem to be restricted to infection of *Sulfitobacter dubius* SH24-1b by phage ICBM5. *Microviridae*-like episomes have been predicted in *Mesorhizobium, Neorizobium, Prevotella, Aphanizomenon, Gramella, Mammalicoccus,* and *Acinetobacter*, hosts which belong to various phyla. It is of high probability that these phages apply a carrier state infection strategy to survive within their host culture without having a dramatic influence on the culture's growth. In addition, a similar carrier state has recently been described for a gokushovirus, which was "revived" from its prophage state in the host genome via molecular cloning techniques (Kirchberger and Ochman 2020). Taken together, this indicates that such a carrier state occurs in various *Microviridae* phages and it is very likely that they co-exist with their host in environmental samples.

6.4 Characterization of the lytic infection cycle of phage ICBM5 via phage targeted direct-geneFISH

In previous studies, fluorescence *in situ* hybridization targeting phage genes was used to characterize the lytic infection cycle of PSA-HP1, which is a dsDNA phage infecting *Pseudoalteromonas* (Allers et al. 2013). Within this method, multiple probes were labelled with dioxygenin which was then followed by a signal amplification step mediated by antibody binding and enzymatic tyramide deposition. In this study, a modified and updated version of the direct-geneFISH technique was applied which is known as phage targeted direct-geneFISH (Barrero-Canosa et al. 2017; Barrero-Canosa and Moraru 2021). This procedure relies on the usage of multiple probes that are directly labelled with fluorochromes and thereby making the signal amplification step unnecessary. The phage targeted direct-geneFISH method was recently applied for the intracellular virus detection of archaeal and pico-eukaryotic dsDNA viruses of environmental samples and pure cultures (Castillo et al. 2020; Castillo et al. 2021; Rahlff et al. 2021). Here, due to the use of phage-targeted genome-wide probes, the lytic infection cycle of a new ssDNA phage could be characterized. The moment for bacterial cell lysis has been identified at 110 min after infection starts within the tested conditions. In comparison, this is shorter than the previously determined ~ 3 h for the vB_Cib_ssDNA_P1 phage which infects *Citromicrobium* sp. (Zheng et al. 2018) and the vB_RpoMi_Mini infecting *Ruegeria pomeroyi* DSS-3 (Zhan and Chen 2019b). Conversely, the time frame for the lytic cycle of ΦX174, the to-date best-characterized *microvirus*, was only 20 min on *Escherichia coli* (Hutchison and Sinsheimer 1963). The time variations are probably due to genetic differences between the phages but also due to differences in the physiological characteristics of the two hosts. In addition, by measuring the total phage signal intensity values per cell and normalizing them to the intensity of a single phage, the quantification of the per-cell genome numbers of ICBM5 could be calculated. In the final stage of the lytic cycle, the majority of cells harboured up to 125 ICBM5 genome copies and some cells even reached more than 300 ICBM5 genome copies. However, measurements of the signal quantification are not capable to differentiate if these genomes were indeed packed into mature virions and released from the cell. Nevertheless, the burst size was additionally calculated by the PFU measurements and approximately 250 phages per cell were released during cell lysis. In comparison, ΦX174 and vB_Cib_ssDNA_P1 have a burst size of ~170 phages per cell (Hutchison and Sinsheimer 1966), and vB_RpoMi_Mini of only ~8 phages per cell (Zhan and Chen 2019).

6.5 Additional genomic and proteomic features of phage ICBM5

The *in silico* analysis regarding uncharacterized and undetected genomic features on the ICBM5 genome has allowed for additional insights and raised a set of new questions. Regarding the genomic architecture and the way genes are transcribed in other *Microviridae,* the best-studied candidate is ΦX174 which infects *Escherichia coli* C. This particular bacteriophage encodes at least 11 genes, so far all in the same reading direction and some of them are overlapping each other. Phage ICBM5 only has six identified ORFs. However, apart from the identified genes that are shared with phage ICBM5 (pilot (H), MCP (F), lysis (E) and Rep (A)), ΦX174 utilizes additional genes responsible for shutting off the host DNA synthesis (A*), internal scaffolding (B), DNA maturation (C), external scaffolding (D), major spike protein (G), DNA binding protein (J) and an additional gene with unknown function (K). Two additional ORFs with coding capacity have been predicted on the genome of ΦX174 but their function remains unknown as well (Cherwa and Fane 2011). Nevertheless, it is quite likely that the hypothetical proteins of phage ICBM5 (HP1 and HP2) fulfil some of the functions that appear in phage ΦX174. When comparing gene length, the gene for HP2 is of very similar size to gene G of ΦX174, which encodes for the major spike protein (527 nts to 528 nts, respectively). For the well-investigated coliphage, the major spike protein seems to be of crucial importance for the interaction with host cell components during attachment, which is commonly regulated by the recognition of lipopolysaccharides (LPS) on the outer cell surface (reviewed in Doore and Fane 2016). Not much is known about specific adsorption parameters on a molecular level for *Microviridae* infecting *Alphaproteobacteria,* but members of this group are also known to produce LPS structures on their outer membrane, suggesting that they could be involved in an attachment process as well (Gatsos et al. 2008). However, *Gokushovirinae* do not produce the major spike protein G but instead form “mushroom-like” protrusions, indicating that their recognition process is differently regulated (Chipman et al. 1998; Everson et al. 2003). Nevertheless, gene length is rather an estimate of similarity than a measure and more information is required to identify the remaining hypothetical proteins of ICBM5. It could be shown that even without sequence homology between related groups of phages, some of the viral proteins are still conserved at the structural level (reviewed in Dion et al. 2020). This suggests that comparative structural analysis of protein folding patterns could be another important element in terms of identifying unknown protein functions and help to improve our understanding of the viral proteomic inventory.

Furthermore, three RBSs have been predicted which allows for additional insight into the phage ICBM5 translation machinery. The first RBS is located in front of the gene translating for HP1, while the pilot gene is without an RBS. This indicates a polycistronic translation for the first two genes of the phage ICBM5 genome, which could mean that HP1 and the pilot protein are translated together and probably cleaved into separate proteins afterwards. The second RBS is located in front of the MCP gene, the gene for HP2 and the lysis gene are without an RBS. This indicates again a polycistronic translation and would mean that all three genes are translated together. Directly after the gene for the lysis protein, a Rho-independent terminator was predicted indicating an ending mechanism for transcription. Finally, the last RBS is located in front of the REP gene, which translates for the replication initiation protein. When considering the different steps of a lytic infection cycle, after the genome has been inserted into the host cell cytoplasm, the phage genome replicates itself first. Therefore, the REP would participate in the early stage of infection and initiate replication via rolling circle. Next, the ribosome would bind to RBS_1 and initiate the translation of HP1 and the pilot, respectively. HP1 is of unknown function and the pilot gene translates for the pilot protein, which is involved in DNA transport and DNA translocation (Roznowski and Fane 2016). The later stages of infection, protein synthesis of the MCP, assembly and lysis would all be regulated by RBS_2. This ribosomal binding site initiates the translation for the MCP, HP2 and finally the lysis protein. According to the genomic order of phage ΦX174, the gene for the major spike protein (G) is located directly after the gene for the MCP (F). This is an indication that the gene for HP2 of phage ICBM5 could translate for the major spike protein. Nevertheless, more work will be required to elucidate the function of the two remaining hypothetical proteins and to get a better understanding of the translation machinery of phage ICBM5 and *Microviridae* viruses in general.

6.6 Additional transcriptomic features of phage ICBM5

In addition, the question was raised if ssDNA phages from the family *Microviridae* are capable of utilizing RNA-based regulatory elements like sRNAs. The majority of described phage-encoded sRNAs are found in phages infecting *Gammaproteobacteria* like *Salmonella, Escherichia* or *Yersinia.* Only one phage was mentioned infecting a *Cyanobacteria* which encodes a sRNA involved in regulating the phage AMG *psbA* (Millard et al. 2010). The *in silico* analysis on the phage ICBM5 genome performed by presRAT has generated three potential

sRNA candidates showing scores which compare to already described sRNAs (Kumar et al. 2020). One of them seemed to be capable of folding properly after applying additional secondary structure prediction. The putative sRNA was found within the pilot gene of phage ICBM5 and a possible mRNA target could be predicted as well. The small regulator shows the potential to repress the expression of the host *murF* gene. This particular gene is involved in bacterial cell wall synthesis and inhibiting its translation would presumably be of detrimental effect on the host organism given that cell division would be restricted. Since sRNA_18 is located within the pilot gene, suppression of the *murF* gene would presumably take place after genome replication and during protein synthesis. An interaction between sRNA_18 and the *murF* gene could result in a destabilization of the cell wall, maybe as a preparation mechanism for the following cell lysis. The phage ICBM5 lysis protein is translated later during infection and could benefit from a destabilized cell wall. In other *Microviridae* strains, the lysis protein functions by inhibiting cell wall synthesis. For example in ΦX174, the lysis protein acts by targeting the *mraY* gene. This particular gene encodes for the translocase I, which catalyses the formation of a lipid-linked intermediate within cell wall biosynthesis (Bernhardt et al. 2000). The phage lysis protein of phage ICBM5 does not cluster together with the ΦX174 lysis protein, as shown by the VirClust analysis (see Fig. 23). Instead, the lysis gene of phage ICBM5 is similar to the M15 metallopeptidase that has been observed in other *Microviridae* (Zhan and Chen 2019). This particular gene is rather associated with canonical lysis genes, since it shows structural identity to other M15 family proteases which are involved in the degradation of the bacterial cell wall at the end of an infection cycle (Roux et al. 2012). Therefore, sRNA_18 and the phage ICBM5 lysis protein could work together towards bacterial cell lysis. In this scenario, the small ribonucleic acid based regulator would inhibit the cell wall synthesis which would lead to a destabilized membrane structure. This would be followed by a degradation of cell wall components induced by the viral lysis protein and ultimately causing lysis of the host bacterium (see Fig. 37).

