Carl von Ossietzky
Universität
Oldenburg

Lytic bacteriophages of the Roseobacter group

Lytische Bakteriophagen der Roseobacter Gruppe

Dissertation

Zur Erlangung des akademischen Grades und Titels eines Doktors der Naturwissenschaften

- Dr. rer. nat. -

der Fakultät für Mathematik und Naturwissenschaften

der Carl-von-Ossietzky Universität Oldenburg

Vorgelegt von

Vera Bischoff

Geboren am 13. Oktober 1991 in Georgsmarienhütte

Oldenburg, Oktober 2023

Die vorliegende Dissertation wurde in der Zeit von Januar 2018 bis Oktober 2023 an der Carl von Ossietzky Universität Oldenburg am Institut für Chemie und Biologie des Meeres (ICBM) in der Arbeitsgruppe Biologie geologischer Prozesse angefertigt.

Die Arbeit wurde von der Deutschen Forschungsgemeinschaft (DFG) im Rahmen des Sonderforschungsbereiches „Roseobacter“ (SFB TRR 51) gefördert.

Erstgutachter: **Prof. Dr. Meinhard Simon (ICBM)**

Zweitgutachterin: **Assoc. Prof. Dr. Karin Holmfeldt (LNU)**

Betreuerin: **Dr. Cristina Moraru (ICBM)**

Disputationsdatum: 11.12.2023

Summary

Viruses are the most abundant biological entities in the ocean. Bacteriophages, or phages, are viruses that infect bacteria and quite often outnumber their hosts by an order of magnitude. Replicating via infection and lysis of bacteria, phages have an immense impact on bacterial cell abundances and community structures. In addition, phages influence their hosts´ metabolism and evolution through the expression of auxiliary metabolic genes and horizontal gene transfer. Therefore, and because of the release of bacterial cell debris upon viral lysis, phages are major players in global ocean biogeochemical cycles.

Members of the family *Roseobacteraceae*, operationally also termed "the Roseobacter group", are important heterotrophic bacteria in the marine environment, being most abundant in coastal waters and during algal blooms. They have been found worldwide in diverse aquatic habitats and great advances have been made in the recent years in uncovering their enormous genomic and metabolic diversity. In contrast, comparably little information has been gained so far about the phages infecting these bacteria, the so-called roseophages. Until now, phages have been isolated for only twelve out of more than 130 *Roseobacteraceae* genera.

The overall objective of this dissertation was to investigate the roseophage diversity in the North Sea by means of isolation, cultivation and whole genome sequencing. Seawater samples taken from the southern North Sea during spring and summer as well as from a mesocosm experiment with an artificially induced algal bloom served for phage isolation using two approaches: enrichment cultures and direct plating of seawater concentrated via tangential flow filtration. More than one hundred new roseophages infecting the ecologically relevant genera *Sulfitobacter*, *Lentibacter*, and *Octadecabacter* have been isolated and genome sequenced, revealing an impressive diversity on several levels.

In the first study presented in this dissertation, two dsDNA phages infecting *Lentibacter* sp. SH36 were isolated from enrichment cultures. Using database mining, we found related phages infecting the genera *Sulfitobacter* and *Celeribacter* and we showed that these cobaviruses are distributed worldwide in marine environments. We assigned these viruses to the new family *Zobellviridae* in the class *Caudoviricetes,* which was officially accepted by the International Committee on Taxonomy of Viruses (ICTV). Further on, a large-scale isolation campaign with direct plating yielded 128 new roseophage isolates with unique dsDNA genomes. These were grouped into twelve genus-level genomic clusters. Whole genome-based

classification of 28 representative phages assigned them to eight families in the class *Caudoviricetes*, comprising four already existing and four here newly proposed families. These new roseophages are diverse in terms of genome size, morphology and lifestyle, as predicted by the presence of lysogeny-related genes. The majority of the new isolates belonged to one large group of highly similar, potentially species-level related *Sulfitobacter* phages. Host range determination against their original isolation host strains revealed a complex infectivity network. Initial steps have been made to investigate closer the dynamics between these sulfiviruses and their hosts. Genome analysis of the phages as well as of thirty whole-genome sequenced host strains revealed a high degree of microdiversity on both sides. Finally, a small ssDNA phage infecting *Sulfitobacter* sp. SH24-1b was isolated, also from North Sea water. A preliminary characterization revealed its allocation to the *Microviridae* family. We found prophages related to this phage in the genomes of various taxa within the phyla *Proteobacteria* and *Bacteroidetes*, indicating that this type of prophages is widespread.

Overall, the phages isolated in this dissertation enrich the known diversity of roseophages in the North Sea and worldwide. With only a few host genera, an enormous diversity of roseophages was isolated, suggesting how versatile and complex the interaction with their hosts, the roseobacters, and their impact on marine ecosystems must be.

Zusammenfassung

Viren sind die am häufigsten vorkommenden biologischen Einheiten im Meer. Solche Viren, die Bakterien infizieren, werden Bakteriophagen, oder auch kurz Phagen genannt. Ihre Zahl ist im Meer zehnmal höher als die ihrer Wirte, der Bakterien. Da sich Phagen durch Infektion und die anschließende Lyse von Bakterien vermehren, haben sie einen immensen Einfluss auf die Häufigkeit bestimmter Bakterienzellen und die Struktur bakterieller Gemeinschaften. Durch die Freisetzung von Zellmaterial bei der viralen Lyse sind sie wichtige Akteure in den globalen biogeochemischen Kreisläufen der Ozeane. Außerdem beeinflussen Phagen den Stoffwechsel und die Evolution ihrer Wirte durch die Expression zusätzlicher Stoffwechsel-Gene und durch horizontalen Gentransfer.

Die Mitglieder der Familie *Roseobacteraceae*, die auch als "Roseobacter-Gruppe" bezeichnet wird, sind wichtige heterotrophe Bakterien im Meer. Sie sind besonders zahlreich in Küstengewässern vertreten sowie während Algenblüten und kommen weltweit in den verschiedensten aquatischen Lebensräumen vor. In den letzten Jahren wurden große Fortschritte bei der Erforschung ihrer enormen genomischen und metabolischen Diversität erzielt. Vergleichsweise wenig ist hingegen bisher bekannt über die Phagen, die diese Bakterien infizieren, die sogenannten Roseophagen. Bislang wurden nur für zwölf von mehr als 130 *Roseobacteraceae*-Gattungen Phagen isoliert.

Das übergeordnete Ziel dieser Dissertation war die Untersuchung der Roseophagen-Vielfalt in der Nordsee mittels deren Isolierung, Kultivierung und Genom-Sequenzierung. Meerwasserproben aus der südlichen Nordsee sowie von einem Mesokosmen-Experiment mit einer künstlich induzierten Algenblüte dienten zur Phagen-Isolierung. Dabei wurden zwei verschiedene Ansätze angewendet: Anreicherungskulturen und die direkte Ausplattierung von mittels Tangentialflussfiltration aufkonzentriertem Meerwasser. Auf diese Weise wurden mehr als einhundert neue Roseophagen isoliert, die die ökologisch relevanten Gattungen *Sulfitobacter*, *Lentibacter* und *Octadecabacter* infizieren. Die Sequenzierung ihrer Genome offenbarte eine beeindruckende Vielfalt in mehrfacher Hinsicht.

In der ersten Studie, die in dieser Dissertation vorgestellt wird, wurden zwei dsDNA-Phagen aus Anreicherungskulturen mit *Lentibacter* sp. SH36 isoliert. Durch eine Datenbankrecherche identifizierten wir verwandte Phagen, die die Gattungen *Sulfitobacter* und *Celeribacter* infizieren, und konnten zeigen, dass diese Cobaviren weltweit in marinen Lebensräumen verbreitet sind. Wir ordneten sie einer neuen Familie innerhalb der Klasse

Caudoviricetes zu, der Familie *Zobellviridae*, die vom International Committee on Taxonomy of Viruses (ICTV) offiziell anerkannt wurde. Des Weiteren wurden im Rahmen einer groß angelegten Isolierungskampagne mittels direkter Ausplattierung 128 neue dsDNA Roseophagen-Isolate gewonnen. Diese konnten in zwölf verschiedene Gattungen eingruppiert werden. Eine genombasierte Klassifizierung von 28 stellvertretenden Phagen ordnete diese acht verschiedenen Familien der Klasse *Caudoviricetes* zu, darunter vier bereits bekannte und vier neue Familien. Die neuen Roseophagen unterscheiden sich in Genomgröße, Morphologie und Vermehrungszyklus, wie das Vorhandensein lysogener Gene vermuten lässt. Die Mehrheit der neuen Isolate gehörte zu einer großen Gruppe sehr ähnlicher *Sulfitobacter*-Phagen, die möglicherweise als eine Art angesehen werden kann. Die Bestimmung des Wirtsspektrums in Bezug auf die ursprünglichen Isolierungswirtsstämme ergab ein komplexes Bild mit stark variierenden Wirtsspektren. Erste Schritte wurden unternommen, um die Dynamik zwischen diesen Sulfiviren und ihren Wirten näher zu untersuchen. Die Analyse der Phagen-Genome sowie der Genome von dreißig sequenzierten Wirtsstämmen zeigte ein hohes Maß an Mikrodiversität auf beiden Seiten. Das abschließende Kapitel dieser Dissertation beschreibt die Isolierung eines kleinen ssDNA-Phagen aus der Nordsee, der *Sulfitobacter* sp. SH24-1b infiziert. Eine vorläufige Charakterisierung ergab seine Zugehörigkeit zur Familie der *Microviridae*. Zudem identifizierten wir verwandte Prophagen in den Genomen verschiedenster Taxa innerhalb der Phyla *Proteobacteria* und *Bacteroidetes*, was darauf hinweist, dass diese Art von Prophagen weit verbreitet ist.

Insgesamt bereichern die in dieser Dissertation isolierten Phagen die bekannte Vielfalt der Roseophagen in der Nordsee und weltweit. Mit nur wenigen Wirtsgattungen wurde eine enorme Diversität an Roseophagen isoliert. Dies verdeutlicht, wie vielseitig und komplex die Interaktionen mit ihren Wirten, der Roseobacter-Gruppe, und damit ihr Einfluss auf marine Ökosysteme sein müssen.

Publication record

Peer-reviewed publications

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), S. 1404–1421. DOI: 10.1038/s41396-019-0362-7. (Included in this dissertation, see chapter 2)

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), Artikel veac070. DOI: 10.1093/ve/veac070. (Shared first-authorship; partly included in this dissertation, see chapter 6)

Book chapters

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

Phage taxonomic proposals

Bischoff, Vera; Adriaenssens, Evelien M.; Kropinski, Andrew M.; Duhaime, Melissa B.; Moraru, Cristina (2020): ICTV-proposal 2020.187B, *Zobellviridae*. Create one new family (*Zobellviridae*) including one new subfamily (*Cobavirinae*), seven new genera and 12 new species (*Caudovirales*). (Included in this dissertation, see chapter 3)

Contributions to national and international conferences

Bischoff, Vera; Bunk, Boyke; Meier-Kolthoff, Jan P.; Spröer, Cathrin; Poehlein, Anja; Dogs, Marco et al. (2018): Cobaviruses - a phage group infecting marine *Rhodobacteraceae* is found in highly productive marine regions. Talk, Annual Conference of the Association for General and Applied Microbiology (VAAM), Wolfsburg.

Bischoff, Vera; Bunk, Boyke; Meier-Kolthoff, Jan P.; Spröer, Cathrin; Poehlein, Anja; Dogs, Marco et al. (2018): Cobaviruses - a phage group infecting marine *Rhodobacteraceae* is found in highly productive marine regions. Poster, EMBO Workshop Viruses of Microbes, Wrocław, Poland.

Bischoff, Vera; Heyerhoff, Benedikt; Poehlein, Anja; Zucker, Falk; Alejandre-Colomo, Carlota; Heins, Anneke; Pradella, Silke; Daniel, Rolf; Simon, Meinhard; Moraru, Cristina (2019): Large scale isolation of bacteriophages infecting marine *Rhodobacteraceae*. Poster, Marine Microbiota, Oldenburg.

Bischoff, Vera; Heyerhoff, Benedikt; Poehlein, Anja; Zucker, Falk; Alejandre-Colomo, Carlota; Heins, Anneke; Pradella, Silke; Daniel, Rolf; Simon, Meinhard; Moraru, Cristina (2019): Large scale isolation of bacteriophages infecting marine *Rhodobacteraceae*. Poster, Symposium of Aquatic Microbial Ecology (SAME), Potsdam.

List of abbreviations

Abi	Abortive infection
AMG	Auxiliary metabolic gene
ANI	Average nucleotide identity
ASW	Artifical saltwater
BC	Baltimore Classes
BLAST	Basic Local Alignment Search Tool
Bp	Base pairs
BSA	Bovine serum albumin
BSC	Biological Species Concept
CRISPR/Cas	Clustered regularly interspaced short palindromic repeats / cascade
CsCl	Cesium chloride
dDDH	Digital DNA-DNA hybridization
DJR-MCP	Double-jelly-role major capsid protein
DMSP	Dimethylsulfoniopropionate
DOC	Dissolved organic carbon
DOM	Dissolved organic matter
dsDNA	Double-stranded DNA
DSMZ	Deutsche Sammlung für Mikroorganismen und Zellkulturen
dsRNA	Double-stranded RNA
DTR	Direct terminal repeat
EDTA	Ethylenediaminetetraacetic acid
ENA	European Nucleotide Archive
FISH	Fluorescence in situ hybridization
HAB	Harmful algal bloom
HGT	Horizontal gene transfer
HK97-MCP	HK97-like major capsid protein
ICBM	Institute of Chemistry and Biology of the Marine Environment, Oldenburg, Germany
ICTV	International Committee on Taxonomy of Viruses
IMEDEA	Mediterranean Institute for Advanced Studies, Esporles, Spain
IPP	Intact/active prophage
ITS	Internal transcribed spacer
IVP	Internal virion proteins
kb	Kilobases
LPSN	List of Prokaryotic names
MB	Marine broth
ML phylogeny	Maximum likelihood phylogeny
MPI	Max-Planck-Institute for Marine Microbiology, Bremen, Germany
NBII	Nucleotide-based intergenomic identity

NC	Negative control
NCBI	National Center for Biotechnology Information
OD_{600}	Optical density at 600 nm
OMV	Outer membrane vesicles
ORF	Open reading frame
pAgo	Prokaryotic argonaute
PC	Positive control (chapter 2+3), protein cluster (chapter 4+5)
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PFGE	Pulsed-field gel electrophoresis
PHROGS	Prokaryotic Virus Remote Homologous Groups
PSCs	Protein super clusters
pVOGs	Prokaryotic Virus Orthologous Groups
RAPD	Randomly amplified polymorphic DNA
RBP	Receptor binding protein
Rcf	Relative centrifugal field
RM	Restriction-modification
RNR	Ribonucleotide reductase
rpm	Rounds per minute
rRNA	Ribosomal RNA
SAR endolysin	Signal anchor release endolysin
SM buffer	Sodium chloride Magnesium sulphate buffer
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
TEM	Transmission electron microscopy
VGCs	Viral genome clusters
VHG	Virus hallmark gene
VLP	Virus-like particle
VOGs	Virus Orthologous Groups

List of figures

Fig. 1: Cartoon illustrating the ways in which phages influence the composition of marine bacterial communities 2
Fig. 2: Major morphological groups of bacteriophages 3
Fig. 3: Taxonomy of the class *Caudoviricetes*, which contains all tailed phages. 5
Fig. 4: The lytic and lysogenic life cycle of bacteriophages. 7
Fig. 5: Viral DNA structure inside the virion before infection, circularization, and packaging mechanism of different phages with different DNA packaging strategies. 11
Fig. 6: a. Enrichment of phages specific for *Lentibacter* sp. SH36 from North Sea water. 37
Fig. 7: Phylogenetic positioning of the *Lentibacter* sp. SH36 viruses and their relatives within the *Podoviridae*. 41
Fig. 8: Genome map of cultured and environmental cobaviruses. 46
Fig. 9: Phylogenetic analysis of the spanin **(a)**, glutaredoxin **(b)** and ribonucleotide reductase **(c–e)** genes from cobaviruses. 48
Fig. 10: Global distribution of cobaviruses 51
Fig. 11: Whole genome phylogenetic tree depicting the new *Zobellviridae* family. 64
Fig. 12: Phylogenetic positioning of the terminase proteins belonging to the phages in the *Zobellviridae* family. 65
Fig. 13: Nucleotide based, intergenomic similarities between the members of the here newly proposed *Zobellviridae* family. 66
Fig. 14: Transmission electron micrograph of molybdenum stained, cell debris bound Lentibacter virus vB_LenP_ICBM1 and uranyl acetate stained, free Lentibacter virus vB_LenP_ICBM2. 67
Fig. 15: Transmission electron micrograph of Pseudoalteromonas phage HP1 67
Fig. 16: Workflow of phage isolation 70
Fig. 17: Scheme of the streak assay for phage purification. 76
Fig. 18: Map of the german bight showing sampling sites for the isolation of bacterial strains (red) and phages (blue). 91
Fig. 19: Neighbor-joining tree based on 16S rRNA gene sequence similarity showing the phylogenetic affiliation of the bacterial strains used for phage isolation in this study 92
Fig. 20: Gel electrophoresis of RAPD-PCR products for 13 phage isolates. 93
Fig. 21: Nucleotide-based intergenomic identities of 94 unique and complete roseophage genomes, calculated with VIRIDIC. 103

Fig. 22: Whole-genome based proteomic tree of 965 dsDNA phages. 107
Fig. 23: Origins of the new phage isolates from different North Sea water samples 112
Fig. 24: Section of the whole-genome based proteomic tree showing the *Autographiviridae* family. 115
Fig. 25: Nucleotide-based intergenomic identities of phages in clusters A and B of the *Autographiviridae* family (VGC_4), calculated with VIRIDIC. 119
Fig. 26: PC-based intergenomic similarities of phages in clusters A and B of the *Autographiviridae* family (VGC_4), calculated with VirClust..... 120
Fig. 27: Genome map of members of the *Autographiviridae* family. 121
Fig. 28: Section of the whole-genome based proteomic tree showing VGC_9..... 125
Fig. 29: Nucleotide-based intergenomic identities of phages in the *Casjensviridae* family (VGC_9), calculated with VIRIDIC. 129
Fig. 30: PC-based intergenomic similarities of phages in the *Casjensviridae* family (VGC_9), calculated with VirClust..... 130
Fig. 31: Genome map Octadecabacter phage ICBM156, member of VGC_9. 130
Fig. 32: Section of the whole-genome based proteomic tree showing VGC_11..... 135
Fig. 33: Nucleotide-based intergenomic identities of phages in VGC_11, calculated with VIRIDIC..... 139
Fig. 34: PC-based intergenomic similarities of phages in VGC_11, calculated with VirClust 140
Fig. 35: Genome map of *Sulfitobacter* phages ICBM129 and ICBM153. 141
Fig. 36: Section of the whole-genome based proteomic tree showing VGC_25..... 148
Fig. 37: Internal clustering of VGC_25. 150
Fig. 38: Genome maps of VGC_25. 151
Fig. 39: Section of the whole-genome based proteomic tree showing VGC_19..... 160
Fig. 40: Nucleotide-based intergenomic identities of the phages in VGC_19 calculated with VIRIDIC..... 163
Fig. 41: PC-based intergenomic similarities of the phages in VGC_19, calculated with VirClust 164
Fig. 42: Genome maps of VGC_19. 165
Fig. 43: Section of the whole-genome based proteomic tree showing VGC_6..... 172
Fig. 44: Nucleotide-based intergenomic identities of the phages in VGC_6, calculated with VIRIDIC..... 176

Fig. 45: PC-based intergenomic similarities of the phages in VGC_6, calculated with VirClust 177
Fig. 46: Genome map of *Sulfitobacter* phages ICBM111, ICBM117, ICBM118, ICBM130 and ICBM55, members of the "Viktorvirus" genus. 178
Fig. 47: Section of the whole-genome based proteomic tree showing the *Zobellviridae* family. 181
Fig. 48: Nucleotide-based intergenomic identities of the phages in the *Zobellviridae* family, calculated with VIRIDIC. 185
Fig. 49: PC-based intergenomic similarities of the phages in the *Zobellviridae* family, calculated with VirClust 186
Fig. 50: Genome map of the *Lentibacter* phages belonging to the *Zobellviridae* family. 186
Fig. 51: Host range determination. 198
Fig. 52: Neighbor-joining tree based on the 16S rRNA gene and ITS region of 46 sulfivirus host strains 200
Fig. 53: Neighbor-joining tree based on the 16S rRNA gene of 46 sulfivirus host strains .. 201
Fig. 54: Hierachical clustering tree based on whole genome distances of 30 *Sulfitobacter* host strains plus 17 related strains. 202
Fig. 55: TEM image of uranyl-acetate stained virions of *Sulfitobacter* phages ICBM16 and ICBM18. 204
Fig. 56: Nucleotide-based intergenomic identities within the "Sulfivirus" genus, as calculated with VIRIDIC. 209
Fig. 57: Whole genome-based phylogeny of the 89 sulfiviruses and reference genomes 210
Fig. 58: Hierarchical clustering of the sulfivirus genomes when using relaxed parameters for protein clusterings. 213
Fig. 59: Genome map of Sulfitobacter phage ICBM128. 214
Fig. 60: Hierarchical clustering of the sulfivirus genomes when the intergenomic distances were calculated using protein clusters containing identical proteins 220
Fig. 61: Reconstructed phylogenetic tree of sulfiviral core regions and sites of recombination and polymorphism detected with ClonalFrameML. 221
Fig. 62: Distribution of number of infected hosts among sulfiviruses. 222
Fig. 63: Host ranges of 89 sulfiviruses tested against their original isolation hosts (60 strains). 224
Fig. 64: Host ranges clustered by ClonalFrameML phylogenetic tree. 224

Fig. 65: Host ranges of 89 sulfiviruses tested against their original isolation hosts (60 strains), clustered based on isolation hosts. 225
Fig. 66: Susceptibility of *Sulfitobacter* strains to sulfivirus infection clustered based on the presence/absence of plasmid clusters. 231
Fig. 67: Nucleotide-based intergenomic identity of the active prophages predicted in the genomes of 30 *Sulfitobacter* strains and representatives of sulfiviruses, calculated with VIRIDIC. 232
Fig. 68: (A) ICBM5 morphology determined by TEM of uranyl-acetate-stained virions. **(B)** Agarose gel shows enzymatic digestion of ICBM5 ssDNA phage. 243
Fig. 69: 16S rRNA phylogenetic analysis of the (pro)-phage hosts. 252

List of tables

Table 1: Abundance of cobaviruses in the S1 and S2 phage enrichments. 40
Table 2: Environmental cobaviruses, retrieval from databases. 40
Table 3: Seasonal occurrence of cobaviruses in different locations. 52
Table 4: Parameters used by VICTOR to calculate a whole genome tree based on amino acid sequences and to demarcate taxa. 63
Table 5: Newly proposed taxonomy of the *Zobellviridae* family. 68
Table 6: Bacterial strains used for phage isolation in this study. 72
Table 7: Primers used for 16S - 23s rRNA fragment amplification and sequencing. 73
Table 8: Seawater sources for phage isolation. 75
Table 9: Previously described dsDNA roseophages collected from literature for whole-genome based classification. 85
Table 10: Pairing of the phage water samples with the bacterial hosts during the phage isolation procedures. 90
Table 11: Counts of phage isolates, sequenced and final unique phage genomes. 95
Table 12: Sizes of complete phage genomes comparing PFGE and sequencing. 95
Table 13: Total unique phages isolated and genome sequenced in this study. 98
Table 14: Clustering of 94 unique and complete roseophage genomes into genera and species clusters. 103
Table 15: PhageTerm results of genome end determination for the representative ICBM phages. 108
Table 16: Overview of the newly described and classified ICBM phages. 109
Table 17: Phages of cluster A and B in the *Autographiviridae* family. 116
Table 18: Gene annotations of Lentibacter phage ICBM157. 121
Table 19: Core proteins of the newly proposed subfamily “Incetivirinae” in the family *Autographiviridae*. 123
Table 20: Phages of VGC_9. 126
Table 21: Gene annotations of Octadecabacter phage ICBM156. 131
Table 22: Phages of VGC_11. 136
Table 23: tRNAs found in the genome of Sulfitobacter phage ICBM153. 140
Table 24: Gene annotations of Sulfitobacter phage ICBM129. 141
Table 25: Core proteins of the newly proposed subfamily “Maresulfivirinae” in the family *Mesyanzhinoviridae*. 144

Table 26: Phages of VGC_25. 149
Table 27: tRNAs found in the genomes of "Sulfivirus" phages. 151
Table 28: Gene annotations of Lentibacter phage ICBM16. 151
Table 29: Gene annotations of Lentibacter phage ICBM7. 154
Table 30: Core proteins of the newly proposed families "Hayaniviridae" and "Schlingloffviridae". 156
Table 31: Phages of VGC_19. 161
Table 32: Gene annotations of Lentibacter phage ICBM121. 165
Table 33: Gene annotations of Lentibacter phage ICBM165. 167
Table 34: Gene annotations of Lentibacter phage ICBM8. 168
Table 35: Core proteins of the newly proposed family "Diferiteviridae". 169
Table 36: Phages of VGC_6. 173
Table 37: Gene annotations of Sulfitobacter phage ICBM55. 178
Table 38: Core proteins of the newly proposed family "Woolleyviridae". 179
Table 39: Phages of the *Zobellviridae* family. 182
Table 40: Gene annotations of Lentibacter phage ICBM6. 187
Table 41: Core proteins of the subfamily *Cobavirinae* in the family *Zobellviridae*. 188
Table 42: Phages of the "Sulfivirus" genus. 205
Table 43: Core proteins of the "Sulfivirus" genus. 211
Table 44: Non-core protein clusters of the "Sulfivirus" genus. 212
Table 45: tRNAs predicted in the genomes of the "Sulfivirus" genus. 217
Table 46: Gene annotations of Lentibacter phage ICBM128. 218
Table 47: Genome characteristics of 30 sulfivirus host strains. 228
Table 48: Plasmids of 30 sulfivirus host strains. 229
Table 49: Intact prophages predicted in the plasmids and chromosomes of 30 sulfiviruses strains. 230
Table 50: List of *Rhodobacteraceae* strains used for the host range assay. 244
Table 51: Bacterial strains containing predicted ICBM5-like regions. 246

Table of contents

Summary V
Zusammenfassung VII
Publication record IX
List of abbreviations XI
List of figures XIII
List of tables XVII
Table of contents XIX
1. Introduction 1
1.1. Importance of (marine) bacteriophages 1
1.2. Classification of bacteriophages 2
1.2.1. Traditional classification of bacteriophages 2
1.2.2. The megataxonomy of viruses 4
1.3. Lifestyles of bacteriophages 6
1.3.1. Lytic life cycle 6
1.3.2. Lysogenic life cycle 8
1.3.3. Pseudolysogeny and chronic infections 8
1.3.4. DNA packaging strategies 9
1.4. Phage-host interactions 12
1.4.1. Host range of bacteriophages 12
1.4.2. The arms-race between phages and bacteria 14
1.5. Phages infecting marine *Roseobacteraceae* 16
1.6. How to investigate phage diversity 21
1.7. Aims and outline of this dissertation 22
2. Cobaviruses - a new globally distributed phage group infecting *Rhodobacteraceae* in marine ecosystems 25
2.1. Chapter summary 26
2.2. Materials and methods 27
2.2.1. Cultivation media 27
2.2.2. Phage enrichments and isolation of ICBM1 and ICBM2 phages 27
2.2.3. Preparation of phage ICBM1 and ICBM2 glycerol stocks 29
2.2.4. ICBM1 and ICBM2 phage high titer lysates 29
2.2.5. Transmission electron microscopy 29
2.2.6. Isolation and purification of phage DNA for sequencing 30
2.2.7. Genome sequencing and assembly 32

2.2.8. Retrieval of phage genomes related to ICBM1 and ICBM2 33
2.2.9. Genome-based phylogeny and classification 34
2.2.10. Genome annotation and protein clustering 34
2.2.11. Phylogenetic analyses of single proteins 35
2.2.12. Phylogenetic analysis of the host rRNA 35
2.2.13. Biogeographic distribution of cobaviruses and read mapping 35
2.3. Results and discussion 36
2.3.1. Isolation and host range of two *Lentibacter* sp. SH36 viruses 36
2.3.2. Sequencing the phage isolates and enrichments 37
2.3.3. Retrieval of similar phage genomes and phylogenetic positioning 38
2.3.4. Genomic organization 42
2.3.5. Cobaviral hosts 47
2.3.6. Environmental distribution of the cobaviruses 49
2.3.7. Protists as habitat for the cobaviral hosts 52
2.3.8. Conclusions 54
3. Taxonomic proposal of the *Zobellviridae* family 55
3.1. Short description 56
3.2. Abstract 56
3.3. Text of proposal 56
3.3.1. Zobellviridae family 58
3.3.2. Cobavirinae subfamily 59
3.3.2.1. Genus Siovirus 59
3.3.2.2. Veravirus genus 60
3.3.2.3. Genomic organization of cobaviruses 61
3.3.3. Other genera in the *Zobellviridae* family 61
3.3.3.1. Melvirus genus 61
3.3.3.2. Paundecimvirus genus 62
3.3.3.3. Citrovirus genus 62
3.3.3.4. Salinovirus genus 62
3.3.3.5. Vipivirus genus 63
3.3.3.6. Icepovirus genus 63
3.4. Supporting evidence 63
4. Isolation and classification of novel dsDNA roseophages 69
4.1. Chapter summary 69
4.2. Materials and methods 71
4.2.1. Cultivation media 71

4.2.2. Origin and cultivation of host strains 71
4.2.3. 16S sequencing and phylogenetic analysis of host strains 72
4.2.4. Water samples for phage isolation 73
4.2.5. Isolation of phages from direct plating and their further characterization 75
4.2.5.1. Direct phage isolation by plaque assay 75
4.2.5.2. Preparation of fresh lysates and glycerol stocks 76
4.2.5.3. PFGE for phage genome size determination/prediction 77
4.2.5.4. Phage DNA extraction 78
4.2.5.5. Screening for unique phages with RAPD-PCR 79
4.2.6. Isolation of phages from enrichments 79
4.2.7. Phage genome sequencing, assembly and genome end determination 82
4.2.8. Clustering at species and genus level - VIRIDIC analysis 82
4.2.9. Retrieval of related phage genomes and of other roseophages from sequence databases 82
4.2.10. VirClust analysis – hierarchical clustering and genome annotation 83
4.3. Results 90
4.3.1. Strains from three *Roseobacteraceae* genera served for the isolation of 277 phages.... 90
4.3.2. Quick screening for unique phages by RAPD-PCR 93
4.3.3. Genome sequencing and assembling of selected phage isolates 94
4.3.4. The new roseophage isolates belong to 12 genus-level genomic clusters 102
4.3.5. Taxonomic classification and genomic characterization of selected new phage isolates…. 105
4.3.5.1. First roseophage isolates to cluster within the *Autographiviridae* family 113
4.3.5.2. A novel roseophage infecting *Octadecabacter* falls within the *Casjensviridae* family…… 123
4.3.5.3. Two new *Sulfitobacter* infecting roseophages form a new subfamily within the *Mesyanzhinoviridae* family 133
4.3.5.4. “Hayaniviridae” and “Schlingloffviridae” - two new families of lytic roseophages 145
4.3.5.5. The “Diferiteviridae” - a new family of lytic and potentially temperate roseophages infecting *Lentibacter* and *Sulfitobacter* 157
4.3.5.6. The “Woolleyviridae” - a new family of potentially temperate roseophage infecting *Sulfitobacter* 170
4.3.5.7. The *Zobellviridae* family is extended by a new roseophage species 180
4.4. Discussion 188
4.4.1. Thresholds for phage taxonomic classification 189
4.4.2. Diversity of the discovered phage families 190

4.4.3. Genetic particularities of the new roseophages 190
4.4.4. Conclusions 191
5. Microdiversity of the sulfiviruses and their hosts (preliminary results) 193
5.1. Chapter summary 193
5.2. Materials and methods 194
5.2.1. 16S / ITS phylogenetic analysis of sulfivirus host strains 194
5.2.2. Whole genome sequencing of sulfivirus host strains 194
5.2.3. Phylogenomic analysis of the sulfivirus host strains 195
5.2.4. Initial comparative analysis of host genomes 195
5.2.5. Phage genomic analysis 196
5.2.5.1. Clustering of sulfivirus genomes on species and genus level 196
5.2.5.2. Protein-based hierarchical clustering of the sulfivirus genomes and gene annotations using VirClust 196
5.2.5.3. Detection of recombination events 197
5.2.6. Host range determination 197
5.2.7. TEM of sulfiviruses ICBM16 and ICBM18 198
5.3. Results and discussion 199
5.3.1. A micro-diverse bacterial collection served for isolation of Sulfivirus phages 199
5.3.2. 89 phage isolates of one genus 203
5.3.2.1. Composition of sulfiviral genomes 214
5.3.3. Sulfivirus genomes are strongly influenced by recombination 221
5.3.4. Sulfiviruses display wide range of host specificity 222
5.3.5. Genome analysis of 30 sulfivirus host strains 225
5.3.6. Conclusions 232
6. Isolation and preliminary characterization of ICBM5, a ssDNA phage infecting *Sulfitobacter* sp. SH24-1b 235
6.1. Chapter summary 235
6.2. Materials and Methods 236
6.2.1. Cultivation media 236
6.2.2. Isolation of phage ICBM5 236
6.2.3. Host range of phage ICBM5 237
6.2.4. Purification of phage ICBM5 via CsCl gradient ultracentrifugation 237
6.2.5. Transmission electron microscopy of phage ICBM5 238
6.2.6. Testing the ssDNA nature of the ICBM5 phage genome 238
6.2.7. Sequencing of the ICBM5 phage genome via Illumina sequencing 239
6.2.8. Assembly and annotation of the ICBM5 phage genome 240

6.2.9. Detection of ICBM5-like regions in bacterial genomes 240
6.2.10. Phylogenetic analysis of all host 16S rRNA genes and species assignment for *Sulfitobacter dubius* SH24-1b 240
6.3. Results and discussion 241
7. General discussion and outlook 253
7.1. The complex and eternally evolving field of phage classification and taxonomy 254
7.2. The role of roseophages in the marine environment, more specifically during algal blooms 256
7.3. Methodological considerations 260
7.4. Outlook 262
Appendix 265
a) Supplementary tables 265
b) Supplementary figures 280
c) Supplementary text 307
d) List of supplementary files 310
Bibliography 313
Acknowledgements 355
Erklärung 357

1. Introduction

1.1. Importance of (marine) bacteriophages

Viruses that infect bacteria are termed “bacteriophages” (or short “phages”). In marine systems, phages are the most abundant biological entities. They usually outnumber their bacterial hosts by an order of magnitude, with abundances of about 10^7 virus-like particles per milliliter (Cochlan et al. 1993; Suttle 2005). However, they are not only present in marine habitats. Phages have been found in all types of ecosystems, ranging from deserts (Fancello et al. 2013) to Antarctic lakes (Potter et al. 2018), from the human gut (Gantzer et al. 2002) to wastewater (Göller et al. 2021). They show an extreme morphological and genomic diversity. Their size ranges from very small, as for example single-stranded DNA (ssDNA) phage phiX174 (*Microviridae*) with a capsid diameter of 26 nm and a genome of 5.386 kilobases (kb) (McKenna et al. 1992), to very large, for example “jumbo phages”, with capsid sizes above 200 nm and genomes larger than 200 kb (Malone et al. 2020). A virus is not considered a living organism, because it requires the metabolism and DNA replication machinery of its host to produce its own progeny. Isolation of phages in the laboratory is dependent on cultivation of the respective host bacterium. Since many bacteria are difficult to cultivate under laboratory conditions, the majority of phages remains uncultivated and detailed knowledge about their lifestyle and interactions with hosts is limited to phages infecting comparably few bacterial taxa.