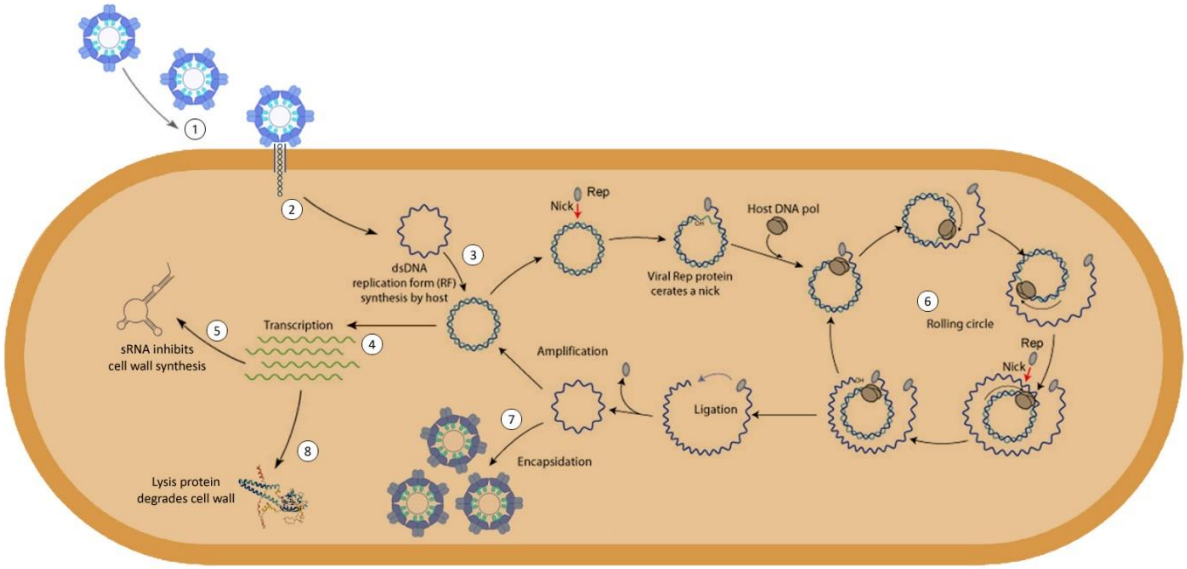

Fig. 37: Shows the different steps of a lytic infection cycle of phage ICBM5 including the predicted host gene regulation by sRNA_18. 1) Attachment takes place and 2) the phage injects its genomic material across the periplasmic space into the cytoplasm. 3) The replicative form is generated as preparation for the rolling circle replication, 4) to transcribe mRNAs and 5) putative sRNAs. The putative sRNA would inhibit the bacterial cell wall synthesis and thereby leaving it prone for later lysis. Simultaneously, 6) the phage encoded replication protein initiates rolling circle replication and 7) the newly synthesized ssDNA molecules are packaged into assembled capsids. During the end of the infection cycle 8) the lysis protein degrades the host cell wall until it finally bursts. Replication cycle overview was taken from ViralZone, SIB Swiss Institute of Bioinformatics (https://viralzone.expasy.org/1941). The image was created using BioRender (https://biorender.com/).

Nevertheless, these findings can be no more than an indication of potential regulatory RNA-based interaction mechanisms between phage and host. Predicting sRNAs reliably is very difficult only by *in silico* investigation and further wet-lab-based methods have to be applied for a better understanding. Recent studies even demonstrated the potential for utilizing fluorescence *in situ* hybridization techniques to detect sRNAs within human and murine tissues, but the principle could probably be applied to prokaryotic organisms as well (Urbanek et al. 2015). The described procedure could be applied to an infected cell and thereby help to identify additional factors involved in the regulation of phage-host interaction. Small RNAs of temperate phages are also known to be involved in maintaining the state of lysogeny as recently revealed by transcriptomic analysis (Owen et al. 2020)

7. Conclusion

In recent years, much progress has been made regarding our knowledge of tailless, icosahedral, ssDNA viruses. Within this study, many new *Microviridae*-like prophages as well as episomes have been found. Their discovery raises the question of whether they are more spread in bacterial genomes than previously anticipated. The new prophages infect members of the order *Hyphomicrobiales* and *Rhodobacterales*, both within the class of *Alphaproteobacteria*. The two bacterial groups are known for their broad occurrence around the globe occupying a variety of different ecological niches. Therefore, the two newly identified subfamilies “Tainavirinae” and “Occultatumvirinae” help to improve our understanding of the interaction between ssDNA phages and environmentally relevant bacteria. The relatively unsophisticated procedure for predicting prophages demonstrated that, at the date of our search, most of the detectable prophage and episome diversity has been recovered. Furthermore, the life cycle investigations showed that members of the family *Microviridae* can employ different lifestyles and not only lyse their respective host bacterium. They have adapted to alternative interaction strategies – like the carrier state – which can be considered as symbiotic by coexisting with their host while also conferring resistance to other infections. The additional findings from the bioinformatical analysis regarding transcriptomic features have shown promising indications in terms of regulating the genetic inventory of their host. However, further wet-lab-related work will be required to reveal if these interactions can indeed take place during infection. Nevertheless, they can give an idea about the vast unknown landscape of ribonucleic acid-related regulatory elements and that there is still much to reveal. Undoubtedly, an improved understanding of the regulation regarding gene expression could further help to identify the mechanisms involved in switching between the different lifestyle strategies.

8. Future perspectives

The possibility of visualizing the different stages of viral infection via epifluorescence microscopy opens up a variety of interesting opportunities. Combining the phage targeted direct-geneFISH technique for phages with protocols for mRNA-FISH or even sRNA-FISH would allow for additional insights into the regulation of certain genes, not only during lytic infection cycles but also in more dormant lifestyles like the carrier state or the lysogenic cycle. Targeting a certain gene via mRNA-FISH or a certain ribonucleotide regulator via sRNA-FISH could also allow to quantify the signal intensity of the respective target. Thereby, this would indicate when those molecules are up- or down-regulated and thus improve our understanding of the phage-host interaction. Furthermore, recent bioinformatics developments – for example new algorithms for protein folding – could improve the structural identity analysis on the level of three-dimensional peptide folding patterns. These advances can be of enormous advantage while trying to identify viral protein function, which still is one of the biggest challenges for modern phage biology.

Viruses are capable of regulating their surroundings on a multidimensional level: they regulate molecular processes through the expression of sRNAs and they mediate evolutionary processes by horizontal gene transfer. They also coordinate ecological processes by influencing host metabolism through the expression of AMGs or by shaping community compositions in different habitats through bacterial lysis. Still, our understanding of virus-host interactions is quite often associated with a militaristic terminology. “The virus kills the cell”, or “the host needs to defend against a viral invasion” are just examples of regularly appearing descriptions of naturally occurring phenomena. This type of language, especially if applied repetitively, might distort our perception of such biological processes and it could be worth to consider an adjustment. Maybe viruses can rather be regarded as organic regulatory entities and carriers of biological information which are important keystones of life.

9. References

Publication bibliography

Ackermann, Hans-W. (1998): Tailed Bacteriophages: The Order Caudovirales. In Karl Maramorosch, Frederick A. Murphy, Aaron J. Shatkin (Eds.): Advances in Virus Research, vol. 51: Academic Press, pp. 135–201.

Allers, Elke; Moraru, Cristina; Duhaime, Melissa B.; Beneze, Erica; Solonenko, Natalie; Barrero-Canosa, Jimena et al. (2013): Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. In *Environmental microbiology* 15 (8), pp. 2306–2318. DOI: 10.1111/1462-2920.12100.

Angly, Florent; Youle, Merry; Nosrat, Bahador; Srinagesh, Shailaja; Rodriguez-Brito, Beltran; McNairnie, Patrick et al. (2009): Genomic analysis of multiple Roseophage SIO1 strains. In *Environmental microbiology* 11 (11), pp. 2863–2873. DOI: 10.1111/j.1462-2920.2009.02021.x.

Bailly-Bechet, Marc; Vergassola, Massimo; Rocha, Eduardo (2007): Causes for the intriguing presence of tRNAs in phages. In *Genome Research* 17 (10), pp. 1486–1495. DOI: 10.1101/gr.6649807.

Balasubramanian, Divya; Ragunathan, Preethi T.; Fei, Jingyi; Vanderpool, Carin K. (2016): A Prophage-Encoded Small RNA Controls Metabolism and Cell Division in Escherichia coli. In *mSystems* 1 (1). DOI: 10.1128/mSystems.00021-15.

Balch, W. E.; Fox, G. E.; Magrum, L. J.; Woese, C. R.; Wolfe, R. S. (1979): Methanogens: reevaluation of a unique biological group. In *Microbiological Reviews* 43 (2), pp. 260–296.

Baltimore, D. (1971): Expression of animal virus genomes. In *Bacteriological Reviews* 35 (3), pp. 235–241. DOI: 10.1128/br.35.3.235-241.1971.

Barrero-Canosa, Jimena; Moraru, Cristina (2021): Linking microbes to their genes at single cell level with direct–geneFISH. In Nuno F. Azevedo, Carina Almeida (Eds.): Fluorescence In-Situ Hybridization (FISH) for Microbial Cells. Methods and Concepts: Springer US (Methods in Molecular Biology), pp. 169–205.

Barrero-Canosa, Jimena; Moraru, Cristina; Zeugner, Laura; Fuchs, Bernhard M.; Amann, Rudolf (2017): Direct-geneFISH. A simplified protocol for the simultaneous detection and quantification of genes and rRNA in microorganisms. In *Environmental microbiology* 19 (1), pp. 70–82. DOI: 10.1111/1462-2920.13432.

Bartlau, Nina; Wichels, Antje; Krohne, Georg; Adriaenssens, Evelien M.; Heins, Anneke; Fuchs, Bernhard M. et al. (2021): Highly diverse flavobacterial phages isolated from North Sea spring blooms. In *ISME J*. DOI: 10.1038/s41396-021-01097-4.

9. References

Batinovic, Steven; Wassef, Flavia; Knowler, Sarah A.; Rice, Daniel T. F.; Stanton, Cassandra R.; Rose, Jayson et al. (2019): Bacteriophages in Natural and Artificial Environments. In *Pathogens (Basel, Switzerland)* 8 (3). DOI: 10.3390/pathogens8030100.

Bednarz, Michael; Halliday, Jennifer A.; Herman, Christophe; Golding, Ido (2014): Revisiting bistability in the lysis/lysogeny circuit of bacteriophage lambda. In *PloS one* 9 (6), e100876. DOI: 10.1371/journal.pone.0100876.

Bergh; Knut Yngve BØrsheim; Gunnar Bratbak & Mikal Heldal (1989): High abundance of viruses found in aquatic environments 340, pp. 467–468.

Bernhardt, T. G.; Roof, W. D.; Young, R. (2000): Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis. In *Proceedings of the National Academy of Sciences of the United States of America* 97 (8), pp. 4297–4302. DOI: 10.1073/pnas.97.8.4297.

Bernhardt, Thomas G.; Wang, Ing-Nang; Struck, Douglas K.; Young, Ry (2002): Breaking free: "Protein antibiotics" and phage lysis. In *Research in microbiology* 153 (8), pp. 493–501. DOI: 10.1016/S0923-2508(02)01330-X.

Bertozzi Silva, Juliano; Storms, Zachary; Sauvageau, Dominic (2016): Host receptors for bacteriophage adsorption. In *FEMS Microbiology Letters* 363 (4). DOI: 10.1093/femsle/fnw002.

Bin Jang, Ho; Bolduc, Benjamin; Zablocki, Olivier; Kuhn, Jens H.; Roux, Simon; Adriaenssens, Evelien M. et al. (2019): Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. In *Nature Biotechnology* 37 (6), pp. 632–639. DOI: 10.1038/s41587-019-0100-8.

Bischoff, Vera; Bunk, Boyke; Meier-Kolthoff, Jan P.; Spröer, Cathrin; Poehlein, Anja; Dogs, Marco et al. (2019): Cobaviruses - a new globally distributed phage group infecting Rhodobacteraceae in marine ecosystems. In *ISME J* 13 (6), pp. 1404–1421. DOI: 10.1038/s41396-019-0362-7.

Bischoff, Vera; Zucker, Falk; Moraru, Cristina (2021): Marine Bacteriophages. In Dennis Bamford, Mark A. Zuckerman (Eds.): Encyclopedia of virology. Fourth edition. Amsterdam: Academic Press, pp. 322–341.

Blasdel, Bob G.; Chevallereau, Anne; Monot, Marc; Lavigne, Rob; Debarbieux, Laurent (2017): Comparative transcriptomics analyses reveal the conservation of an ancestral infectious strategy in two bacteriophage genera. In *The ISME journal* 11 (9), pp. 1988–1996. DOI: 10.1038/ismej.2017.63.

9. References

Bloch, Sylwia; Lewandowska, Natalia; Węgrzyn, Grzegorz; Nejman-Faleńczyk, Bożena (2021): Bacteriophages as sources of small non-coding RNA molecules. In *Plasmid* 113, p. 102527. DOI: 10.1016/j.plasmid.2020.102527.

Bordet, Jules (1925): Le problème de l'autolyse microbienne transmissible ou du bactériophage. In *Annales de l'Institut Pasteur*, pp. 1–47.

Børsheim, K. Y.; Bratbak, G.; Heldal, M. (1990): Enumeration and biomass estimation of planktonic bacteria and viruses by transmission electron microscopy. In *Applied and environmental microbiology* 56 (2), pp. 352–356.

Bradley, D. E. (1967): Ultrastructure of bacteriophage and bacteriocins. In *Bacteriological Reviews* 31 (4), pp. 230–314. DOI: 10.1128/br.31.4.230-314.1967.

Bragg, Jason G.; Chisholm, Sallie W. (2008): Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. In *PloS one* 3 (10), e3550. DOI: 10.1371/journal.pone.0003550.

Breitbart, Mya (2012): Marine viruses: truth or dare. In *Annual review of marine science* 4, pp. 425–448. DOI: 10.1146/annurev-marine-120709-142805.

Breitbart, Mya; Bonnain, Chelsea; Malki, Kema; Sawaya, Natalie A. (2018): Phage puppet masters of the marine microbial realm. In *Nature microbiology*. DOI: 10.1038/s41564-018-0166-y.

Brentlinger, Karie L.; Hafenstein, Susan; Novak, Christopher R.; Fane, Bentley A.; Borgon, Robert; McKenna, Robert; Agbandje-McKenna, Mavis (2002): Microviridae, a family divided: isolation, characterization, and genome sequence of phiMH2K, a bacteriophage of the obligate intracellular parasitic bacterium Bdellovibrio bacteriovorus. In *Journal of bacteriology* 184 (4), pp. 1089–1094.

Brinkhoff, Thorsten; Giebel, Helge-Ansgar; Simon, Meinhard (2008): Diversity, ecology, and genomics of the Roseobacter clade: a short overview. In *Archives of microbiology* 189 (6), pp. 531–539. DOI: 10.1007/s00203-008-0353-y.

Brinkman, Nichole E.; Villegas, Eric N.; Garland, Jay L.; Keely, Scott P. (2018): Reducing inherent biases introduced during DNA viral metagenome analyses of municipal wastewater. In *PloS one* 13 (4), e0195350. DOI: 10.1371/journal.pone.0195350.

Brum, Jennifer R.; Schenck, Ryan O.; Sullivan, Matthew B. (2013): Global morphological analysis of marine viruses shows minimal regional variation and dominance of non-tailed viruses. In *The ISME journal* 7 (9), pp. 1738–1751. DOI: 10.1038/ismej.2013.67.

9. References

Brussaard, Corina P. D. (2004): Optimization of procedures for counting viruses by flow cytometry. In *Applied and environmental microbiology* 70 (3), pp. 1506–1513. DOI: 10.1128/AEM.70.3.1506-1513.2004.

Brüssow, Harald; Canchaya, Carlos; Hardt, Wolf-Dietrich (2004): Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. In *Microbiology and molecular biology reviews: MMBR* 68 (3), 560-602, table of contents. DOI: 10.1128/MMBR.68.3.560-602.2004.

Buchan, Alison; González, José M.; Moran, Mary Ann (2005): Overview of the marine roseobacter lineage. In *Applied and environmental microbiology* 71 (10), pp. 5665–5677. DOI: 10.1128/AEM.71.10.5665-5677.2005.

Buchholz, Holger H.; Michelsen, Michelle L.; Bolaños, Luis M.; Browne, Emily; Allen, Michael J.; Temperton, Ben (2021): Efficient dilution-to-extinction isolation of novel virus-host model systems for fastidious heterotrophic bacteria. In *The ISME journal* 15 (6), pp. 1585–1598. DOI: 10.1038/s41396-020-00872-z.

Casjens, Sherwood R.; Hendrix, Roger W. (2015): Bacteriophage lambda: Early pioneer and still relevant. In *Virology* 479-480, pp. 310–330. DOI: 10.1016/j.virol.2015.02.010.

Castillo, Yaiza M.; Forn, Irene; Yau, Sheree; Morán, Xosé Anxelu G.; Alonso-Sáez, Laura; Arandia-Gorostidi, Néstor et al. (2021): Seasonal dynamics of natural Ostreococcus viral infection at the single cell level using VirusFISH. In *Environmental microbiology* 23 (6), pp. 3009–3019. DOI: 10.1111/1462-2920.15504.

Castillo, Yaiza M.; Sebastián, Marta; Forn, Irene; Grimsley, Nigel; Yau, Sheree; Moraru, Cristina; Vaqué, Dolors (2020): Visualization of Viral Infection Dynamics in a Unicellular Eukaryote and Quantification of Viral Production Using Virus Fluorescence in situ Hybridization. In *Frontiers in microbiology* 11, Article 1559, p. 2306. DOI: 10.3389/fmicb.2020.01559.

Chang, Tzu-Hao; Huang, Hsi-Yuan; Hsu, Justin Bo-Kai; Weng, Shun-Long; Horng, Jorng-Tzong; Huang, Hsien-Da (2013): An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. In *BMC bioinformatics* 14 Suppl 2, S4. DOI: 10.1186/1471-2105-14-S2-S4.

Cherwa, James E.; Fane, Bentley A. (2011): Microviridae: Microviruses and Gokushoviruses. DOI: 10.1002/9780470015902.a0000781.pub2.

Chevallereau, Anne; Pons, Benoît J.; van Houte, Stineke; Westra, Edze R. (2021): Interactions between bacterial and phage communities in natural environments. In *Nature reviews. Microbiology*. DOI: 10.1038/s41579-021-00602-y.

9. References

Chipman, Paul R.; Agbandje-McKenna, Mavis; Renaudin, Joël; Baker, Timothy S.; McKenna, Robert (1998): Structural analysis of the spiroplasma virus, SpV4: implications for evolutionary variation to obtain host diversity among the Microviridae. In *Structure* 6 (2), pp. 135–145. DOI: 10.1016/S0969-2126(98)00016-1.

Clokie, Martha R. J.; Millard, Andrew D.; Letarov, Andrey V.; Heaphy, Shaun (2011): Phages in nature. In *Bacteriophage* 1 (1), pp. 31–45. DOI: 10.4161/bact.1.1.14942.

Collins, Courtney L.; DeNardo, Dale F.; Blake, Mellecha; Norton, Jessica; Schmidlin, Kara; Fontenele, Rafaela S. et al. (2021): Genome Sequences of Microviruses Identified in Gila Monster Feces. In *Microbiol Resour Announc* 10 (11). DOI: 10.1128/MRA.00163-21.

Colombo, Stefano; Arioli, Stefania; Neri, Eros; Della Scala, Giulia; Gargari, Giorgio; Mora, Diego (2017): Viromes As Genetic Reservoir for the Microbial Communities in Aquatic Environments: A Focus on Antimicrobial-Resistance Genes. In *Frontiers in microbiology* 8, p. 1095. DOI: 10.3389/fmicb.2017.01095.

Creasy, Alexandria; Rosario, Karyna; Leigh, Brittany A.; Dishaw, Larry J.; Breitbart, Mya (2018): Unprecedented Diversity of ssDNA Phages from the Family Microviridae Detected within the Gut of a Protochordate Model Organism (Ciona robusta). In *Viruses* 10 (8). DOI: 10.3390/v10080404.

D'Hérelle, F. (1917): On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917. In *Research in microbiology* 158 (7), pp. 553–554. DOI: 10.1016/j.resmic.2007.07.005.

D'Hérelle, F. (1926): The Bacteriophage and its Behaviour. In *Nature* 118 (2962), pp. 183–185. DOI: 10.1038/118183a0.

Darling, Aaron E.; Mau, Bob; Perna, Nicole T. (2010): progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. In *PloS one* 5 (6), e11147. DOI: 10.1371/journal.pone.0011147.

Desnues, Christelle; Rodriguez-Brito, Beltran; Rayhawk, Steve; Kelley, Scott; Tran, Tuong; Haynes, Matthew et al. (2008): Biodiversity and biogeography of phages in modern stromatolites and thrombolites. In *Nature* 452 (7185), pp. 340–343. DOI: 10.1038/nature06735.

Dion, Moïra B.; Oechslin, Frank; Moineau, Sylvain (2020): Phage diversity, genomics and phylogeny. In *Nat Rev Microbiol* 437, p. 356. DOI: 10.1038/s41579-019-0311-5.

Dishaw, Larry J.; Flores-Torres, Jaime; Lax, Simon; Gemayel, Kristina; Leigh, Brittany; Melillo, Daniela et al. (2014): The gut of geographically disparate Ciona intestinalis harbors a core microbiota. In *PloS one* 9 (4), e93386. DOI: 10.1371/journal.pone.0093386.

9. References

Doore, Sarah M.; Fane, Bentley A. (2016): The microviridae: Diversity, assembly, and experimental evolution. In *Virology* 491, pp. 45–55. DOI: 10.1016/j.virol.2016.01.020.

Duarte, Carlos M. (2015): Seafaring in the 21St Century: The Malaspina 2010 Circumnavigation Expedition. In *Limnology and Oceanography Bulletin* 24 (1), pp. 11–14. DOI: 10.1002/lob.10008.

Eddy, Sean R. (2011): Accelerated Profile HMM Searches. In *PLoS computational biology* 7 (10), e1002195. DOI: 10.1371/journal.pcbi.1002195.