Not only are bacteriophages abundant in the marine environment, they also play a major role in ecosystem functioning. They shape microbial communities simply by lysing their host, but also by altering the host´s metabolism, for example through expression of auxiliary metabolic genes (AMGs), and by maintaining microbial genetic diversity through horizontal gene transfer (HGT) (Fig. 1) (Fuhrman 1999; Suttle 2005; Brum et al. 2015). Furthermore, phages are major drivers of marine biogeochemical cycles, because the killing of their bacterial hosts and the subsequent release of nutrients creates a short cut in the flow of carbon, which is known as the viral shunt (Wilhelm and Suttle 1999; Suttle 2005, 2007). Therefore, it is fundamental for a profound understanding of the ocean´s ecosystem to investigate the diversity of marine bacteriophages and their interaction with the bacterial hosts.

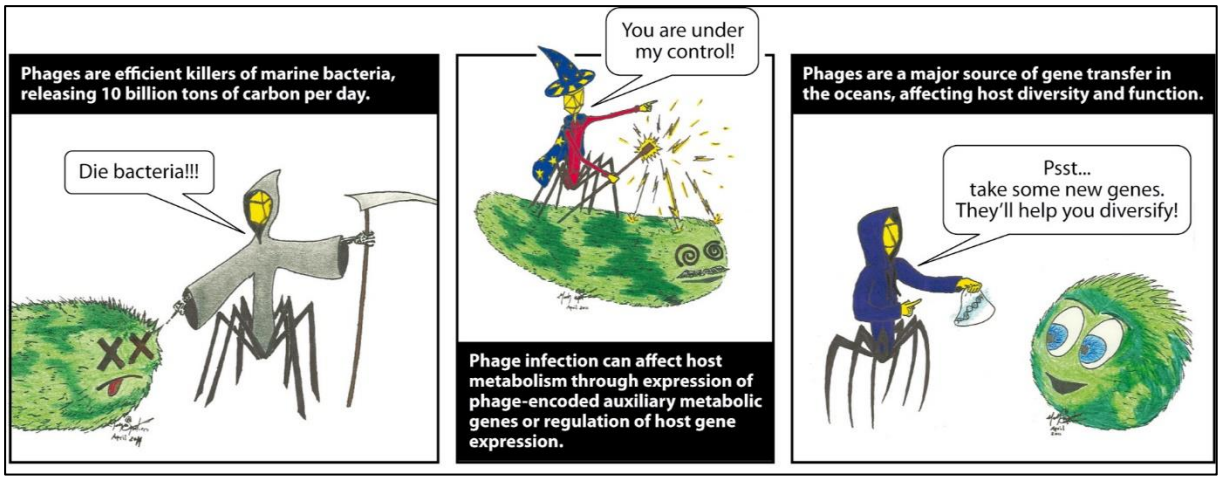


Fig. 1: Cartoon illustrating the ways in which phages influence the composition of marine bacterial communities: by cell lysis, host metabolism manipulation and gene exchange. Figure from Breitbart (2012).

For human health, phages can have both negative and positive effects. For example, some bacteria that are usually harmless for humans can be turned into virulent pathogens by certain temperate phages (Waldor and Mekalanos 1996). At the same time, phages are part of our immune system as they reduce bacterial colonization of mucosal surfaces (Barr et al. 2013). Since its discovery in the 1920's, the therapeutic potential of phages has been mainly utilized in Eastern Europe. However, due to the urge to fight antibiotic-resistant pathogens, phage therapy has been (re-)discovered in the last decades also in the Western hemisphere (Kutter et al. 2010; Vandamme and Mortelmans 2019). Phage-based control of bacterial contamination and (phyto-)pathogens is already applied in aquaculture (Culot et al. 2019; Ramos-Vivas et al. 2021), agriculture and food production (Goodridge and Bisha 2011; Torres-Acosta et al. 2019; Kazantseva et al. 2021). This way, phages and the exploitation of their antimicrobial action will most likely have an increasing impact on human health, nutrition, and economy in the future.

1.2. Classification of bacteriophages

1.2.1. Traditional classification of bacteriophages

The first attempts to classify bacteriophages were based on their morphological differences observed by transmission electron microscopy. In 1967, Bradley described six groups of bacteriophages taking their basic morphology and also the type of nucleic acid into account. These categories were adopted by the International Committee on Viral Taxonomy (ICTV) and served as a basis for the description of the first phage families (Fenner 1976; Calendar and Abedon 2006). The majority of phages harbors the nucleic acid in a capsid. The capsid consists of several copies of the major capsid protein and its form can vary, often being icosahedral,

sometimes prolate or even helical (Bradley 1967). In rare cases, the icosahedral capsid is coated with a lipid layer (Lundström et al. 1979).

Three of Bradley´s morphotypes have a tail structure attached to an icosahedral capsid and are equipped with linear double-stranded DNA (dsDNA) genomes. They were grouped in the now disbanded order Caudovirales (Ackermann 1998). The tail is either long, rigid and with a contractile sheath (myoviruses), or long, flexible, and non-contractile (siphoviruses), or short and non-contractile (podoviruses) (Fig. 2). At the tail tip there are fibers, which are important for attachment to the host (Nobrega et al. 2018). Phages of the fourth group have small icosahedral capsids with small, circular ssDNA genomes (*Microviridae*). *Inoviridae* also have small ssDNA genomes, but have a very different shape, with long, filamentous virions. The sixth group turned into the now abolished *Leviviridae* family, consisting of phages with small, icosahedral capsids and ssRNA genomes (Bradley 1967; Calendar and Abedon 2006).

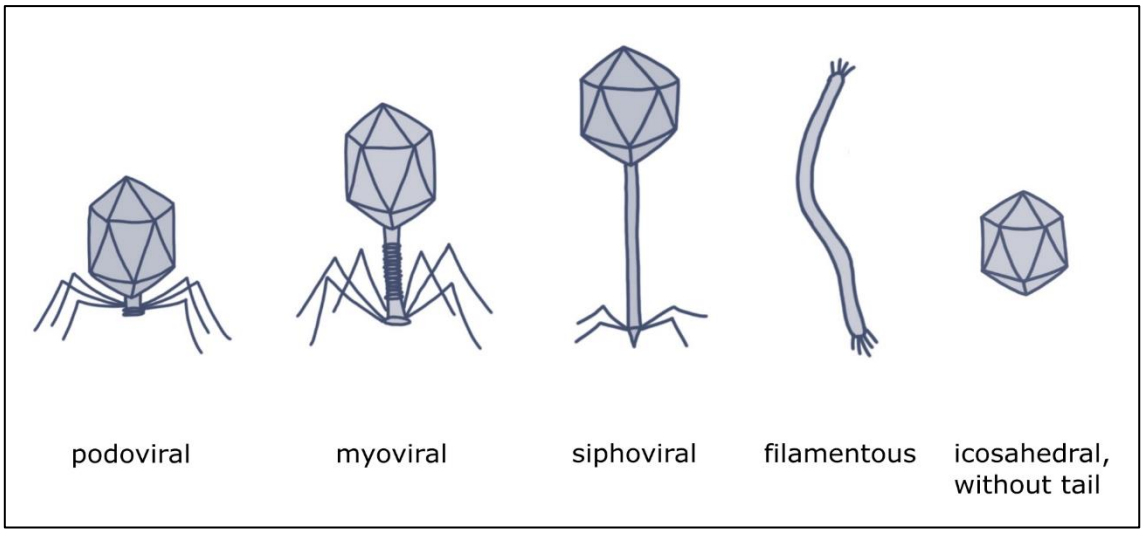


Fig. 2: Major morphological groups of bacteriophages (simplified scheme).

Later on and in parallel to phage classification based on morphology, Baltimore (1971) classified viruses in general into six (later seven) Baltimore Classes (BC), based on nucleic acid type and replication strategies: dsDNA (I), ssDNA (II), dsRNA (III), positive-sense ssRNA (IV), negative-sense ssRNA (V), positive-sense ssRNA with reverse transcription (VI), and ssDNA with reverse transcription (VII). For decades, the Baltimore Classes and the six morphotypes have been the basis for phage taxonomy. With the development of high-throughput sequencing techniques and bioinformatic tools, viral genomic information became accessible for a molecular-based classification and large-scale viral metagenomes and phage isolation campaigns revealed the true diversity of phages.

1.2.2. The megataxonomy of viruses

The majority of cultivated phages are tailed, with dsDNA genomes. Therefore, it has been believed for a long time that this phage type is prevailing in nature (Ackermann 2007; Dion et al. 2020). Recently, improved methodologies showed that non-tailed phages are more abundant than previously thought, and in some environments they can even be dominant (Brum et al. 2013). Moreover, new viruses are continuously being detected by viral metagenomics, viruses showing no genomic homology to known reference phages (Paez-Espino et al. 2016; Gregory et al. 2019; Benler and Koonin 2021). As a result, it became obvious that morphological features and nucleic acid type are not sufficient for a comprehensive classification of this vast viral diversity. Furthermore, the traditional classification system does not sufficiently reflect evolutionary relationships and thus lacks taxonomic relevance (Koonin et al. 2020a). In recent years, major efforts have been put into rethinking the classification of viruses. In 2020, Koonin and collegues proposed the "Megataxonomy of the Virus World", based on which a new global organization with a hierarchical taxonomy of viruses has been developed and officially accepted by the ICTV (Koonin et al. 2020a).

Unlike *Bacteria* and *Archaea*, viruses do not have a single common ancestor and are thus missing a universal marker gene comparable to the 16S rRNA gene, which has been traditionally used to investigate the phylogenetic relationships between cellular microbes (Krupovic and Koonin 2017; Numberger et al. 2019). This lack of a marker gene makes viral taxonomy much more difficult and a different approach needs to be taken to create the new megataxonomy, which is a combination of single/multiple gene phylogenetic methods and gene-sharing networks (Koonin et al. 2020a). For a bipartite gene-genome network approach, homologous viral proteins are first grouped into protein clusters, and then the viral genomes are grouped based on their shared protein clusters. Bipartite networks have two classes of nodes, one class representing the viral genome clusters, and the other the protein clusters connected to the viral genomes (Iranzo et al. 2016). During this analysis, the core proteins of certain phage groups are determined, which in turn can be used for phylogenetic tree building for the respective phage group. Despite the lack of a single universal gene for all viral genomes, with this approach it was possible to detect viral hallmark genes (VHGs), genes that are present in many different viral groups and can be used to differentiate them (Koonin et al. 2006).

These VHGs have been used for the new definition of six viral realms: *Adnaviria, Duplodnaviria, Monodnaviria, Riboviria, Ribozyviria* and *Varidnaviria* (Koonin et al. 2020a; Koonin et al. 2022). Two of these realms currently correspond to two distinct groups of dsDNA

viruses, which differ in the structure of their major capsid protein. The realm *Varidnaviria* comprises those dsDNA viruses with a double-jelly-role major capsid protein (DJR-MCP), the realm *Duplodnaviria* those with a HK97-like major capsid protein (HK97-MCP) (Koonin et al. 2020a). Within each realm, viruses are organized in eight major hierarchical ranks (kingdom, phylum, class, order, family, genus and species) and seven sub-ranks, inspired by the Linnaean taxonomy (Koonin et al. 2020a; Gorbalenya et al. 2020). The order Caudovirales and the morphology-based families Siphoviridae, Myoviridae and Podoviridae were dissolved, as they turned out not to be monophyletic (Turner et al. 2021). Instead, the class *Caudoviricetes* within the realm *Duplodnaviria* (Fig. 3) was created, comprising all tailed phages. The classification of the phages into genome-based, monophyletic families is currently in progress. The terms "siphoviral", "myoviral", and "podoviral" are still valid to describe the respective morphology. Within the class *Caudoviricetes*, there are currently four orders (*Crassvirales*, *Kirjokansivirales*, *Methanobavirales* and *Thumleimavirales*), and several families and subfamilies, many of them not yet assigned to any order (Walker et al. 2022).

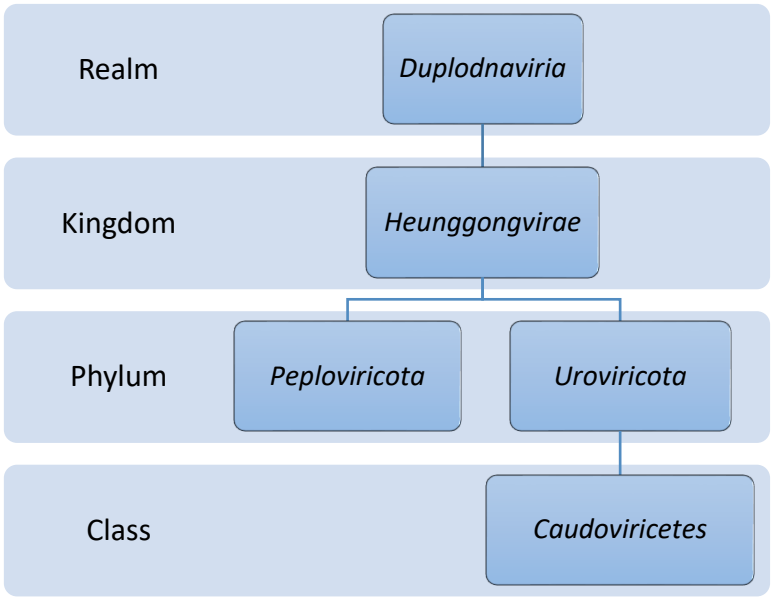


Fig. 3: Taxonomy of the class *Caudoviricetes*, which contains all tailed phages.

While the six viral realms have been defined based on VHGs, the lower ranks (genus and species) should be defined based on intergenomic sequence identity, which can be calculated with tools like VIRIDIC (Moraru et al. 2020). According to ICTV recommendations, viruses of the same species should have a sequence identity higher than 95%, and for the genus level higher than 70% (Turner et al. 2021; Moraru 2023). For the definition of intermediate ranks such as family and order, whole-proteome based clustering and comparison of protein family profiles should be used (Turner et al. 2021; Simmonds et al. 2023). For this kind of whole-

proteome based phage classification, several tools have been developed, such as ViPTree (Nishimura et al. 2017b), VICTOR (Meier-Kolthoff and Göker 2017), GRAViTy (Aiewsakun and Simmonds 2018) and vConTACT (Bolduc et al. 2017a). One of the most recently developed tools is VirClust (Moraru 2023). This is the main tool used for virus classification in this dissertation, because it performs a hierarchical classification of viruses based on their shared protein clusters (PCs). VirClust also provides protein annotation and core protein identification (Moraru 2023).

1.3. Lifestyles of bacteriophages

To produce progeny, a phage needs a bacterial host cell to replicate its DNA and synthesize new virus particles. Viruses use multiple infection strategies and phage lifestyles to achieve this, which have been investigated to varying degrees. Best known so far are the canonical lifestyles “lytic” and “lysogenic”, which at the same time represent the opposing extremes. While in the lytic lifestyle, infection immediately leads to the production of phage progeny and lysis of the host cell, in the lysogenic life cycle the viral nucleic acid is integrated into the host genome and is automatically replicated during cell division until the lytic production cycle is eventually induced (Fig. 4). More recently, additional intermediate strategies like “pseudolysogeny” and “chronic infection” were detected and by now, phage infection instead of being seen as a matter of distinct categories, rather seems to be a continuum of different strategies (Chevallereau et al. 2021).

1.3.1. Lytic life cycle

Irrespective of the type of infection cycle, it always starts with recognition and attachment of the virion to the host cell, followed by injection of the viral genomic material into the host cytoplasm (Fig. 4). In the lytic life cycle, reproduction of the virus is immediately started by reprogramming the host metabolism to replicate the phage genome and produce new phage particles (Weinbauer 2004). In the second phase, packaging of the new viral genomes into the capsids takes place, and this can be achieved by a number of different mechanisms (see section 1.3.4.). Finally, upon lysis of the host cell, the virion progeny is released. The time from adsorption to cell lysis is termed “latent period” (Weinbauer 2004). Its duration varies dramatically between different phage-host systems and depends on the bacterial host growth rate (Weinbauer 2004; Middelboe 2000). The same is true for the burst size, the number of new

virions released per cell, which is reduced under unfavorable growth conditions (Kokjohn and Sayler 1991).

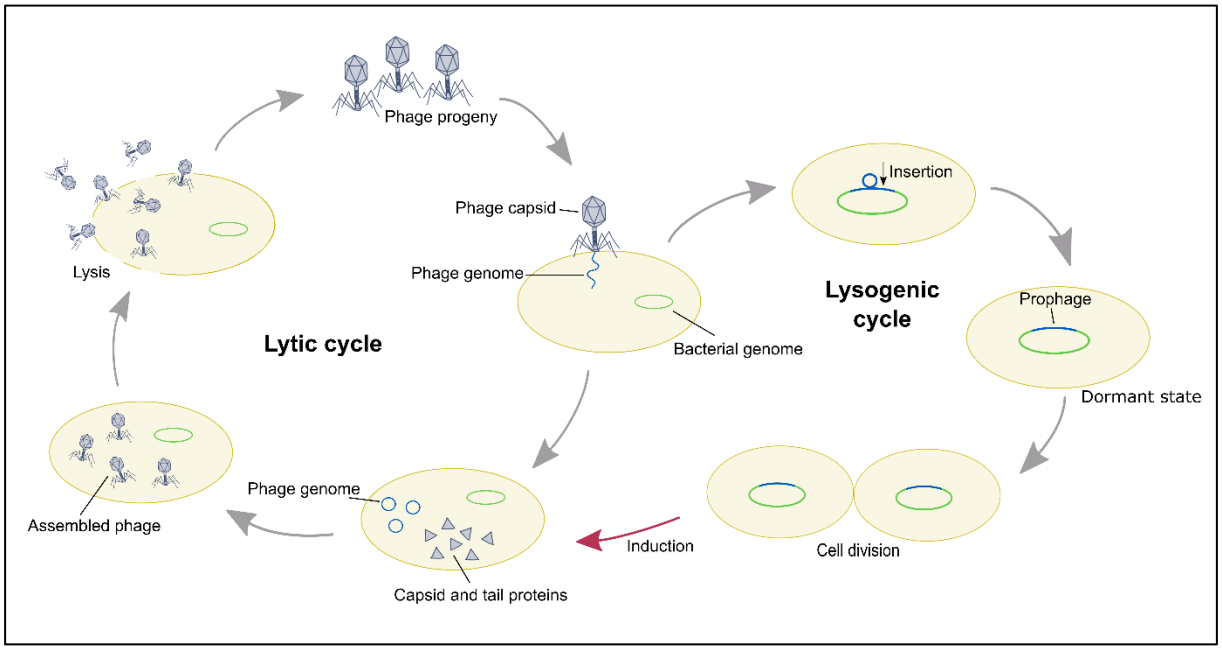


Fig. 4: The lytic and lysogenic life cycle of bacteriophages. Both cycles start with the attachment to the host cell and insertion of the viral genome. In the lytic cycle, the phage genome is immediately replicated, and capsid and tail proteins are produced. Virions are assembled and the phage genome is packaged. Finally, the host cell is lysed, and progeny phages are released into the environment. In the lysogenic cycle, the phage genome integrates into the host chromosome and becomes a prophage. During the latent period (dormant state), no phage progeny is produced. The prophage is replicated together with the host chromosome via cell division, until an external trigger leads to induction of the lysogenic cycle. The figure has been modified from Feiner et al. (2015).

Phage-mediated cell lysis is a carefully regulated process, with the mechanisms being diverse amongst different kinds of phages. For dsDNA phages of Gram-negative bacteria, the lysis pathway comprises three steps corresponding to the three parts of the bacterial cell wall (Young 2014). For the first step, two different systems are known: (i) Holins (phage-encoded proteins) accumulate within the cytoplasmic membrane and form micron-scale holes in the inner membrane. These holes allow endolysins to pass the membrane and degrade the peptidoglycan in the periplasm. (ii) Pinholins form small channels, which leads to depolarization of the membrane. This causes the activation of “signal anchor release” (SAR) endolysins that degrade the peptidoglycan (Cahill and Young 2019). The degradation of peptidoglycan is regarded as the second step of the lysis pathway. Finally, in the third step of lysis, the outer membrane is disrupted by spanins, proteins that form a bridge connecting outer and inner membrane. This can either be achieved by a two-component spanin complex or by a single protein that spans the periplasm (Young 2014; Cahill and Young 2019).

1.3.2. Lysogenic life cycle

Temperate phages undergo the lysogenic life cycle, in which the viral DNA enters a dormant state after injection into the host cytoplasm. Throughout this dormant state, the viral genome exists either as part of the host chromosome, or as an extrachromosomal element (Fig. 4) (Feiner et al. 2015). During this state, the viral genome is referred to as "prophage" and is replicated together with the host chromosome and thereby vertically transmitted between host generations. In this dormant or "latent" state, no production of new viral particles and no cell lysis takes place (Feiner et al. 2015). The bacteria carrying the prophage are called "lysogens" (Abedon 2022). The integration of the phage genome is often mediated by virus-encoded DNA integrases and either takes place at a specific attachment sites in the host genome, as for example in the case of *Escherichia coli* phage λ, or at random positions (e.g., phage Mu) (Bukhari and Zipser 1972; Shimada et al. 1972; Feiner et al. 2015). During the latent state, production of new virions is prevented by repression of viral lytic genes until entering of the lytic life cycle is induced (Bednarz et al. 2014; Feiner et al. 2015). This "switch" can be spontaneous or induced by phage regulatory genes or external triggers of the bacterial DNA damage stress response (e.g., temperature change, nutrient change or oxidative stress) (Howard-Varona et al. 2017). The expression of viral lytic genes then leads to DNA replication, assembly of phage particles, packaging and ultimately lysis of the host cell (Feiner et al. 2015).

1.3.3. Pseudolysogeny and chronic infections

Another, yet less understood life cycle with a latent state is "pseudolysogeny" or "carrier state life cycle". In contrast to lysogeny, the viral genome is not integrated into the host genome but asymmetrically transferred to one cell of the daughter generation. It remains in the host cytoplasm as an episome (or preprophage) (Miller and Day 2008; Feiner et al. 2015). This life strategy mostly occurs under nutrient limited conditions. This way, long-term coexistence of phage and host is ensured until conditions are better and the phage enters either the lytic or the lysogenic pathway (Miller and Ripp 2002).

The feature of performing chronic infections is not limited to, but typical for members of the *Inoviridae* family, which are filamentous ssDNA phages (Zeng et al. 2021). They follow a productive chronic pathway, in which they manipulate the host cellular mechanisms to replicate their viral genome and produce new phage particles that are continuously released from the cell by extrusion without lysing it (Chevallereau et al. 2021; Mäntynen et al. 2021).

1.3.4. DNA packaging strategies

In the course of the infection cycle, after new viral capsids have been produced, the genome of a tailed phage needs to be packed as a linear DNA molecule into the new capsid. This can be performed by various packaging techniques involving different types of physical genome ends (Merrill et al. 2016). Determining the physical genome start and end is crucial for proper arrangement of viral genomes and thereby ensuring an easy comparison of multiple genomes. Instead of laborious wet lab techniques, it is possible to use software-based methods operating on raw sequencing data to predict the genome ends and DNA packaging strategy (Merrill et al. 2016; Garneau et al. 2017).

Upon infection of a new host cell, the linear phage DNA is injected and circularized inside the cytoplasm (for most phages). There are different mechanisms of circularization, depending on the packaging strategy and the genome termini (Casjens and Gilcrease 2009; Merrill et al. 2016). If homologous recombination is used for circularization, identical sequence regions are necessary at both ends of the genome. This can be achieved through various ways, one of them being exact direct terminal repeats (DTRs), which can be short or long (Merrill et al. 2016). After circularization, there is one copy of the DTR in the genome, which is then multiplied by rolling circle replication resulting in linear concatemers with one copy of the DTR between each concatemer (Fig. 5). During packaging, the terminase cuts after the DTR and the DTRs are duplicated so that each capsid gets a molecule of exactly one genome-length with identical repeats at both ends. An example for this packaging strategy is *Escherichia coli* phage T7 (Merrill et al. 2016; Garneau et al. 2017).

Another type of genome termini are cohesive ends with either 3’ overhangs (e.g., *Escherichia coli* phage HK97) or 5’ overhangs (e.g., *Escherichia coli* phage Lambda) (Merrill et al. 2016). For circularization, the complementary overhangs or “sticky ends” are ligated (Fig. 5). Through rolling circle replication concatemers of the genome are created, separated by a specific *cos* site, which upon packaging is recognized by the terminase and used as packaging start and cutting spot. With the terminase cutting precisely at the *cos* site, it is ensured that each viral capsid contains a molecule of exactly one genome-length (Merrill et al. 2016).

In contrast to the above-described packaging strategies that use strictly defined genome ends, phages using headful packaging have terminal repeats that vary amongst progeny virions (Fig. 5) (Merrill et al. 2016). The genome is again circularized by homologous recombination and linear concatemers are created by rolling circle replication. For packaging, the terminase

recognizes a specific *pac* site for the first cut. However, the following cuts are made when the respective capsid is full (headful packaging), so after slightly more than one genome length. This leads to repeated sequences at each end, which are different for each virion. The chromosomes in the progeny virions are circularly permuted (Merrill et al. 2016). An example for this packaging strategy is *Escherichia coli* phage P1 (Garneau et al. 2017). There are also phages using headful packaging without any specific packaging (*pac*) site, which thus have random termini on both genome ends and are circularly permuted as well, such as *Escherichia coli* phage T4 (Garneau et al. 2017).

Apart from these four categories of packaging and genome termini, there are further strategies known, which do not involve repeated sequence regions and thus cannot be easily identified by analysis of raw sequencing data (Merrill et al. 2016). *Escherichia coli* phage Mu for example integrates into the host chromosome and before it is packaged, the phage genome is cut out of the host chromosome and thereby extended by segments of host DNA at both ends. These ends differ amongst progeny virions depending on the former position of the respective phage genome within the host chromosome (Bukhari and Taylor 1975; Merrill et al. 2016). Bacillus phage phi29 on the other hand has a protein that is covalently bound to each genome end and initiates DNA replication (Ortĭn et al. 1971; Merrill et al. 2016).

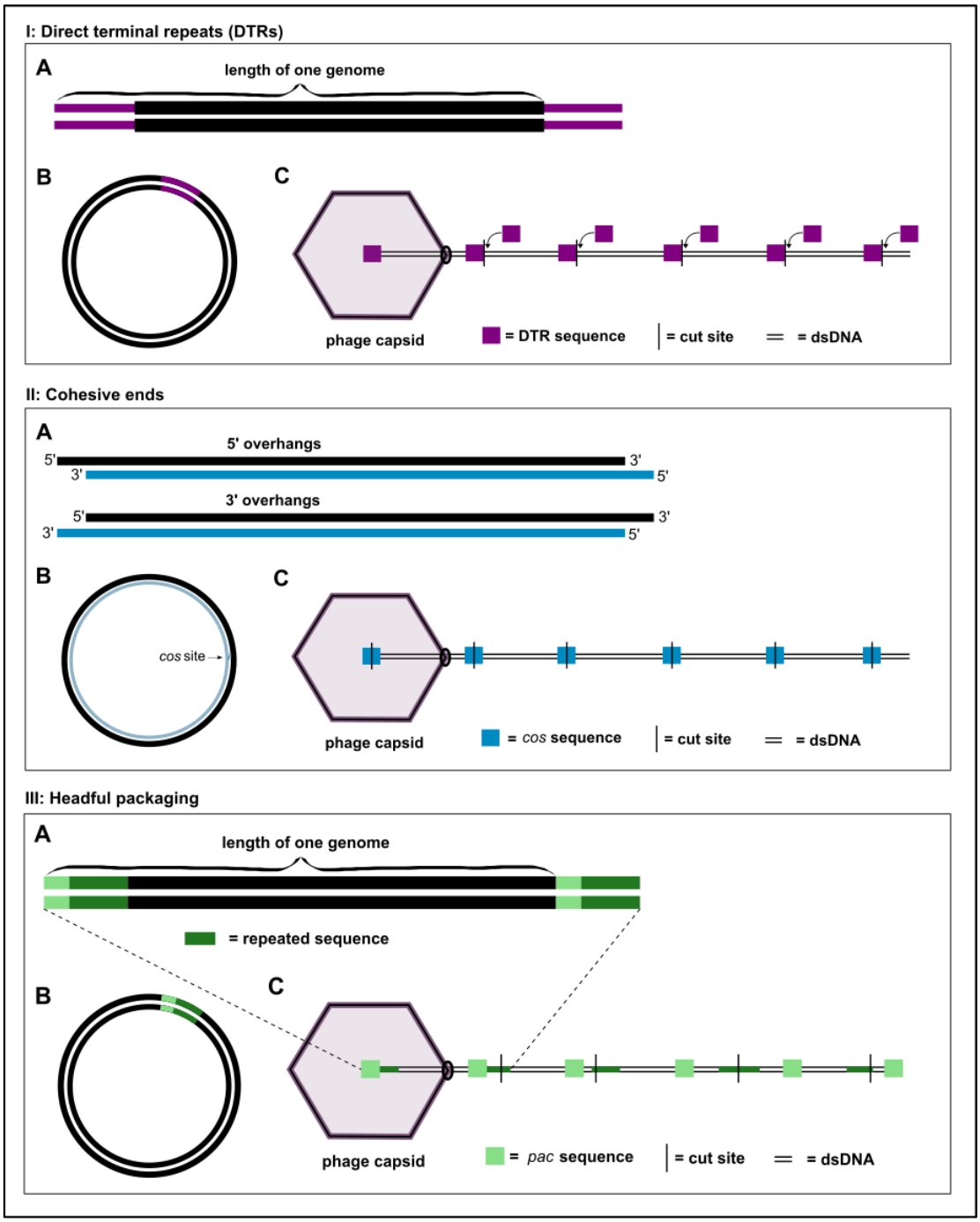


Fig. 5: Viral DNA structure inside the virion before infection, circularization, and packaging mechanism of different phages with different DNA packaging strategies. **I** Exact direct terminal repeats (DTR). **I-A** Before infection, phage DNA inside the virion has identical sequences at both ends. All virions have the same end sequences. **I-B** Formation of a circular DNA molecule by homologous recombination. **I-C** A linear concatemer is formed by rolling circle replication. During DNA packaging, the DTR sequences are duplicated. Each virion gets the same repeats at both ends. **II** Cohesive ends. **II-A** The linear genome can have 3' or 5' overhangs. **II-B** Sticky ends are ligated after infection. **II-C** Rolling circle replication creates a linear concatemer. 3' or 5' overhangs are formed with the terminase cutting directly at the *cos* site. Into each virion, exactly one genome length is packaged. **III** Headful packaging. **III-A** The viral genome prior to infection has similar sequences at both ends, varying between virions. **III-B** By homologous recombination a circular DNA molecule of exactly one genome length is generated containing one *pac* site. **III-C** Rolling circle replication creates a linear concatemer. The terminase starts packaging at the *pac* site and cuts after slightly more than one genome length. This way, a repeated sequence at both ends is created and the position of the *pac* site differs between each virion. The figure has been modified from (Merrill et al. 2016).

1.4. Phage-host interactions

1.4.1. Host range of bacteriophages

All the above-mentioned effects that bacteriophages have on microbial communities and on their evolution, on biogeochemical cycles and on human health are critically dependent on how and with which hosts they interact. Thus, the host range, i.e. the taxonomic diversity of hosts a specific bacteriophage can infect, is one of the crucial traits to investigate (de Jonge et al. 2019). To discuss and compare the host ranges of different phages, first one needs to define what a successful infection is and what a broad or a narrow host range is. Just as for viral taxonomy, the fact that the discovery of the true viral diversity has only just begun, makes a clear delineation of "broad" and "narrow" difficult. Furthermore, possibilities to determine the host range experimentally and computationally are limited. Host range determination in the laboratory is restricted by the collection of bacterial strains tested and by the requirement to detect a successful infection. In addition, phages can have high mutation rates and the host range can evolve rapidly (Meyer et al. 2016). Furthermore, with one point mutation being sufficient to change the host range of a phage, alterations become reversible and repeatable (Le et al. 2013; de Jonge et al. 2019).