Ellis, E. L.; Delbrück, M. (1939): THE GROWTH OF BACTERIOPHAGE. In *The Journal of General Physiology* 22 (3), pp. 365–384. DOI: 10.1085/jgp.22.3.365.

Enav, Hagay; Béjà, Oded; Mandel-Gutfreund, Yael (2012): Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts. In *The ISME journal* 6 (3), p. 619. DOI: 10.1038/ismej.2011.146.

Everson, J. S.; Garner, S. A.; Fane, B.; Liu, B.-L.; Lambden, P. R.; Clarke, I. N. (2002): Biological Properties and Cell Tropism of Chp2, a Bacteriophage of the Obligate Intracellular Bacterium Chlamydophila abortus. In *Journal of bacteriology* 184 (10), pp. 2748–2754. DOI: 10.1128/JB.184.10.2748-2754.2002.

Everson, J. S.; Garner, S. A.; Lambden, P. R.; Fane, B. A.; Clarke, I. N. (2003): Host range of chlamydiaphages phiCPAR39 and Chp3. In *Journal of bacteriology* 185 (21), pp. 6490–6492. DOI: 10.1128/JB.185.21.6490-6492.2003.

Falkowski, Paul G.; Fenchel, Tom; DeLong, Edward F. (2008): The microbial engines that drive Earth's biogeochemical cycles. In *Science (New York, N.Y.)* 320 (5879), pp. 1034–1039. DOI: 10.1126/science.1153213.

Fane, B.; Brentlinger, Karie L.; Burch, A. D.; Chen, M.; Hafenstein, S.; Moore, E.; et al. (2006): *ØX174 et al., the Microviridae*. In *The Bacteriophages, 2nd Edn. New York, NY: Oxford University Press*, pp. 129–145.

Feng, J. N.; Model, P.; Russel, M. (1999): A trans-envelope protein complex needed for filamentous phage assembly and export. In *Molecular microbiology* 34 (4), pp. 745–755. DOI: 10.1046/j.1365-2958.1999.01636.x.

Fenner, F. (1976): The Classification and Nomenclature of Viruses. In *Journal of General Virology* 31 (3), pp. 463–470. DOI: 10.1099/0022-1317-31-3-463.

Finn, Robert D.; Attwood, Teresa K.; Babbitt, Patricia C.; Bateman, Alex; Bork, Peer; Bridge, Alan J. et al. (2017): InterPro in 2017-beyond protein family and domain annotations. In *Nucleic acids research* 45 (D1), D190-D199. DOI: 10.1093/nar/gkw1107.

Forcone, Kathryn; Coutinho, Felipe H.; Cavalcanti, Giselle S.; Silveira, Cynthia B. (2021): Prophage Genomics and Ecology in the Family Rhodobacteraceae. In *Microorganisms* 9 (6). DOI: 10.3390/microorganisms9061115.

Fuhrman, J. A. (1999): Marine viruses and their biogeochemical and ecological effects. In *Nature* 399 (6736), pp. 541–548. DOI: 10.1038/21119.

Gasol, Josep M.; Kirchman, David L. (Eds.) (2018): Microbial ecology of the oceans. Third edition. Hoboken NJ: John Wiley & Sons/Blackwell.

Gatsos, Xenia; Perry, Andrew J.; Anwari, Khatira; Dolezal, Pavel; Wolynec, P. Peter; Likić, Vladimir A. et al. (2008): Protein secretion and outer membrane assembly in Alphaproteobacteria. In *FEMS microbiology reviews* 32 (6), pp. 995–1009. DOI: 10.1111/j.1574-6976.2008.00130.x.

Gorbalenya, Alexander E.; Krupovic, Mart; Mushegian, Arcady; Kropinski, Andrew M.; Siddell, Stuart G.; Varsani, Arvind et al. (2020): The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. In *Nat Microbiol* 5 (5), pp. 668–674. DOI: 10.1038/s41564-020-0709-x.

Gottesman, Susan; Storz, Gisela (2011): Bacterial small RNA regulators: versatile roles and rapidly evolving variations. In *Cold Spring Harbor perspectives in biology* 3 (12). DOI: 10.1101/cshperspect.a003798.

Grazziotin, Ana Laura; Koonin, Eugene V.; Kristensen, David M. (2017): Prokaryotic virus orthologous groups (pVOGs). A resource for comparative genomics and protein family annotation. In *Nucleic acids research* 45, D491-D498. DOI: 10.1093/nar/gkw975.

Gregory, Ann C.; Zayed, Ahmed A.; Conceição-Neto, Nádia; Temperton, Ben; Bolduc, Ben; Alberti, Adriana et al. (2019): Marine DNA viral macro- and microdiversity from Pole to Pole. In *Cell*. DOI: 10.1016/j.cell.2019.03.040.

Gruber, Andreas R.; Lorenz, Ronny; Bernhart, Stephan H.; Neuböck, Richard; Hofacker, Ivo L. (2008): The Vienna RNA websuite. In *Nucleic acids research* 36 (Web Server issue), W70-4. DOI: 10.1093/nar/gkn188.

Gurevich, Alexey; Saveliev, Vladislav; Vyahhi, Nikolay; Tesler, Glenn (2013): QUAST: quality assessment tool for genome assemblies. In *Bioinformatics (Oxford, England)* 29 (8), pp. 1072–1075. DOI: 10.1093/bioinformatics/btt086.

Hahnke, Sarah; Brock, Nelson L.; Zell, Claudia; Simon, Meinhard; Dickschat, Jeroen S.; Brinkhoff, Thorsten (2013): Physiological diversity of *Roseobacter* clade bacteria co-occurring during a phytoplankton bloom in the North Sea. In *Systematic and applied microbiology* 36 (1), pp. 39–48. DOI: 10.1016/j.syapm.2012.09.004.

Han, Li-Li; Yu, Dan-Ting; Zhang, Li-Mei; Shen, Ju-Pei; He, Ji-Zheng (2017): Genetic and functional diversity of ubiquitous DNA viruses in selected Chinese agricultural soils. In *Scientific reports* 7, p. 45142. DOI: 10.1038/srep45142.

Hendrix, R. W.; Smith, M. C.; Burns, R. N.; Ford, M. E.; Hatfull, G. F. (1999): Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. In *Proceedings of the National Academy of Sciences of the United States of America* 96 (5), pp. 2192–2197. DOI: 10.1073/pnas.96.5.2192.

Holmfeldt, Karin; Odić, Duško; Sullivan, Matthew B.; Middelboe, Mathias; Riemann, Lasse (2012): Cultivated Single-Stranded DNA Phages That Infect Marine Bacteroidetes Prove Difficult To Detect with DNA-Binding Stains. In *Applied and environmental microbiology* 78 (3), pp. 892–894. DOI: 10.1128/AEM.06580-11.

Hopkins, Max; Kailasan, Shweta; Cohen, Allison; Roux, Simon; Tucker, Kimberly Pause; Shevenell, Amelia et al. (2014): Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. In *The ISME journal* 8 (10), pp. 2093–2103. DOI: 10.1038/ismej.2014.43.

Howard-Varona, Cristina; Hargreaves, Katherine R.; Abedon, Stephen T.; Sullivan, Matthew B. (2017): Lysogeny in nature: mechanisms, impact and ecology of temperate phages. In *The ISME journal* 11 (7), pp. 1511–1520. DOI: 10.1038/ismej.2017.16.

Huang, Xingyu; Jiao, Nianzhi; Zhang, Rui (2021): The genomic content and context of auxiliary metabolic genes in roseophages. In *Environmental microbiology*. DOI: 10.1111/1462-2920.15412.

Hurwitz, Bonnie L.; Brum, Jennifer R.; Sullivan, Matthew B. (2015): Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. In *ISME J* 9 (2), pp. 472–484. DOI: 10.1038/ismej.2014.143.

Hutchison, Clyde A.; Sinsheimer, Robert L. (1963): Kinetics of bacteriophage release by single cells of ϕX174-infected E. coli. In *Journal of molecular biology* 7 (2), pp. 206–208. DOI: 10.1016/S0022-2836(63)80046-7.

Hutchison, Clyde A.; Sinsheimer, Robert L. (1966): The process of infection with bacteriophage ΦX174. X. Mutations in a phiX Lysis Gene. In *Journal of molecular biology* 18 (3), 429-IN2. DOI: 10.1016/S0022-2836(66)80035-9.

Ignacio-Espinoza, Julio Cesar; Fuhrman, Jed A. (2018): A non-tailed twist in the viral tale. In *Nature* 554 (7690), pp. 38–39. DOI: 10.1038/d41586-018-00923-8.

Ilag, L. L.; McKenna, R.; Yadav, M. P.; BeMiller, J. N.; Incardona, N. L.; Rossmann, M. G. (1994): Calcium ion-induced structural changes in bacteriophage phi X174. In *Journal of molecular biology* 244 (3), pp. 291–300. DOI: 10.1006/jmbi.1994.1730.

9. References

Iranzo, Jaime; Krupovic, Mart; Koonin, Eugene V. (2016): The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. In *mBio* 7 (4). DOI: 10.1128/mBio.00978-16.

Jian, Huahua; Yi, Yi; Wang, Jiahua; Hao, Yali; Zhang, Mujie; Wang, Siyuan et al. (2021): Diversity and distribution of viruses inhabiting the deepest ocean on Earth. In *The ISME journal* 15 (10), pp. 3094–3110. DOI: 10.1038/s41396-021-00994-y.

Jones, Philip; Binns, David; Chang, Hsin-Yu; Fraser, Matthew; Li, Weizhong; McAnulla, Craig et al. (2014): InterProScan 5: genome-scale protein function classification. In *Bioinformatics (Oxford, England)* 30 (9), pp. 1236–1240. DOI: 10.1093/bioinformatics/btu031.

Karsenti, Eric; Acinas, Silvia G.; Bork, Peer; Bowler, Chris; Vargas, Colomban de; Raes, Jeroen et al. (2011): A holistic approach to marine eco-systems biology. In *PLoS biology* 9 (10), e1001177. DOI: 10.1371/journal.pbio.1001177.

Kazlauskas, Darius; Varsani, Arvind; Koonin, Eugene V.; Krupovic, Mart (2019): Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. In *Nature communications* 10 (1), p. 3425. DOI: 10.1038/s41467-019-11433-0.

Kearse, Matthew; Moir, Richard; Wilson, Amy; Stones-Havas, Steven; Cheung, Matthew; Sturrock, Shane et al. (2012): Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. In *Bioinformatics (Oxford, England)* 28 (12), pp. 1647–1649. DOI: 10.1093/bioinformatics/bts199.

Kiening, Michael; Ochsenreiter, Roman; Hellinger, Hans-Jörg; Rattei, Thomas; Hofacker, Ivo; Frishman, Dmitrij (2019): Conserved Secondary Structures in Viral mRNAs. In *Viruses* 11 (5). DOI: 10.3390/v11050401.

King, Andrew M.Q. (2012): Family - Leviviridae. In: Virus Taxonomy: Elsevier, pp. 1035–1043. Available online at https://www.sciencedirect.com/science/article/pii/B9780123846846000896.

Kirchberger, Paul C.; Ochman, Howard (2020): Resurrection of a global, metagenomically defined gokushovirus. In *eLife* 9. DOI: 10.7554/eLife.51599.

Koonin, Eugene V.; Dolja, Valerian V.; Krupovic, Mart; Varsani, Arvind; Wolf, Yuri I.; Yutin, Natalya et al. (2020): Global Organization and Proposed Megataxonomy of the Virus World. In *Microbiol. Mol. Biol. Rev.* 84 (2). DOI: 10.1128/MMBR.00061-19.

Koonin, Eugene V.; Krupovic, Mart; Dolja, Valerian V. (2022): The global virome: How much diversity and how many independent origins? In *Environ Microbiol*. DOI: 10.1111/1462-2920.16207.

Koonin, Eugene V.; Senkevich, Tatiana G.; Dolja, Valerian V. (2006): The ancient Virus World and evolution of cells. In *Biology direct* 1, p. 29. DOI: 10.1186/1745-6150-1-29.

Krupovic, Mart; Forterre, Patrick (2011): Microviridae goes temperate. Microvirus-related proviruses reside in the genomes of Bacteroidetes. In *PloS one* 6 (5), pp. e19893. DOI: 10.1371/journal.pone.0019893.

Krupovic, Mart; Koonin, Eugene V. (2017): Multiple origins of viral capsid proteins from cellular ancestors. In *Proceedings of the National Academy of Sciences of the United States of America* 114 (12), E2401-E2410. DOI: 10.1073/pnas.1621061114.

Kuhn, Jens H. (2019): Virus Taxonomy. In: Reference Module in Life Sciences: Elsevier. Available online at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7157452/pdf/main.pdf, checked on 8/12/2020.

Kumar, Krishna; Chakraborty, Abhijit; Chakrabarti, Saikat (2020): PresRAT: A server for identification of bacterial small-RNA sequences and their targets with probable binding region.

Labonté, Jessica M.; Suttle, Curtis A. (2013a): Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. In *Frontiers in microbiology* 4, p. 404. DOI: 10.3389/fmicb.2013.00404.

Labonté, Jessica M.; Suttle, Curtis A. (2013b): Previously unknown and highly divergent ssDNA viruses populate the oceans. In *ISME J* 7 (11), pp. 2169–2177. DOI: 10.1038/ismej.2013.110.

Letarov, A. V. (2020): History of Early Bacteriophage Research and Emergence of Key Concepts in Virology. In *Biochemistry. Biokhimiia* 85 (9), 1093-1010. DOI: 10.1134/S0006297920090096.

Lindell, Debbie; Jaffe, Jacob D.; Johnson, Zackary I.; Church, George M.; Chisholm, Sallie W. (2005): Photosynthesis genes in marine viruses yield proteins during host infection. In *Nature* 438 (7064), pp. 86–89. DOI: 10.1038/nature04111.

Loeb, T. (1960): Isolation of a Bacteriophage Specific for the F+ and Hfr Mating Types of Escherichia coli K-12. In *Science* 131 (3404), pp. 932–933. DOI: 10.1126/science.131.3404.932.

López-Bueno, Alberto; Tamames, Javier; Velázquez, David; Moya, Andrés; Quesada, Antonio; Alcamí, Antonio (2009): High diversity of the viral community from an Antarctic lake. In *Science (New York, N.Y.)* 326 (5954), pp. 858–861. DOI: 10.1126/science.1179287.

9. References

Louten, Jennifer (2016): Virus Structure and Classification. In *Essential Human Virology*, pp. 19–29. DOI: 10.1016/B978-0-12-800947-5.00002-8.

Lwoff, André Michel (1953): Lysogeny. In *Bacteriological Reviews* 17 (4), pp. 269–337. DOI: 10.1128/br.17.4.269-337.1953.

Madsen, Eugene L. (2005): Identifying microorganisms responsible for ecologically significant biogeochemical processes. In *Nature reviews. Microbiology* 3 (5), pp. 439–446. DOI: 10.1038/nrmicro1151.

Mäntynen, Sari; Laanto, Elina; Oksanen, Hanna M.; Poranen, Minna M.; Díaz-Muñoz, Samuel L. (2021): Black box of phage-bacterium interactions: exploring alternative phage infection strategies. In *Open biology* 11 (9), p. 210188. DOI: 10.1098/rsob.210188.

Martinez-Hernandez, Francisco; Fornas, Oscar; Lluesma Gomez, Monica; Bolduc, Benjamin; La Cruz Peña, Maria Jose de; Martínez, Joaquín Martínez et al. (2017): Single-virus genomics reveals hidden cosmopolitan and abundant viruses. In *Nature communications* 8, p. 15892. DOI: 10.1038/ncomms15892.

Middelboe, Mathias; Chan, Amy M.; Bertelsen, Sif K. (2010): Isolation and life cycle characterization of lytic viruses infecting heterotrophic bacteria and cyanobacteria. In Steven Wilhelm, Markus Weinbauer, Curtis Suttle (Eds.): Manual of Aquatic Viral Ecology: American Society of Limnology and Oceanography, pp. 118–133.

Millard, Andrew D.; Gierga, Gregor; Clokie, Martha R. J.; Evans, David J.; Hess, Wolfgang R.; Scanlan, David J. (2010): An antisense RNA in a lytic cyanophage links psbA to a gene encoding a homing endonuclease. In *The ISME journal* 4 (9), pp. 1121–1135. DOI: 10.1038/ismej.2010.43.

Moraru, Cristina (2021a): Gene-PROBER - a tool to design polynucleotide probes for targeting microbial genes. In *Systematic and applied microbiology* 44 (1), p. 126173. DOI: 10.1016/j.syapm.2020.126173.

Moraru, Cristina (2021b): VirClust, a tool for hierarchical clustering, core gene detection and annotation of (prokaryotic) viruses. In *BioRXiv*. DOI: 10.1101/2021.06.14.448304.

Moraru, Cristina; Varsani, Arvind; Kropinski, Andrew M. (2020): VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. In *Viruses* 12 (11). DOI: 10.3390/v12111268.

Mushegian, A. R. (2020): Are There 10(31) Virus Particles on Earth, or More, or Fewer? In *Journal of bacteriology* 202 (9). DOI: 10.1128/JB.00052-20.

Naville, Magali; Ghuillot-Gaudeffroy, Adrien; Marchais, Antonin; Gautheret, Daniel (2011): ARNold: a web tool for the prediction of Rho-independent transcription terminators. In *RNA biology* 8 (1), pp. 11–13. DOI: 10.4161/rna.8.1.13346.

Noble, R. T.; Fuhrman, J. A. (2000): Rapid virus production and removal as measured with fluorescently labeled viruses as tracers. In *Applied and environmental microbiology* 66 (9), pp. 3790–3797. DOI: 10.1128/AEM.66.9.3790-3797.2000.

Noguchi, Hideki; Taniguchi, Takeaki; Itoh, Takehiko (2008): MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. In *DNA research: an international journal for rapid publication of reports on genes and genomes* 15 (6), pp. 387–396. DOI: 10.1093/dnares/dsn027.

Orton, Joseph P.; Morales, Matheo; Fontenele, Rafaela S.; Schmidlin, Kara; Kraberger, Simona; Leavitt, Daniel J. et al. (2020): Virus Discovery in Desert Tortoise Fecal Samples: Novel Circular Single-Stranded DNA Viruses. In *Viruses* 12 (2). DOI: 10.3390/v12020143.

Owen, Siân V.; Canals, Rocío; Wenner, Nicolas; Hammarlöf, Disa L.; Kröger, Carsten; Hinton, Jay C. D. (2020): A window into lysogeny: revealing temperate phage biology with transcriptomics. In *Microbial genomics* 6 (2). DOI: 10.1099/mgen.0.000330.

Paez-Espino, David; Chen, I-Min A.; Palaniappan, Krishna; Ratner, Anna; Chu, Ken; Szeto, Ernest et al. (2017): IMG/VR. A database of cultured and uncultured DNA Viruses and retroviruses. In *Nucleic acids research* 45 (D1), pp. D457-D465. DOI: 10.1093/nar/gkw1030.

Paez-Espino, David; Eloe-Fadrosh, Emiley A.; Pavlopoulos, Georgios A.; Thomas, Alex D.; Huntemann, Marcel; Mikhailova, Natalia et al. (2016): Uncovering Earth's virome. In *Nature* 536 (7617), pp. 425–430. DOI: 10.1038/nature19094.

Paul, J. H. (1999): Microbial gene transfer: an ecological perspective. In *Journal of molecular microbiology and biotechnology* 1 (1), pp. 45–50.

Proctor, Lita M.; Fuhrman, Jed A. (1990): Viral mortality of marine bacteria and cyanobacteria. In *Nature* 343 (6253), pp. 60–62. DOI: 10.1038/343060a0.

Quaiser, Achim; Dufresne, Alexis; Ballaud, Flore; Roux, Simon; Zivanovic, Yvan; Colombet, Jonathan et al. (2015): Diversity and comparative genomics of Microviridae in Sphagnum-dominated peatlands. In *Frontiers in microbiology* 6, p. 375. DOI: 10.3389/fmicb.2015.00375.

Rahlff, Janina; Turzynski, Victoria; Esser, Sarah P.; Monsees, Indra; Bornemann, Till L. V.; Figueroa-Gonzalez, Perla Abigail et al. (2021): Lytic archaeal viruses infect abundant primary producers in Earth's crust. In *Nature communications* 12 (1), p. 4642. DOI: 10.1038/s41467-021-24803-4.

Rakhuba, D. V.; Kolomiets, E. I.; Dey, E. Szwajcer; Novik, G. I. (2010): Bacteriophage Receptors, Mechanisms of Phage Adsorption and Penetration into Host Cell. In *Pol J Microbiol* 59 (3), Article 10.33073/pjm-2010-023, pp. 145–155. DOI: 10.33073/pjm-2010-023.

Rand, Arthur C.; Jain, Miten; Eizenga, Jordan M.; Musselman-Brown, Audrey; Olsen, Hugh E.; Akeson, Mark; Paten, Benedict (2017): Mapping DNA methylation with high-throughput nanopore sequencing. In *Nature methods* 14 (4), pp. 411–413. DOI: 10.1038/nmeth.4189.

Ravin, Nikolai V. (2015): Replication and Maintenance of Linear Phage-Plasmid N15. In *Microbiology spectrum* 3 (1), PLAS-0032-2014. DOI: 10.1128/microbiolspec.PLAS-0032-2014.

Rohwer, Forest; Segall, Anca; Steward, Griegx; Seguritan, Victor; Breitbart, Mya; Wolven, Felise; Farooq Azam, Farooqx (2000): The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with nonmarine phages. In *Limnol. Oceanogr.* 45 (2), pp. 408–418. DOI: 10.4319/lo.2000.45.2.0408.