What defines successful infection? De Jonge et al. (2019) specified it as completion of the viral life cycle. For strictly lytic phages, this corresponds to the lysis of the host cell and the release of viral progeny, which is comparably easy to detect. On solid medium (double-layer agar), cell lysis becomes visible by clearing zones in the bacterial lawn (plaques), in liquid cultures by a drop in optical density (OD), decrease of bacterial cell numbers and increase of phage particles. The latter can be measured also with flow cytometry or fluorescence in situ hybridization (FISH) combined with epifluorescence microscopy, for example. For host range tests, most often agar-overlay techniques like spot assay or plaque assay are used (Holmfeldt et al. 2007; Synnott et al. 2009; Zhan et al. 2016; Li et al. 2016a; Sonnenschein et al. 2017).

In their review, de Jonge et al. (2019) defined that phages being capable to complete their life cycle in only one host are narrow-host-range phages. If a phage can complete its life cycle in multiple hosts that are taxonomically distinct, even different strains of the same species, they refer to it as broad-host-range phage. A distinction can be made between two cases. Either the individual phage particle can infect multiple hosts (phenotypical mechanisms), or the individual particle can only infect one host, but the phage "quasispecies" is able to infect multiple hosts (genotypical mechanisms). The latter phenomenon is called "host switching" (de Jonge et al.

2019). Despite this categorization, host range can rather be seen as a continuum, especially at the population level.

As described in section 1.3, phage infection involves different phases. To be capable to infect a broad range of hosts, phages need to have the required adaptations in all stages of their life cycle (de Jonge et al. 2019). The first (and most obviously host range-relevant) level is host recognition, involving surface-adhesion mechanisms. The binding of the phage to the host receptor is mediated by the phage receptor-binding protein (RBP), which for tailed phages is located at the tip of the tail or at the baseplate (Tremblay et al. 2006; Legrand et al. 2016; Klumpp et al. 2023). One phage particle can have multiple RBPs (polyvalent), or the quasispecies can consist of individual particles with only one RBP (monovalent), but host switching is enabled through targeted genetic diversification of the RBP. It could also be that a phage particle has multiple RBPs, but only one is expressed under certain conditions (Chow and Bukhari 1976; Liu et al. 2002; de Jonge et al. 2019). After cell entry, the challenges for the phage continue. To replicate its own genome, the phage needs to adapt to the host´s replication machinery, for example in terms of codon usage, and to withstand the host´s defense mechanisms (see the following chapter). The integration of temperate phages into the host´s chromosome is often mediated by integrases. With those being highly specific, it requires a host chromosome with the corresponding integration target site (Smith et al. 2010; de Jonge et al. 2019). Also for the final step, the lysis of the host cell, the equipment of the phage (e.g., endolysins, holins) needs to be compatible with the host´s cell wall and defense mechanisms (reviewed in de Jonge et al. 2019).

Phages with very different host ranges have been described so far, from highly specific ones infecting only one strain, to broad-host-range phages infecting several genera (Holmfeldt et al. 2007; Born et al. 2011; Ahern et al. 2014). However, most isolated and model phages have a rather narrow host range and so do most of the described roseophages (Zhan and Chen 2019a). In contrast, isolation of new broad-host-range phages like the non-tailed *Autolykiviridae* (Kauffman et al. 2018) shows that broad-host-range phages might be more abundant than previously assumed (Roux et al. 2016; Paez-Espino et al. 2016). This contradiction could partly be explained by the interplay of host range and infection efficiency. For some broad-host-range phages, it has been shown that their ability to infect many hosts comes along with a decrease in virulence (Keen 2014; Ford et al. 2014; Kauffman et al. 2018). Since isolation techniques require a certain level of virulence, e.g., to obtain a visible plaque for picking, there is a bias in favor of phages that have a high infection efficiency on the specific

host and thus likely a narrow host range. The same bias is of course also to be taken into account when evaluating host range assays. What holds true for culture-dependent experiments in general, applies here in particular: the determined results only reflect the host range under these exact conditions (e.g., host abundance, host diversity, temperature, time). Hyman and Abedon (2010) suggested to always specify the experimental technique when talking about host range results and call it for example a "plaquing" host range or "spotting" host range.

1.4.2. The arms-race between phages and bacteria

As indicated in the previous chapter, for successful infection phages need to overcome a number of obstacles in each stage of their life cycle. Bacteria exhibit a plethora of defense mechanisms to prevent infection and/or killing, acting on the different viral infection stages. These together form the "immune-system" of bacteria (Bernheim and Sorek 2020; Millman et al. 2022). Likewise, phages have developed an arsenal of counteracting anti-defense systems. The arms-race between phages and bacteria leads to a huge and rapidly-evolving variety of defense and anti-defense systems (Hampton et al. 2020). In bacterial genomes, defense genes are often located on genomic islands (defense islands) (Makarova et al. 2011). Hussain et al. (2021) investigated a large collection of closely related *Vibrio* strains and their phages (Nahant collection) and found phage susceptibility of the bacteria to be mediated by large mobile genetic elements, which are highly diverse, exchanged at great evolutionary speed and make up the majority of the flexible genome. The authors argue that being localized on mobile genetic elements, defense genes can be exchanged rapidly between clonal strains. This paves the way for a high diversity of coexisting defense systems at population level, lowering prey concentration and increasing the chance of species survival. At the same time, the core genome and the encoded physiological and metabolic adaptations are independent from phage defense and can be maintained on population level even upon phage predation. In addition, it makes sense that defense genes are on mobile genetic elements, because they also imply a fitness-cost for the host (Koonin et al. 2020b).

Bacterial defense systems

Already the very first step of infection, the adsorption of the phage, can be prevented by the host by multiple means, such as receptor mutations, phase variation or receptors being masked by exopolysaccharide capsules (reviewed in Hampton et al. 2020). In addition, bacteria can use outer membrane vesicles (OMVs) with receptors on the surface to entice phages away (Reyes-Robles et al. 2018). After cell entry, major types of known defense systems can be roughly classified in three categories (reviewed in Labrie et al. 2010; Bernheim and Sorek 2020; Tal

and Sorek 2022). The first category comprises systems based on the degradation of viral nucleic acids and include restriction-modification (RM) systems and some CRISPR-Cas systems. Most commonly, a specific sequence is methylated on adenine or cytosine bases in the bacterial genome. Upon infection, the respective un-methylated sequence in the phage genome is recognized and cleaved (Oliveira et al. 2014; Mruk and Kobayashi 2014). Other modifications are known, also systems in which the viral DNA is modified (Bair and Black 2007; Wang et al. 2007; Thiaville et al. 2016). Further DNA degrading strategies include the CRISPR-Cas systems working with adaptive immune memory (Hille et al. 2018; Cady et al. 2012), as well as prokaryotic argonautes (pAgo) (Lisitskaya et al. 2018). The second category comprises abortive infection (Abi) systems, which recognize a phage infection and sacrifice the infected cell to prevent formation of phage progeny and thereby protect the community. Various mechanisms are known, including signaling systems, retron systems and toxin-antitoxin systems. Also type III CRISPR-Cas systems and many others lead to cell death or growth inhibition (Lopatina et al. 2020; Millman et al. 2020). In a third category, phage DNA and RNA synthesis can be inhibited via chemical defense involving different molecules such as prokaryotic viperins or anthracyclines (Bernheim et al. 2021; Hardy et al. 2023). Additionally, defensive enzymes (e.g., dGTPases) can cause nucleotide depletion and this way hinder viral genome replication (Tal et al. 2022). Apart from the systems described above, a large number of new systems with yet unknown mechanisms has been discovered. Since defense systems are often co-located in defense islands, analysis of the genomic surrounding of known defense genes has led to the discovery of new systems (Doron et al. 2018; Gao et al. 2020; Rousset et al. 2022; Millman et al. 2022). Still, most probably there are many more to explore, not to mention the fact that they are constantly evolving.

One bacterial genome encodes multiple defense systems (also several of the same kind) to evade infection of different types of phages. Possessing several lines of defense counteracts the fact that phages can become resistant. However, due to the fitness burden implied, a single bacterial strain cannot encode all the defense systems that may ever be needed (Bernheim and Sorek 2020). Thus, similar to the principle of “host switching” described by de Jonge et al. (2019) and in accordance with the considerations of Hussain et al. (2021) and Koonin et al. (2020b) mentioned above, Bernheim and Sorek (2020) suggested to see the available defense systems in a population of bacteria as a shared pool and proposed the “pan-immune system model”.

This shared pool is maintained and expanded by prophages (and plamids). They can encode additional defense systems to prevent the lysogen from being infected by related phages, a phenomenon called “superinfection exclusion” (Taylor et al. 2019). The genomic regions of prophages containing these non-essential transcribed genes are termed “morons”. They can also harbor other fitness factors for the host, such as antibiotic resistance, increased early biofilm formation or additional metabolic functions (Wang et al. 2010; Taylor et al. 2019). In general, this alteration of the host´s phenotype is described as “lysogenic conversion” (Little 2005).

Phage anti-defense strategies

As mentioned above, phages have a battery of anti-defense systems to counteract the bacterial immune system (reviewed in Hampton et al. 2020). Apart from improved adsorption by multiple and modified RBPs, there are multiple ways how phages evade the bacterial defense systems after cell entry. To escape RM systems, phage DNA can become methylated itself either by methyltransferases of its own or by manipulation of the respective host enzymes (Murphy et al. 2013). In addition, the target sequence of RM or CRISPR-Cas systems can be removed or modified (Pleška and Guet 2017; Vlot et al. 2018). Furthermore, phages can use anti-defense proteins, which act against restriction endonucleases after being either expressed early after infection or injected into the cell together with the DNA (Atanasiu et al. 2002; Piya et al. 2017). Anti-CRISPR (Acr) proteins inactivating CRISPR-Cas systems have also been discovered (Bondy-Denomy et al. 2013). To evade from toxin-antitoxin systems, phages can have several antitoxins encoded themselves or they produce other proteins interfering with the host´s toxin production or antitoxin degradation (reviewed in Hampton et al. 2020). In correspondence with the multiple defense systems encoded by the host, phages need to have multiple counter-defense systems. Thus, Bernheim and Sorek (2020) predict that among phage proteins with so far unknown function, more, yet unknown anti-defense genes will be identified in the future.

1.5. Phages infecting marine *Roseobacteraceae*

“Roseobacter group” or “roseobacters” is an informal term that has been used for a long time to describe a subgroup of the *Rhodobacteraceae* family comprising mainly marine members. Just recently, the classification has changed and the roseobacter group has been moved into its own family, the *Roseobacteraceae* fam. nov., because the members are clearly distinct from other *Rhodobacteraceae* on genomic, phylogenetic and phenotypic level (Liang et al. 2021). In

this dissertation, both classifications will be used. In the main parts of the thesis, the new classification as *Roseobacteraceae* will be applied, while in the chapters corresponding to already published work, the group is still referred to as part of the *Rhodobacteraceae*. Roseobacters represent a predominant lineage of heterotrophic bacteria in the marine environment (Simon et al. 2017) and have been found in diverse marine habitats, in the pelagic, in sediments and surface-associated, ranging from coastal regions to the deep-sea (Giebel et al. 2011; Kanukollu et al. 2016; Tang et al. 2016). They are most abundant in coastal areas, from temperate to polar regions, especially during phytoplankton blooms (Selje et al. 2004; Lamy et al. 2009; Lenk et al. 2012). Roseobacters have a high metabolic diversity and are capable to metabolize a large variety of organic compounds, perform anoxygenic photosynthesis and produce different secondary metabolites (Buchan et al. 2005; Newton et al. 2010; Simon et al. 2017). They have been shown to be phytoplankton-associates, utilizing algal exudates such as dimethylsulfoniopropionate (DMSP) and potentially engaging in both mutualistic and pathogenic interactions with the algae (Buchan et al. 2014; Amin et al. 2015). Accordingly, roseobacters typically increase in abundance during phytoplankton blooms, where they dominate the active bacterial community (Buchan et al. 2014; Wemheuer et al. 2015; Teeling et al. 2016; Bakenhus et al. 2017).

Although the roseobacters are such a diverse group comprising more than 130 genera (LPSN List of Prokaryotic names, Parte et al. (2020), accessed 15.06.2023), roseophages (phages that infect roseobacters) have been isolated from only twelve genera (*Celeribacter*, *Dinoroseobacter*, *Loktanella*, *Paracoccus*, *Pelagibaca*, *Rhodobacter*, *Rhodovulum*, *Roseobacter*, *Roseovarius*, *Ruegeria*, *Sulfitobacter* and *Thiobacimonas*). Nevertheless, the described roseophages show a remarkable genomic diversity and in recent years, scientific interest and the description of roseophage isolates have increased impressively. Roseophage isolates originate mainly from marine, coastal environments. However, for several of them searches in metagenomics databases have revealed that they are cosmopolitan. They occur mostly in coastal areas, but also in the open ocean, the deep sea and freshwater environments (Chan et al. 2014; Zhan et al. 2016; Qin et al. 2022). Moreover, the exact habitats from which roseophages have been isolated are divers and can have extreme conditions, ranging from Antarctic sea ice to tidal flat sediments, shallow-sea hydrothermal systems and soda lakes (van Zyl et al. 2016; Lin et al. 2016; Luhtanen et al. 2018; Hwang et al. 2020). When Zhan and Chen reviewed our knowledge about roseophages in 2019, 32 roseophages had been described. Until 2010, it had been only four (Zhan and Chen 2019a). Now (in 2023) the number of described roseophage isolates has increased to over 90, not including this work. The majority of them

have a dsDNA genome and a podo- or siphoviral morphology, with a high morphological variety in terms of capsid size and tail length (Zhan and Chen 2019a).

The SIO1 phage infecting *Roseobacter* sp. SIO67 was the first described roseophage and the first sequenced marine phage (Rohwer et al. 2000). Highly similar phages (>96% nucleotide identity) were isolated twelve years later from the same geographic area, providing evidence that marine viruses can exist as discrete populations over long periods of time (Angly et al. 2009). The SIO1 phage is a podovirus and is distantly related to the T7 phage (Rohwer et al. 2000), but has no RNA polymerase. A close relative of SIO1 is the roseophage P12053L infecting *Celeribacter marinus* IMCC12053 (Kang et al. 2012; Hardies et al. 2016). More than a third of the described roseopodoviruses belong to the N4-like group, named after *Escherichia coli* phage N4. The first roseophages of this kind were identified as phage DSS3_P2 infecting *Ruegeria pomeroyi* DSS-3 and phage EE36_P1 infecting *Sulfitobacter* sp. EE-36 (Zhao et al. 2009). In the last decade, 13 more N4-like roseophages have been isolated from the host genera *Ruegeria* (phages vB_RpoP-V12, V13, V14, V17, V21), *Sulfitobacter* (phage ϕCB2047-B), *Roseovarius* (phage RLP1), *Roseobacter* (phages RPP1, RD-1410W1-01, RD-1410Ws-07), and *Dinoroseobacter* (phages vB_DshP-R1, vB_DshP-R2C and DS-1410Ws-06) (Zhan and Chen 2019a; Ankrah et al. 2014a; Chan et al. 2014; Ji et al. 2015; Cai et al. 2015; Li et al. 2016a). The N4-like roseophages show genome sizes similar to *E. coli* phage N4 (73 to 75 kb) and they all possess the N4-like characteristic large virion-encapsidated RNA polymerase gene (vRNAP). Based on genomic approaches and with the new viral megataxonomy at hand, the N4-like phage group was recently transformed into the new family *Schitoviridae* (Wittmann et al. 2020). “HMO-2011-type phages” describes a podoviral group that was first detected by metagenomics studies to be highly abundant in the ocean and named after their first cultivated representative Puniceispirillum phage HMO-2011, infecting SAR116 strain IMCC1322 (Kang et al. 2013). Nine roseopodophages infecting *Roseobacter* RCA strains (phages CRP-1, CRP-2 and CRP-3) and *Roseobacter* sp. strains (CRP-207, CRP-212, CRP-235, CRP-345, CRP-603 and CRP-738) belong to this group (Zhang et al. 2019a; Qin et al. 2022). A characteristic feature of these viruses is a unique domain architecture of the DNA polymerase gene. Furthermore, they have the potential for a lysogenic life cycle, as they possess an integrase gene (Zhang et al. 2019a). Six more *Roseobacter* RCA podophages have been described (phages CRP-4, CRP-5, CRP-6, CRP-7, CRP-9 and CRP-13), clustering into four separate groups based on genomic comparison (Zhang et al. 2019a; Zhai et al. 2021). Moreover, one of the first phage isolates from Antarctic sea ice is a roseophage. The podovirus Antarctic DB virus 2 (OANV2) was isolated infecting *Octadecabacter* sp. IceBac430 (Luhtanen et al. 2018). Apart from (strictly)

lytic roseopodoviruses, three temperate ones have been described. All of them infect members of the *Sulfitobacter* genus (phages ϕCB2047-A, ϕCB2047-C and NYA-2014a) (Ankrah et al. 2014b).

Roseosiphoviruses are even more diverse. Chi-like roseosiphophages (related to phages of the *Chivirus* genus) infect *Ruegeria pomeroyi* DSS-3 (phages DSS3ϕ1, vB_RpoS-V7, V11, V16 and V18) (Zhan et al. 2018; Zhan and Chen 2019a). Phage pCB2051-A, infecting *Loktanella* sp. CB2051, had earlier been described as Chi-like as well. It was recently classified in its own genus *Broinstvirus* in the *Casjensviridae* family (Tolstoy et al. 2021). Being closely related to CbK-like phages (which infect *Caulobacter crescentus*, a freshwater bacterium), phages DSS3ϕ8 and vB_RpoS-V10 infecting *Ruegeria pomeroyi* DSS-3 and phage MD18 infecting *Phaeobacter inhibens* are the roseophages with the largest genomes (about 146 kb) described so far (Zhan et al. 2016; Zhan and Chen 2019a; Urtecho et al. 2020). Phages of the *Xiamenvirus* genus infect *Roseobacter denitrificans* OCh114 (phages RDJLϕ1 and RDJLϕ2) and *Ruegeria* sp. AU67, a bacterial sponge symbiont (phage Tedan) (Zhang and Jiao 2009; Huang et al. 2011; Liang et al. 2016; Baum et al. 2021). Paracoccus phage Shpa, isolated from the sediment of a soda lake in the East African Rift valley, infects *Paracoccus* sp. HS3 (van Zyl et al. 2016). Another lytic *Paracoccus* siphophage (vB_RmaS-R3) was isolated from the South China Sea (Xu et al. 2015). In addition, a second *Octadecabacter* phage isolated from the Antarctic sea ice, Antarctic DB virus 1 (OANV1), showed a siphoviral morphology (Luhtanen et al. 2018). More recently, 26 new phage isolates infecting *Rhodobacter* capsulatus have been described, which all have dsDNA genomes and a siphoviral morphology (Bollivar et al. 2016; Rapala et al. 2021). Based on genomic comparisons, these phages form six clusters, separate from other known phages. Two of these clusters have been described as the genera *Cronusvirus* and *Titanvirus* (Rapala et al. 2021). Siphophage vB_Dsh-R4C was isolated infecting *Dinoroseobacter shibae* DFL12 and was shown to be distantly related to the phages of the *Cronusvirus* genus (Cai et al. 2019). Phage vB_DshS-R5C also infects *Dinoroseobacter shibae* DFL12, but so far clusters alone amongst the described roseophages (Yang et al. 2017; Zhan and Chen 2019a). It was recently assigned to its own genus *Nanhaivirus* (Kropinski et al. 2018). Roseosiphophages DSS3_VP1 and DSS3_PM1 have been isolated infecting *Ruegeria pomeroyi* DSS-3. They represent their own family *Naomiviridae* (Rihtman et al. 2021). Their DNA has an unusual substitution of deoxythymidine by deoxyuridine. As this substitution makes them unavailable for common methods of library preparation for metagenome sequencing, an underestimation of their abundance is suspected (Rihtman et al. 2021). Amongst the described roseosiphoviruses, the majority is strictly lytic. Only a few of them possess

integrase genes indicating lysogenic potential. Exceptions are the prophages of *Rhodobacter capsulatus* (phages RcapMu and RC1), *Thiobacimonas profunda* JLT2016 (phage vB_ThpS-P1) and *Pelagibaca abyssi* JLT2014 (phage vB PeaS-P1) as well as *Rhodovulum* sp. P5 (phage vB_RhkS_P1), which are capable of transposition. These siphoviruses belong to the "Mu-like head phage group" as they share homologous head element sequences (Fogg et al. 2011; Lin et al. 2016; Tang et al. 2017; Rapala et al. 2021). Furthermore, Decewicz et al. (2019) described five temperate phages induced from different *Paracoccus* strains, one of them being the first isolated roseomyovirus (phage vB_PyeM_Pyei1). The others have a siphoviral morphology (phages vB_PbeS_Pben1, vB_PkoS_Pkon1, vB_PsuS_Psul1 and vB_PthS_Pthi1). *In silico* search of publicly available *Paracoccus* genomes has let to identification of 53 more prophages, revealing a large diversity of *Paracoccus* (pro)-phages, distinct from other known phages (Decewicz et al. 2019).

Finally, only two roseophage isolates have ssDNA genomes. The two unclassified *Microviridae* phages vB_RpoMi-Mini and vB_RpoMi-V15 have been isolated from *Ruegeria pomeroyi* DSS-3 and have the smallest genomes (4.2 kb) amongst known ssDNA phages (Zhan and Chen 2019b, 2019a).

In this dissertation, strains of three distinct *Roseobacteraceae* genera have been used as hosts for phage isolation. They will shortly be introduced in the following. Among the roseobacters, *Lentibacter* is a genus relevant in coastal and estuarine waters, where its relative abundance can be up to 30% of the bacterial community (Wallace et al. 2018), and has been repeatedly isolated from algal blooms in different geographical locations (Li et al. 2012; Hahnke et al. 2013). So far, the *Lentibacter* genus is rather small, with only one described species (*Lentibacter algarum*) and few unclassified strains listed in the NCBI taxonomy browser at the time of writing (22.03.2023). Prior to the work described in this dissertation, no phage infecting the *Lentibacter* genus had been described. Members of the *Sulfitobacter* genus are often algae-associated (Ivanova et al. 2004; Fukui et al. 2015; Wang et al. 2021). It was shown to be amongst the most abundant genera on the surface of *Fucus spiralis* and is thought to promote algal growth by provision of vitamin B12 and siderophores (Dogs et al. 2017). In addition, *Sulfitobacter* strains associated with the dinoflagellate *Alexandrium minutum* are also known to promote growth of the dinoflagellate (Yang et al. 2021). At the same time, sulfitobacters display algicidal effects during harmful algal blooms (Zhang et al. 2020b). The genus *Sulfitobacter* is large, with 27 species and a large number of unclassified strains at the time of writing (NCBI taxonomy browser, 18.01.2023). Except this work, currently six

Sulfitobacter phages have been isolated from three strains, comprising two N4-like podoviruses (phages EE36_P1 and ϕCB2047-B), one lytic siphovirus (phage GT1) and three temperate siphoviruses (phages ϕCB2047-A, ϕCB2047-C and NYA-2014a), as mentioned above (Zhao et al. 2009; Ankrah et al. 2014a, 2014b; Hwang et al. 2020). *Octadecabacter* is a diverse genus as well, currently comprising eight species and several unclassified strains (NCBI taxonomy browser, 18.01.2023). The first representatives *Octadecabacter arcticus* and *Octadecabacter antarcticus* have been isolated from the sea ice of the Arctic and Antarctic, respectively (Gosink et al. 1997). Others were found free-living in seawater or associated to marine red algae (Billerbeck et al. 2015; Park et al. 2016; Jin et al. 2023). Furthermore, members of the *Octadecabacter* genus were shown to dominate the subcuticular bacterial community of brittle star *A. squamata* (Morrow et al. 2018). At the moment of writing, only two *Octadecabacter* infecting viruses have been described in literature, so far both described as unclassified members of the class *Caudoviricetes*. These are *Octadecabacter* Antarctic BD virus 1 (siphovirus) and *Octadecabacter* Antarctic BD virus 2 (podovirus), isolated from Antarctic sea ice (Luhtanen et al. 2018; Demina et al. 2021).

1.6. How to investigate phage diversity

There are two categories of methods to study phage diversity: (i) culture-dependent methods, which are based on isolation of phages in pure cultures and (ii) culture-independent methods, which are working directly on environmental samples, for example sequencing of phage genomes or marker genes. Both types of methods have their advantages. Having pure cultures at hand allows an in-depth analysis of specific phage-host systems, e.g., the determination of the host range of the specific phage, the burst size, and the temporal course of the infection cycle. The virion morphology can be visualized by electron microscopy. Culture-independent methods on the other hand investigate only on the molecular level, but allow for a much higher throughput and are less biased by laboratory conditions. Quantification of virus numbers by flow cytometry or epifluorescence microscopy can help to understand the viral impact on bacterial communities (Brussaard 2004; Heinrichs et al. 2020). Viral diversity in environmental samples can be captured relatively easy by metagenome sequencing (Gregory et al. 2019) and single phage genomes can be obtained directly from the environmental sample without isolation (Breitbart et al. 2002). Nevertheless, in such approaches, one very fundamental piece of information about the phage is missing and challenging to predict – the corresponding host (Roux et al. 2018; Roux et al. 2021). Bioinformatic prediction of the potential host of

uncultivated phages can be attempted with programs like HostPhinder (Villarroel et al. 2016), using BLAST searches against databases and sequence composition analyses. The predictions, however, are dependent on database knowledge and thus severely restricted by the fact that bacterial diversity is unevenly represented in those (Coclet and Roux 2021). Furthermore, Dion et al. (2020) suggested that, despite the advantages of metagenomics for the investigation of viral communities, phage populations with high microdiversity could be overlooked, as closely related genomes cannot be differentiated. Even though new approaches as single-virus genomics and viral tagging metagenomics have been developed (Deng et al. 2014; Martinez-Hernandez et al. 2017), the authors point out, that for a comprehensive picture of phage communities, culture-independent methods need to be combined with phage isolations and culture work (Dion et al. 2020).

1.7. Aims and outline of this dissertation

Despite the progress made in recent years in the isolation and description of new roseophages, the vast majority of phages infecting roseobacters still needs to be discovered. To date, phages for only twelve roseobacter genera have been described in literature, in sharp contrast with the more than 130 roseobacter genera. Therefore, the major goal of this thesis was to uncover further roseophage diversity. As outlined above, isolation and cultivation of phages is a valuable tool, as it lays the foundation for in-depth analysis of the phage´s life style and interaction with its host(s). Thus, it was chosen as the main approach for discovery of new roseophage diversity in this dissertation. The focus was placed on three environmentally relevant roseobacter genera: *Lentibacter*, *Sulfitobacter* and *Octadecabacter*.

This dissertation is written in a monographic style. Nevertheless, its chapters reflect the division into four projects, which are or will be published independently:

Chapter 2 focuses on the isolation of two novel roseophage species infecting *Lentibacter* sp. SH36, and the delineation of a new viral group, the “cobaviruses”. The phylogenetic classification, genomic organization and host range of these viruses were investigated, as well as their biogeographical distribution. This work was published in the *The ISME journal* (Bischoff et al. 2019). Chapter 3 contains the further taxonomic classification of the two roseophages and the description of the *Zobellviridae* family, published as an ICTV-accepted taxonomic proposal (Bischoff et al. 2020).

Chapter 4 describes a large-scale roseophage isolation campaign that resulted in the isolation and genome sequencing of 128 new dsDNA roseophages, infecting mainly *Sulfitobacter*, but also *Lentibacter* and *Octadecabacter* strains. The chapter focuses on the classification of the new phages into seven viral genome clusters and describes their genomic organization and characteristics. The corresponding publication to this chapter is in preparation.

Chapter 5 is dedicated to a subgroup of the above-mentioned new roseophages, the sulfiviruses. This is a collection of highly similar phages infecting a group of almost identical *Sulfitobacter* host strains. Phage-host interactions were investigated using host range assays and the results are further discussed in the light of phage and host genomic microdiversity.

At last, in chapter 6, a short excursion into the ssDNA bacteriophage world is undertaken, describing the isolation and the characterization of Sulfitobacter phage ICBM5, to which I contributed with the laboratory work of determining morphology, host range and host phylogeny. This work served as basis for a larger study that has been published in *Virus Evolution* (Zucker, Bischoff et al. 2022).

2. Cobaviruses - a new globally distributed phage group infecting *Rhodobacteraceae* in marine ecosystems

This chapter corresponds to an article published in The *ISME journal* in 2019 (without the introduction). Thus, the described and proposed viral taxonomy, also the spelling of viral taxa (e.g., in italics or with quotation marks), corresponds to the classification at that time, which is now partially invalid. Subsequent changes in the taxonomy of the described viruses are further discussed in chapter 3 of this dissertation.

Authors of the publication:

Vera Bischoff[1], Boyke Bunk[2], Jan P. Meier-Kolthoff[2], Cathrin Spröer[2], Anja Poehlein[3], Marco Dogs[1], Mary Nguyen[4], Jörn Petersen[2], Rolf Daniel[3], Jörg Overmann[2], Markus Göker[2], Meinhard Simon[1], Thorsten Brinkhoff[1], Cristina Moraru[1]

[1]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Carl-von-Ossietzky-Str. 9 -11, D-26111 Oldenburg, Germany

[2]Leibniz-Institut DSMZ—Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Inhoffenstraße 7 B, D-38124 Braunschweig, Germany

[3]Georg-August-University Göttingen, Institute of Microbiology and Genetics, Department of Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Grisebachstr. 8, D-37077 Göttingen, Germany

[4]Centre for Marine Bio-Innovation, School of Biological, Earth and Environmental Sciences, The University of New South Wales, Kensington, NSW 2052, Australia

Reference: ISME J 13 (6), P. 1404–1421. DOI: 10.1038/s41396-019-0362-7.

Author contributions:

CM designed the research, contributed to data analysis and drafted the manuscript. VB performed the laboratory phage work, contributed to data analysis and to manuscript drafting. MN and VB isolated ICBM2 phage. BB, CS and JO conducted PacBio sequencing, assembly, annotation and DTR recognition. AP and RD conducted Illumina sequencing and assembly. JMK and MG performed the VICTOR analysis. MD and TB calculated the 16 S rRNA trees. JP contributed to the host range tests. TB and MS advised through the course of research and contributed to data interpretation. All authors contributed to manuscript finalization.

2.1. Chapter summary

Bacteriophages are widely considered to influence bacterial communities, however most phages are still unknown or not studied well enough to understand their ecological roles. We have isolated two phages infecting *Lentibacter* sp. SH36, affiliated with the marine Roseobacter group, and retrieved similar phage genomes from publicly available metagenomics databases. Phylogenetic analysis placed the new phages within the Cobavirus group, in the here newly proposed genus Siovirus and subfamily Riovirinae of the *Podoviridae*. Gene composition and presence of direct terminal repeats in cultivated cobaviruses point toward a genome replication and packaging strategy similar to the T7 phage. Investigation of the genomes suggests that viral lysis of the cell proceeds via the canonical holin-endolysin pathway. Cobaviral hosts include members of the genera *Lentibacter, Sulfitobacter* and *Celeribacter* of the Roseobacter group within the family *Rhodobacteraceae* (*Alphaproteobacteria*). Screening more than 5,000 marine metagenomes, we found cobaviruses worldwide from temperate to tropical waters, in the euphotic zone, mainly in bays and estuaries, but also in the open ocean. The presence of cobaviruses in protist metagenomes as well as the phylogenetic neighborhood of cobaviruses in glutaredoxin and ribonucleotide reductase trees suggest that cobaviruses could infect bacteria associated with phototrophic or grazing protists. With this study, we expand the understanding of the phylogeny, classification, genomic organization, biogeography and ecology of this phage group infecting marine *Rhodobacteraceae*.

2.2. Materials and methods

2.2.1. Cultivation media

Liquid cultures of the bacterial host, *Lentibacter* sp. SH36, were grown in artificial saltwater medium (1x ASW) (24.32 g/l NaCl, 10 g/l $MgCl_2x6H_2O$, 1.5 g/l $CaCl_2x2H_2O$, 0.66 g/l KCl, 4 g/l Na_2SO_4, 2.38 g/l HEPES, 0.6 g/l peptone, 0.3 g/l yeast extract, 84 mM KBr, 40 mM H_3BO_3, 15 mM $SrCl_2$, 40 mM NH_4Cl, 4 mM KH_2PO_4, 7 mM NaF, pH 7.5), which was autoclaved and completed before use with 1 ml/l of sterile filtered multi vitamin solution (after (Balch et al. 1979)), 0.25 ml/l of sterile filtered trace element solution A (1.5 g $FeCl_2x4H_2O$ in 10 ml 25% HCl and 250 ml MilliQ water) and 0.1 ml/l of autoclaved trace element solution B (19 mg/l $CoCl_2x6H_2O$, 10 mg/l $MnCl_2x2H_2O$, 7 mg/l $ZnCl_2$, 3.6 mg/l $Na_2MoO_4x2H_2O$, 2.4 mg/l $NiCl_2x6H_2O$, 0.6 mg/l H_3BO_3, 0.2 mg/l $CuCl_2x2H_2O$). The solid medium used for plaque assays was Marine Broth (MB). This media had the following recipe. 5.0 g/l peptone, 1.0 g/l yeast extract, 0.1 g/l $C_6H_8FeO_7$, 12.6 g/l $MgCl_2x6H_2O$, 3.24 g/l Na_2SO_4, 19.45 g/l NaCl, 2.38 g/l $CaCl_2x2H_2O$, 0.55 g/l KCl, 0.16 g/l $NaHCO_3$, 0.01 g/l $Na_2HPO_4x2H_2O$, 0.008 g/l KBr, 0.034 g/l $SrCl_2x6H_2O$, 0.022 g/l H_3BO_3, 0.007 g/l $Na_2SiO_3x3H_2O$, 0.0024 g/l NaF, 0.0016 g/l NH_4NO_3. To prepare MB agar plates, the medium was supplemented with 18 g/l Bacto Agar (BD Biosciences) prior to autoclavation.