Rosario, Karyna; Dayaram, Anisha; Marinov, Milen; Ware, Jessica; Kraberger, Simona; Stainton, Daisy et al. (2012): Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). In *The Journal of general virology* 93 (Pt 12), pp. 2668–2681. DOI: 10.1099/vir.0.045948-0.

Roux, Simon; Brum, Jennifer R.; Dutilh, Bas E.; Sunagawa, Shinichi; Duhaime, Melissa B.; Loy, Alexander et al. (2016a): Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. In *Nature* 537 (7622), pp. 689–693. DOI: 10.1038/nature19366.

Roux, Simon; Brum, Jennifer R.; Dutilh, Bas E.; Sunagawa, Shinichi; Duhaime, Melissa B.; Loy, Alexander et al. (2016b): Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. In *Nature* 537 (7622), pp. 689–693. DOI: 10.1038/nature19366.

Roux, Simon; Krupovic, Mart; Daly, Rebecca A.; Borges, Adair L.; Nayfach, Stephen; Schulz, Frederik et al. (2019): Cryptic inoviruses revealed as pervasive in bacteria and archaea across Earth's biomes. In *Nature microbiology* 4 (11), pp. 1895–1906. DOI: 10.1038/s41564-019-0510-x.

Roux, Simon; Krupovic, Mart; Poulet, Axel; Debroas, Didier; Enault, François (2012): Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. In *PloS one* 7 (7), e40418. DOI: 10.1371/journal.pone.0040418.

Roux, Simon; Páez-Espino, David; Chen, I-Min A.; Palaniappan, Krishna; Ratner, Anna; Chu, Ken et al. (2021): IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. In *Nucleic acids research* 49 (D1), D764-D775. DOI: 10.1093/nar/gkaa946.

Roux, Simon; Solonenko, Natalie E.; Dang, Vinh T.; Poulos, Bonnie T.; Schwenck, Sarah M.; Goldsmith, Dawn B. et al. (2016c): Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. In *PeerJ* 4, e2777. DOI: 10.7717/peerj.2777.

Roznowski, Aaron P.; Fane, Bentley A. (2016): Structure-Function Analysis of the φX174 DNA-Piloting Protein Using Length-Altering Mutations. In *Journal of virology* 90 (17), pp. 7956–7966. DOI: 10.1128/JVI.00914-16.

Ruska, H. (1940): Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop. In *Naturwissenschaften* 28 (3), pp. 45–46. DOI: 10.1007/BF01486931.

Sanger, F.; Air, G. M.; Barrell, B. G.; Brown, N. L.; Coulson, A. R.; Fiddes, J. C. et al. (1977a): Nucleotide sequence of bacteriophage ϕX174 DNA. In *Nature* 265 (5596), pp. 687–695. DOI: 10.1038/265687a0.

Sanger, F.; Nicklen, S.; Coulson, A. R. (1977b): DNA sequencing with chain-terminating inhibitors. In *Proceedings of the National Academy of Sciences* 74 (12), pp. 5463–5467. DOI: 10.1073/pnas.74.12.5463.

Sertic, V.; Boulgakov, N. (1935): Classification et identication des typhi-phages. In *C. R. Soc. Bio.* 119, pp. 1270–1272.

Sharma, Sonika; Chatterjee, Soumya; Datta, Sibnarayan; Prasad, Rishika; Dubey, Dharmendra; Prasad, Rajesh Kumar; Vairale, Mohan G. (2017): Bacteriophages and its applications: an overview. In *Folia microbiologica* 62 (1), pp. 17–55. DOI: 10.1007/s12223-016-0471-x.

Shivam, Shashwat; Li, Guanlin; Lucia-Sanz, Adriana; Weitz, Joshua S. (2022): Timescales modulate optimal lysis-lysogeny decision switches and near-term phage reproduction. In *Virus evolution* 8 (1), veac037. DOI: 10.1093/ve/veac037.

Simon, Meinhard; Scheuner, Carmen; Meier-Kolthoff, Jan P.; Brinkhoff, Thorsten; Wagner-Döbler, Irene; Ulbrich, Marcus et al. (2017): Phylogenomics of *Rhodobacteraceae* reveals evolutionary adaptation to marine and non-marine habitats. In *The ISME journal* 11 (6), pp. 1483–1499. DOI: 10.1038/ismej.2016.198.

Simpson, Jared T.; Workman, Rachael E.; Zuzarte, P. C.; David, Matei; Dursi, L. J.; Timp, Winston (2017): Detecting DNA cytosine methylation using nanopore sequencing. In *Nature methods* 14 (4), pp. 407–410. DOI: 10.1038/nmeth.4184.

Sinsheimer, Robert L. (1959): A single-stranded deoxyribonucleic acid from bacteriophage ϕX174. In *J. Mol. Biol.* 1, pp. 43–53, checked on 8/6/2018.

Steinegger, Martin; Meier, Markus; Mirdita, Milot; Vöhringer, Harald; Haunsberger, Stephan J.; Söding, Johannes (2019): HH-suite3 for fast remote homology detection and deep protein annotation. In *BMC bioinformatics* 20 (1), p. 473. DOI: 10.1186/s12859-019-3019-7.

Sullivan, Matthew B.; Coleman, Maureen L.; Weigele, Peter; Rohwer, Forest; Chisholm, Sallie W. (2005): Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. In *PLoS biology* 3 (5), e144. DOI: 10.1371/journal.pbio.0030144.

Sun, Lei; Young, Lindsey N.; Zhang, Xinzheng; Boudko, Sergei P.; Fokine, Andrei; Zbornik, Erica et al. (2014): Icosahedral bacteriophage ΦX174 forms a tail for DNA transport during infection. In *Nature* 505 (7483), pp. 432–435. DOI: 10.1038/nature12816.

Suttle, Curtis A. (2007): Marine viruses--major players in the global ecosystem. In *Nat Rev Micro* 5 (10), pp. 801–812. DOI: 10.1038/nrmicro1750.

Székely, Anna J.; Breitbart, Mya (2016): Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. In *FEMS Microbiology Letters* 363 (6). DOI: 10.1093/femsle/fnw027.

Tisza, Michael J.; Pastrana, Diana V.; Welch, Nicole L.; Stewart, Brittany; Peretti, Alberto; Starrett, Gabriel J. et al. (2020): Discovery of several thousand highly diverse circular DNA viruses. In *eLife* 9. DOI: 10.7554/eLife.51971.

Touchon, Marie; Moura de Sousa, Jorge A.; Rocha, Eduardo Pc (2017): Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. In *Current opinion in microbiology* 38, pp. 66–73. DOI: 10.1016/j.mib.2017.04.010.

Townsend, Eleanor M.; Kelly, Lucy; Muscatt, George; Box, Joshua D.; Hargraves, Nicole; Lilley, Daniel; Jameson, Eleanor (2021): The Human Gut Phageome: Origins and Roles in the Human Gut Microbiome. In *Frontiers in cellular and infection microbiology* 11, p. 643214. DOI: 10.3389/fcimb.2021.643214.

Turner, Dann; Kropinski, Andrew M.; Adriaenssens, Evelien M. (2021): A Roadmap for Genome-Based Phage Taxonomy. In *Viruses* 13 (3). DOI: 10.3390/v13030506.

Twort, F. W. (1915): AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. In *The Lancet* 186 (4814), pp. 1241–1243. DOI: 10.1016/S0140-6736(01)20383-3.

Urbanek, Martyna O.; Nawrocka, Anna U.; Krzyzosiak, Wlodzimierz J. (2015): Small RNA Detection by in Situ Hybridization Methods. In *International journal of molecular sciences* 16 (6), pp. 13259–13286. DOI: 10.3390/ijms160613259.

van Cauwenberghe, Jannick; Santamaría, Rosa I.; Bustos, Patricia; Juárez, Soledad; Ducci, Maria Antonella; Figueroa Fleming, Trinidad et al. (2021): Spatial patterns in phage-Rhizobium coevolutionary interactions across regions of common bean domestication. In *The ISME journal* 15 (7), pp. 2092–2106. DOI: 10.1038/s41396-021-00907-z.

Waldor, Matthew; Mekalanos, John (1996): Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin.

Wang, Shuai; Yang, Yu; Jing, Jiaojiao (2022): A Synthesis of Viral Contribution to Marine Nitrogen Cycling. In *Frontiers in microbiology* 13, p. 834581. DOI: 10.3389/fmicb.2022.834581.

Weinbauer, Markus G. (2004): Ecology of prokaryotic viruses. In *FEMS microbiology reviews* 28 (2), pp. 127–181. DOI: 10.1016/j.femsre.2003.08.001.

Weitz, Joshua S.; Wilhelm, Steven W. (2012): Ocean viruses and their effects on microbial communities and biogeochemical cycles. In *F1000 biology reports* 4, p. 17. DOI: 10.3410/B4-17.

Wen, Kevin; Ortmann, Alice C.; Suttle, Curtis A. (2004): Accurate estimation of viral abundance by epifluorescence microscopy. In *Applied and environmental microbiology* 70 (7), pp. 3862–3867. DOI: 10.1128/AEM.70.7.3862-3867.2004.

Whitman, W. B.; Coleman, D. C.; Wiebe, W. J. (1998): Prokaryotes: the unseen majority. In *PNAS* 95 (12), pp. 6578–6583. DOI: 10.1073/pnas.95.12.6578.

Wickham, H. (2016): ggplot2: Elegant Graphics for Data Analysis: Springer International Publishing (Use R!). Available online at https://books.google.de/books?id=XgFkDAAAQBAJ.

Wildy, Peter (1971): Classification and nomenclature of viruses: Karger Publishers.

Wilhelm, Steven W.; Suttle, Curtis A. (1999): Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the structure and function of aquatic food webs. In *BioScience* 49 (10), pp. 781–788. DOI: 10.2307/1313569.

Wommack, K. E.; Colwell, R. R. (2000): Virioplankton: viruses in aquatic ecosystems. In *Microbiology and molecular biology reviews: MMBR* 64 (1), pp. 69–114.

Wright, Patrick R.; Georg, Jens; Mann, Martin; Sorescu, Dragos A.; Richter, Andreas S.; Lott, Steffen et al. (2014): CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. In *Nucleic acids research* 42 (Web Server issue), W119-23. DOI: 10.1093/nar/gku359.

Xiaoyu Sun; Serban L. Ilca; Juha T. Huiskonen; Minna M. Poranen (2018): Dual Role of a Viral Polymerase in Viral Genome Replication and Particle Self-Assembly. In *mBio* 9 (5), e01242-18. DOI: 10.1128/mBio.01242-18.

Young, Ry (2013): Phage lysis: do we have the hole story yet? In *Current opinion in microbiology* 16 (6), pp. 790–797. DOI: 10.1016/j.mib.2013.08.008.