2.2.2. Phage enrichments and isolation of ICBM1 and ICBM2 phages

Surface seawater was collected from multiple stations (53.978 8.059; 53.937 7.806; 53.896 7.535; 53.840 7.255; 53.793 6.997) in the southern North Sea, during a phytoplankton bloom in March 2015, on board of the cruise ship RV Heincke. Further, the seawater from each station was filtered on board through 0.7 μm filters, 47 mm in diameter (GTTP filters, Millipore). To prevent clogging, the filters were exchanged every 2 liters. The seawater from all stations was pooled, transported to the laboratory and stored at 4 °C in the dark.

Two phage enrichments (S1 and S2), each of 100 ml, were set up by mixing 90 ml of freshly filtered (Nalgene rapid-flow, 0.2 μm, PES membrane, Thermo-Scientific) seawater with 10 ml of 10x ASW (see chapter 2.2.1.) and 2.1 ml of exponentially growing host culture *Lentibacter* sp. SH36 (final OD_{600} = 0.006). Two controls were prepared in parallel. The first, a positive control (PC) for host growth, consisted of 100 ml 1x ASW and 2.1 ml of exponentially growing host culture (final OD_{600} = 0.006). The second, a negative control (NC) for growth of seawater bacteria contaminants, which might have passed through the 0.2 μm

filter, consisted of 90 ml freshly filtered (0.2 µm) seawater and 10 ml 10x ASW. Bacterial growth was monitored by measuring the optical density at 600 nm (Beckmann DU520, USA). The cultures were incubated at 20 °C and 100 rpm overnight, until the enrichment cultures showed signs of cell lysis. Lysis was indicated by decreasing optical density at 600 nm (OD_{600}) in S1 and S2 cultures compared to the positive control and by the presence of cell debris in S1 and S2. Bacterial cells and their debris were removed from the phage enrichments by a centrifugation step (15 min, 4000 x g, 20 °C), followed by 0.2 µm filtration (Rotilabo-syringe filters, Carl Roth) of the supernatant. The cell free phage lysates from the S1 and S2 enrichments were used to isolate the ICBM1 and ICBM2 phages, respectively, by plaque assays and single plaque picking.

To obtain single plaques, serial dilutions (10^0, 10^{-1}, etc.) were prepared from the phage fractions by mixing with ASW base (see chapter 2.2.1.). 100 µl of phage dilution were mixed with 280 µl of exponentially growing host culture (OD_{600} = 0.2-0.3) and incubated for 15 min on ice. Afterwards, the mixture was transferred to 3 ml MB-soft agar (0.6% low melting point Biozym Plaque GeneticPure agarose, Biozym, kept warm at 37 °C), mixed by brief vortexing and poured onto the bottom MB agar layer (1.8 % agar). After drying of the top layer, the plates were incubated for three days at 20 °C. For isolation, when phage plaques were observed as clearing zones within the grown bacterial lawn, they were picked with sterile Pasteur pipettes and incubated overnight in 500 µl ASW base at 4 °C. After subsequent centrifugation (10 min, 10000 x g, 4 °C), the supernatant was used for a next round of plaque assays. This procedure of plaque assay, picking of plaques and re-plating was repeated three times to ensure purity of the newly isolated phages. The ICBM1 and ICBM2 phages were stored either as phage lysate at +4 °C or as glycerol stock of free phages or infected cells at -80 °C (for details, see chapter 2.2.3.).

To determine the host range, 94 strains (Table S1) covering the phylogenetic diversity of *Rhodobacteraceae* were challenged with ICBM1 and ICBM2 phages, at three different temperatures (15, 20, and 28 °C), using first the spot assay technique and then the plaque assay technique for confirmation of positive results. For the spot assay, serial dilutions (10^0, 10^{-1}, etc.) were prepared from the phage fractions by mixing with ASW base. 280 µl of exponentially growing host culture (OD_{600} = 0.2 - 0.3) were mixed with 3 ml MB-soft agar (kept warm at 37 °C) and poured onto the bottom MB agar layer. After drying of the top layer, 10 µl of each phage fraction dilution were pipetted on top as droplets. The plates were incubated at 20 °C and regularly checked for plaque formation.

2.2.3. Preparation of phage ICBM1 and ICBM2 glycerol stocks

In two Erlenmeyer flasks, 20 ml 1x ASW medium (see chapter 2.2.1.) was inoculated with an exponentially growing *Lentibacter* sp. SH36 culture (final OD_{600} = 0.006). One culture was infected with 300 µl of the phage ICBM1 lysate, the other was not infected and regarded as control. Both cultures were incubated at 20 °C and 100 rpm overnight, until bacterial lysis in the infected culture was indicated by low OD_{600} (compared with the control culture) and disrupted cell particles. The phage fraction was obtained by removing cells and debris by centrifugation (15 min, 4000 x g, 20 °C), followed by 0.22 µm filtration of the supernatant. The phage fraction was stored at +4 °C. For long term storage, two types of glycerol stocks were prepared: i) stock of free phage particles (1 part phage fraction and 1 part MB media with 50% glycerol) and ii) stock of infected host cells (1 part infected cells – 375 µl phage fraction added to 375 µl host culture, 15 min on ice for absorption – and 1 part MB media with 50% glycerol).

2.2.4. ICBM1 and ICBM2 phage high titer lysates

To obtain a high amount of ICBM1 and ICBM2 phage biomass for transmission electron microscopy (TEM) and genome sequencing, two subsequent infection cultures of *Lentibacter* sp. SH36 with phage ICBM1 or ICBM2 were performed. For the first round of infection, 1x ASW was inoculated with exponentially growing *Lentibacter* sp. SH36 to a final OD_{600} of 0.006 and with phage ICBM1 or ICBM2 stock. After an overnight incubation at 20 °C and shaking at 100 rpm, lysis was observed, indicated by a decrease in the OD_{600} (in comparison with the control, non-infected culture) and cellular debris. The remaining cells and cell debris were removed by centrifugation (15 min, 4000 x g, 20 °C) and 0.22 µm filtration. For the second round of infection, a highly concentrated phage-host mixture was obtained by pelleting an exponentially growing culture of *Lentibacter* sp. SH36 and re-suspending the cell pellet in the phage fraction from the first infection culture. After the phage-host mixture was incubated on ice for 15 min to facilitate phage absorption, an equal volume of 2x ASW was added to it and the infection culture incubated overnight at 20 °C and 100 rpm. After lysis, the phage fraction was obtained by centrifugation (15 min, 4000 x g, 20 °C) and 0.22 µm filtration to remove remaining cells and cell debris.

2.2.5. Transmission electron microscopy

ICBM1 and ICBM2 phage lysates were further concentrated for TEM by polyethylene glycol (PEG) precipitation and purified by cesium chloride gradient ultracentrifugation. 150 ml phage fraction resulting from two subsequent infections (see above) were incubated for 2 h at 4 °C

with PEG (final concentration 10%) and NaCl (final concentration 0.6 mM). After centrifugation for 2 h at 7197 x g and 4 °C, the supernatant was discarded and the pellet resuspended in 500 µl SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.4) in total. For resuspension of the phages, 30 min incubation at 4 °C followed. PEG was removed by mixing with an equal volume of 100% chloroform, shaking for 5 min and incubation on ice for 5 min. After centrifugation (10 min, 3000 x g, 4 °C), the upper layer was collected.

For purification by ultracentrifugation, a density gradient was set up in UltraClear™ centrifuge tubes from cesium chloride solutions with different densities (from bottom up): 1.5 ml of 1.65 g/ml, 2 ml of 1.5 g/ml, 2 ml of 1.4 g/ml, 1 ml of 1.2 g/ml. The PEG concentrated phage fraction was transferred on top. Ultracentrifugation was run for 4 h at 20 °C and 25000 rpm (Beckman, SW 41 Ti). Afterwards, the visible band corresponding to the phages was collected with syringe and needle through the sidewall of the ultracentrifuge tube (~500 µl). Removal of cesium chloride was done by dialysis with Slide-A-Lyzer® G2 Dialysis Cassettes 10 K MWCO (ThermoScientific) against ASW base for 21 h with buffer exchange after 3 h and 18 h.

Two staining procedures were performed for each phage prior to transmission electron microscopy (TEM): (1) ammonium molybdate staining and (2) uranyl acetate staining. 30 µl of phage ICBM1 or ICBM2 concentrate were pipetted on top of a carbon coated grid (Formvar 162, 200 mesh) and phages were allowed to absorb for 3 min, followed by staining with 30 µl uranylacetate (2%) or ammonium molybdate (2%) for 45 sec and gentle removal of the liquid with filter paper. After air drying for 15 min, the grids were visualized with the transmission electron microscope Zeiss EM902A. Images were documented with the Proscan High Speed SSCCD camera and analyzed using the software ImageSP viewer (Version 1.2.5.16). Phages negatively stained by uranyl acetate were used for capsid size measurements.

2.2.6. Isolation and purification of phage DNA for sequencing

Phage isolates – extraction of DNA from virions

Phage DNA was extracted from cell free phage lysates obtained by infecting *Lentibacter* sp. SH36 with ICBM1 or ICBM2. Phages were concentrated from the phage lysates by precipitation with polyethylene glycol. For this purpose, 4 x 25 ml phage lysate prepared as above (with one exception: to avoid phage loss, cells and debris were removed only by centrifugation, and not by filtration) were mixed with 50% PEG (final concentration 10%) and 5 M sodium chloride (final concentration 0.6 M) and incubated for 2 h at 4 °C. After

centrifugation (2 h, 7197 x g, 10 °C) the phage pellets were resuspended in 500 µl SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.4) each.

Extracellular DNA was removed by incubating the phage concentrates with 0.04 units/µl of Turbo DNase (Invitrogen, Ambion) for 30 min at 37 °C, followed by enzyme inactivation by incubating for 10 min at 75 °C with 15 mM EDTA. Further, the phage DNA was extracted using the ChargeSwitch gDNA Mini Bacteria Kit (ThermoFisher Scientific), according to the instructions manual, including the with the RNase digestion, but with the exception of no lysosyme in the first step. The DNA was finally eluted in 1 ml elution buffer. The concentration and quality of the obtained DNA was checked fluorometrically with Qubit 2.0 and the Qubit® dsDNA HS Assay, spectrophotometrically with Nanodrop 2000 spectrophotometer and visually by regular gel electrophoresis (0.7% agarose gel, 50 V, SYBR Gold staining).

Phage enrichments – extraction of DNA from the intracellular phage fraction

To extract the intracellular phage fraction from the phage enrichment, when lysis was observed, the cells from the enrichments and the positive control were retrieved by centrifugation (15 min, 4000 x g, 20 °C). The cell pellets were embedded in agarose plugs by mixing with SeaKem Gold Agarose for PFGE, Lonza Rockland Inc. (final concentration 0.8%), distributing the mixture into 100 µl molds and allowing it to solidify for 30 min at 4 °C. Plugs were collected in a 50 ml Falcon tube and incubated overnight at 50 °C in 2 ml ESP buffer (1% N-laurylsarcosine, 1 mg/ml proteinase K, 0.5 M EDTA pH 9.0). Afterwards, the ESP was discarded, the plugs were washed three times with TE buffer (10 mM Tris-HCl pH 8.0, 2 mM EDTA pH 9.0) and stored in TE buffer at 4 °C until further use. The DNA from the agarose plugs was separated during agarose gel electrophoresis (1% SeaPlaque GTG Agarose, Lonza Rockland Inc., TAE buffer - 40 mM Tris-acetate, 1 mM EDTA, pH 8.3, migration 2 h at 60 V). Afterwards, the gel was cut into two halves. One half was stained with Ethidium bromide bath (1 µg/ml, 30 min) and documented with BioDocAnalyze system (Biometra). The distance from the loading pocket to the phage DNA band was measured and used to localize the phage DNA band in the unstained half and to cut it out from the gel. The agarose pieces were stored overnight in TE buffer (10 mM Tris-HCl pH 8.0, 5 mM EDTA pH 9.0), followed by agarose digestion with 1 unit of β-agarase (New England Biolabs Inc.), according to the manufacturer's protocol. Undigested agarose was removed by centrifugation for 10 min at 20000 x g. The obtained supernatants were concentrated to 100 µl each using 100 kDa Amicon Ultra centrifugal filters (0.5 ml volume, Merck Millipore) and then stored at -20 °C. Afterwards, an additional purification step was performed using the ChargeSwitch gDNA Mini Bacteria Kit (ThermoFisher Scientific) for DNA extraction. This was done according to the kit manual

(including the RNase digestion, but without lysozyme) and the DNA was finally eluted in 100 µl elution buffer. Concentration and quality of the obtained DNA were checked with Qubit 2.0 fluorometer, Nanodrop 2000 spectrophotometer and by regular gel electrophoresis (0.7% agarose gel, 50 V, Ethidium bromide staining).

2.2.7. Genome sequencing and assembly

The ICBM1 phage and the S2 enrichment were sequenced using both Illumina (paired-end technology 2 × 300 bp) and PacBio technologies. The ICBM2 phage and the S1 enrichment were sequenced only by Illumina. The Illumina and the PacBio assemblies were performed separately and they resulted in identical phage genomes (Table S2). Error free assembly of the PacBio samples was possible due to the high coverage obtained (>4000 x). The phage genomes are available in the NCBI GenBank database under the following accession numbers: MF431617 (ICBM1), MF431616 (ICBM2) and MF431615 (ICBM3, assembled from the S2 phage enrichment).

Illumina genome sequencing

The extracted DNA from the ICBM1 and ICBM2 phages and the phage enrichments was used to generate Illumina NexteraXT shotgun paired-end sequencing libraries, which were sequenced with a MiSeq instrument and the MiSeq reagent kit version 3, as recommended by the manufacturer (Illumina, San Diego, CA, USA). For quality-filtering, Trimmomatic version 0.35 (Bolger et al. 2014) or Bbduk from the BBTools package (BBTools; https://jgi.doe.gov/data-and-tools/bbtools/) were used. The assembly was performed with the SPAdes genome assembler software version 3.9.0 (Bankevich et al. 2012) and the read coverage of the whole assembly determined with QualiMap version 2.1 (García-Alcalde et al. 2012). In addition, the read mapping of enrichments reads on individual phage genomes was done BBMap from BBTools package.

PacBio library preparation, sequencing and assembly

SMRTbell template library was prepared according to the instructions from Pacific Biosciences, Menlo Park, CA, USA, following the Procedure & Checklist Greater than 10 kb Template Preparation and Sequencing. Briefly, for preparation of 10kb libraries ~4 µg of each phage DNA was sheared using a Covaris S220, Woburn, MA, USA according to the manufacturer´s instructions. DNA was end-repaired and ligated overnight to barcoded SMRTbell adapters applying components from the DNA/Polymerase Binding Kit P6 from Pacific Biosciences (Menlo Park, CA, USA). Reactions were carried out according to the

instructions of the manufacturer. One part VB-1 SMRTbell template was combined with 2.5 parts Ex53-3 SMRTbell template. BluePippin Size-Selection to greater than 4 kb was performed according to the manufacturer´s instructions (Sage Science, Beverly, MA, USA). Conditions for annealing of sequencing primers and binding of polymerase to purified SMRTbell template were assessed with the Calculator in RS Remote, PacificBiosciences, Menlo Park, CA, USA. SMRT sequencing was carried out on the PacBio RSII (PacificBiosciences, Menlo Park, CA, USA) taking one 240-minutes movie for one SMRT cell. Long read genome assemblies of all three phages have been performed using the HGAP4 Whitelisting protocol (https://github.com/PacificBiosciences/Bioinformatics-Training/wiki/HGAP-Whitelisting-Tutorial) within SMRTPipe 2.3.0 applying a genome size of 100kb and a minimum subread length of 1 kb after demultiplexing using the RS_Subreads.1 protocol contained within SMRT Portal 2.3.0.

2.2.8. Retrieval of phage genomes related to ICBM1 and ICBM2

The following datasets were queried for sequences related to ICBM1 and ICBM2: (i) the Tara Ocean Viromes (Brum et al. 2015), (ii) the Earth Virome (Paez-Espino et al. 2016), (iii) the Global Ocean Virome (Roux et al. 2016), (iv) the IMG/VR (Paez-Espino et al. 2017) and (v) the Environmental Viral Genomes (Nishimura et al. 2017a). The Tara Oceans Viromes (assembled DNA contigs and predicted proteins) and Global Ocean Virome datasets were downloaded from the iVirus (Bolduc et al. (2017b), http://ivirus.us/) using the CyVerse platform and its Discovery Environment (Merchant et al. (2016), https://de.cyverse.org/de/). The Earth Virome dataset (assembled DNA contigs) was downloaded from http://portal.nersc.gov/dna/microbial/prokpubs/EarthVirome_DP/. The downloaded datasets were imported in Geneious 9.1.5, transformed in BLAST databases and queried by megaBLAST (e-value 1e−5), using the ICBM1 and ICBM2 genomes and by BLASTp (e-value 1e−5) using the portal and terminase proteins of the ICBM1, ICBM2, SIO1 (only the terminase protein was used) and P12053L phages. The IMG-VR viral sequence database was queried by BLASTn (e-value 1e−5) webservice offered at https://img.jgi.doe.gov/cgi-bin/vr/main.cgi, using the ICBM1 and ICBM2 genomes. The proteins retrieved by Blastp were added to the databases of terminase or portal proteins from known phages, followed by multiple alignment with Muscle and calculation of phylogenetic trees with FastTree v 2.1.5 (Price et al. 2010). Further, the proteins in the vicinity of ICBM1, ICBM2, SIO1 and P12053L were selected, and their corresponding contigs retrieved. These contigs were pooled with all those retrieved by

nucleotide Blast. All contigs smaller than 34 kbps (~85% of the ICBM1 genome length) were considered incomplete and removed.

2.2.9. Genome-based phylogeny and classification

To reconstruct the whole genome-based phage phylogenetic tree, a set of genomes comprising all podoviral genomes recognized by the International Committee of Taxonomy of Viruses (ICTV) was supplemented with the cobavirus-related genomes retrieved from the different public sequence datasets (see above). For consistency, open reading frames (ORFs) for the complete set of genomes were detected using MetaGeneAnnotator (Noguchi et al. 2008), which was implemented in the VirSorter program (Roux et al. 2015). Using the Virus Classification and Tree Building Online Resource (VICTOR (Meier-Kolthoff and Göker 2017), available at https://victor.dsmz.de), all pairwise phage comparisons of the amino acid sequences were conducted via the underlying Genome-BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff et al. 2013) under settings recommended for prokaryotic viruses (Meier-Kolthoff and Göker 2017). The resulting intergenomic distances (including 100 replicates each) were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR postprocessing (Lefort et al. 2015) for each of the formulas d0, d4, and d6. The trees were rooted at the midpoint (Farris 1972) and visualized with iTOL (Letunic and Bork 2016). Taxon boundaries at the species, genus, subfamily and family level were estimated with the OPTSIL program (Göker et al. 2009) using the recommended clustering thresholds (Meier-Kolthoff et al. 2013) and an F value (fraction of links required for cluster fusion) of 0.5 (Meier-Kolthoff et al. 2014).

2.2.10. Genome annotation and protein clustering

All phage genomes compared in this study, including the already published ones, were re-annotated using the same procedure to eliminate differences resulted from different annotation pipelines. Initially, ORFs were detected using MetaGeneAnnotator (Noguchi et al. 2008) implemented in VirSorter (Roux et al. 2015). Proteins were then annotated by comparing them with several databases and manually deciding the final annotations. The NR database (http://ncbi.nlm.nih.gov/) was queried using Protein-Protein BLAST 2.6.0+, the InterPro database v66.0 (Finn et al. 2017) was queried using Inter-ProScan 5.27-66.0 tool (Jones et al. 2014), and the prokaryotic viruses orthologous groups database (Grazziotin et al. 2017) was queried using hmmscan command from HMMER 3.1b2 package (Eddy 2011). The proteins were clustered by first performing an all against all BlastP, with an e-value threshold of 1e−5

and a bitscore threshold of 50, and the results were inputted into the mcl program, with the parameters "-I 2 --abc". The online tool tRNAscan-SE v. 2.0 (Lowe and Chan (2016), http://lowelab.ucsc.edu/tRNAscan-SE/index.html) was used for tRNA prediction. Rho-independent terminators were predicted with ARNold http://rna.igmors.u-psud.fr/toolbox/arnold/index.php. Only the terminators with deltaG higher than 10.5 were considered. Details of genome features, protein clusters and DNA sequences for all cobaviruses identified are listed in SI file S2-2 and S2-3. The comparative genome map was generated using the genoPlotR package (Guy et al. 2010) from the R programming environment (https://www.rproject.org/).

2.2.11. Phylogenetic analyses of single proteins

Phylogenetic trees were constructed for the terminase protein, to gain insights about the genome ends. Phylogenetic analysis of spanin, glutaredoxin and cobalamin dependent ribonucleotide reductase (RNR) proteins was conducted to find insights about the hosts of the environmental cobaviruses and their habitat. Proteins were aligned with Muscle, and then phylogenetic trees were constructed using the FastTree v 2.1.5 program (Price et al. 2010) integrated as a plugin in Geneious v 9.1.5 (Kearse et al. (2012), http://www.geneious.com), using default parameters. Phylogenetic trees were visualized using FigTree v1.4.3. (Rambaut (2006), available at http://tree.bio.ed.ac.uk/software/figtree/).

2.2.12. Phylogenetic analysis of the host rRNA

The 16 S rRNA gene phylogenetic tree was constructed using the ARB software package (Ludwig et al. (2004), www.arb-home.de/ version arb-6.0.2). Sequences of the type material (>1,300 bp) were used for the backbone-tree using the neighbor joining method with 1500 replicates. Shorter sequences used in this study were added afterwards by parsimony interactive without using a filter.

2.2.13. Biogeographic distribution of cobaviruses and read mapping

Metagenomic data sets, download and preprocessing

The unassembled datasets used for read mapping were downloaded from the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena). A complete list of datasets used is given in SI file S2-4. The Tara Ocean survey datasets have been cleaned before their deposition at ENA (Alberti et al. 2017), and thus, we used them as such for read mapping. We cleaned the remaining datasets using BBDuk from the BBTools package (BBTools

(https://jgi.doe.gov/data-and-tools/bbtools/), as follows: (i) reads corresponding to the *Enterobacteria* phage phiX174 were filtered out; (ii) sequences of Illumina adapters and primers as provided in the BBTools package were removed (ktrim = r k = 21 rcomp = t mink = 11 hdist = 1 tpe tbo); (iii) low quality (quality value lower than 20) nucleotides from both read ends were removed and reads with low average quality (<20) or short length (<30 bases) were also removed (qtrim = rl trimq = 20 ftm = 5 maq = 20 minlen = 30). Quality control of the cleaned samples was performed on a subset of random samples, using FastaQC. The metadata associated with the metagenomes were retrieved from the NCBI site, BioSamples databases. In specific cases, if metadata were missing, we received them by direct contact with the principle investigators for the respective projects.

Read mapping

BBMap from the BBTools package was used to map the reads from the unassembled datasets to the cobaviral genomes. The output was sent to Samtools View and then to Samtools Sort to produce a sorted bam file. A phage was considered to be present in a particular sample when at least 75% of its genome was covered by reads with at least 90% identity, as previously determined (Roux et al. 2017). The relative abundance of a phage genome in a sample was calculated by the following formula: "number of bases at ≥90% identity aligning to the genome / genome size in bases / library size in gigabases (Gb)". All code used for read mapping and data analysis are available in SI files S2-7a-d.

2.3. Results and discussion

2.3.1. Isolation and host range of two *Lentibacter* sp. SH36 viruses

Two strictly lytic bacteriophages, Lentibacter virus vB_Len-P_ICBM1 (ICBM1) and Lentibacter virus vB_LenP_ICBM2 (ICBM2) were isolated to pure cultures from phage enrichments S1 and S2, respectively (Fig. 6a). The phage source in the enrichments was surface seawater collected during a March 2015 algal bloom in the southern North Sea. The host was *Lentibacter* sp. SH36, which was isolated from a seawater sample taken on 12 May 2007 in the southern North Sea during a phytoplankton bloom (Hahnke et al. 2013). ICBM1 and ICBM2 phages negatively stained with uranyl acetate had isometric capsids with hexagonal cross-sections of 58.7 ± 3.7 nm (sample size = 100 phages) and 59.2 ± 2.8 nm (sample size = 100 phages), respectively, and short tails (Fig. 6b). Assessment on 94 *Rhodobacteraceae* strains

(Table S1) showed that ICBM1 and ICBM2 have a narrow host range, infecting only *Lentibacter* sp. SH36.

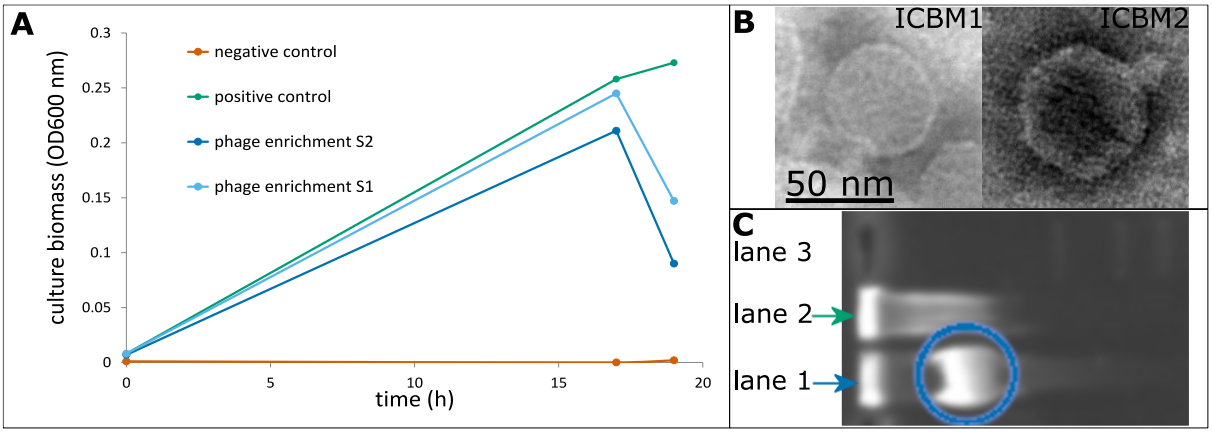


Fig. 6: a. Enrichment of phages specific for *Lentibacter* sp. SH36 from North Sea water. The experimental setup consists of: (i) Phage enrichments (blue lines) – logarithmic phase cells added to nutrient amended, 0.2 µm filtered seawater, (ii) Positive control for cell growth (green line) – logarithmic phase cells added to artificial seawater and (iii) Negative control (orange line) – only nutrient amended, 0.2 µm filtered seawater, no cells added. In the S1 and S2 enrichments the decrease in OD at 18 h is most likely due to phage cell lysis. No growth was detected in the negative control. **b.** Transmission electron micrograph of molybdenum stained, cell debris bound Lentibacter virus vB_LenP_ICBM1 and uranyl acetate stained, free Lentibacter virus vB_LenP_ICBM2. Scale bar: 100 nm. **c.** Agarose gel electrophoresis of cellular DNA from i) *Lentibacter* sp. SH36 phage infected cells, S2 enrichment (lane 1, blue arrow) and ii) not infected cells, positive control for cell growth (lane 2, green arrow). Blue circle: intracellular phage DNA. Lane 3: 1 kbp Plus DNA Ladder.

2.3.2. Sequencing the phage isolates and enrichments

To gain insights into the diversity of the phage enrichments, we sequenced both the purified phages (ICBM1 and ICBM2) and the intracellular phage fraction of the enrichments (Fig. 6c, Table S2). From the S2 enrichment two complete phage genomes were assembled, that of ICBM2 and of a third phage. The latter had 99.6% sequence similarity at nucleotide level with ICBM1 (under VICTOR formula d0 see SI file S2-6) and was named Lentibacter virus vB_LenP_ICBM3 (ICBM3). Both ICBM1 and ICBM3 have been assembled twice, once from Illumina and once from PacBio reads, with identical results (Table S2). Therefore, differences between them were real and not due to sequencing errors. According to the VICTOR (Meier-Kolthoff and Göker 2017) results, ICBM1 and ICBM3 formed a species cluster, whereas ICBM2 represented a distinct species (see section below and Fig. 7).

From the S1 enrichment we retrieved an ICBM3-like genome (99.9% identical with ICBM3, differences potentially due to sequencing errors, see Fig. S1 and S2). Read mapping with a cutoff of 100% read identity showed that both ICBM1 and ICBM3 were present in S1 (Table 1, Fig. S2). The presence of both ICBM1 and ICBM3 phages in the S1 enrichment is strengthened by the isolation of ICBM1 from this enrichment and it indicates microdiversity.

Microdiversity in phage enrichments have been previously reported (Villamor et al. 2018) and it potentially reflects the situation in the original seawater.

Using a 95% read identity cutoff for mapping, all reads in the enrichments recruited either to the ICBM1/ICBM3 or to the ICBM2 genomes (Table 1). This indicates that, without considering microdiversity, most likely no other phage was present and our isolation efforts retrieved the complete phage diversity in the enrichments at the species level.

2.3.3. Retrieval of similar phage genomes and phylogenetic positioning

Cultivated and environmental phage genomes similar to ICBM1 and ICBM2 were found in public sequence data sets. The related cultivated phages were P12053L infecting *Celeribacter marinus* IMCC12053, SIO1 infecting *Roseobacter* sp. SIO67 and four other SIO1 related strains, infecting *Roseobacter* sp. SIO67 and *Roseobacter* sp. GAI-101 (Angly et al. 2009). The last four phages had incomplete genomes, with several regions of uncertainty (long N stretches). Therefore, they were included in the phylogenetic analysis as draft genomes, but excluded from further genomic analysis. From the environmental genomes, only those bigger than 35 kbp were considered for further analysis. One of these circularized due to terminal redundancies, indicating genome completeness - EnvX (40752 bp), and five of them were incomplete, but close in size to the complete genomes and contiguous - Env9 (41607 bp), EnvY (36003 bp), EnvZ (35824 bp), Env8 (38447 bp) and Env14 (35006 bp). EnvX, EnvY, EnvZ, Env8 and Env9 were retrieved from IMG/VR /Earth Virome datasets. Env14 was retrieved from the GOV dataset (see Table 2).

The VICTOR method (Meier-Kolthoff and Göker 2017) for phage phylogeny and classification was used because it is universal and allows for an informed decision on the evolutionary relationships between prokaryotic viruses. The method was thoroughly optimized against a large reference dataset of genome-sequenced taxa recognized by the International Committee on Taxonomy of Viruses (ICTV) and showed a high agreement with the classification, particularly at the species and genus level.

The genome-based VICTOR (Meier-Kolthoff and Göker 2017) phylogeny combined with taxon boundaries prediction based on OPTISIL (Göker et al. 2009) showed that the *Lentibacter* sp. SH36 phages (ICBM1, ICBM2 and ICBM3), together with SIO1, P12053L and some of the environmental genomes formed a highly supported genus level clade (Fig. 7). This proposed genus was tentatively named here as Siovirus (from the SIO1 phage) (Fig. 7). Most of the sioviruses had a class II, cobalamin dependent RNR and were placed within one cluster,

which we called the Cobavirus (cobalamin-dependent) group. Two of the environmental sioviruses had a class I RNR and formed a separate clade. RNRs are used to convert host ribonucleotides in deoxyribonucleotides necessary for phage replication. Because the RNR class is predictive of the phage habitat (Sakowski et al. 2014) and class II RNRs point toward an association with phototrophic protists, we focused further on the Cobavirus group, which included all cultivated and part of the environmental sioviruses (Fig. 7).

In agreement with previous findings for the SIO1 and P12053L phages (Hardies et al. 2016), the cobaviruses clustered within the RIO-1 subgroup (Fig. 7). The OPTISIL based taxon boundaries reported by VICTOR (Meier-Kolthoff and Göker 2017) suggested that the RIO-1 subgroup forms a maximally supported group, which we propose to define as a new subfamily in the *Podoviridae*, and tentatively named here Riovirinae (from the RIO-1 phage).

We have excluded the ICBM3 phage from further analysis, due to its high similarity with ICBM1 (Fig. 7 and S6) and the phages SIO1_2001, OS, MB, SBRSIO67, because their genomes contained several regions of sequence uncertainty (long stretches of Ns).

Table 1: Abundance of cobaviruses in the S1 and S2 phage enrichments. *Abundance expressed in % from total bases.

Enrichments		**ICBM1**				**ICBM2**				**ICBM3**			
		95% read identity		100% read identity		95% read identity		100% read identity		95% read identity		100% read identity	
Name	Mbps	% genome covered	Abundance*	% genome covered	Abundance	% genome covered	Abundance	% genome covered	Abundance	% genome covered	Abundance	% genome covered	Abundance
S1	232.2	100.0	95.3	100.0	54.1	5.3	n.d.	0.0	0.0	100.0	97.5	100.0	66.7
S2	472.6	100.0	56.2	96.8	26.9	100.0	39.7	100.0	28.1	100.0	58.9	100.0	41.4

Table 2: Environmental cobaviruses, retrieval from databases.

Genome	**Datasets**	**(meta)genome accession in IMG/VR / GOV datasets**	**Contig name in IMG/VR / GOV datasets.**	**NCBI BioProject**	**NCBI BioSample**	**NCBI Run**	**Coordinates**	**Location**
EnvX	IMG/VR	3300003263	JGI26117J46588_1000006	PRJNA366974	SAMN06267889	SRR5268549	36.25 N 122.2099 W	Monterey Bay
EnvY	IMG/VR	3300003264	JGI26119J46589_1000020	PRJNA366976	SAMN06267891	SRR5268667	36.25 N 122.2099 W	Monterey Bay
EnvZ	IMG/VR	3300006637	Ga0075461_10000015	PRJNA375611	SAMN06343913	SRR5600317	39.283 N 75.3633 W	Delaware Bay
Env8	IMG/VR	3300000117	DelMOWin2010_c10000155	PRJNA336873	SAMN05518585	missing	39.0042816 N 77.1012173 W	Delaware Coast
Env9	IMG/VR	3300002483	JGI25132J35274_1000025	PRJNA366059	SAMN06268330	SRR5251700	18.9200 N 104.8900 W	Pacific Coast of Mexico
Env14	GOV	124_MIX	Tp1_124_DCM_0_0d2_scaffold2117_1	PRJEB4419	TARA_R100000700	ERR599367	-9.0714 N -140.5973 E	Marquesas Islands

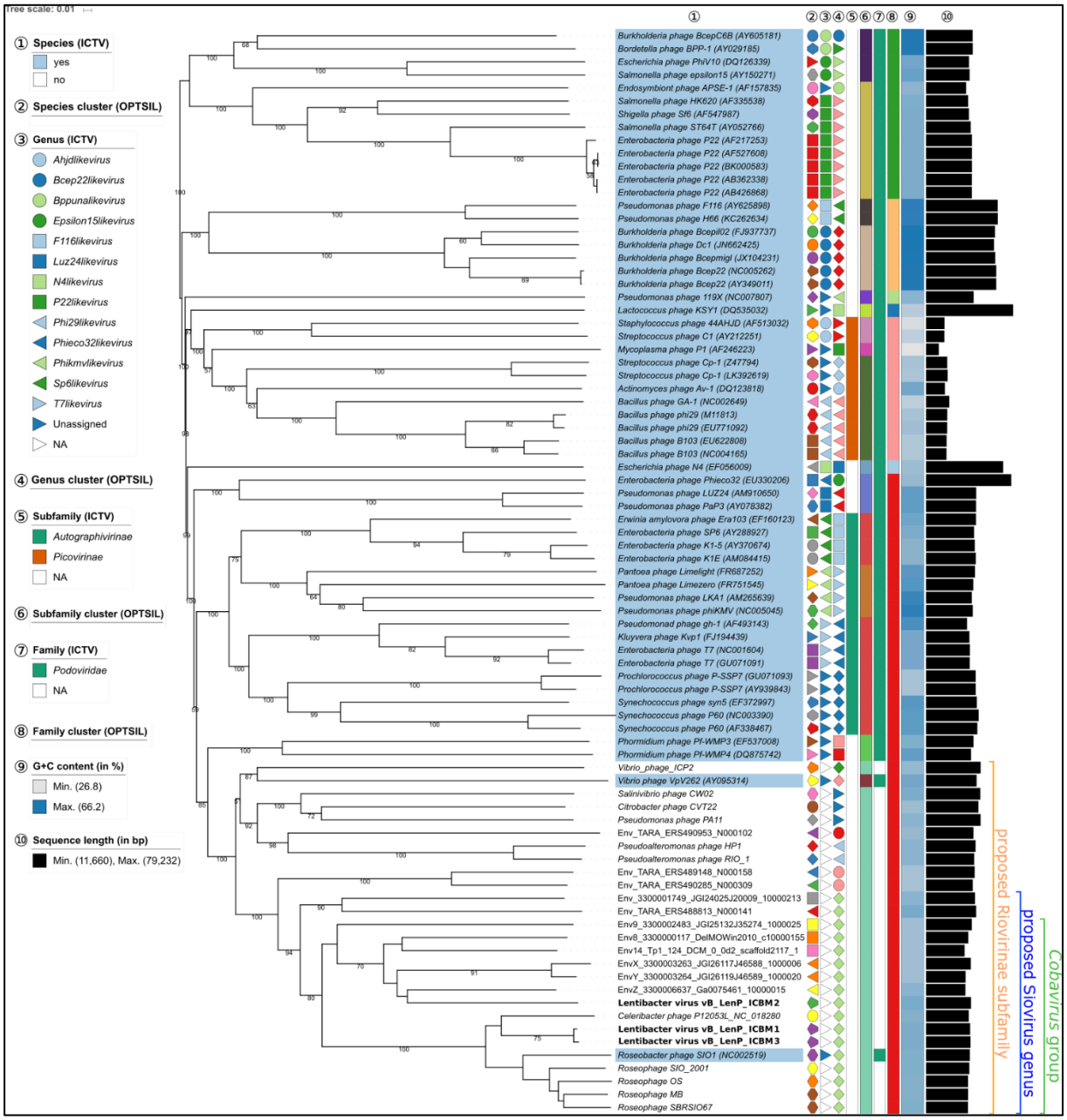


Fig. 7: Phylogenetic positioning of the *Lentibacter* sp. SH36 viruses and their relatives within the *Podoviridae*. The whole-genome-based phylogeny was inferred using the Genome-BLAST Distance Phylogeny method implemented in the VICTOR web service, using the amino acid data. Internal branch labels represent pseudo bootstrap support values if larger than 50%. The proposed subfamily Riovirinae, the proposed genus Siovirus and the Cobavirus group (sioviruses with cobalamin-dependent RNR) are annotated at the right-hand side. Further information regarding the affiliation of phages to ICTV taxa and OPTSIL clusters as well as G + C content and genome sizes is described within the figure legend (circled numbers). "Viruses annotated as "Unassigned" in legend "Genus (ICTV)" have been assigned to both an ICTV species and family but not to a genus level, whereas "NA" refers to viruses which have not been recognized as a taxon by the ICTV. The affiliation of one or more viruses to a distinct species, genus, subfamily or family cluster is indicated by a specific symbol of same shape and color.

2.3.4. Genomic organization

Genome termini

Within the proposed subfamily Riovirinae, the genome ends of the VpV262 and RIO-1 phages have been characterized and consist of direct terminal repeats (DTRs) (Hardies et al. 2003; Hardies et al. 2013). The SIO1 phage was reported to have inverted repeats (Rohwer et al. 2000), and no information was available for P12053L. We used coverage information and read structure from genome sequencing to determine the termini of the ICBM1, ICBM2 and ICBM3 phages. With Illumina 194 sequencing, a sharp drop in coverage was noticed for ICBM1 and ICBM2 genomes, which could indicate linear, non-circularly permuted genomes, with cohesive ends (Merrill et al. 2016), (Fig. S3). On the other hand, for both genomes, we noticed a region of low coverage and high G+C content (Fig. S3). The low coverage could be due to the library preparation with the NexteraXT kit, which is known to have low performance at the genome ends (Illumina 2015). Furthermore, the high G+C content could contribute to the drop in coverage (Aird et al. 2011). To elucidate the genome ends, we turned to PacBio single molecule real time (SMRT) sequencing, because native DNA is used and thus, no PCR bias is observed. Furthermore, much longer read lengths can be retrieved, which facilitates assembly, especially important for mixed samples, for example the phage enrichments. Thus, the artificial redundancies produced by the assembler at the end of the three phage contigs were larger, having a size of ~5 kb, which is equal to the mean subread length achieved with our PacBio sequencing approach (Fig. S4). In all three phage genomes, short DTRs of 159 - 173 bp were easily recognized as spikes in coverage (Fig. S4a). To delineate the final genome structure of the three phages, artificial redundancies were removed and the phage genome was adjusted to their direct terminal repeats (Fig. S4c). The exact genome start and stop positions were derived from long read mappings by a detailed inspection of the respective regions in Integrative Genome Viewer (Robinson et al. 2011) (Fig. S5). The sharp drop in coverage in the Illumina assemblies corresponded to a GC rich region of the DTRs (Fig. S3), and thus, explained the apparent discrepancy between the Illumina and PacBio read coverage data.

We further investigated the ends of the other cobaviral genomes. The ends of SIO1 were originally determined after whole genome sequencing through a combination of shotgun cloning and Sanger technology. Inverted repeats (251 to 637 bases) detected at the ends were presumed to be involved in replication (Rohwer et al. 2000). Our own analysis indicated that the ends were most likely placed incorrectly, probably due to low read coverage. Several facts supported our conclusion. First, in phylogenetic trees for the terminase gene, the phages ICBM1, ICBM3 and SIO1 grouped closely (Fig. S7), indicating that they likely have similar

genome packaging strategies (Merrill et al. 2016). Second, re-sequencing of the SIO1 genome did not retrieve the complete region of the inverted repeats (Angly et al. 2009). This was initially attributed to difficulties in PCR amplification of the repeats. On the other hand, the lack of retrieval can also suggest misassembly of the original genome sequence in this region. Third, from the three inverted repeats, none were placed at the exact ends of the genome and two of the inversions were located at the same end (Fig. S8). Inverted terminal repeats at the genome ends are found in viruses which replicate by a protein-primed mechanism, where they are positioned at the exact ends of the genomes (Escarmis et al. 1985; Savilahti and Bamford 1993). Hence, it is unlikely that the three inverted repeats of the SIO1 phage have a role in replication. Fourth, in its original order the SIO1 genome shows an ORF free region exactly in between two gene modules (Fig. S8), a region which shares high sequence similarity (~80% identity) with the DTRs from ICBM1 and ICBM3 phages (Fig. S9). An inspection of podovirus genomes from public databases revealed that related phages can have DTRs with a nucleotide identity within the 70 - 100% range (Table S5). Therefore, we used the DTRs from ICBM1 to find the genome termini and rearrange the gene order accordingly, not only for the SIO1 phage, but also for the P12053L and environmental cobaviruses. A search with the ICBM1 DTR in the SIO1 genome revealed the presence of a similar region (80% nucleotide identity) at position 8716 - 8891. Based on this approach, base 8716 from the original SIO1 genome became base 1 in the reordered genome, with the left side being concatenated at the end of the right side, and the ICBM1 phage DTR homologous region being added also at the right end (Fig. S5). Similar regions were found in all four SIO1-related phages isolated by Angly et al. (2009), having 90% to 99% nucleotide identity with the SIO1 DTR. Likewise, a search with the ICBM1 DTR in the P12053L genome found a similar region (~94% identity) at position 244 - 415. Therefore, the genome was rearranged in a similar way to SIO1, with base 244 becoming base 1. No gene rearrangement was necessary in this case. The DTR sequence was determined by homology to ICBM1 DTR and added at the right end of the genome, as well. For two of the environmental genomes (EnvX and Env9) we determined the genome start by finding regions with high identity with the left region of the ICBM2 DTR (>70% over 49 nt, see Fig. S9). However, the complete sequence of the DTRs could not be established, because of the low similarity over the remaining alignment (~50% identity). Env9 was not circular, but because we established the start at position 3404, we plotted the position 1 - 3403 at the end of the genome in figure 8. The remaining environmental genomes showed no regions of similarity with the ICBM1 or ICBM2 phage DTRs, presumably due to their incompleteness. The phylogenetic positioning both in the

GBPD and terminase trees strongly support the presence of DTRs at the genome termini for all the environmental genomes.

Our results show that the cultivated cobaviruses have DTRs. The presence of DTRs indicate that cultivated cobaviruses, similar to the T7 phage, most likely use long concatemeric DNA molecules as intermediates in replication and packaging, concatemers formed by the annealing of 3′ single strands resulted at the DTR level during replication (Serwer 2005; Kulczyk and Richardson 2016). The 5′ ends of all cobaviral DTRs have a conserved, G+C rich region (Fig. S9), underlining a potentially more important role of this region in genome circularization or replication, for example as enzyme binding site. The phylogenetic positioning in the GBDP-based VICTOR tree (Fig. 7) as well as in the terminase tree (Fig. S7) suggests that the environmental cobaviruses also have DTRs and thus potentially the same DNA replication strategy.

Gene composition and modular organization

ICBM1 and ICBM2 phages had linear genomes of ~40 kb, a G + C content of ~47% (Table S3) and 58 and 55 ORFs, respectively. More than half of the ORFs coded for hypothetical proteins. No tRNAs were found. The genes were organized in two genomic arms, with opposite transcriptional directions and separated by a bidirectional, rhoindependent transcriptional terminator (Fig. 8, Table S4). We found protein-encoding genes for replication and nucleotide metabolism on the left arm: two nucleases, a DNA polymerase, a dual primase/helicase, a cobalamin dependent RNR, a glutaredoxin, a ThyX thymidylate synthase (ThyX), a guanosine 3′, 5′-bispyrophosphate (ppGpp) hydrolase (MazG) and a P-loop containing nucleoside triphosphate hydrolase (PhoH) (Leduc et al. 2004; Clokie and Mann 2006; Gross et al. 2006; Dwivedi et al. 2013; Sengupta and Holmgren 2014; Smet et al. 2016). On the right genomic arm, we found genes for lysis and virion structure and morphology. Both phages had spanins, which were easy to recognize due to their specific architecture. At the N terminus the spanins had a lipoprotein domain for binding to the outer membrane. At the C terminus they had a transmembrane domain for binding to the inner membrane (Summer et al. 2007a). For endolysins, ICBM1 had a lysozyme-like protein and ICBM2 had an N-acetylmuramoyl-L-alanine amidase. The lysis genes were followed by genes for the internal virion proteins (IVP) B and D, a Gcn5-related N-acetyltransferase (GNAT), the tubular proteins A and B, a major capsid protein, a scaffolding protein, a portal protein, a large terminase subunit, two tail fibers and three tail assembly chaperone proteins (see Fig. 8). With the exception of the endolysins, all other genes have been previously annotated in SIO1 or P12053L phages (Rohwer et al. 2000; Summer et al. 2007a; Angly et al. 2009; Kang et al. 2012; Hardies et al. 2016). A previous

study (Hurwitz et al. 2013) annotated the gene for the pc40 protein from SIO1 as a long-chain fatty acid transporter (FadL) and thus listed it as AMG. However, our BlastP and InterProScan searches identified pc40 as a Gcn5-related N-acetyltransferases (GNAT). Based on the GNAT domain, pc40 could correspond to gp13 from T7, which is also positioned next to the internal virion proteins and has been suggested to play a role in virion morphogenesis (Kemp et al. 2005).

Within the Cobavirus group, the genetic composition and synteny was mostly conserved (see Fig. 8). All genomes were organized in two arms, with genes for replication and nucleotide metabolism on the left and lysis and virion structure and morphogenesis on the right. This genomic organization was not previously reported for the SIO1 and P12053L phages, but it became evident once the genomes were rearranged according to the DTR positions (see section above). Furthermore, it appears in other members of the proposed Riovirinae subfamily, although the contained modules can vary (Seed et al. 2011; Hardies et al. 2013). Most cobaviruses had a bidirectional, rho-independent transcriptional terminator in between the two genomic arms, indicating a likely transcriptional separation (see Fig. 8). This type of terminator was shown to be functional in vitro for the Pf-WMP3 phage (Liu et al. 2017).

Most of the genes with a functional annotation in ICBM1 and ICBM2 phage were also found in all other cobaviruses, with the exception of glutaredoxin, ThyX and PhoH, which were not found in some of the environmental cobaviruses (see Fig. 8). The endolysins were found in all cobaviruses, with the exception of EnvX and EnvY. They were free of membrane anchoring domains, indicating that cell lysis most likely proceeds via the canonical holin-endolysin pathway (Young 2013; Young 2014). The endolysins were diverse both in sequence and enzymatic function, encoding either lysozyme-like domains, or N-acetylmuramoyl-L-alanine amidase or peptidase domains. The spanin was found in all cultured cobaviruses, and only in two of the environmental genomes, Env9 and Env8. In the vicinity of the spanin and endolysins genes we found several genes encoding one or two transmembrane domains, representing potential holins and antiholins (see Fig. 8).

In agreement with cobavirus phylogenetic positioning and virion morphology revealed by TEM (Fig. 6b), the genes present in the virion structure and morphogenesis module most likely indicated a podoviral, T7-like virion structure (Hu et al. 2013; Cuervo et al. 2013; Guo et al. 2014). A conserved genetic composition and synteny characterized the genomic region between the lysis module and the terminase gene (Fig. 8). The genomic region between the terminase and the 3′ end of the genome was variable both in gene count and composition and it

encoded the proteins required for tail fibers, fiber connectors or tail assembly proteins. Most proteins were unique to a single phage or shared by a few. Some proteins (pc53, pc56, pc311) were similar to tail fibers or fiber connectors from myoviruses or siphoviruses, as noticed for other phages in the RIO-1 subgroup (Hardies et al. 2016). For example, pc53 resembled the short tail fiber protein from the T4 phage, a myovirus (Leiman et al. 2010). The pc56 protein was similar with the L-shaped tail fiber protein from the T5 phage and the T5-like siphoviruses DT57C and DT571/2 (Golomidova et al. 2016). Therefore, the tail fibers of the cobaviruses likely depart from the simplicity of T7-like fibers, which are formed from a single protein (gp17) directly connected to the tubular protein A.

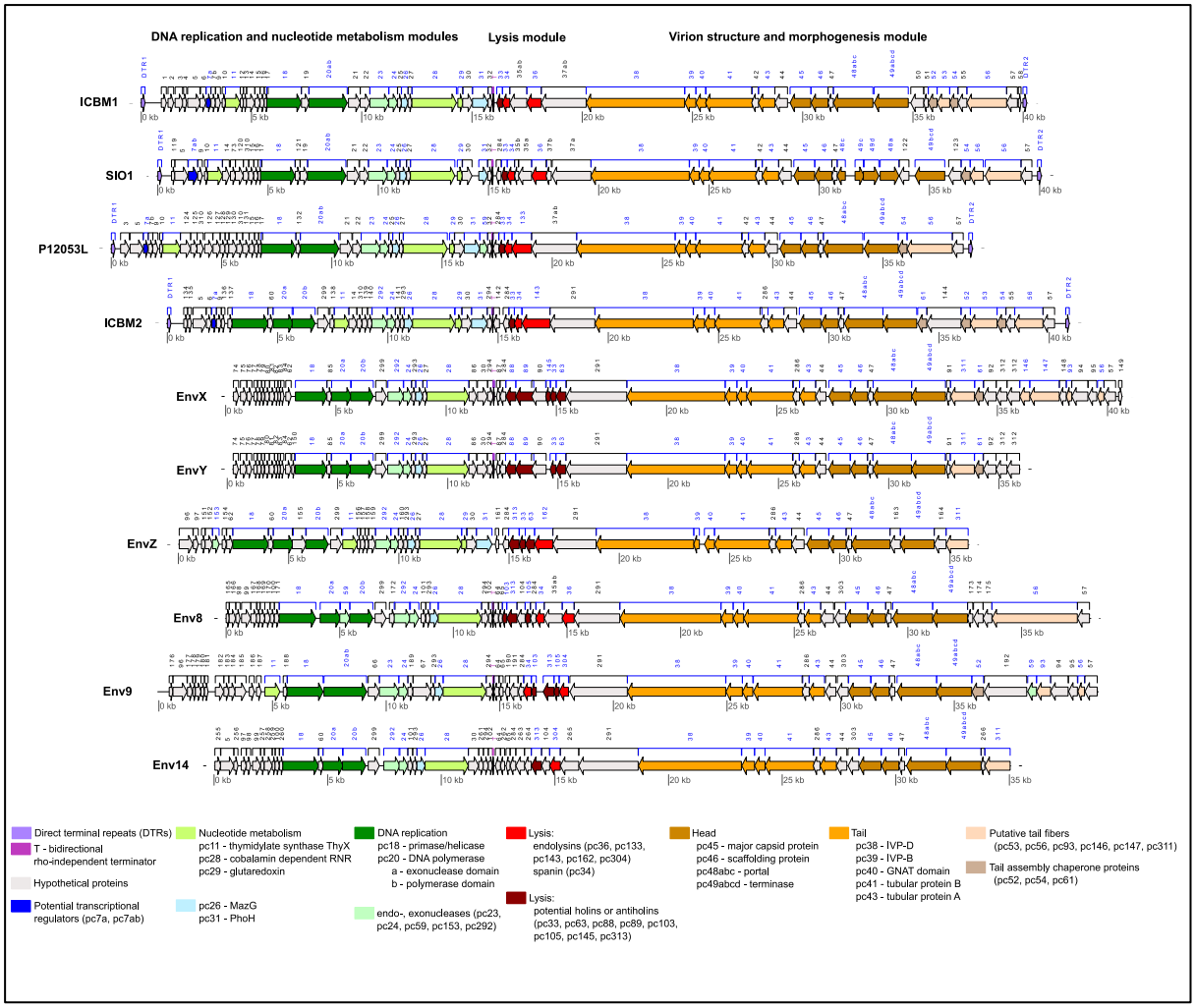


Fig. 8: Genome map of cultured and environmental cobaviruses. The genomes are centered in the bidirectional rho-independent terminator. With the exception of EnvX and Env9, all other environmental genomes are incomplete, with sequence information missing at the two ends of the genomes (the host interaction and tail fibers modules). Blue numbers indicate protein clusters with functional annotation.

2.3.5. Cobaviral hosts

The Cobavirus group contained both cultivated phages, with known hosts, and environmental phages, with unknown hosts. To have an up to date phylogeny of the hosts of cultivated cobaviruses, we built a 16 S rRNA gene-based tree (Fig. S10). Our results showed that *Roseobacter* sp. SI067 belongs to the *Lentibacter* genus (>99% nucleotide identity with the type species) and *Roseobacter* sp. GAI-101 to the *Sulfitobacter* genus (>98% nucleotide identity). Therefore, hosts of cultivated cobaviruses comprise members of the *Lentibacter*, *Sulfitobacter* and *Celeribacter* genera, within the *Rhodobacteraceae* family.

Furthermore, we searched for clues linking the environmental cobaviruses to potential hosts. A search in the CRISPR spacer database from IMG/VR returned no results, and no tRNAs where found within the cobavirus genomes. We found, however, three lines of evidence that point to environmental cobaviruses infecting members of the *Rhodobacteraceae* family. First, cobaviruses clustered into one genus, with nine out of 15 representatives known to infect *Rhodobacteraceae* members. According to Meier-Kolthoff and Göker (Meier-Kolthoff and Göker 2017), phage genera usually infect hosts within the same family. Second, all cobaviruses had a cobalamin-dependent RNR gene, encoding an enzyme used to reroute host resources toward phage replication. These phages need to infect bacteria able to synthesize cobalamin, and this ability is widespread within marine *Rhodobacteraceae* (Sañudo-Wilhelmy et al. 2014). Genes involved in vitamin B12 synthesis are present in the two publicly available genomes from the cobaviral hosts. Additionally, in phylogenetic trees the RNRs from environmental cobaviruses clustered closely with ICBM2 (Fig. 9), whose host is *Lentibacter* sp. SH36. Third, all cultivated and two environmental cobaviruses (Env8 and Env9) had a spanin gene characteristic for roseophages. BLAST searches in the NR database from NCBI with the cobaviral spanins returned hits only from roseophages or members of *Rhodobacteraceae*, with the exception of one *E. coli* phage hit, which had very low similarity (Fig. 9a). This is not surprising, considering that spanins have little sequence homology to each other and Summer et al. (Summer et al. 2007a) found no homolog for the SIO1 spanin. Using prophage prediction (PHASTER, Arndt et al. (2016)), we determined that the spanins from *Rhodobacteraceae* genomes were present in putative prophage regions (Table S6). Therefore, phages infecting *Rhodobacteraceae* have similar spanins, another evidence that Env8 and Env9 most likely infect *Rhodobacteraceae*.

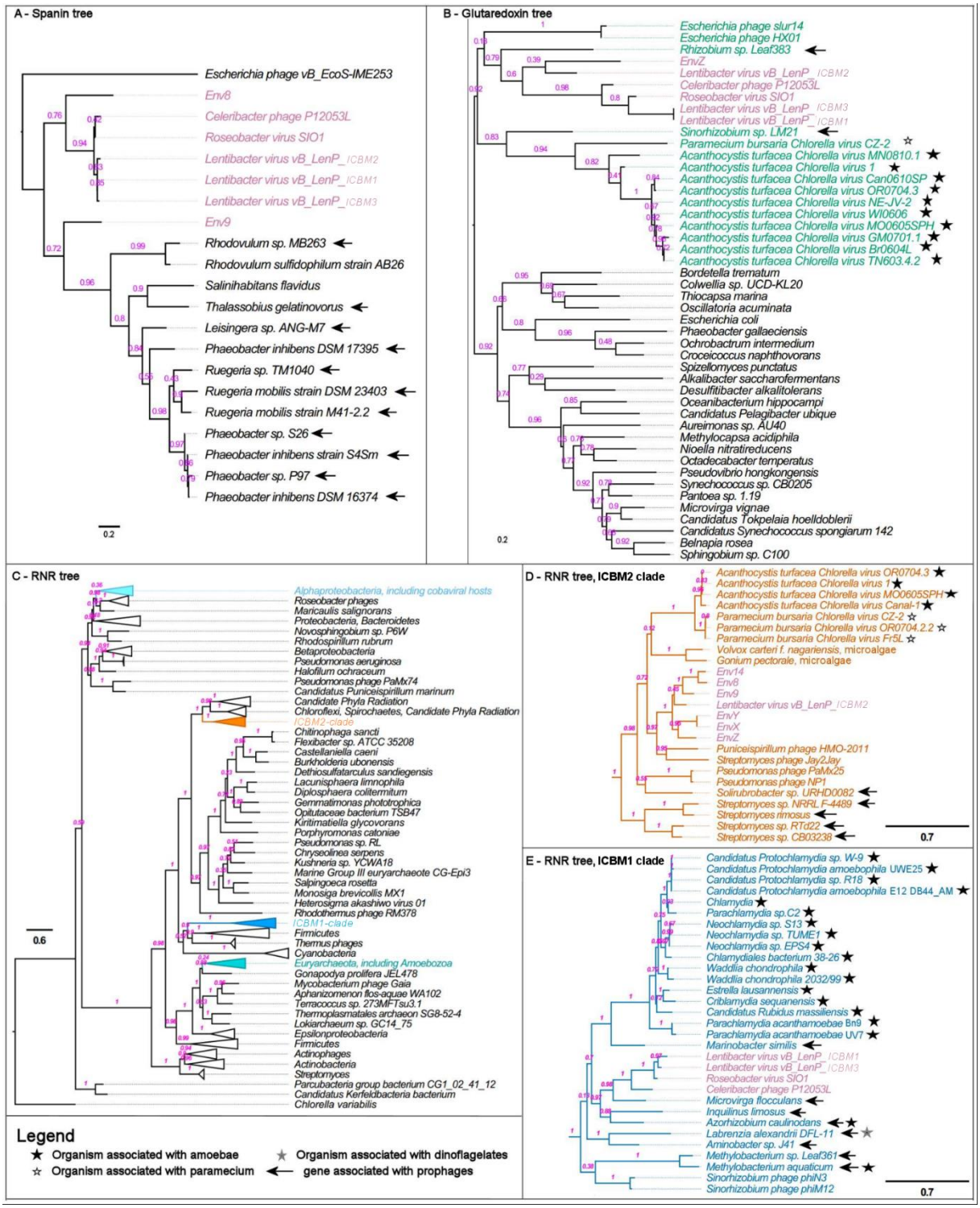


Fig. 9: Phylogenetic analysis of the spanin **(a)**, glutaredoxin **(b)** and ribonucleotide reductase **(c–e)** genes from cobaviruses. The evolutionary history was inferred using the approximately-maximumlikelihood method implemented in FastTree 2.1.5. The node labels represent Fast Tree support values. The tree is drawn to scale, with branch lengths measured in number of amino acid substitutions per site. Association of the bacteria to eukaryotic organisms is indicated by stars. Location of the spanin, glutaredoxin or RNR genes in prophage regions, predicted with PHASTER (Arndt et al. 2016), is indicated by arrows and further detailed in the appendix (Tables S6, S11 and S12).

2.3.6. Environmental distribution of the cobaviruses

Cobaviruses have been isolated from three distinct coastal locations in the Northern Hemisphere: SIO1 from the American coast of the Pacific Ocean (Scripps Pier, California) (Rohwer et al. 2000; Angly et al. 2009), P12053L from the Yellow Sea, South Korea (Kang et al. 2012) and ICBM1 and ICBM2 from the North Sea, Germany (this study). Sequences related to the SIO1 and P12053L phages were previously reported in predominantly coastal viromes from the North Pacific USA coast (Scripps Pier, British Columbia), the Gulf of Mexico, the Arctic Ocean, the North Atlantic (Chesapeake Bay and Sargasso Sea) and the Yellow Sea (Goseong Bay) (Breitbart et al. 2002; Bench et al. 2007; Angly et al. 2009; Hwang et al. 2016; Hwang et al. 2017). To further assess the environmental distribution, we queried for the presence of cobaviruses in more than 5,000 publicly available marine metagenomes, by mapping unassembled reads to cobaviral genomes. The queried metagenomes covered a wide range of marine environments, from coastal to open oceans, and from water column, to benthic, sediment and animal associated samples. All metagenomes from the Tara Ocean Expeditions (Alberti et al. 2017) were included in the dataset, as well as the viromes from Malaspina expeditions (Duarte 2015), along other marine datasets available in ENA in November 2017 (see SI file S2-4 for a complete list of all datasets used). We found cobaviruses in bonafide viromes and in metagenomes from cellular fractions, mostly in the prokaryotic range, but also in the small protist range (Fig. 10 and SI file S2-5). The presence of cobaviruses in cellular fractions could be explained by i) active infections at the time of sample collection, or ii) free phage particles retained on the large pore size filters by unspecific binding to the filter membrane or cell debris. A third explanation, the integration of cobaviruses in bacterial genomes as prophages, is unlikely, because, firstly, no cobaviral genes with functional annotations indicated a temperate life style. Secondly, although in phylogenetic trees using spanin cobaviruses were placed close to prophage regions from roseobacter genomes (Fig. 9a, Table S6), in whole genome trees cobaviruses were distant from these prophages (Fig. S11).

Cobaviruses were detected in the euphotic water column, mainly close to coastal areas but also in the open ocean of the Pacific, Atlantic and Indian Oceans, as well as in the North Sea, theMediterranean, the Adriatic, the Red Sea, the Arabian Sea, the Yellow Sea, the Salish Sea and in several estuaries (see Fig. 10a for an overview and SI file S2-5 for the list of coordinates). These waters span temperate to tropical regions. Hot spots for cobaviruses were in bays or estuaries, with several cobaviruses being detected in these locations, for example the Goseong Bay, Delaware Estuary and Chesapeake Bay (Fig. 10b, Table 3). This is consistent with a 16 S rRNA based survey, which retrieved known cobaviral hosts mainly from coastal

areas (Fig. S12, Tables S4-10). Generally, abundance of cobaviruses was low. However, it increased markedly in the Port of Los Angeles samples (Fig. 10b, SI file S2-5), where roseophage SIO1, its related phages, and their respective host have been isolated (Angly et al. 2009).

Specific cobavirus strains are cosmopolitan, as revealed by the finding of specific genomes across distant geographical locations. For example, the ICBM1 and ICBM2 phages have been isolated from the North Sea, but similar phages have been found by read mapping in metagenomes from the Australian Coast (ICBM1), and from the Goseong Bay, Yellow Sea and the Port of Los Angeles (ICBM2, Fig. 10). Similarly, environmental cobaviruses have been found by read mapping not only in the metagenomes from which they were originally assembled (Fig. 10a, Table 2), but also in many other locations (Fig. 10b). The biogeographic distribution of the cobaviruses could be explained by passive transport by oceanic currents and local selection by environmental factors shaping host communities, as proposed for marine viruses by Brum et al. (2015). In addition, considering that many positive locations for cobaviruses are also harbor areas, ship ballast water could contribute to virus transport across the oceans, in line with the findings by Kim et al. (2016).

A few of the metagenomes positive for cobaviruses were part of sampling time series, allowing us to catch a glimpse of the cobaviral seasonality (see Table 3). In the North Sea, in metagenomic samples focused on spring/early summer algal blooms (Teeling et al. 2016), EnvZ and Env9 were present in successive years, mostly post-bloom, but also before and during the blooms. In Goseong Bay (Hwang et al. 2017) and Delaware Estuary, cobaviruses where present in early spring, late summer, fall and winter. This suggests that cobaviruses persist throughout the years in coastal environments.

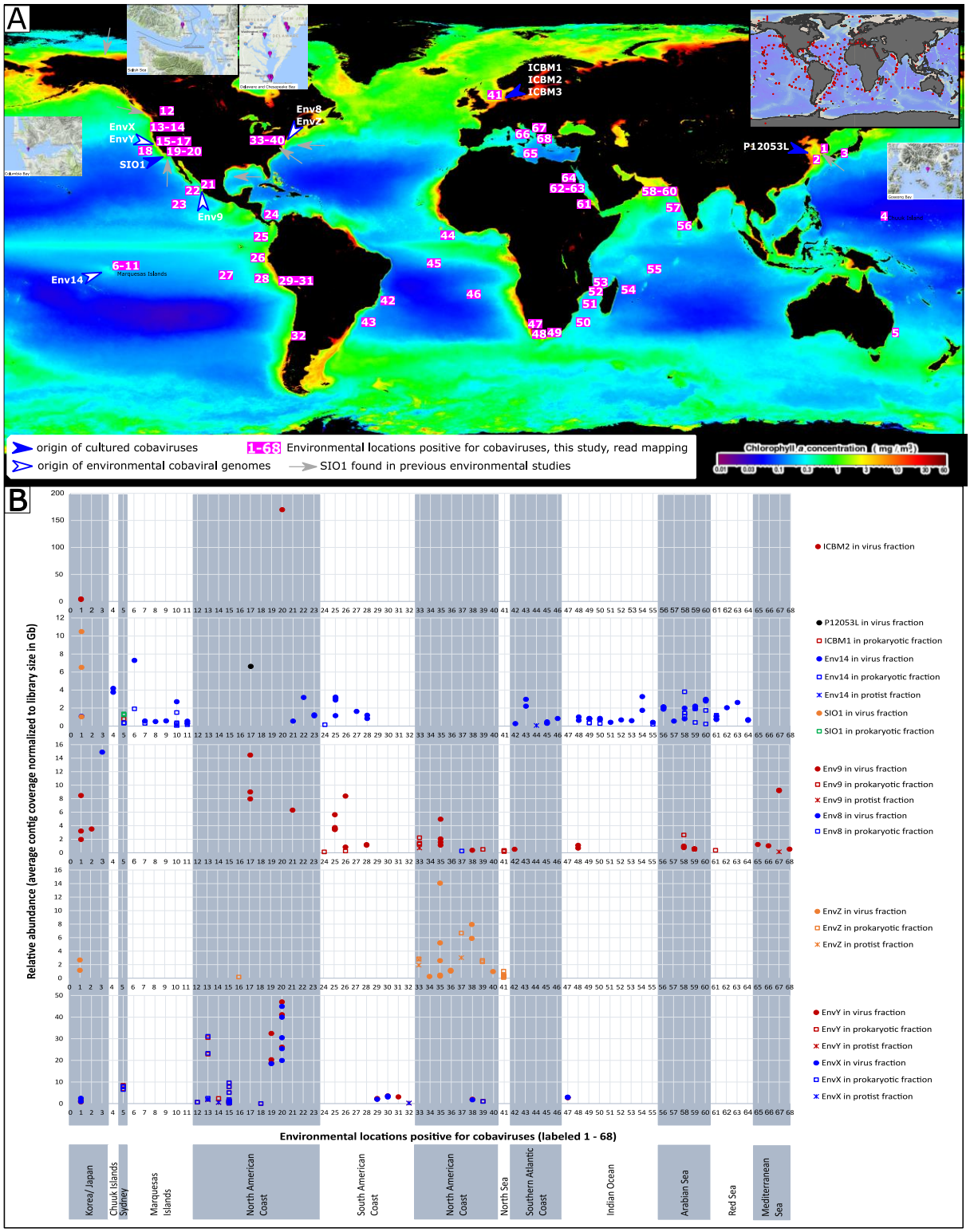


Fig. 10: Global distribution of cobaviruses **(a)** and their abundance (average contig coverage per Gb metagenome) in metagenomics samples from marine environments **(b)**. **a (i)** main map – each location were cobaviruses were found by read mapping in this study is labeled with a number, from 1 to 68; **(ii)** inset upper right corner – locations of all metagenomes searched in this study. Locations superimposed on an ocean chlorophyll concentration map (Aqua MODIS mission, 2010 annual composite, https://oceancolor. gsfc.nasa.gov/cgi/l3 – NASA Goddard Space Flight Center, Ocean Biology Processing Group, 2014).