Young, Ryland (2014): Phage lysis: Three steps, three choices, one outcome. In *Journal of Microbiology* 52 (3), pp. 243–258. DOI: 10.1007/s12275-014-4087-z.

Zayed, Ahmed A.; Lücking, Dominik; Mohssen, Mohamed; Cronin, Dylan; Bolduc, Ben; Gregory, Ann C. et al. (2021): efam: an expanded, metaproteome-supported HMM profile database of viral protein families. In *Bioinformatics (Oxford, England)*, Article btab451. DOI: 10.1093/bioinformatics/btab451.

Zhan, Yuanchao; Chen, Feng (2019): The smallest ssDNA phage infecting a marine bacterium. In *Environ Microbiol* 21 (6), pp. 1916–1928. DOI: 10.1111/1462-2920.14394.

Zheng, Qiang; Chen, Qi; Xu, Yongle; Suttle, Curtis A.; Jiao, Nianzhi (2018): A Virus Infecting Marine Photoheterotrophic Alphaproteobacteria (Citromicrobium spp.) Defines a New Lineage of ssDNA Viruses. In *Front. Microbiol.* 9, p. 403. DOI: 10.3389/fmicb.2018.01418.

Zucker, Falk; Bischoff, Vera; Olo Ndela, Eric; Heyerhoff, Benedikt; Poehlein, Anja; Freese, Heike M. et al. (2022): New Microviridae isolated from Sulfitobacter reveals two cosmopolitan subfamilies of single-stranded DNA phages infecting marine and terrestrial Alphaproteobacteria. In *Virus evolution* 8 (2), Article veac070. DOI: 10.1093/ve/veac070.

10. Appendix

A) Calculation of generation time for *Sulfitobacter dubius* SH24-1b

$$\text{G (Generation time)} = \frac{\text{t (time)}}{\text{n (number of generations)}}$$

$T_4 = 2^n \times T_3$

$1.55 \times 10^7 = 2^n \times 7.67 \times 10^6$ | : 7.67×10^6

$2.0208605 = 2^n$ | (log)

$0.305536334 = 0.301029996 \times n$ | : 0.301029996

n = 1.014969733

G = t/n

G = 2/1.014969733

G = 1.9705021

$T_6 = 2^n \times T_5$

$4.37 \times 10^7 = 2^n \times 2.17 \times 10^7$ | : 2.17×10^7

$2.013824885 = 2^n$ | (log)

$0.304021703 = 0.301029996 \times n$ | : 0.301029996

n = 1.009938236

G = t/n

G = 2/1.009938236

G = 1.980319121

Tab. A1: Overview of CFU/ml over time, values have been used to create the curve shown in Fig. 10. T2 was removed from the curve, the value has probably resulted from a sampling error.

Time point	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Time in min	0	120	240	360	480	600	720	840	960	1080	1200
CFU/ml	1,67E+06	4,00E+06	9,00E+06	7,67E+06	1,55E+07	2,17E+07	4,37E+07	7,93E+07	1,19E+08	1,10E+08	1,11E+08

Tab A2: Overview of OD_{600} over time in min, values have been used to create the curve shown in Fig. 10

Time point	T0	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Time in min	0	120	240	360	480	600	720	840	960	1080	1200
OD_{600}	0.02	0.04	0.065	0.110	0.190	0.334	0.625	0.760	0.887	0.898	1.000

B) Calculating the base to dye ratio for polynucleotides used in phage targeted direct-geneFISH

$$A_{base} = A_{260} - (A_{dye} \text{ x } CF_{260})$$

A_{260} = 0.824

A_{dye} = 0.743

CF_{260} = 0.43

A_{base} = 0.824 – (0.743 x 0.43) = > **0.50451**

$$\text{base dye} = (A_{base} \text{ x } \varepsilon_{dye}) \ / \ (A_{dye} \text{ x } \varepsilon_{base})$$

A_{base} = 0.50451

ε_{dye} = 80400

A_{dye} = 0.743

ε_{base} = 6600

base:dye = 0.50451 * 80400 / 0.743 * 6600 => **8.2716**

base:dye = 8.2716

C) ICBM5 protein sequences

>ICBM5_Gene_1
MADKATSPTTVSKKDRDDFELHVRLLSRAGYKTAEARTIAWLRGPAGLQEMLAGVKVAS

>ICBM5_Gene_2
MSAIAGALISGGASLLGGLFGRSSASKQQARQNEYNKPINIRKRAEEGGFNPLLWAGQGNIQMQPGPSGIMGSAIANAGLALA
DGMSEQRQLDLERTKLKQDQERLDALIEKQTIRPKVGGIYAGSQQTPSVARAPGRPLMNGAPQPGSAPVFNPPTEYNPIPDDG
PRLQTKVMRSDGMTSADPENPAEMEGDWWTWAREGTFWQNNNEILRRNTPETLHYKGRDALFPKMIDGARKAHKKAQED
FEKNPPKRRKLKGVNPNLNDKKW

>ICBM5_Gene_3
MVMNMSKYQRPTNTRRESRTIAGRFRGGKLAPVMASAFRESESAILSQQVTYELDPIAGRMITPIMAELISVYVPVQAIDALKN
PEEAYAGNTEVVRDKLLSGTPLFGLEDESEISKRLGVNPISVGGVKKVNEAARLAHNCAVNFLRQRKYVNTVKLLADNMNVTPA
LISQTVLDRLNAVLDPEDRVNGAVQLDLGNVRMPVEGVGVEYNAVPVTTPGWNDTENDRAADPHYFSDHAAGIRLAIASENG
RPAVYAVGEGQNAGGISLTDFYNAELMDSLVRQMRQIVDDNPEYGEEMVTRWAHGLSVDNGKTPWILHQSQQMFGNQFR
RAMDGANLDVAQSDMMQSLEFTVPVPPTELGGVVITFASVKPDETLGSQPHPFLSTTWEATNYVSDELARDPEAVTMRHLNS
DVPMVDEDTRALYIGHNGLKKAYISYGFNRHLDPTTVEAKTAIWQLEVPMSVTPESVIYPEDLDHYPFADQLAEVCTYQVSSTAT
VRTPMVFGPTPVEELAQIETDNVFEDV

>ICBM5_Gene_4
MQVNDAKRSVLLVGLNGGEVSVISADGEIIATEGVTAGRHKCSSWVPFMSNEGDELSFSGDVVPMVPNGGRVRPMAYGPGQ
FESGANPDFVVTSADRMARELDHKIRGLDQTAKKVEARIAQLNNLAERAKTPVEVAKEEKVDVIDDDDVSDVRSSGDDSDDTG
PDFGAEKVGAE

>ICBM5_Gene_5
MNDAPKKSNSGRKQSWLRAYARGLKGGTPKTLCESLRLWSIPQEQAQDVTKKTYLNAADGPMIWLGEQIVKRMEKAGYPSRI
FCGYRSPEQQDKEFAEGDSKARAYQSPHQFYEAVDIIHKTKAWNVSQDYWDTLAAIVRVVEREYAIDLVHGYDWGWDSAHIEI
ALWRQVPKRQIGKTGSNYPPSPWEREQRFKELLPAVYAAQHRP

>ICBM5_Gene_6
MCSDLIHIDGQQFACRKCNECITARKNGWVARAMAEKAVTAETFSVTLTYNDATQESRDAAKTFEYRHVKNWIKNLRRQIEYT
TGQTGLLRYLVAGERGSDKGRCHWHVILFCNADILTLGKMTHWPSGKLAADRAEIITTGKRKKRINWSLWPYGFVSFQEPDQ
WGMEYALAYALKDQFNIVSAAGTAREAHVSRTSAGMFRMSKKPPIGFPFLERKLNALDARGQLPVDLKIRVPDYKGYWYPTGA
IREYMLDRLRISNELYKAQHGRNAPQWTSLTRSVEQNEKDWERLIHGTEAQEEEQVEDFEEWQRSILLRTKEIRQQQIDRDTRK
RCGGLSACSRCLNSLTPQDFDAAARWAERQARKHGGYDAAEKHYRSENRCNPYCGYRELPTQKRAFKKGA