Table 3: Seasonal occurrence of cobaviruses in different locations. *n.d. = not determined.

Location		Phage	Year	Date (day.month)	Bloom situation
Helgoland		EnvZ	2010	11.05	post bloom
				18.05	post bloom
			2011	28.04	in between blooms
			2012	05.04	post bloom
				24.05	post bloom
				31.05	post bloom
				07.06	post bloom
		Env9	2012	08.03	pre bloom
			2014	20.06	n.d.*
Goseong Bay		SIO1	2014	10.03	n.d.
				06.12	n.d.
		VB2	2014	10.03	n.d.
		EnvX	2014	10.03	n.d.
				06.12	n.d.
		EnvY	2014	10.03	n.d.
				06.12	n.d.
		EnvZ	2014	10.03	n.d.
				06.12	n.d.
		Env9	2014	10.03	n.d.
				08.09	n.d.
				20.09	n.d.
				06.12	n.d.
		Env14	2014	10.03	n.d.
Delaware Estuary	station 36	EnvZ	2015	11.04	n.d.
	station 37	EnvZ	2015	13.04	n.d.
		Env8			
	station 38	EnvX	2015	11.04	n.d.
		EnvY			
		EnvZ			
		Env9			
	station 39	EnvX	2015	15.04	n.d.
		EnvY			
		EnvZ			
		Env9			
	station 40	EnvZ	2015	17.08	n.d.
Chesapeake Bay	station 33	EnvZ	2014	01.11	n.d.
				03.11	n.d.
		Env9	2014	22.03	n.d.
				spring	n.d.
				03.11	n.d.
	station 34	EnvZ	2014	30.08	n.d.
	station 35	EnvZ	2014	22.03	n.d.
				02.11	n.d.
				03.11	n.d.
		Env9	2014	22.03	n.d.
				03.11	n.d.

2.3.7. Protists as habitat for the cobaviral hosts

We used the search for cobaviruses in microbial metagenomes (see section above), as well as glutaredoxin and RNR trees, to find indications regarding the habitat of the cobaviral hosts. Cobavirus genomes were present in several metagenomes from the protist size fractions (Fig. 10b, SI file S2-5), suggesting that cobaviruses infect protist-associated bacteria. Most often cobaviruses were present in the 0.8 - 5 µm fraction, which could arguably be contaminated with free-living bacteria, but also in the >3 µm fraction (SI file S2-5), which makes it more likely that the bacterial cells present there were attached to or consumed by

protists. The small protist size fraction is dominated by *Alveolata*, including dinoflagellates, followed by *Rhizaria* and *Stramenopila* (Vargas et al. 2015), thus consisting of phagotrophic, parasitic and phototrophic species. Previous research (Sakowski et al. 2014) proposed that class II RNR-containing phages are infecting vitamin B12-producing bacteria associated with phototrophic protists. This was based on the phylogenetic positioning of phage class II RNRs, including that of SIO1, next to chloroviruses (viruses of the single cell green alga *Chlorella*) and microalgae, and on the cobalamin requirement by the RNR. Other studies showed that marine *Rhodobacteraceae* can be associated with protists (Green et al. 2010; Guannel et al. 2011; Fiebig et al. 2013; Chen et al. 2014), including close relatives of the cobaviral hosts (Tables S7 and S8).

In our own analysis, the phylogenetic neighborhood of the cobaviral glutaredoxin and RNR (Fig. 9) points toward a relationship of the cobaviral hosts not only with phototrophic protists, but also with phagotrophic/mixotrophic protists, as detailed further. Interactions with phagotrophs/mixotrophs, especially amoeba, but also paramecium and dinoflagellates, are a recurring theme in the RNR and glutaredoxin trees (Fig. 9). For example, several organisms found in the vicinity of cobaviruses in both glutaredoxin and RNR trees are resistant to amoeba (Pagnier et al. 2008; Delafont et al. 2013; Pagnier et al. 2015; Paquet and Charette 2016) and, most significantly, the *Chlamydiae* are well known endosymbionts or lytic parasites of amoebae (Taylor-Brown et al. 2015). Even the chloroviruses point towards amoeba or paramecium interactions, because they infect only *Chlorella* strains that form endosymbioses with amoebae or paramecium (Hoshina et al. 2010; Quispe et al. 2017). Amoebae themselves have a functional cobalamin-dependent RNR (Crona et al. 2013) (Fig. 9a) and therefore, they need partners such as the cobaviral hosts, able to synthesize vitamin B12. Many dinoflagellates are mixotrophic or heterotrophic, being able to ingest diverse prey, including bacteria (Jeong et al. 2010), and their dependence on external sources of vitamin B12 has been documented previously (Tang et al. 2010; Wagner-Döbler et al. 2010; Cruz-López and Maske 2016). Taking all this into consideration, we propose that at least some of the cobaviral hosts are frequently interacting with phagotrophic/mixotrophic protists, beyond just being grazed upon. It is possible that the cobaviral hosts, associated or not with phototrophic algae, have developed mechanisms to escape digestion in food vacuoles of predatory protists, in a similar way to amoebae-resistant bacteria (Pagnier et al. 2008; Schmitz-Esser et al. 2008; Bertelli and Greub 2012; Pagnier et al. 2015; Schulz et al. 2015; Paquet and Charette 2016).

In their interactions with phototrophic and mixotrophic protists, marine *Rhodobacteraceae* form both mutualistic and pathogenic relationships, the latter resulting in

protist lysis (Wang et al. 2014; Segev et al. 2016). Therefore, by exerting control on their host populations, cobaviruses could have roles in biogeochemical cycling that go beyond the release of bacterial cellular components. They could indirectly affect both marine phytoplankton growth, and thus carbon fixation, and its lysis, and thus release of the fixed organic matter in the environment. Future studies are necessary to understand the roles that cobaviruses play in the environment and their impact on roseobacter populations.

2.3.8. Conclusions

This study significantly extends our knowledge of phages infecting organisms of the Roseobacter group, a key player in the cycling of organic matter in marine ecosystems. Using an approach that combines phage isolation with database mining for environmental phage genomes we have delineated the new Cobavirus group. Our biogeography survey included marine metagenomes from the viral, prokaryotic and protist fractions and is to date one of the largest surveys applied for a specific phage group. Cobaviruses impact Roseobacter populations at a global scale, from temperate to tropical marine waters, especially in coastal areas, and thus could have an influence on the biogeochemical cycling in these environments.

3. Taxonomic proposal of the *Zobellviridae* family

When the study described in the previous chapter 2 was published in *The ISME journal* in 2019, the newly proposed viral taxa needed to be submitted to the International Committee on Taxonomy of Viruses (ICTV) to be officially recognized. Meanwhile, the ICTV had announced that the families Podoviridae, Siphoviridae and Myoviridae would soon be dissolved, because they had been shown to be polyphyletic (Adriaenssens et al. 2021). Thus, the subfamily "Riovirinae", described in the ISME publication and chapter 2 of this thesis, was officially recognized as the new family *Zobellviridae* in the order Caudovirales (which was later dissolved as well). The following chapter corresponds to the taxonomic proposal of the *Zobellviridae* family (Bischoff et al. 2020). Again, the spelling of viral taxa (e.g., in italics or with quotation marks), corresponds to the classification at the time of publishing and might now be partially invalid. For a tabular overview of the newly proposed taxonomy, see Table 5, submitted with the proposal.

Authors of the proposal:

Vera Bischoff[1], Evelien M. Adriaenssens[2], Andrew M. Kropinski[3], Melissa Duhaime[4], Cristina Moraru[1]

[1]Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Germany

[2]Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

[3]University of Guelph, Canada

[4]Department of Ecology and Evolutionary Biology, University of Michigan, USA

Publicly available at: https://ictv.global/ictv/proposals/2020.187B.R.Zobellviridae.zip.

List of the ICTV Study Group(s) that have seen this proposal: Bacterial and Archaeal Viruses Subcommittee Caudovirales Study Group

3.1. Short description

One (1) new family (*Zobellviridae*) including one (1) new subfamily (*Cobavirinae*), eight (8) new genera (*Siovirus, Veravirus, Melvirus, Paundecimvirus, Citrovirus, Salinovirus, Vipivirus, and Icepovirus)* and nine (9) new species will be created in the order *Caudovirales*.

3.2. Abstract

We propose here a new family of the order *Caudovirales,* called *Zobellviridae.* As members of *Zobellviridae* we propose one new subfamily, called *Cobavirinae,* several new genera and several species, most of which are new. We propose that the *Cobavirinae* subfamily comprises the genera *Siovirus* and *Veravirus.* The genus *Siovirus* comprises the species *Lentibacter virus ICBM1, Celeribacter virus P12053L* and *Roseobacter virus SIO1*. The genus *Veravirus* has only one species, *Lentibacter virus ICBM2*. The other genera we propose to belong to *Zobellviridae*, but do not assign to any subfamily are: i) *Melvirus,* with two species, *Pseudoalteromonas virus HP1* and *Pseudoalteromonas virus RIO1,* ii) *Paundecimvirus,* with the species *Pseudomonas virus PA11,* iii) *Citrovirus,* with the species *Citrobacter virus CVT22,* iv) *Salinovirus,* with the species *Salinivibrio virus CW02,* v) *Vipivirus,* with the species *Vibrio virus VpV262,* and vi) *Icepovirus,* with the species *Vibrio virus ICP2.* From the above species, all are newly proposed here, with the exception of *Roseobacter virus SIO1* and *Vibrio virus VpV262.* All species have cultured representatives.

3.3. Text of proposal

This taxonomic proposal is based on phylogenetic calculations performed by Bischoff et al. (2019) and additional calculations of the intergenomic similarities at nucleotide level. Shortly, the isolation of two new phages, Lentibacter virus vB LenP ICBM1 (ICBM1) and Lentibacter virus vB LenP ICBM2 (ICBM2), and the enrichment of a third, Lentibacter virus vB LenP ICBM3 (ICBM3), all infecting *Lentibacter* sp. SH36, has led to the recruitment of related phages from publicly available databases. Then, two type of phylogenetic trees were constructed: i) a whole genome phylogenetic tree, based on protein sequences (see Fig. 11) and ii) a single protein tree, based on the terminase protein (see Fig. 12).

For genome-based phylogenetic analysis, Bischoff et al. (2019) have built a dataset containing all ICTV-recognized genomes of podoviruses, ICBM1 and ICBM2 phages and their relatives retrieved by similarity searches in publicly available sequence databases. Pairwise phage comparisons of the amino acid sequences were performed with the Virus Classification and Tree Building Online Resource (VICTOR) (Meier-Kolthoff and Göker 2017) using the Genome-BLAST Distance Phylogeny (GBDP) (Meier-Kolthoff et al. 2013) method with settings recommended for prokaryotic viruses (Meier-Kolthoff and Göker 2017). Intergenomic distances were calculated with 100 replicates each. A balanced minimum evolution tree with branch support was constructed using FASTME including SPR postprocessing (Lefort et al. 2015) for the d6 formula. The trees were rooted at the midpoint (Farris 1972) and visualized with iTOL (Letunic and Bork 2016).

Taxon boundaries at the species, genus, subfamily and family level were estimated with the OPTSIL program (Göker et al. 2009) using the recommended clustering thresholds (Meier-Kolthoff and Göker 2017) and an F value (fraction of links required for cluster fusion) of 0.5 (Meier-Kolthoff et al. 2014). The distance thresholds used by VICTOR (Meier-Kolthoff and Göker 2017) for species, genus and subfamily demarcation can be found in Table 4 of this proposal. These thresholds have been shown to yield the highest agreement with the ICTV 2014 classification regarding the investigated taxonomic ranks (Meier-Kolthoff and Göker 2017). The respective thresholds for the analysis of amino acid datasets at the species, genus and subfamily level are 0.118980, 0.749680 and 0.888940, respectively. These distance thresholds were applied to matrices of accurate intergenomic distances calculated using the Genome BLAST Distance Phylogeny (GBDP) approach (Meier-Kolthoff et al. 2013) under settings optimized for the comparison of phage genomes (Meier-Kolthoff and Göker 2017).

In addition, for this proposal we used VIRIDIC (viridic.icbm.de, Moraru et al. (2020)) to calculate nucleotide based intergenomic similarities (Fig. 13). VIRIDIC first calculated all possible pairwise alignments based on BLASTN. For one genome pair, the number of identical nucleotide matches reported by BLASTN were summed up for all aligned genomic regions. In the case of overlapping alignments, the overlapping parts were removed from one of the aligned regions, such that, at the end, the different genome regions were represented only once in the alignments. The intergenomic similarity were then calculated based on a formula previously proposed by Meier-Kolthoff and Göker (2017).

$$simAB = ((idAB + idBA) * 100)/(lA + lB),$$

where

$$idAB = identical\ bases\ when\ genome\ A\ is\ aligned\ to\ genome\ B$$

$$idBA = identical\ bases\ when\ genome\ B\ is\ aligned\ to\ genome\ A$$

$$lA = length\ genome\ A$$

$$lB = length\ genome\ B$$

$$simAB = intergenomic\ similarity\ between\ genomes\ A\ and\ B$$

3.3.1. *Zobellviridae* family

In the whole genome tree (Fig. 11) prepared in Bischoff et al. (2019), ICBM1, ICBM2 and ICBM3 phages were placed in a maximally supported clade together with the RIO-1 phage and several other phages, in agreement with previous findings (Hardies et al. 2016). This clade is referred here as the RIO clade. Phages in the *Autographiviridae* family (formerly *Autographivirinae* subfamily) were placed in a closely related clade, of similar ranking to RIO clade. A third clade of similar rank was formed by the two *Phormidium* phages (see Fig. 11). VICTOR based taxon boundaries (see Table 4 for criteria used for taxonomic delineation) suggested that each of the three clades form subfamilies on their own (see Fig. 11). Therefore, Bischoff et al. (2019) suggested to give the RIO clade the rank of subfamily and to name it *Riovirinae*, a subfamily of the family *Podoviridae*. However, meanwhile the *Caudovirales* taxonomy is being reshaped, the family *Podoviridae* will be dissolved in time and the *Autographivirinae* subfamily has been upgraded to a family. Taking these changes in consideration, we propose here to declare the RIO clade as a new family within the order *Caudovirales*, and to name it *Zobellviridae* (from Claude Zobell, the first to isolate marine phages).

In the *Zobellviridae*, we propose i) one new subfamily, the *Cobavirinae*, made of two new genera, *Siovirus* and *Veravirus*, and ii) six new genera, *Icepovirus*, *Vipivirus*, *Salinovirus*, *Citrovirus*, *Paundecimvirus* and *Melvirus*, not affiliated yet to any subfamily. The intergenomic similarities used to delimitate the above genera are found in figure 13. The phylogenetic tree based on the Terminase protein (found in all phages from *Zobellviridae*) confirms the phylogenetic assignments based on the whole genome tree (see Fig. 12).

3.3.2. *Cobavirinae* subfamily

In the whole genome tree, ICBM1, ICBM2 and ICBM3 formed a highly supported clade together with Roseobacter virus SIO1 and its relatives, Celeribacter phage P12053L and several environmental phage genomes. This clade is referred here as the old SIO clade (see Fig. 11). VICTOR based taxonomic thresholds suggested this clade represents a genus. Therefore, Bischoff et al. (2019) suggested to call this clade the *Siovirus* genus. Within the old SIO clade, all phages with the exception of two environmental genomes (Env 3300001749 JGI24025J20009 10000213 and Env TARA ERS488813 N000141) had a cobalamin dependent ribonucleotide reductase and formed a highly supported clade. Bischoff et al. (2019) suggested to call this clade the "Cobavirus Group", part of the *Siovirus* genus. In the light of current changes in ICTV classification rules, we propose to classify the "Cobavirus group" as a subfamily, named *Cobavirinae* (from the presence of cobalamin dependent ribonucleotide reductase in the phage genomes). Because their genomes are incomplete, the two environmental genomes which were part of the SIO clade, but not of the Cobavirus group, are not considered here. The *Siovirus* genus (the old SIO clade), as defined by Bischoff et al. (2019) is further disregarded.

Based on the 70% nucleotide similarity threshold currently recommended by ICTV for the determination of phage genus borders, we propose three genera within the *Cobavirinae* subfamily (see Fig. 13 for intergenomic similarity values): i) *Siovirus* (not to be mistaken with the old *Siovirus* genus defined by Bischoff et al. (2019), which corresponds to the whole SIO clade), with the species *Lentibacter virus ICBM1*, *Celeribacter phage P12053L* and *Roseobacter virus SIO1*, and ii) *Veravirus*, with the species *Lentibacter virus ICBM2*. With the exception of *Lentibacter virus ICBM1*, all other species are represented by single strains. *Lentibacter virus ICBM1* contains two strains, ICBM1 and ICBM3, having an intergenomic similarity at nucleotide level of 98.67% (see Fig. 13). ICBM2 is the sole representative of the *Lentibacter virus ICBM2* species. An additional phylogenetic analysis, based on the terminase protein, supports the conclusions from the whole genome phylogeny (see Fig. 12).

3.3.2.1. Genus *Siovirus*

Genus Siovirus consists of three species: *Lentibacter virus ICBM1*, *Celeribacter phage P12053L* and *Roseobacter virus SIO1*. The genus name derives from Roseobacter phage SIO1, the first phage described from this genus and the first sequence marine bacteriophage. As type species we propose *Lentibacter virus ICBM1*, because its representatives are actively

maintained in the laboratory of Cristina Moraru and hopefully will be deposited at the DMSZ culture collection.

***Lentibacter virus ICBM1* species**

Two phages belong to the *Lentibacter virus ICBM1* species: Lentibacter virus vB LenP ICBM1 (ICBM1) and Lentibacter virus vB LenP ICBM3 (ICBM3). They were enriched from southern North Sea water, collected during an algal bloom (Bischoff et al. 2019). ICBM1 was further purified to a single isolate. ICBM3 is known only as a phage genome from the enrichment. Transmission electron microscopy revealed that ICBM1 phage had an isometric capsid with hexagonal cross-sections of 58.7 ± 3.7 and a short tail (sample size = 100 phages each; negative staining with uranyl acetate) (Fig. 14). The host range of ICBM1 was determined on 94 *Rhodobacteraceae* strains. ICBM1 infected only *Lentibacter* sp. SH36, the original isolation host. ICBM1 has a genome size of 40.163 kb and a G+C content of 47.0% (GenBank Accession MF431617). ICBM3 was obtained by sequencing of the phage enrichment. ICBM3 has a genome size of 40.498 kb and a G+C content of 47.30% (GenBank accession MF431615). Both phages have direct terminal repeats (DTRs) as genome termini.

***Roseobacter virus SIO1* species**

Roseobacter virus SIO1 is currently described in ICTV as an unclassified *Podoviridae* species. Here we proposed to move it into the new subfamily *Cobavirinae*, *Siovirus* genus. Bischoff et al. (2019) have determined the genome ends as DTRs, and reordered the genome accordingly. Other SIO1 related phages (SIO 2001, OS, MB, SBRSIO67) have incomplete genomes and therefore are not included in this proposal.

***Celeribacter phage P12053L* species**

The phage belonging to *Celeribacter phage P12053L* species was previously described (Kang et al. 2012). It has a dsDNA genome (GenBank accession JQ809650) of 35.889 kb length, with a G+C content of 46.1 (Kang et al. 2012) and DTRs as genome termini (Bischoff et al. 2019). It was isolated from the Yellow Sea in South Korea with the original host *Celeribacter marinus* IMCC12053 (Kang et al. 2012; Yang et al. 2016).

3.3.2.2. *Veravirus* genus

Veravirus genus consists of one species, which is also the type species, namely *Lentibacter virus ICBM2*. The name of the genus comes from the researcher who isolated the Lentibacter virus vB LenP ICBM2 (ICBM2).

***Lentibacter virus ICBM2* species**

One phage belongs to the *Lentibacter virus ICBM2* species: Lentibacter virus vB LenP ICBM2. It was isolated from southern North Sea water collected during an algal bloom (Bischoff et al. 2019). Transmission electron microscopy revealed that ICBM2 phage had an isometric capsid with hexagonal cross-sections of 59.2 ± 2.8 and a short tail (sample size = 100 phages each; negative staining with uranyl acetate) (Fig. 14). The host range of ICBM2 was determined on 94 *Rhodobacteraceae* strains. ICBM2 infected only *Lentibacter* sp. SH36, the original isolation host. ICBM2 has a genome size of 40.907 kb and a G+C content of 47.8% (GenBank Accession MF431616). As genome termini, ICBM2 has DTRs.

3.3.2.3. Genomic organization of cobaviruses

The phages of the proposed subfamily *Cobavirinae* have a conserved genomic organization. All have DTRs as genomic ends (Bischoff et al. 2019). This suggests that cobaviruses might use a packaging strategy similar to the T7 phage. The genes are organized in two genomic arms with opposite transcriptional directions, separated by a bidirectional, rho-independent transcriptional regulator. On the left genomic arm, genes for replication and nucleotide metabolism are encoded, while the right arm encodes genes for lysis and virion structure and morphology (Bischoff et al. 2019).

3.3.3. Other genera in the *Zobellviridae* family

3.3.3.1. *Melvirus* genus

Melvirus genus consists of two species, *Pseudoalteromonas virus HP1* and *Pseudoalteromonas virus RIO1* (see Fig. 13 for the intergenomic similarities used to delimitate this genus). As type species we propose the *Pseudoalteromonas virus HP1*, because it is actively maintained in the laboratory of Melissa Duhaime. The genus name comes from the researcher who has characterized it.

***Pseudoalteromonas virus HP1* species**

The *Pseudoalteromonas virus HP1* species has one phage – Pseudoalteromonas phage HP1. This phage, infecting two closely related strains of *Pseudoalteromonas* sp. (strain H-100 and strain 13-15) was isolated from seawater samples from the North Sea, near Helgoland, Germany (Duhaime et al. 2017). It has a genome of 45.035 kb in size and a G+C content of 44.67 (GenBank accession KF302037.1). It has a podoviral morphology with an icosahedral capsid and a short tail (Fig. 15) (Duhaime et al. 2017). Its host range is narrow among strains tested, infecting two out of seven strains of *Pseudoalteromonas* sp.

***Pseudoalteromonas virus RIO1* species**

The *Pseudoalteromonas virus RIO1* species has one phage – Pseudoalteromonas phage RIO-1. This phage was isolated from seawater samples from the East Sea, South Korea, together with its host *Pseudoalteromonas marina* CL-E25 (Hardies et al. 2013). The phage infects only its original host (Hardies et al. 2013). It has an icosahedral head of 51 nm and a short tail. Its genome is 43.882 kb large, with 39.6% G+C content and direct terminal repeats (Hardies et al. 2013) (GenBank accession KC751414).

3.3.3.2. *Paundecimvirus* genus

Paundecimvirus genus has only one species, which is also the type species, namely *Pseudomonas virus PA11*. The genus name comes from the phage name.

***Pseudomonas virus PA11* species**

Pseudomonas virus PA11 species comprises the PA11 phage. This phage infects *Pseudomonas aeruginosa* and has a genome size of 49.639 kb, with 44.8% G+C content (GenBank accession DQ163915) (Kwan et al. 2006).

3.3.3.3. *Citrovirus* genus

Citrovirus genus has one species, which is also the type species, namely *Citrobacter virus CVT22*. The genus name comes from the phage host name.

***Citrobacter virus CVT22* species**

Citrobacter virus CVT22 species comprises the Citrobacter phage CVT22. This phage infects *Citrobacter* sp. strain TM1552. Together with its host it was isolated from the gut of the Formosan subterranean termite *Coptotermes formosanus* (Tikhe et al. 2015). It has a podoviral morphology and a circular permuted genome of 47.636 kb size and 41.6% G+C content (GenBank accession KP774835) (Tikhe et al. 2015).

3.3.3.4. *Salinovirus* genus

Salinovirus genus has a single species, which is also the type species, namely *Salinivibrio virus CW02*. The genus name comes from the phage host name.

***Salinivibrio virus CW02* species**

Salinivibrio virus CW02 species consists of the Salinivibrio phage CW02. This phage infects the *Salinivibrio costicola*-like bacterium SA50 (99% 16S rRNA sequence identity with *S. costicola* subsp. *costicola* strain ATCC 33508) and it was, like its host, isolated from the

Great Salt Lake, USA (Shen et al. 2012). CW02 has an icosahedral capsid of ~60 nm in diameter and a short tail. The genome of phage CW02 is 40.547 kb in size with 47.67% G+C content and has no terminal repeats (GenBank accession JQ446452) (Shen et al. 2012).

3.3.3.5. *Vipivirus* genus

Vipivirus genus has only one species, which is also the type species, namely *Vibrio virus VpV262*. The genus name comes from the phage name.

Vibrio virus VpV262 species

The *Vibrio virus VpV262* species includes the phage Vibrio phage VpV262, and it is currently described in ICTV as unclassified *Podoviridae*. We propose here to move this species to the *Zobellviridae* family, *Vipivirus* genus.

3.3.3.6. *Icepovirus* genus

Icepovirus genus has one species, which is also the type species, namely *Vibrio virus ICP2*. The genus name comes from the phage name.

Vibrio virus ICP2 species

Vibrio virus ICP2 species consists of Vibrio phage ICP2. This phage infects different *Vibrio cholerae* strains. It was isolated from stool-samples of cholera patients in Bangladesh, with the original host *V. cholerae* O1 El Tor. It can also infect *V. cholera* O139 strain M010. It has a genome of 49.675 kb and 42.7% G+C content (GenBank accession HQ641345) (Seed et al. 2011). ICP2 has an icosahedral capsid of 60 nm in diameter and a short tail (13 nm long, 8 nm wide) (Seed et al. 2011).

3.4. Supporting evidence

Table 4: Parameters used by VICTOR to calculate a whole genome tree based on amino acid sequences and to demarcate taxa.

Parameter	Value
Word length	3
E-value filter	0.1
Algorithm	Greedy-with-trimming
Formula	d6
Distance threshold, species	0.118980
Distance threshold, genus	0.749680
Distance threshold, subfamily	0.888940

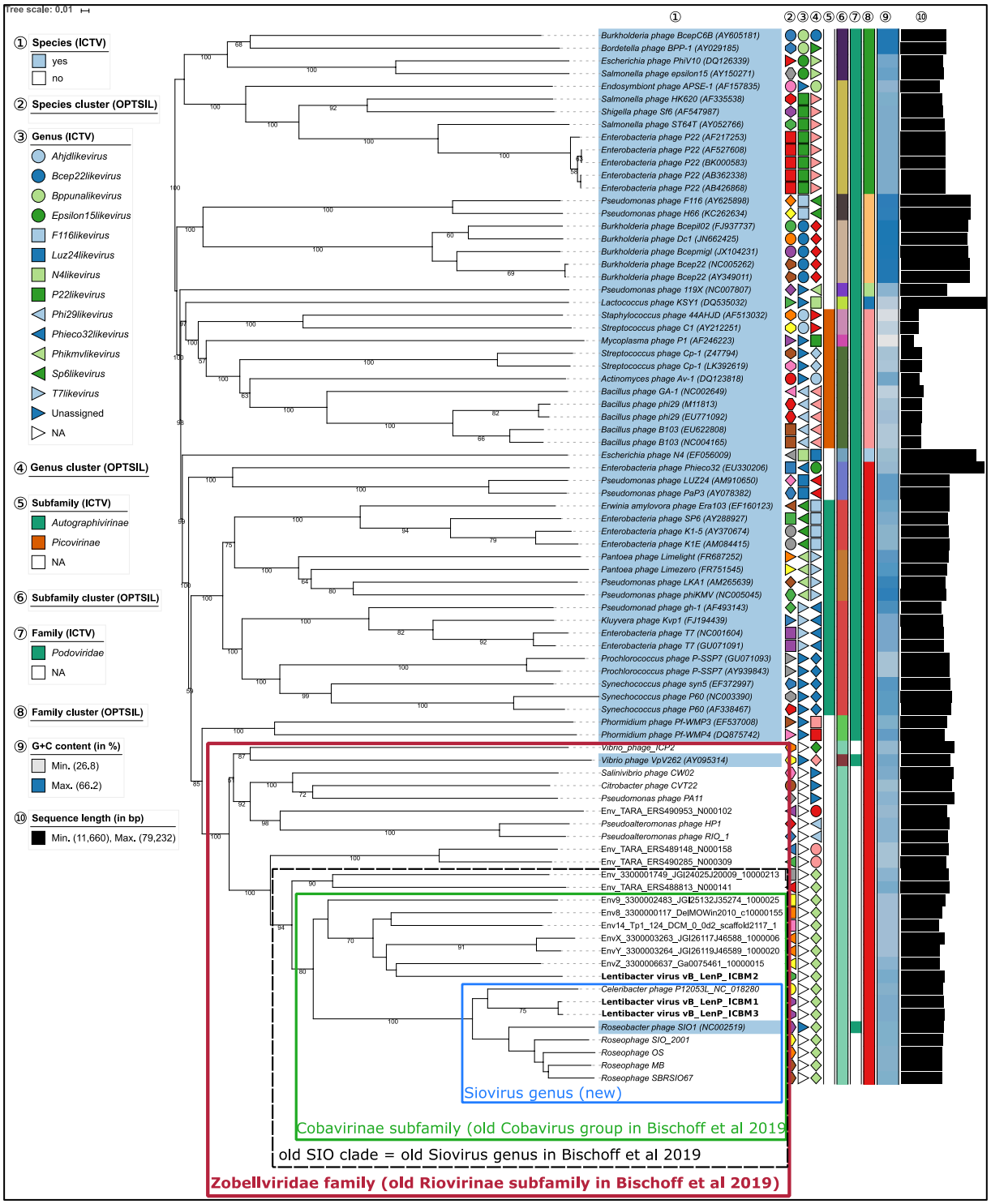


Fig. 11: Whole genome phylogenetic tree depicting the new *Zobellviridae* family. The whole-genome-based phylogeny was inferred using the Genome-BLAST Distance Phylogeny method implemented in the VICTOR web service, using the amino acid data. Internal branch labels represent pseudo bootstrap support values if larger than 50%. The proposed family *Zobellviridae* and the proposed subfamily *Cobavirinae* are annotated with colored rectangles. Further information regarding the current affiliation of phages to ICTV taxa and OPTSIL clusters, as well as G+C content and genome sizes is described within the figure legend (circled numbers). "Viruses annotated as "Unassigned" in legend "Genus (ICTV)" have been assigned to both an ICTV species and family but not to a genus level, whereas "NA" refers to viruses which have not been recognized as a taxa by the ICTV." The affiliation of one or more viruses to a distinct species, genus, subfamily or family cluster is indicated by a specific symbol of same shape and color. Adapted from Bischoff et al. (2019). Used under CC BY (http://creativecommons.org/licenses/by/4.0/).