>ICBM5_Gene_7
MAKYFDYTRNYPGSERQRETFEEQNDKTPCYRRLLRPPYWAEKVRKLLDEWLGLVPR

>ICBM5_Gene_8
VPRQYAFRTAEQVKKLWWVSLSQLRICEALERKQKLRLVAARTLLEQCLLESQWPYSEESQKIPQLANLALYMSLFKAVLAALVC
RALTQFKYPLNNAFLT

>ICBM5_Gene_9
LSGRPVSSRRKYGPALFAEDIAREGQRANTYREELGRDAARNWAVASCWVRGYYRERTFAFTTHRYRTGAA

D) ICBM5 genome sequence

>Ascunsovirus_oldenburgi_ICBM5

TTCAGACGAAAGGATGCGCGACGATAAAATTAGCTCTTGCAAACCCCGTCCAAATTCGACTAAGGTCAATTTAGGGAATGGTTCCCGGCGCGGGTTTATCC
CGTGGTGCTAACAAAGGAAAAACACATGGCTGATAAAGCAACTTCCCCGACTACCGTCTCTAAAAAAGATCGTGATGATTTTGAGCTTCATGTTCGGCTCTT
GAGCCGGGCAGGTTACAAAACTGCCGAAGCTCGGACAATTGCATGGCTGCGCGGCCCTGCTGGGCTGCAAGAAATGCTTGCGGGCGTCAAGGTCGCAAG
CTAACATAGGTTGAGCGGTTCGCCGCTCAACATTAACCGCAAAAAAGGAGAGAAATGTGAGTGCATTGAAACACTTTCTTACAAACCGTTTGACAGTCGTG
GTTCTTACCTCGCTGGCAACGGCAGTAGGTACTGCGCTTGCGACGGAGTTTCCGTCCATTTATAGCGCAGTCTGTGCCTAATGTCGGCTATCGCTGGTGCTC
TTATCTCAGGTGGCGCAAGCCTCCTAGGTGGTCTGTTCGGGCGTTCTTCTGCTAGCAAGCAGCAAGCCCGACAGAATGAGTACAACAAGCCGATTAACATT
CGCAAGCGGGCCGAGGAGGGGGGGTTCAACCCCCTGCTCTGGGCCGGTCAAGGCAACATCCAAATGCAGCCGGGTCCGTCCGGCATCATGGGTTCCGCT
ATTGCGAATGCTGGCCTAGCTCTTGCCGATGGCATGAGCGAACAACGCCAGCTCGACCTTGAGCGTACCAAGCTCAAGCAAGATCAAGAGCGTCTCGACG
CTCTGATCGAAAAACAGACCATCCGGCCAAAGGTCGGCGGCATTTATGCCGGGTCGCAACAAACGCCTTCTGTAGCGCGCGCTCCCGGTCGCCCGCTTATG
AATGGCGCTCCTCAACCCGGCTCTGCGCCGGTCTTTAACCCGCCAACGGAGTACAACCCAATCCCCGATGATGGCCCTCGCCTGCAAACGAAAGTGATGCG
TAGCGATGGCATGACCTCGGCGGATCCTGAAAATCCCGCCGAAATGGAGGGCGATTGGTGGACGTGGGCCAGAGAGGGAACTTTCTGGCAAAACAACAA
CGAAATTCTGCGGCGTAATACGCCGGAGACGTTGCACTACAAGGGTCGCGATGCCTTGTTCCCGAAAATGATTGACGGGGCAAGGAAAGCGCATAAGAA
GGCTCAAGAGGACTTCGAGAAAAACCCGCCAAAACGCCGCAAGCTTAAAGGCGTCAACCCTAACCTAAACGACAAGAAATGGTAATGAACATGTCAAAGT
ATCAACGTCCTACAAACACACGCCGCGAAAGCCGGACCATCGCTGGCCGGTTCCGTGGCGGCAAGTTGGCTCCTGTTATGGCGTCCGCGTTCCGTGAGAG
CGAAAGTGCAATCCTTTCGCAACAAGTTACCTATGAACTTGACCCAATCGCGGGCCGTATGATTACGCCGATCATGGCGGAACTTATCTCTGTATATGTTCC
GGTCCAAGCGATCGACGCCCTAAAAAACCCTGAGGAGGCTTATGCCGGTAACACTGAGGTTGTCCGTGACAAGCTCCTCTCAGGAACGCCGCTGTTCGGC
CTCGAAGACGAAAGCGAGATTTCGAAGCGTCTAGGCGTTAACCCTATCTCTGTGGGTGGTGTCAAAAAAGTGAACGAGGCGGCACGCCTTGCGCATAACT
GCGCCGTGAACTTCCTACGTCAGCGCAAGTACGTCAACACGGTCAAGCTATTGGCCGATAATATGAACGTAACGCCTGCGCTGATCTCTCAAACGGTTCTC
GACCGGCTGAACGCCGTGCTCGATCCTGAGGATCGTGTCAACGGTGCTGTGCAGCTTGACCTTGGGAACGTGCGTATGCCGGTCGAAGGCGTCGGCGTTG
AATACAATGCCGTTCCAGTGACTACCCCCGGTTGGAACGATACAGAGAATGACCGTGCTGCTGATCCTCACTATTTCTCGGACCACGCCGCAGGCATTCGC
CTTGCGATTGCCTCGGAGAATGGTCGTCCAGCGGTCTACGCTGTTGGCGAGGGTCAAAATGCCGGCGGCATTTCCCTTACTGACTTCTACAACGCCGAACT
GATGGACAGTCTTGTGCGCCAGATGCGGCAGATTGTCGATGATAACCCTGAGTATGGCGAAGAAATGGTCACACGCTGGGCGCATGGTCTTTCTGTTGAT
AACGGCAAAACTCCTTGGATTTTGCACCAAAGCCAGCAAATGTTCGGCAATCAATTCCGCCGTGCCATGGATGGTGCAAACCTCGATGTTGCTCAATCCGA
CATGATGCAAAGCCTTGAATTCACTGTCCCCGTGCCTCCAACGGAATTGGGCGGGGTCGTGATTACCTTTGCATCTGTGAAGCCTGACGAAACACTAGGCT
CTCAGCCGCACCCGTTCCTGTCGACAACATGGGAGGCGACTAACTACGTCTCCGACGAGTTGGCACGCGACCCTGAGGCCGTGACCATGCGTCACCTGAAC
AGTGACGTTCCTATGGTCGATGAGGACACTCGTGCTCTCTACATCGGTCACAACGGTCTGAAAAAAGCCTACATCTCATATGGCTTTAACCGGCATCTCGAC
CCTACTACGGTCGAAGCGAAAACTGCCATCTGGCAGCTTGAGGTGCCAATGTCCGTAACGCCTGAAAGCGTTATCTATCCCGAGGACTTGGACCATTATCC
GTTTGCGGATCAGCTTGCAGAGGTCTGCACCTATCAGGTTAGTTCGACCGCTACAGTGCGGACGCCTATGGTCTTCGGTCCGACACCAGTTGAGGAACTGG
CCCAGATTGAGACAGACAACGTCTTCGAAGACGTATAAAAAAAGGGGGGCGGTTATGCCGCCCCTCAATAAACCTTAAAAAACTGGAAAGATGAAATGCA
AGTTAATGATGCAAAACGCTCCGTTCTTCTCGTCGGCCTTAATGGCGGCGAAGTTTCTGTAATCTCCGCAGATGGTGAGATTATCGCGACCGAAGGTGTAA
CCGCTGGTCGGCATAAATGCTCCTCTTGGGTGCCATTTATGTCCAATGAGGGCGATGAATTGAGCTTCTCCGGCGATGTGGTCCCAATGGTGCCAAATGGC
GGTCGTGTCCGGCCTATGGCCTACGGCCCCGGTCAATTTGAAAGCGGTGCCAATCCCGATTTCGTCGTTACTTCGGCGGATCGGATGGCTCGTGAGCTTGA
CCATAAAATTAGGGGTCTTGACCAGACCGCTAAAAAAGTTGAGGCCCGTATTGCGCAGTTGAACAATCTAGCAGAACGCGCTAAAACACCTGTAGAAGTA
GCAAAGGAGGAAAAAGTCGATGTTATTGATGATGACGATGTTTCTGACGTTCGTAGCTCTGGCGACGATAGTGACGATACGGGACCTGATTTTGGCGCTG
AAAAAGTGGGGGCAGAATGAATGATGCCCCGAAAAAAAGCAATTCGGGTCGCAAACAAAGCTGGCTCCGCGCGTATGCGCGGGGCCTCAAAGGCGGCA
CCCCGAAAACGCTCTGCGAAAGCCTGCGCCTGTGGTCAATCCCACAAGAGCAGGCGCAAGACGTGACAAAAAAAACTTACCTCAACGCCGCAGATGGCCC
CATGATCTGGCTCGGTGAGCAGATTGTCAAACGCATGGAAAAGGCGGGTTATCCCTCCCGCATTTTCTGCGGCTATCGTTCCCCAGAACAGCAAGACAAGG
AGTTTGCAGAGGGCGACAGCAAGGCTAGGGCCTATCAAAGCCCACATCAGTTTTATGAAGCGGTAGACATAATTCATAAGACAAAAGCATGGAACGTCTC
TCAAGATTATTGGGATACACTCGCCGCAATTGTGCGAGTTGTAGAGCGTGAATATGCTATAGACCTCGTGCATGGTTATGATTGGGGATGGGACAGTGCCC
ATATTGAGATTGCTTTGTGGCGTCAAGTTCCAAAGCGTCAAATTGGCAAAACTGGCTCTAATTATCCCCCGTCACCTTGGGAACGTGAACAGCGTTTCAAGG
AGTTGCTTCCCGCTGTTTATGCGGCGCAACATCGGCCATAGACTTATGGCAATCTATGCAAACTAACAAGCGGGCGGCCTCTATGCGAGAGCCGCCCGTTT
TGCATCCCTGCGGGCCAGTCGGAGACTGGCCAGCGGCAACTATATTCCGGGCCAAATCTTTTGGACGCGCCTTAGCGCCCCGGATGGCTACGCAGCGTTA
AGCTAGTGCCTAACAATGTGTCAGTAAGCTCGTCAACGTAAACCTCTCTCAGAATTGAGGACGGTTTGCGTTGATCCCCCTAGCAGGCCCCCCTTGTTCCTG
TATACACTTACTGACACAAAAAACAGGAGTTAAAGAATATGTGCAGTGATTTAATCCATATCGACGGGCAGCAATTTGCCTGTCGTAAGTGTAATGAGTGC
ATCACCGCTAGGAAAAATGGTTGGGTTGCTCGTGCGATGGCTGAAAAGGCCGTCACTGCTGAAACTTTCAGCGTAACCCTAACTTATAATGATGCTACTCA
GGAAAGCCGCGATGCGGCAAAGACGTTCGAGTATCGACACGTCAAAAACTGGATTAAGAACCTTAGGCGTCAAATAGAGTACACCACCGGCCAGACTGG
CCTTCTCCGCTACCTTGTAGCGGGTGAGCGCGGTTCCGATAAGGGCCGCTGCCACTGGCATGTAATCTTGTTCTGTAACGCCGATATTTTAACGCTAGGAAA
AATGACGCACTGGCCCTCTGGCAAACTTGCTGCCGATAGGGCCGAAATTATAACTACAGGAAAAAGGAAAAAACGTATTAACTGGTCTCTCTGGCCGTATG
GCTTCGTCAGCTTTCAAGAGCCTGACCAGTGGGGCATGGAGTACGCCTTAGCGTATGCCCTCAAAGATCAGTTTAACATTGTCTCCGCTGCCGGTACGGCG
CGGGAGGCTCACGTTTCCCGCACTTCTGCGGGTATGTTCCGCATGTCCAAAAAACCCCCAATCGGTTTCCCCTTTCTTGAGCGCAAGCTCAACGCGCTAGAT
GCGCGTGGTCAGCTTCCAGTTGACCTAAAAATAAGGGTTCCCGATTACAAAGGGTACTGGTATCCTACGGGAGCTATCCGCGAGTACATGCTCGACCGCTT
GCGGATCAGCAATGAGCTTTACAAAGCTCAGCATGGGCGTAATGCGCCACAATGGACCTCGCTAACGCGAAGCGTTGAGCAAAACGAAAAAGATTGGGA
AAGGTTAATCCATGGCACCGAAGCGCAAGAAGAGGAGCAAGTCGAAGACTTCGAGGAGTGGCAACGCTCAATCCTCCTCCGTACAAAAGAAATACGCCA
ACAACAGATCGACAGAGACACGCGAAAGCGATGTGGCGGGCTTTCTGCTTGCTCACGATGCCTTAACAGCCTCACGCCGCAAGATTTCGACGCGGCGGCG
AGATGGGCGGAACGCCAAGCTCGAAAACACGGCGGCTATGACGCCGCCGAAAAACACTACCGCAGCGAAAACCGCTGCAATCCCTATTGTGGCTACAGG
GAACTACCAACCCAGAAAAGAGCCTTCAAAAAAGGAGCATGACGACAGGAA

11. Declaration

I hereby assure that so far I have not taken any final examination or parts thereof in the above-mentioned PhD degree program - neither at a German nor at any other European institution of higher education and that I am not currently undergoing an examination procedure in the same or any other program at another institution.

I declare that I have developed and written the enclosed PhD Thesis by myself, and have not used sources or means without declaration in the text. Any thoughts from others or literal quotations are clearly marked.

Oldenburg, XX.XX.2022 Signature