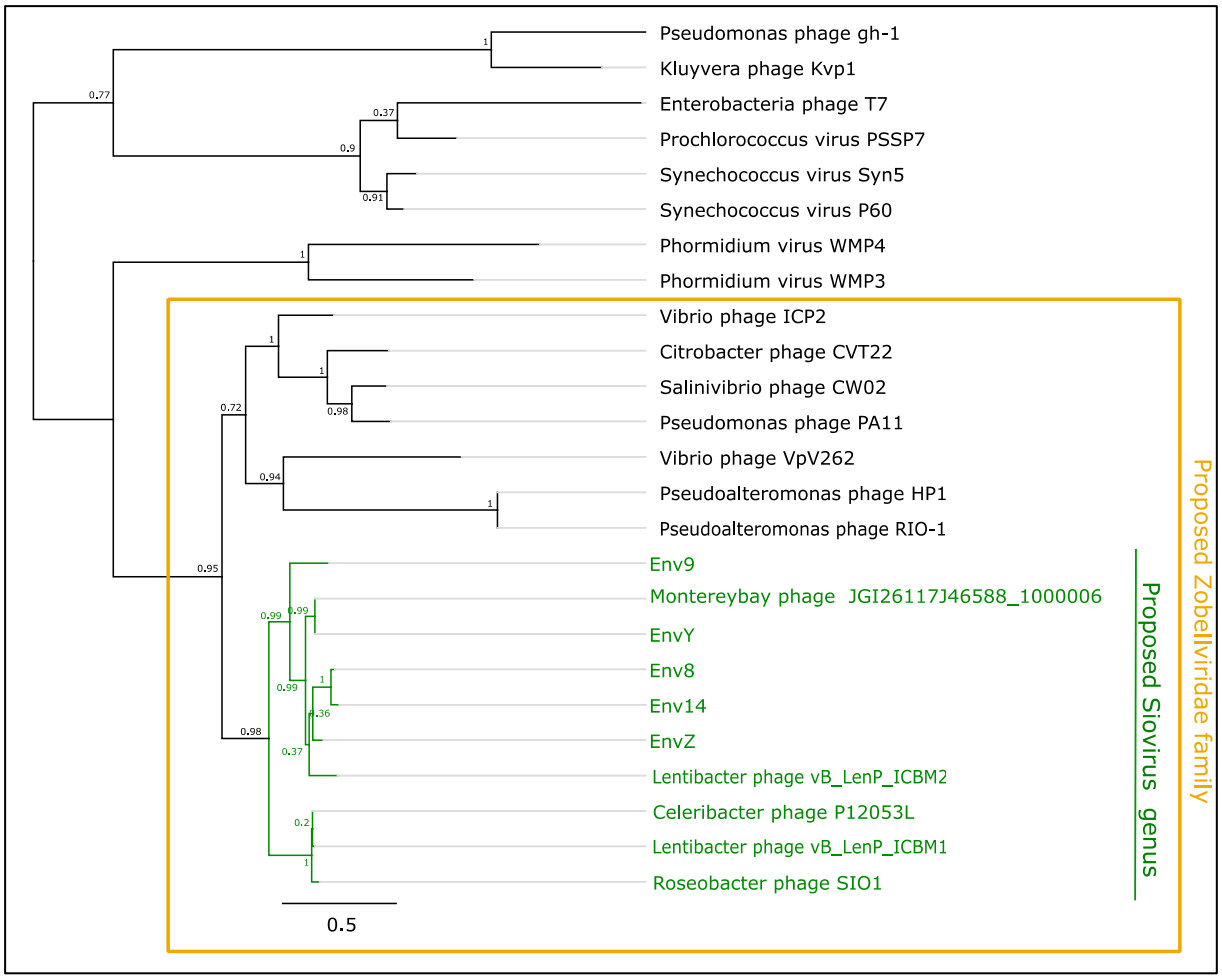


Fig. 12: Phylogenetic positioning of the terminase proteins belonging to the phages in the *Zobellviridae* family. The terminase protein phylogeny was inferred using the approximately-maximum-likelihood method implemented in FastTree 2.1.5 in Geneious. The node labels represent Fast Tree support values. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. The tree is unrooted. The affiliation of viruses to the proposed *Cobavirinae* subfamily or the proposed *Zobellviridae* family is indicated in green and orange colour, respectively.

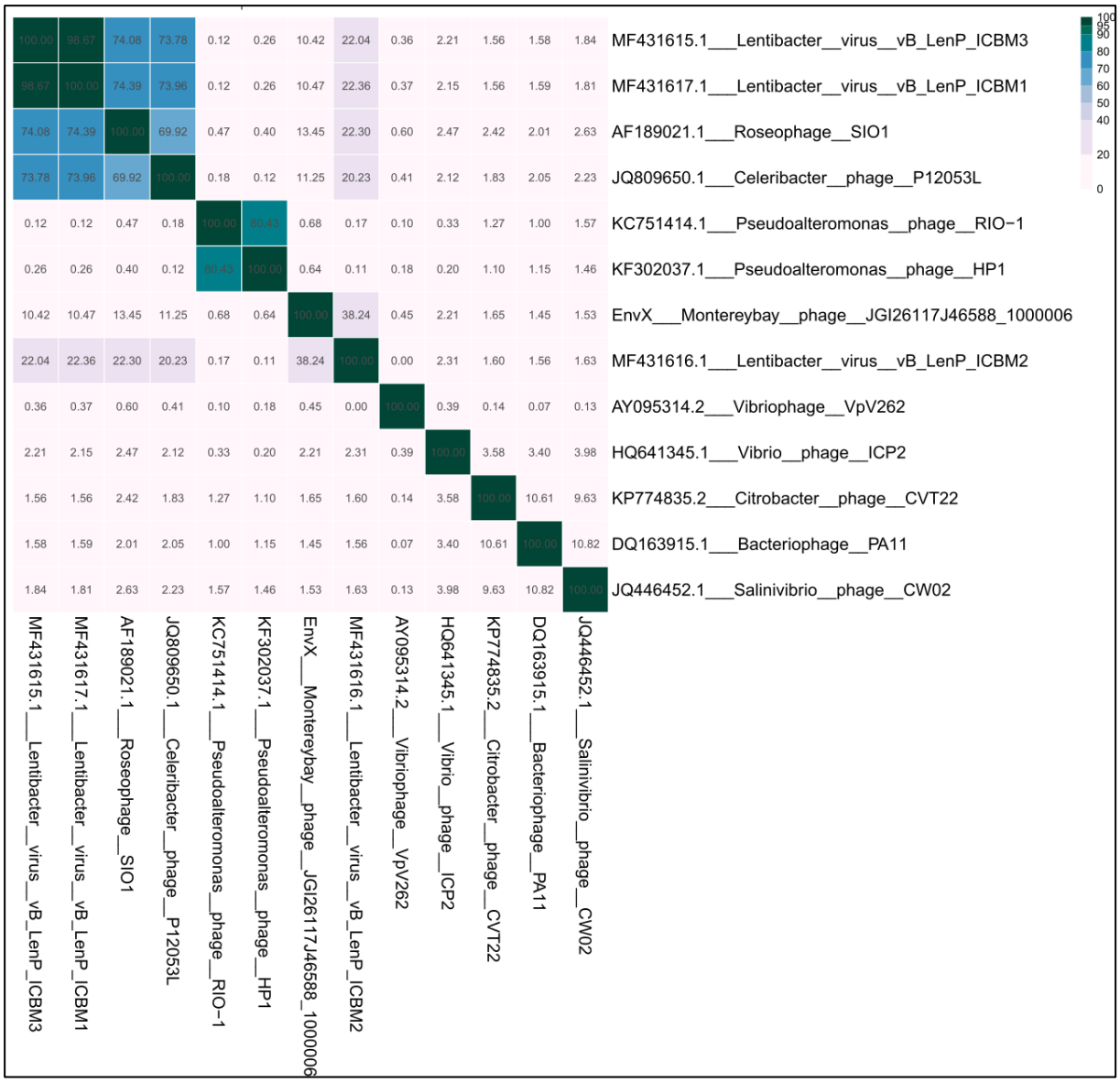


Fig. 13: Nucleotide based, intergenomic similarities between the members of the here newly proposed *Zobellviridae* family. The intergenomic similarities were calculated using the VIRIDIC webservice (viridic.icbm.de), using the formula simAB = ((idAB+idBA)*100) / (lA+lB), where idAB = identical bases when genome A is aligned to genome B, idBA = identical bases when genome B is aligned to genome A, lA = length genome A, lB = length genome B, simAB = intergenomic similarity between genomes A and B.

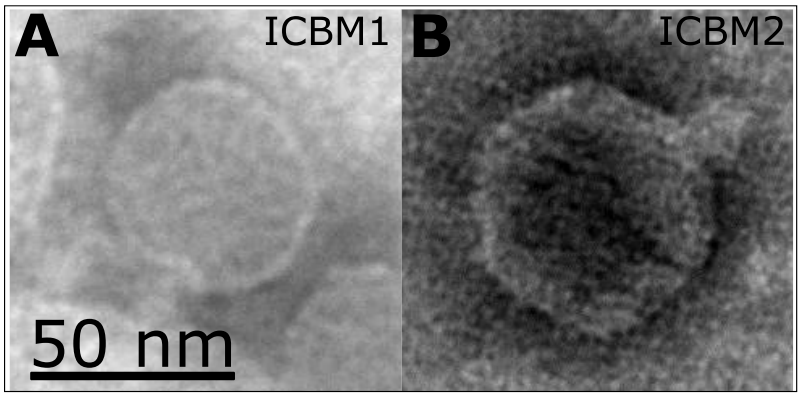


Fig. 14: Transmission electron micrograph of molybdenum stained, cell debris bound Lentibacter virus vB_LenP_ICBM1 and uranyl acetate stained, free Lentibacter virus vB_LenP_ICBM2. Scale bar: 50 nm. From Bischoff et al. (2019). Used under CC BY (http://creativecommons.org/licenses/by/4.0/) No changes made.

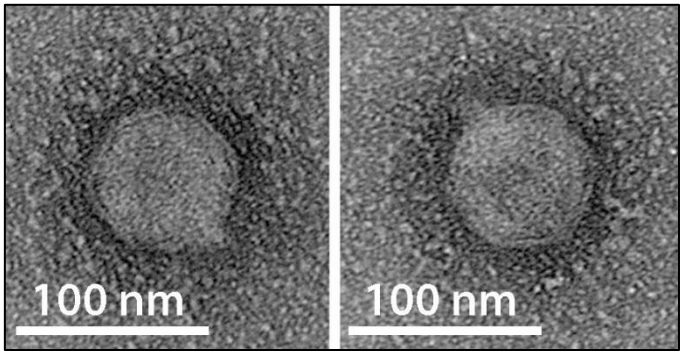


Fig. 15: Transmission electron micrograph of Pseudoalteromonas phage HP1 (N. Solonenko). Scale bar: 100 nm.

Table 5: Newly proposed taxonomy of the *Zobellviridae* family. All proposed families belong to the order *Caudovirales*, in the class *Caudoviricetes*, in the phylum *Uroviricota*, in the kingdom of *Heunggongvirae*, in the realm of *Duplodnoviria*. Newly proposed taxonomy is marked in red. *CG = complete genome

PROPOSED TAXONOMY									**PROPOSED CHANGE**	
Family	**Subfamily**	**Genus**	**Species**	**Type species?** (1/0)	**Exemplar** GenBank Accession Number	**Exemplar** virus name	**Genome coverage**	**Genome composition**	**Change**	**Rank**
Zobellviridae									Create new	family
Zobellviridae	*Cobavirinae*								Create new	subfamily
Zobellviridae	*Cobavirinae*	*Siovirus*							Create new	genus
Zobellviridae	*Cobavirinae*	*Siovirus*	*Lentibacter virus ICBM1*	1	MF431617	Lentibacter phage vB_LenP_ICBM1	CG*	dsDNA	Create new; assign as type species	species
Zobellviridae	*Cobavirinae*	*Siovirus*	*Roseobacter virus SIO1*	0	AF189021	Roseobacter phage SIO1	CG	dsDNA	Move	species
Zobellviridae	*Cobavirinae*	*Siovirus*	*Celeribacter virus P12053L*	0	JQ809650	Celeribacter phage P12053L	CG	dsDNA	Create new	species
Zobellviridae	*Cobavirinae*	*Veravirus*							Create new	genus
Zobellviridae	*Cobavirinae*	*Veravirus*	*Lentibacter virus ICBM2*	1	MF431616	Lentibacter phage vB_LenP_ICBM2	CG	dsDNA	Create new; assign as type species	species
Zobellviridae		*Icepovirus*							Create new	genus
Zobellviridae		*Icepovirus*	*Vibrio virus ICP2*	1	HQ641345	Vibrio phage ICP2	CG	dsDNA	Create new; assign as type species	species
Zobellviridae		*Vipivirus*							Create new	genus
Zobellviridae		*Vipivirus*	*Vibrio virus VpV262*	1	AY095314.2	Vibrio phage VpV262	CG	dsDNA	Move	species
Zobellviridae		*Salinovirus*							Create new	genus
Zobellviridae		*Salinovirus*	*Salinivibrio virus CW02*	1	JQ446452	Salinivibrio phage CW02	CG	dsDNA	Create new; assign as type species	species
Zobellviridae		*Citrovirus*							Create new	genus
Zobellviridae		*Citrovirus*	*Citrobacter virus CVT22*	1	KP774835	Citrobacter phage CVT22	CG	dsDNA	Create new; assign as type species	species
Zobellviridae		*Paundecimvirus*							Create new	genus
Zobellviridae		*Paundecimvirus*	*Pseudomonas virus PA11*	1	DQ163915	Pseudomonas phage PA11	CG	dsDNA	Create new; assign as type species	species
Zobellviridae		*Melvirus*							Create new	genus
Zobellviridae		*Melvirus*	*Pseudoalteromonas virus RIO1*	0	KC751414	Pseudoalteromonas phage RIO-1	CG	dsDNA	Create new	species
Zobellviridae		*Melvirus*	*Pseudoalteromonas virus HP1*	1	KF302037.1	Pseudoalteromonas phage HP1	CG	dsDNA	Create new; assign as type species	species

4. Isolation and classification of novel dsDNA roseophages

4.1. Chapter summary

In a large isolation campaign, more than 350 bacterial strains from the Roseobacter Group were used as potential hosts and seven different seawater samples from the North Sea were used as virus source. With two different methods, using either direct plating of concentrated seawater or enrichment cultures, 277 phage isolates were obtained (Fig. 16). After screening for unique phages by randomly amplified polymorphic DNA (RAPD) - PCR and genome size estimation using pulsed-field gel electrophoresis (PFGE), sequencing yielded 128 unique dsDNA roseophage genomes. They belong to twelve different genera. We investigated genome characteristics of 28 representative phages and classified them within the context of the current ICTV scheme as well as previously reported roseophages. The new roseophages infect *Sulfitobacter*, *Lentibacter* and *Octadecabacter* strains and belong to eight different families, four already ICTV-recognized and four newly proposed.

The following chapter describes unpublished data. However, much of the experimental data and analysis make part of a manuscript in the final stages of preparation.

Contributions to this work:

Cristina Moraru designed the research and contributed to data analysis (genome assembly, annotation, and taxonomic classification) and manuscript writing. I performed much of the laboratory work and wrote this chapter, with great help from Benedikt Heyerhoff (sampling and preparation of seawater samples, phage isolation, DNA extraction, and RAPD-PCR), Aaron Woolley (purification of bacterial strains and phages), Mary Nguyen (phage isolation), Anne Bögeholz (phage DNA extraction and RAPD-PCR) and Andrea Schlingloff (host 16S rRNA gene and ITS sequencing). PFGE was conducted in the laboratory of our collaborator Silke Pradella at the DSMZ (Deutsche Sammlung für Mikroorganismen und Zellkulturen) in Braunschweig. Sequencing of the phage genomes was performed by our collaborators Anja Poehlein and Mechthild Bömeke at the Göttingen Genomics Laboratory. Carlota Alejandre-Colomo and Anneke Heins from the Max-Planck-Institute for Marine Microbiology (MPI) in Bremen provided the bacterial host strains.

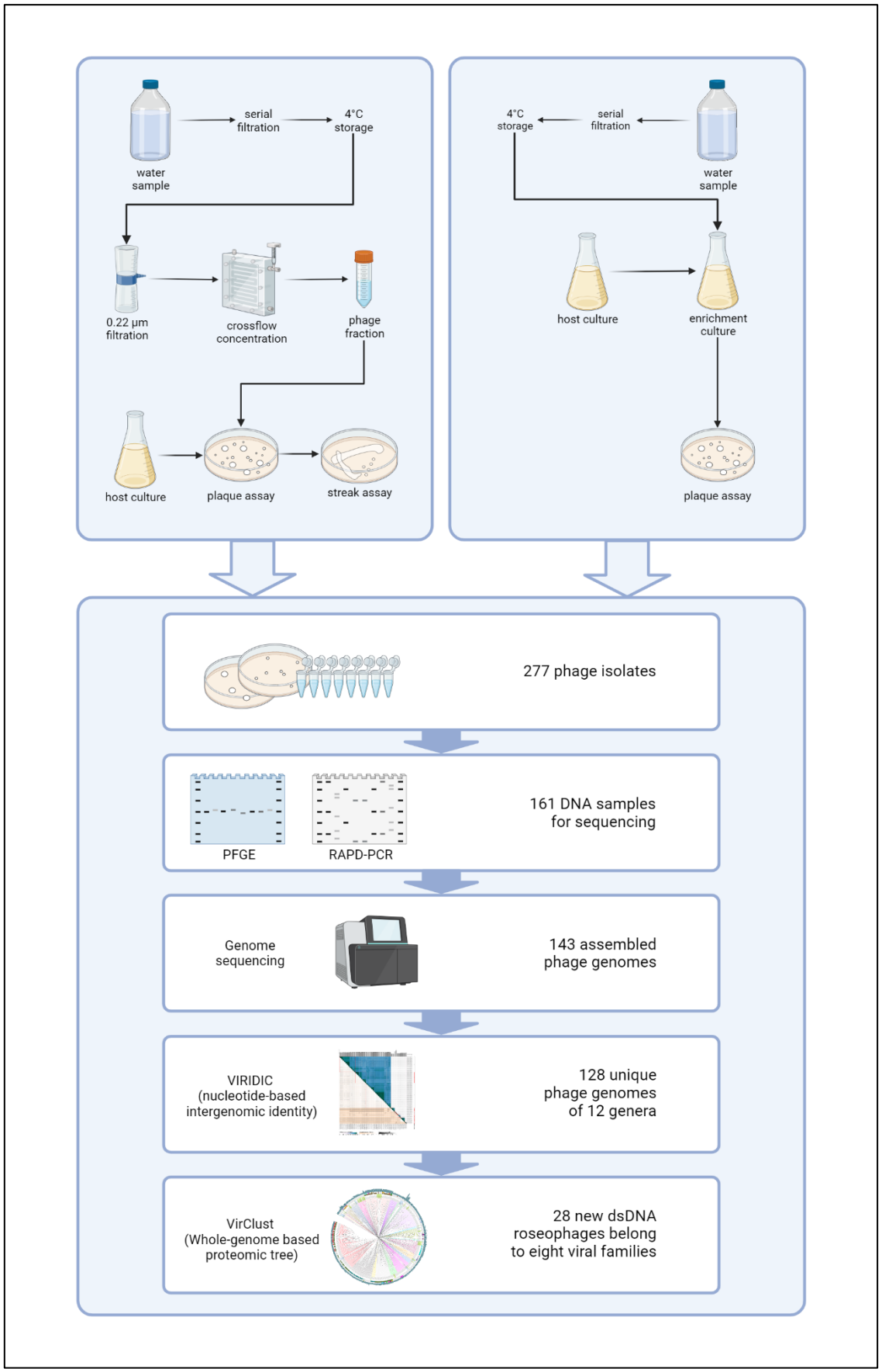


Fig. 16: Workflow of phage isolation by direct plating (left) or enrichment cultures (right) and subsequent genome sequencing and classification (Image created with BioRender.com).

4.2. Materials and methods

4.2.1. Cultivation media

For phage isolation and subsequent cultivation, host strains were grown either in artificial saltwater medium (ASW) or MB50 medium. ASW was prepared using the following recipe. 24.32 g/l NaCl, 10 g/l $MgCl_2x6H_2O$, 1.5 g/l $CaCl_2x2H_2O$, 0.66 g/l KCl, 4 g/l Na_2SO_4, 2.38 g/l HEPES, 0.6 g/l peptone, 0.3 g/l yeast extract, 84 mM KBr, 40 mM H_3BO_3, 15 mM $SrCl_2$, 40 mM NH_4Cl, 4 mM KH_2PO_4, 7 mM NaF. After adjustment to pH 7.5, the solution was autoclaved and completed before use with 1 ml/l of sterile filtered multi vitamin solution (after (Balch et al. 1979)), 0.25 ml/l of sterile filtered trace element solution A (1.5 g $FeCl_2x4H_2O$ in 10 ml 25% HCl and 250 ml MilliQ water) and 0.1 ml/l of autoclaved trace element solution B (19 mg/l $CoCl_2x6H_2O$, 10 mg/l $MnCl_2x2H_2O$, 7 mg/l $ZnCl_2$, 3.6 mg/l $Na_2MoO_4x2H_2O$, 2.4 mg/l $NiCl_2x6H_2O$, 0.6 mg/l H_3BO_3, 0.2 mg/l $CuCl_2x2H_2O$). 10 x ASW medium contained a ten-fold higher amount of yeast extract, peptone and vitamins, respectively. ASWbase medium was prepared without any of the three ingredients.

To prepare MB50 medium, ready-to-use Marine broth medium (MB) Difco™ (BD Biosciences) (5 g/l peptone, 1 g/l yeast extract, 0.10 g/l $C_6H_8FeO_7$, 19.45 g/l NaCl, 5.90 g/l $MgCl_2$, 3.24 g/l Na_2SO_4, 1.80 g/l $CaCl_2$, 0.55 g/l KCl, 0.16 g/l $NaHCO_3$, 0.08 g/l KBr, 0.034 g/l $SrCl_2x6H_2O$, 0.022 g/l H_3BO_3, 0.004 g/l $Na_2SiO_3x3H_2O$, 0.0024 g/l NaF, 0.0016 g/l NH_4NO_3, 0.008 g/l Na_2HPO_4, prepared according to the manufacturer´s instructions) was diluted in a 1:1 ratio with ASWbase before autoclavation. For MB50 agar plates, 18 g/l Bacto Agar (BD Biosciences) were added to liquid MB50 medium before autoclaving. For MB50-soft agar, 6 g/l low melting point Biozym Plaque GeneticPure agarose (Biozym) were added to MB50 medium before autoclaving in 250 ml glass bottles. Prior to usage, the MB50-soft agar was melted in a boiling water bath and cooled down to 37 °C. 1 ml/l sterile filtered multi vitamin solution (after Balch et al. (1979)) was added.

4.2.2. Origin and cultivation of host strains

The bacterial hosts used for phage isolation have been all isolated from the North Sea, but in different years and locations and were provided by various collaborators (Table 6). *Lentibacter* sp. SH36 and *Sulfitobacter* sp. SH24-1b (host of phage ICBM5 described in chapter 6) were isolated from a seawater sample from the southern North Sea (54°42'N, 06°48'E; 36 m depth) taken in May 2007 during a phytoplankton bloom (Hahnke et al. 2013). Strain MPI-62 was

isolated from seawater collected during the spring phytoplankton bloom from 2017 at Helgoland roads time series station (54°11'03"N, 7°54'00"E). Strain MM282 was isolated from a seawater sample taken at high tide at the shore of Harlesiel (53°42'39"N 7°48'28"E) in October 2017. Both strains were provided by Anneke Heins (MPI, Bremen). For more details on the isolation procedure of these two bacterial strains, see appendix.

A large culture collection of 388 strains from the Roseobacter group was provided by Carlota Alejandre-Colomo (MPI, Bremen and IMEDEA, Esporles, Spain) to be used in the phage isolation campaign (see SI file S4-1). The strains had been isolated by direct plating from North Sea surface water samples taken in spring 2016 also at Helgoland roads time series station (54°11'17.88"N, 7°54'0"E) and had been preliminarily assigned to the genus *Sulfitobacter* by a combination of 16S rRNA analysis and whole cell matrix-assisted laser desorption ionization – time offlight mass spectrometry (WC MALDI-TOF MS) (Alejandre-Colomo et al. 2020).

Experiments with *Sulfitobacter* sp. SH24-1b were performed with full MB medium. All other strains were cultivated in MB50 medium and general growth conditions were 20 °C and 100 rpm.

Table 6: Bacterial strains used for phage isolation in this study. *Host of phage ICBM5 described in chapter 6. ** Numbering is from 1 to 388, see SI file S4-1.

Bacterial strains	**Isolated from**	**Isolated by / Reference**	**Strain label**
Lentibacter sp. SH36	Southern North Sea, 54°42'N, 06°48'E	(Hahnke et al. 2013)	SH36
MPI-62	Helgoland Roads, 54°11'17.88"N, 7°54'0"E	Anneke Heins	MPI-62
MM282	Harlesiel, 53°42'39"N 7°48'28"E	Anneke Heins	MM282
Sulfitobacter sp. SH24-1b*	Southern North Sea, 54°42'N, 06°48'E	(Hahnke et al. 2013)	SH24-1b
Roseobacter Group strains labeled with M# **	Helgoland Roads, 54°11'17.88"N, 7°54'0"E	(Alejandre-Colomo et al. 2020).	M#

4.2.3. 16S sequencing and phylogenetic analysis of host strains

The 16S rRNA gene of all bacterial hosts was sequenced to determine their phylogenetic affiliation. A larger DNA fragment comprising the 16S rRNA gene and the internal transcribed spacer (ITS) was amplified and sequenced, because this was needed for a more detailed analysis of the sulfivirus host strains (see chapter 5.3.1.). For each strain, 50 µl of densely grown liquid

culture were centrifuged for 10 min at 12000 x g. The cell pellet was resuspended in 30 µl nuclease free water, frozen at -20 °C and thawed by 10 min ultrasonic treatment in order to open the cells. 1 µl was used to amplify for polymerase chain reaction (PCR) to amplify the 16S rRNA gene and the ITS region. The reaction was performed with 250 µM of each deoxynucleotide triphosphate (dNTP), 0.8 mM $MgCl_2$, 1.2 mg/ml bovine serum albumin (BSA), 10 pmol primer 27F binding to the 16S rRNA gene (Table 7), 10 pmol primer 189R binding to the 23S rRNA gene (Table 7), 0.4 U Phusion High-Fidelity DNA Polymerase (Fisher Scientific, 2 U/µl) and 5 µl 5x reaction buffer containing 1.5 mM $MgCl_2$. The cycling protocol was as follows. After the initial denaturation step with 5 min at 95 °C, 32 cycles of 1 min at 95 °C, 1 min at 42 °C and 2 min at 72 °C were performed. At the transition between denaturation and elongation, temperature was lowered by 1.2 °C per second. A final elongation step of 10 min at 72 °C was performed. Successful amplification was tested by agarose-gel electrophoresis and ethidium bromide staining. PCR products were purified using the pegGold Cycle-Pure Kit (S-line) (PEGLAB Biotechnology) following the instruction manual. Sequencing was performed by GATC Services Eurofins Genomics Germany GmbH (Ebersberg, Germany) with six different primers (Table 7). Contig assembly was done with the DNA Baser software (www.dnabaser.com, version 5.15.0.0BT). Phylogenetic analysis of the 16S rRNA gene was performed with the ARB software package (Ludwig et al. 2004), using the reference data set SSU Ref NR 138.1. A neighbor-joining tree was calculated with Jukes-Cantor correction, 1000 bootstrap replicates and the termini filter. Members of the genus *Psychrobacter* served as an outgroup.

Table 7: Primers used for 16S - 23s rRNA fragment amplification and sequencing.

Primer	Sequence 5' -> 3'	Reference
189R	TACTTAGATGTTTCAGTTC	(Hunt et al. 2006)
27F	AGA GTT TGA TCM TGG CTC AG	(Suzuki and Giovannoni 1996)
341F	CCT ACG GGA GGC AGC AG	modified from Herlemann et al. (2011)
907F	AAA CTC AAA KGA ATT GAC GG	aodified from Muyzer et al. (1995)
GM4F	AAG TCG TAA CAA GGT A	adapted from Muyzer et al. (1995)
GM4R	TAC CTT GTT ACG ACT T	Muyzer et al. (1995)

4.2.4. Water samples for phage isolation

Viruses were isolated from seven different seawater samples originating either from mesocosm experiments (samples P1, P2, and P4) or directly from the seawater column (samples NHS, HE504-33, HE396-6, and HE440-S) (see Table 8). Water samples P1, P2 and P4 were taken from three replicate mesocosms in an experiment mimicking a phytoplankton spring bloom

(Mori et al. 2021; Dlugosch et al. 2023). In each of the mesocosms 600 l artificial seawater had been inoculated with water from the southern coastal North Sea and incubated at natural light and temperature conditions for six weeks. The inoculum had been sampled on board of the research vessel Heincke on March 10th 2018 (54°04'33.0"N 7°37'37.2"E) at 4.2 m below sea surface and was filtered through 100 µm before inoculation (Table 8). Samples for phage isolation were taken at the very end of the experiment, after the emergence and collapse of a phytoplankton bloom had been observed. From each of the replicates (P1, P2 and P4) eight liters were sampled, immediately filtered serially through 8 µm, 5 µm and 0.22 µm polycarbonate filters (Whatman Nuclepore 47mm, Sigma-Aldrich, USA) and stored at 4 °C. Seawater sample NHS originated from the shore in Neuharlingersiel at the coast of the southern North Sea (53°42'12.7"N 7°42'15.0"E) (Table 8). Ten liters were sampled in July 2018 during high tide, filtered through 0.8 µm polycarbonate filters and stored at 4 °C for a few days. Sample HE504-33 was collected on board of the research vessel Heincke in the southern North Sea (53°53'44.5"N 7°32'05.6"E) during a phyloplankton bloom on March 9th 2018 (Table 8). Six liters of seawater were sampled, filtered directly through 0.8 µm polycarbonate filters and stored at 4 °C.

Prior to phage isolation, water samples P1, P2, P4, NHS and HE504-33 were again filtered through 0.22 µm bottle top filters (Nalgene Rapid-Flow, PES membrane, ThermoFisher Scientific). Five liters were concentrated to approximately 50 ml by crossflow filtration (Vivaflow® 200, PES membrane, 30 kDa MWCO, Sartorius). The obtained virus concentrate was stored at 4 °C (for maximum 3 weeks).

Sample HE440-S was the same as used for the isolation of the cobaviruses, i.e. multiple samples from the southern North Sea, taken during a phytoplankton bloom in March 2015 on board of the cruise ship RV Heincke and pooled together (see chapter 2.2.2.). Sample HE396-6 was also taken on board of the cruise ship RV Heincke in the southern North Sea, but in March 2013 (Table 8). It was filtered on board through 0.7 µm filters (GTTP filters, 47 mm in diameter, Millipore), transported to the laboratory and stored at 4 °C in the dark. Both water samples, HE396-6 and HE440-S were filtered again through 0.2 µm (Nalgene rapid-flow, 0.2 µm, PES membrane, Thermo-Scientific) before they were used for phage isolation by enrichment cultures.

Table 8: Seawater sources for phage isolation.

Label	From	Coordinates / References	Processing
HE396-6	southern North-Sea, March 2013	54°20'04.2"N 7°06'58.8"E	0.7 µm → 0.2 µm filtration
HE440-S	southern North Sea, March 2015, pooled multiple stations	53°58'40.8"N 8°03'32.4"E, 53°56'13.2"N 7°48'21.6"E, 53°53'45.6"N 7°32'06.0"E, 53°50'24.0"N 7°15'18.0"E, 53°47'34.8"N 6°59'49.2"E (Bischoff et al. 2019)	
HE504-33	southern North Sea, March 2018	53°53'44.5"N 7°32'05.6"E	0.8 µm → 0.2 µm filtration → 100x concentration of the phage fraction
P1 P2 P4	mesocosms inoculated with water from the southern North Sea, March 2018	54°04'33.0"N 7°37'37.2"E (Mori et al. 2021)	8 µm → 5 µm → 0.2 µm filtration → 100x concentration of the phage fraction
NHS	North Sea shore, July 2018	53°42'12.7"N 7°42'15.0"E	0.8µm → 0.2 µm filtration → 100x concentration of the phage fraction

4.2.5. Isolation of phages from direct plating and their further characterization

4.2.5.1. Direct phage isolation by plaque assay

In the large-scale isolation campaign, phages were isolated by plaque assay and single plaque picking, using methods described before (Kauffman and Polz 2018). Bacterial host strains were challenged with the virus concentrates from the different seawater samples. A volume of 280 µl of exponentially growing host culture (optical density at 600 nm (OD_{600}) = 0.2 – 0.3) were pipetted into the middle of a MB50 agar plate (1.8% agar). The virus concentrate (100 µl) and 3 ml of MB50-soft agar (0.6% low melting point Biozym Plaque GeneticPure agarose, Biozym, kept warm at 37 °C) were pipetted into the middle of the virus-host droplet. The plate was gently shaken to mix and spread the soft agar layer evenly. After drying of the top layer, plates were incubated at 20 °C for 3 - 5 days.

Upon observation of phage plaques, phages were purified by single plaque picking and streak assay. An agar plate was prepared with a top layer of 3 ml MB50-soft agar containing 280 µl of exponentially growing host culture (OD_{600} = 0.2 – 0.3). While the top layer was still liquid, a single phage plaque from the original plate was picked with a sterile pipette tip and

transferred to the still molten top layer of the new plate by streaking through the soft agar in a diluting manner in order to obtain single plaques (Fig. 17). The plate was again incubated for two days at 20 °C and the single plaque picking and streak assay was repeated three times for phage purification.

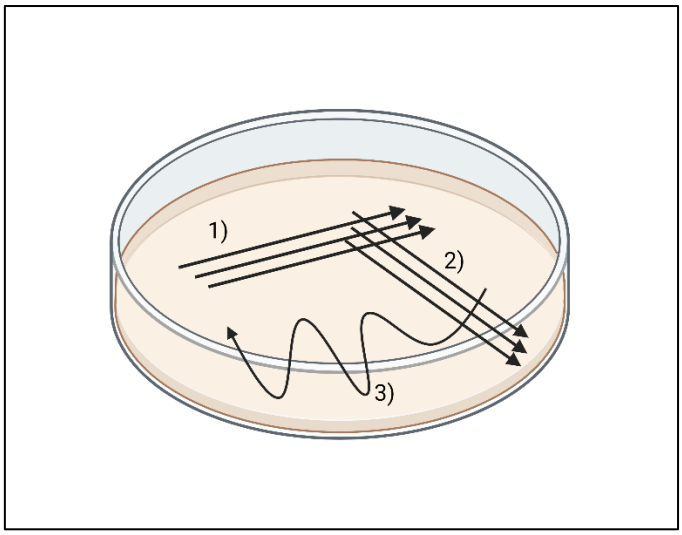


Fig. 17: Scheme of the streak assay for phage purification. Between steps 1, 2 and 3 a new sterile pipette tip was taken. (Image created with Biorender.com).

4.2.5.2. Preparation of fresh lysates and glycerol stocks

Fresh, cell-free phage lysates for subsequent experiments or for glycerol stock preparation were obtained either from a liquid infection culture (method A) or from agar plates with confluent plaques (method B).

Method A: Phage infection cultures were prepared by inoculation of 6 ml MB50% medium with an exponentially growing host culture to an OD_{600} of 0.006 and infection with one single phage plaque picked directly from an agar plate. In parallel, a control culture was set up for every host, containing only the medium and the host inoculum with an OD_{600} of 0.006 without phage. After incubation overnight at 20 °C and 100 rpm bacterial lysis was indicated by disrupted cell particles and low OD_{600} (as in comparison with the control culture). The phage lysate was harvested by centrifugation (15 min, 4000 x g, 4 °C) and 0.22 µm filtration (ROTILABO® syringe filters, CME membrane, CarlRoth®) of the supernatant. Phage lysates were stored at +4 °C.

Method B: For each phage isolate a single purified plaque was picked, resuspended in 500 µl ASWbase and incubated overnight at 4 °C. In order to obtain plates with confluent plaques, serial dilutions (10^0, 10^{-1}, 10^{-3}) of the resuspended plaque were prepared by mixing

with ASW base. 100 µl of phage dilution were pipetted in the middle of an MB50% agar plate (1.8% agar). An aliquot of exponentially growing host culture was added (final OD_{600} in 3 ml = 0.0233). 3 ml of MB50%-soft agar (0.6% low melting Biozym Plaque GeneticPure agarose, Biozym, kept warm at 40 °C) were pipetted into the middle of the phage/host droplet and the plate was shaken for mixing and even distribution of the soft agar layer. Plates were incubated for 1 - 2 days at 20 °C. If confluent plaques were observed, 5 ml of SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.4) were applied on top of the soft agar. After incubation for 1 h at 4 °C the buffer was collected with a syringe and filtered 0.22 µm (ROTILABO® syringe filters, CME membrane, CarlRoth®). The obtained phage lysate was stored at 4 °C.

For long term storage two types of glycerol stocks were prepared: i) stock of free phage particles (1 part phage lysate and 1 part MB50% media with 50% glycerol) and ii) stock of infected host cells (1 part infected cells - 400 µl phage fraction added to 400 µl host culture, 15 min on ice for absorption - and 1 part MB50% media with 50% glycerol). Glycerol stocks were stored at -80 °C.

4.2.5.3. PFGE for phage genome size determination/prediction

To estimate the size of the phage genomes, phages were embedded in agarose plugs and submitted to a pulsed-field gel electrophoresis (PFGE), which is able to separate large DNA fragments. For removal of extracellular DNA and RNA, a 2 ml aliquot of phage lysate (obtained from agar plates with confluent plaques, see chapter 4.2.5.2, method B) was incubated with 1 µl of TurboDNase (2 U/µl; Invitrogen, Ambion) and 1 µl of RNase Cocktail Enzyme Mix (500 U/ml RNase A, 20,000 U/ml RNase T1; Invitrogen) for 30 min at 37 °C. Enzymes were inactivated by incubation for 10 min at 75 °C with 15 mM EDTA. 500 µl of the DNase/RNase treated lysate were used for agarose plug preparation. The aliquot was pre-heated to 37 °C and mixed with 170 µl of melted 3.2% SeaKem® Gold Agarose for PFGE (Lonza). The mixture was then quickly distributed in 100 µl molds (CHEF Disposable Plug Molds, Bio-Rad Laboratories) and allowed to solidify at 4 °C for at least 30 min. Afterwards, the plugs were removed from the molds and all plugs of one phage were collected in the same tube with 1.8 ml ESP buffer (10 mg/ml N-Laurylsarcosine sodium salt (Sigma) in 0.5 M EDTA pH 9.0) containing 1 mg/ml proteinase K (CarlRoth). After incubation overnight at 50 °C, the agarose plugs were washed three times in TE buffer (10 mM Tris-HCl pH 8.0, 2 mM EDTA pH 9.0) with intermediate incubation at room temperature for 15 min and finally stored in TE buffer at 4 °C. PFGE was performed in collaboration with Dr. Silke Pradella (DSMZ, Braunschweig).

For each phage isolate, half an agarose plug was loaded on a 1% agarose gel (PFCA agarose in TBE buffer (45 mM Tris, 45 mM boric acid, 0.1 mM EDTA)). Electrophoresis was run on a CHEF-DR II System (Bio-Rad Laboratories), for 17 h at 14 °C buffer temperature with 120° angle, 1-6 s interval and 6 V/cm. DNA bands were visualized by SYBRgreen staining overnight. Low Range PFG Marker and Mid Range PFG Marker (New England Biolabs) were used as molecular size markers and served as calibration standards for size measurements.

4.2.5.4. Phage DNA extraction

For DNA extraction, a cell-free phage lysate was prepared from a plate with confluent plaques as described above. In contrast, DNA of the *Lentibacter* phages ICBM4, ICBM6 and ICBM7 was isolated from a liquid infection culture (see chapter 4.2.5.2.). To remove extracellular DNA and RNA, a 2 ml aliquot of the phage lysate was incubated with 1 µl of TurboDNase (2 U/µl; Invitrogen, Ambion) and 1 µl of RNase Cocktail Enzyme Mix (500 U/ml RNase A, 20,000 U/ml RNase T1; Invitrogen) for 30 min at 37 °C. Enzymes were inactivated by incubation for 10 min at 75 °C. To remove the viral protein capsid, a treatment with Proteinase K followed. A premix solution was prepared by mixing 100 µl EDTA (0.5 M, pH 8) with 5 µl proteinase K solution (20 mg proteinase K, CarlRoth®, dissolved in 1 ml nuclease free water) and preheated at 50 °C for at least 30 min. The premix solution was added to the phage lysate and incubated overnight at 50 °C. With phages infecting *Sulfitobacter* sp. SW_H+_2_149 this procedure was performed with slightly different enzymes and concentrations. Free DNA and RNA were digested with DNase 1 (Ambion, 0.004 U/µl) and RNase 1 (Ambion, 0.1 U/µl) and the viral capsid was disrupted with Proteinase K (0.05 U/ml) and SDS (0.5%).

Viral DNA was extracted from the DNase, RNase and proteinase K treated lysate by mixing 1 ml of lysate with 1 ml of Wizard® PCR Preps DNA Purification Resin (Promega). A Wizard® Minicolumn was attached to a 5 ml disposable syringe with the plunger removed. The lysate/resin mixture was filled into the syringe and pushed through the minicolumn into a waste collection vessel. In the same way, 2 ml of 80% isopropanol were pushed through to wash the resin. Afterwards, the minicolumn was removed from the syringe and attached to sterile 1.5 ml centrifuge tube. After centrifugation (2 min at 10000 x g) the minicolumn was placed on a new sterile 1.5 ml centrifuge tube, filled with 100 µl 80 °C nuclease free water or TE buffer (10 mM Tris-HCl pH 8.0, 2 mM EDTA pH 9.0) and vortexed gently for ten seconds. After one minute incubation at room temperature it was centrifuged again for 30 seconds at 10000 x g to elute the DNA. The extracted DNA was concentrated to approximately 20 µl using 30 kDa Amicon

Ultra centrifugal filters (0.5 ml volume, Merck Millipore) and finally stored at -20 °C. DNA concentration was determined with the Qubit 2.0 fluorometer and the Qubit® dsDNA HS Assay.

4.2.5.5. Screening for unique phages with RAPD-PCR

To prevent from multiple sequencing of the same phage, a randomly amplified polymorphic DNA (RAPD) PCR was performed on all phage isolates aiming at identification of unique phages. RAPD PCR was either done with extracted phage DNA (for P4 and NHS phages) or on concentrated phage lysates (for P1 and P2 phages). Concentration of P1 and P2 phages was done by polyethylene glycol (PEG) precipitation. 1.6 ml of phage lysate (obtained from liquid infection culture, see chapter 4.2.5.2, method A) were incubated for 2 h (to overnight) at 4 °C with PEG (final concentration 5%) and NaCl (final concentration 0.1 M). After centrifugation (1 h at 7197 x g and 4 °C) the supernatant was discarded and the pellet resuspended in 30 µl nuclease free water (Invitrogen). For resuspension of the phages, the pellet was incubated 1-2 h at 4 °C with occasional gentle vortexing. Phage concentrates were stored at -20 °C. For removal of extracellular DNA and RNA, phage concentrates were incubated with 1 µl of TurboDNase (2 U/µl; Invitrogen, Ambion) and 1 µl of RNase Cocktail Enzyme Mix (500 U/ml RNase A, 20,000 U/ml RNase T1; Invitrogen) for 30 min at 37 °C. Enzyme inactivation was performed by incubation for 10 min at 75 °C with 15 mM EDTA.

The PCR mixture for RAPD PCR consisted of nuclease free water (Invitrogen), reaction buffer (supplied with Taq DNA polymerase), 0.16 mM of each deoxynucleoside triphosphate (Invitrogen), 2.5 U of Taq DNA polymerase (recombinant, Thermo Scientific), 4 µM RAPD primer OPA-9 (5'-ggg taa cgc c-3'; stock concentration 100 pmol/µl, Winget and Wommack (2008)) and 1 µl phage DNA or concentrated lysate. The PCR was run with 10 min at 94 °C for initial denaturation, 30 cycles of 30 s of denaturation at 94 °C, 3 min of annealing at 35 °C, and 1 min of extension at 72 °C, followed by 10 min at 72 °C. For visualization of the PCR products agarose gel electrophoresis was performed (1.8%, 1.5 h, 80 V, ethidium bromide staining). A 1kb Plus DNA ladder (Invitrogen) was used as molecular size marker. Banding patterns of all phages were compared manually based on absence or presence of bands. Phages with a pattern different from all the others were regarded as unique.

4.2.6. Isolation of phages from enrichments

Separate from the large-scale direct phage isolation campaign described above, *Lentibacter* phages ICBM4, ICBM6 and ICBM7 were isolated from enrichment cultures containing the host

bacterium *Lentibacter* sp. SH36 and North Sea water samples. Phages ICBM4 and ICBM7 were isolated from the same water sample as the cobaviruses (HE440-S), phage ICBM6 was isolated from sample HE396-6 (see chapter 4.2.4, Table 8). The enrichment procedure was similar to the one used for the isolation of the cobaviruses. The cultures were set up by mixing 1 part medium (ASW10x or MB10x, containing ten times the amount of peptone and yeast extract in comparison with ASW1x or MB) with 9 parts freshly filtered (Nalgene rapid-flow, 0.2 µm, PES membrane, Thermo-Scientific) seawater (HE440-S or HE396-6, respectively). After addition of exponentially growing cultures of *Lentibacter* sp. SH36 (in ASW1x medium) to a final OD_{600} of 0.006, enrichment cultures were incubated overnight (ICBM4) or for 3 – 7 days (ICBM6 and ICBM7) at 20 °C and 100 rpm. Two control cultures were incubated in parallel: A positive control containing ASWbase instead of seawater sample in order to monitor host growth and a negative control containing medium and seawater, but no host inoculum to check growth of bacterial contaminants in the seawater. Bacterial growth was monitored by optical density measurement at 600 nm (Beckmann DU520, USA). When lysis was indicated by the presence of cell debris and decreasing OD in the enrichment cultures compared to the positive control culture, the enriched phage fraction was obtained by centrifugation (15 min at 4000 x g) and 0.22 µm filtration (0.22 µm, ROTILABO® syringe filters, CarlRoth) and stored at 4 °C or as glycerol stocks containing host bacteria and phages (see chapter 4.2.5.2.).

Later, phage fractions were revived for isolation of single viruses by infecting 5 ml of exponentially growing culture of *Lentibacter* sp. SH36 (OD_{600} = 0.1, in MB medium) with 100 µl of the enrichment glycerol stock or 4 °C – stock, respectively. After incubation overnight at 20 °C and 100 rpm, the fresh phage lysate was harvested by centrifugation (10 min, 10000 x g, 10 °C) and 0.22 µm filtration (syringe filter, Merck Millipore) of the supernatant. Phages were isolated by plaque assay and single plaque picking, using 100 µl of dilutions (10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}) of the fresh phage lysate for the initial plaque assay on MB agar plates. The procedure of single plaque picking was a bit different than described above. A single plaque was picked and dissolved in 1 ml ASWbase overnight at 4 °C. Dilutions were prepared in ASWbase and plated again in another plaque assay obtaining again single plaques to be picked. This step was repeated at least three times in order to obtain pure plaques.

Genomic DNA of *Lentibacter* phages ICBM4, ICBM6 and ICBM7 was extracted after enrichment of phage particles by two subsequent liquid infections as well as polyethylene glycol (PEG) precipitation. A pre-infection culture was set up by mixing 30 ml ASW1x medium with 500 µl of phage glycerol stock and an inoculum of exponentially growing culture of *Lentibacter* SH36 (final OD_{600} = 0.006). A positive control culture was also prepared containing

only medium and the host inoculum. After incubation overnight at 20 °C and 100 rpm, lysis was indicated in the infection culture by cell debris and a decreased OD_{600} compared to the positive control. The fresh phage lysate was harvested by centrifugation (15 min, 4000 x g) and 0.22 µm filtration (ROTILABO® syringe filters, CarlRoth). The first large scale infection culture was set up containing 60 ml ASW1x medium, 1 ml fresh phage lysate and the inoculum from an exponentially growing culture of *Lentibacter* SH36 (final OD_{600} = 0.006). Two positive control cultures were prepared as well containing only medium and host inoculum. After incubation overnight at 20 °C and 100 rpm, lysis was indicated in the infection culture and the phage lysate was harvested by centrifugation (30 min, 4000 x g, 4 °C) and capture of the supernatant. A bacterial cell pellet was prepared by centrifugation of control culture I (30 min, 4000 x g, 20 °C) and resuspended in 50 ml of the phage lysate. The mixture was incubated for 15 min on ice and then transferred to an Erlenmeyer flask. After addition of 50 ml ASW2x medium (containing two times the amount of peptone and yeast extract as compared to ASW1x medium), the thereby obtained second large scale infection culture was incubated overnight at 20 °C and 100 rpm. A positive control culture was prepared by mixing 50 ml of control culture II with 50 ml ASW2x medium and incubated as well. When cell lysis was indicated in the infection culture, the phage lysate was obtained again by centrifugation (30 min, 7000 x g, 4 °C). For further concentration of the phage particles, precipitation with polyethylene glycol (PEG 8000, Molecular Biology Grade, Promega) was performed. 4 x 25 ml phage lysate were mixed with 50% PEG (final concentration 10%) and 5 M sodium chloride (final concentration 0.6 M). After incubation for 2 g at 4 °C and subsequent centrifugation (2 h, 7197 x g, 10 °C), the phage pellets were resuspended in SM buffer, pooling them all together in 500 µl. Extracellular DNA was removed by DNase treatment. The phage concentrate was incubated with 0.04 U/µl Turbo DNase (2 U/µl, Invitrogen, Ambion) for 30 min at 37 °C. Afterwards, the enzyme was inactivated by incubation for 10 min at 75 °C with 15 mM EDTA. The phage DNA was extracted with the ChargeSwitch gDNA Mini Bacteria Kit (ThermoFisher Scientific) following the manufacturer´s instructions. RNase digestion was included in the first step of the protocol, but no lysozyme treatment. In the end, DNA was eluted in 1 ml elution buffer. The extracted DNA was further purified and concentrated to 100 µl using 30 kDa Amicon Ultra centrifugal filters (0.5 ml volume, Merck Millipore) and stored at -20 °C. DNA concentration was determined with the Qubit 2.0 fluorometer and the Qubit® dsDNA HS Assay.

4.2.7. Phage genome sequencing, assembly and genome end determination

Genome sequencing was performed by our collaborators Anja Poehlein and Mechthild Bömeke at the Göttingen Genomics Laboratory (University of Göttingen), using the Illumina technology and the library preparation protocol for dsDNA. They provided us with the raw sequencing reads. The Illumina raw reads were cleaned with BBDuk in two steps. In the first step, the adaptors were removed, using the following parameters for BBDuk: "ktrim=r k=21 mink=8 tbo tpe ftm=5 rcomp=t ordered t=8". In the second step, contaminating reads from phiX174 as well as low quality ends were removed, using the following parameters for BBDuk: "k=21 rcomp=t hdist=1 qtrim=rl trimq=20, maq=20 minlen=30 ordered t=8". The quality of the cleaning was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The cleaned reads were normalized with BBNorm, with the parameters "target=100 min=5". BBDuk and BBNorm are part of the BBTools package (https://jgi.doe.gov/data-and-tools/bbtools/). Afterwards, the normalized reads were assembled with the SPAdes genome assembler software v3.13.0 (Bankevich et al. 2012), with the parameters "-t 8 -k 21,33,55,77,99,127 --sc –careful". The resulted contigs were manually checked with Bandage 0.8.1. The genomes with circular contigs were linearized at export from Bandage (Wick et al. 2015). Further, PhageTerm 1.0.11 (Garneau et al. 2017) was used to determine the genome ends and to reorder the genome sequence accordingly.

4.2.8. Clustering at species and genus level - VIRIDIC analysis

All complete phage genomes from the new isolates were submitted to VIRIDIC (Moraru et al. 2020), for the calculation of pairwise intergenomic identities based on nucleic-acids. VIRIDIC was used with default parameters, that is "-wordsize 7 -reward 2 -penalty -3 -gapopen 5 -gapextend 2" for BLASTn and the "complete" agglomeration method. This procedure was repeated after removal of duplicate genomes (i.e. 100% nucleotide-based intergenomic identity). Species-level and genus-level genome clusters were calculated by cutting the tree at the 95% and 70% identity threshold, respectively.

4.2.9. Retrieval of related phage genomes and of other roseophages from sequence databases

Virus Relative Finder web-app (manuscript in prep., Cristina Moraru) was used to find phage genomes related to our roseophage isolates. In a first step, VirRel Finder predicted proteins for each roseophage isolate using MetaGeneAnnotator (Noguchi et al. 2008), and further translated them using the seqinr R package (Charif and Lobry 2007) and the translation code 11. The

proteins from all roseophages were then used to search using BLASTP a database of all viral genomes downloaded from GenBank in November 2022 (https://ftp.ncbi.nlm.nih.gov/genomes/genbank/viral/). From all search results, the protein hits further considered were only those with their bitscore >30 and coverage >70. In the last step, VirRel Finder selected only those viral genomes that had at least 5% of the proteins similar with those of one of the new roseophage isolates, and a genome length ranging between 50% and 150% the length of its related roseophage isolate. At this point, more than 9000 phage genomes were obtained. After removal of the genomes labeled with "uncultured", the remaining genomes were clustered at the genus level using an identity threshold of 70% with VIRIDIC. To further reduce the dataset, from each genus cluster only one representative was kept.

This dataset was merged with 28 representatives of our newly isolated phages. Furthermore, the genomes of eleven viruses infecting *Sulfitobacter* sp. SW_H+_2_149 (phages Ebeline 1-11) were provided by Nina Bartlau (MPI, Bremen, Germany) and were incorporated into this dataset. They had been isolated from a North Sea water sample (station of "Kabeltonne" (54º 11' 17.88" N, 7º 54' 0" E) at the channel of Helgoland) using enrichment cultures. Additionally, previously published roseophages were added to the dataset (see Table 9) resulting in a collection of 1327 genomes.

The genomes in this dataset were then hierarchically clustered with VirClust (Moraru 2023), using the following parameters: i) code 11 for protein translation; ii) protein clustering based on BLASTp using the log e-values, and matches were kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%; ii) the complete aglomeration method for genome clustering. Further, the genome tree was split into clusters using a 0.995 distance threshold, and only those genome clusters that contained roseophages (either the new isolates, or previously isolated ones) were kept. At this point, we checked the ICTV-sanctioned taxonomic classification of the phages in this dataset, and, where necessary, we supplemented the dataset with further phage genomes from the detected viral families. This was done to ensure a uniform representation of the respective phage families, and resulted in a final dataset of 965 phage genomes, named here Roseo_DB (see SI file S4-7).

4.2.10. VirClust analysis – hierarchical clustering and genome annotation

To enable the classification of the new roseophage isolates, the genomes in Roseo_DB were hierarchically clustered using VirClust (Moraru 2023), as follows. In a first step, the genetic code 11 was used for protein translation, and then proteins were clustered based on BLASTp (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001,

identity >= 0%). Then, the viral genomes were hierarchically clustered with the complete agglomeration method. The tree was then split into viral genome clusters (VGCs) matching most of the ICTV-recognized phage families in the dataset using a distance threshold of 0.995. The tree was further visualized and annotated in iTOL (Letunic and Bork 2021).

And lastly, we annotated the predicted proteins and tRNAs from the genomes of all new roseophage isolates. First, we used VirClust to search for protein homologoues in the following databases: the prokaryotic Virus Orthologous Groups (pVOGs) (Grazziotin et al. 2017) database, the Virus Orthologous Group database (VOGDB, https://vogdb.csb.univie.ac.at, (Kiening et al. 2019)) database, the Prokaryotic Virus Remote Homologous Groups (PHROGS) database (Terzian et al. 2021), and the InterPro database (Finn et al. 2017). Then, the results were manually evaluated and consolidated, to assign a product and a functional category to each protein. And finally, to annotate the tRNAs we used the online tool tRNAscan-SE v. 2.0 (http://lowelab.ucsc.edu/tRNAscan-SE/index.html, Lowe and Chan (2016)). Genome maps of the new roseophage isolates were generated using the genoPlotR (Guy et al. 2010) from the R programming environment (https://www.rproject.org/). Furthermore, VirClust (Moraru 2023) was used to determine the core proteins of all members of a VGC or of smaller phage groups such as potential families using the same parameters as above (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%).

Table 9: Previously described dsDNA roseophages collected from literature for whole-genome based classification. N.d. = not described.

Phage short name	Accession #	Genome length [bp]	Host	Morphology	Previous classification	Reference for previous classification	Reference for phage isolation
vBDshP-R2C	KJ803031	74806	*Dinoroseobacter* shibae DFL12	podoviral	*Caudoviricetes*, *Schitoviridae*, *Rhodovirinae*, various genera	Adriaenssens et al. (2020) Adriaenssens et al. (2021)	Cai et al. (2015)
vBDshP-R1	KJ621082	75028	*Dinoroseobacter* shibae DFL12				Ji et al. (2015)
DS-1410Ws-06	KU885988	76466	*Dinoroseobacter* shibae DFL12				Li et al. (2016a)
RD-1410W1-01	KU885989	72674	*Roseobacter denitrificans* OCh114				
RD-1410Ws-07	KU885990	76298	*Roseobacter denitrificans* OCh114				
RPP1	FR719956	74704	*Roseovarius nubinhibens*				Chan et al. (2014)
RLP1	FR682616	74583	*Roseovarius* sp. 217				
⌽CB2047-B	HQ317387	74485	*Sulfitobacter* sp. 2047				Ankrah et al. (2014a)
EE36_P1	FJ591094	73325	*Sulfitobacter* sp. EE-36				Zhao et al. (2009)
DSS3_P2	FJ591093	74611	*Ruegeria pomeroyi* DSS-3				
vB_RpoP-V12	MH015250	74675	*Ruegeria pomeroyi* DSS-3				Zhan and Chen (2019a)
vB_RpoP-V13	MH015256	74830	*Ruegeria pomeroyi* DSS-3				
vB_RpoP-V14	MH015257	74792	*Ruegeria pomeroyi* DSS-3				
vB_RpoP-V17	MH015259	74665	*Ruegeria pomeroyi* DSS-3				
vB_RpoP-V21	MH015253	74665	*Ruegeria pomeroyi* DSS-3				
⌽CB2047-A	HQ332142	40929	*Sulfitobacter* sp. 2047	n.d.	Caudovirales, Podoviridae, cluster 3	Zhan and Chen (2019a)	Ankrah et al. (2014b)
⌽CB2047-C	HQ317384	40931	*Sulfitobacter* sp. 2047				
P12053L	JQ809650	38889	*Celeribacter* sp. strain IMCC12053	n.d.	*Caudoviricetes*, *Zobellviridae*, *Cobavirinae*, *Siovirus*	Bischoff et al. (2020)	Kang et al. (2012)
SIO1	AF189021	39898	*Roseobacter* sp. SIO67	podoviral			Rohwer et al. (2000)
vB_LenP_ICBM1	MF431617	40163	*Lentibacter* sp. SH36				Bischoff et al. (2019)
vB_LenP_ICBM3	MF431615	40497	*Lentibacter* sp. SH36				
vB_LenP_ICBM2	MF431616	40907	*Lentibacter* sp. SH36	podoviral	*Caudoviricetes*, *Zobellviridae*,	Bischoff et al. (2020)	Bischoff et al. (2019)

					Cobavirinae, *Veravirus*		
CRP-4	MK613346	40768	*Roseobacter* RCA strain FZCC0023	podoviral	Cobavirus-like (Caudovirales, Podoviridae)	Zhang et al. (2019a)	Zhang et al. (2019a)
CRP-5	MK613347	39600	*Roseobacter* RCA strain FZCC0040				
CRP-6	MK613348	44927	*Roseobacter* RCA strain FZCC0042		CRP-6-type (Caudovirales, Podoviridae)		
CRP-7	MK613349	58106	*Roseobacter* RCA strain FZCC0042		CRP-7-type		
CRP-13	MW514247	55022	*Roseobacter* RCA strain FZCC0023	podoviral	HMO-2011-type	Zhai et al. (2021)	Zhai et al. (2021)
CRP-9	MW514246	56157	*Roseobacter* RCA strain FZCC0023				
CRP-1	MK613343	54045	*Roseobacter* RCA strain FZCC0023	podoviral	HMO-2011-like, genus-level cluster I (Caudovirales, Podoviridae)	Qin et al. (2022)	Zhang et al. (2019a)
CRP-2	MK613344	54148	*Roseobacter* RCA strain FZCC0023				
CRP-345	MZ892990	54718	*Roseobacter* sp. FZCC0042	n.d.			Qin et al. (2022)
CRP-603	MZ892991	54551	*Roseobacter* sp. FZCC0012				
CRP-207	MZ892987	54895	*Roseobacter* sp. FZCC0040				
CRP-3	MK613345	52963	*Roseobacter* RCA strain FZCC0040	podoviral	HMO-2011-like, genus-level cluster III (Caudovirales, Podoviridae)	Qin et al. (2022)	Zhang et al. (2019a)
CRP-212	MZ892988	54748	*Roseobacter* sp. FZCC0040	n.d.			Qin et al. (2022)
CRP-235	MZ892989	52729	*Roseobacter* sp. FZCC0040				
CRP-738	MZ892992	53826	*Roseobacter* sp. FZCC0089				
Antarctic DB virus 2	MW805364	39241	*Octadecabacter* sp. IceBac 430	podoviral	*Caudoviricetes*	Demina et al. (2021)	Luhtanen et al. (2018)
RDJL_phi1	HM151342	62668	*Roseobacter denitrificans* OCh114	siphoviral	*Caudoviricetes*, *Xiamenvirus*	Kropinski et al. (2016),	Zhang and Jiao (2009)
RDJL_phi2	KT266805	63513	*Roseobacter denitrificans* OCh114				Liang et al. (2016)
vB_RpoS-V10	MH015255	147480	*Ruegeria pomeroyi* DSS-3	siphoviral	Cbk-like (Siphoviridae, cluster 2)	Zhan and Chen (2019a)	Zhan and Chen (2019a)
DSS3_P8	KT870145	146135	*Ruegeria pomeroyi* DSS-3				Zhan et al. (2016)
MD18	MT270409	149262	*Phaeobacter inhibens*			Urtecho et al. (2020)	(Urtecho et al. 2020)
DSS3_P1	KM581061	59601	*Ruegeria pomeroyi* DSS-3	siphoviral		Zhan and Chen (2019a)	Zhan et al. (2018)

vB_RpoS-V11	MH015254	59549	*Ruegeria pomeroyi* DSS-3		Chi-like (Siphoviridae, cluster 3)		
vB_RpoS-V16	MH015258	61382	*Ruegeria pomeroyi* DSS-3				
vB_RpoS-V18	MH015252	59111	*Ruegeria pomeroyi* DSS-3				
vB_RpoS-V7	MH015249	59573	*Ruegeria pomeroyi* DSS-3				
pCB2051-A	HQ632859	56958	*Loktanella* sp. CB2051	n.d.	*Caudoviricetes, Casjensviridae, Broinstvirus*	Tolstoy et al. (2021)	Unpublished (GenBank 2013)
vBDshS-R5C	KY606587	77874	*Dinoroseobacter* shibae DFL12	siphoviral	*Caudoviricetes, Nanhaivirus*	Kropinski et al. (2018), Turner et al. (2020)	Yang et al. (2017)
Tedan	MT764845	75087	*Ruegeria* sp. AU67	siphoviral	*Xiamenvirus* (Siphoviridae)	Baum et al. (2021)	Baum et al. (2021)
RcCronus	KR935217	35985	*Rhodobacter capsulatus* YW1 C6	siphoviral	*Caudoviricetes, Cronusvirus*	Delesalle et al. (2016b)	Bollivar et al. (2016)
RcRhea	KR935216	36065	*Rhodobacter capsulatus* YW1		Cronusvirus-like genus-level cluster RcA (Caudovirales, Siphoviridae)	Rapala et al. (2021)	
RcSaxon	KT253150	36081	*Rhodobacter capsulatus* YW1				
RcSpartan	KR935215	44194	*Rhodobacter capsulatus* YW1	siphoviral	*Caudoviricetes, Titanvirus*	Delesalle et al. (2016a), Adriaenssens et al. (2021)	
RcTitan	KR935213	44496	*Rhodobacter capsulatus* YW1				
RcThunderbird	MW677526	43941	*Rhodobacter capsulatus* YW1		Titanvirus-like genus-level cluster RcB (Caudovirales, Siphoviridae)	Rapala et al. (2021)	Rapala et al. (2021)
RcHartney	MW677514	43528	*Rhodobacter capsulatus* YW1				
RcOceanus	MW677520	37609	*Rhodobacter capsulatus* YW1	siphoviral			

RcBaka	MW677509	41643	*Rhodobacter capsulatus* YW1		RcC (genus-level cluster in Caudovirales, Siphoviridae)		
RcDormio	MW677510	41640	*Rhodobacter capsulatus* YW1				
RcFrancesLouise	MW677512	42073	*Rhodobacter capsulatus* YW1				
RcHotPocket	MW677515	41765	*Rhodobacter capsulatus* YW1				
RcKemmy	MW677517	41345	*Rhodobacter capsulatus* YW1				
RcGingersnap	MW677513	68225	*Rhodobacter capsulatus* YW1	siphoviral	RcD (genus-level cluster in Caudovirales, Siphoviridae)		
RcIroh	MW677516	68475	*Rhodobacter capsulatus* YW1				
RcMcDreamy	MW677518	68244	*Rhodobacter capsulatus* YW1				
RcMrWorf	MW677519	67929	*Rhodobacter capsulatus* YW1				
RcPutin	MW677522	67605	*Rhodobacter capsulatus* YW1				
RcPescado	MW677521	67494	*Rhodobacter capsulatus* YW1				
RcRios	MW677523	68774	*Rhodobacter capsulatus* YW1				
RcSalem	MW677524	67718	*Rhodobacter capsulatus* YW1				
RcWaterboi	MW677528	38301	*Rhodobacter capsulatus* YW1	siphoviral	RcE (genus-level cluster in Caudovirales, Siphoviridae)		
RcapMu	JN190960	39283	*Rhodobacter capsulatus* SB1003				Fogg et al. (2011)
vB_PmaS-R3	KP162168	42093	*Paracoccus marcusii* JL-65	siphoviral	Caudovirales, Siphoviridae	Xu et al. (2015)	Xu et al. (2015)
vB_PthS_Pthi1	MK291444	39547	*Paracoccus thiocyanatus* JCM 20756	siphoviral	Caudovirales, Siphoviridae	Decewicz et al. (2019)	Decewicz et al. (2019)
Shpa	KR072689	38261	*Paracoccus* sp. HS3	siphoviral	*Caudoviricetes, Vhulanivirus*	van Zyl et al. (2016)	van Zyl et al. (2016)
vB_PsuS_Psul1	MK291443	37901	*Paracoccus sulfuroxidans* JCM 14013	siphoviral	Caudovirales, Siphoviridae	Decewicz et al. (2019)	Decewicz et al. (2019)
vB_PbeS_Pben1	MK291441	39879	*Paracoccus bengalensis* DSM 17099				
vB_ThpS-P1	KT381864	39591	*Thiobacimonas profunda* JLT2016	siphoviral	Caudovirales, Siphoviridae	Tang et al. (2017)	Tang et al. (2017)
vB_PeaS-P1	KT381865	38868	*Pelagibaca abyssi* JLT2014				

vBDshS-R4C	MK882925	36291	*Dinoroseobacter* shibae DFL12	siphoviral	Caudovirales, Siphoviridae	Cai et al. (2019)	Cai et al. (2019)
DSS3_VP1	MN602266	75087	*Ruegeria pomeroyi* DSS-3	siphoviral	*Caudoviricetes*, *Naomiviridae*, *Noahvirus*, *Noahvirus arc*	Rihtman et al. (2021)	Rihtman et al. (2021)
DSS3_PM1	MN602267	70044	*Ruegeria pomeroyi* DSS-3				
Antarctic DB virus 1	MW805363		*Octadecabacter* sp. IceBac 419	siphoviral	*Caudoviricetes*	Demina et al. (2021)	Luhtanen et al. (2018)
RcSimone-Håstad	MW677525	63102	*Rhodobacter capsulatus* SB1003	siphoviral	Caudovirales, Siphoviridae	Rapala et al. (2021)	Rapala et al. (2021)
vB_RhkS_P1	KX077179	38773	*Rhodovulum* sp. P5	siphoviral	unclassified		Lin et al. (2016)
RC1	JF974308	39573	*Rhodobacter capsulatus* E32	siphoviral	unclassified		Engelhardt et al. (2011)
vB_PyeM_Pyei1	MK291445	50161	*Paracoccus yeei* CCUG 32053	myoviral	Caudovirales, Myoviridae	Decewicz et al. (2019)	Decewicz et al. (2019)
RcapNL	JQ066768	40489	*Rhodobacter capsulatus* SB1003	n.d.	unclassified		Hynes (2014)
NYA-2014a	KM233261	42092	*Sulfitobacter* sp. CB2047	n.d.	unclassified		Unpublished (GenBank 2014)
phiGT1	MT584811	40019	*Sulfitobacter* sp. HGT1	n.d.	unclassified		Hwang et al. (2020)
45A6	KT820175	53160	*Ruegeria mobilis* 45A6	n.d.	unclassified		Unpublished (GenBank 2017)
RS1	JF974307	40231	*Rhodovulum* sp. P122A	n.d.	unclassified		Unpublished (GenBank 2013)

4.3. Results

4.3.1. Strains from three *Roseobacteraceae* genera served for the isolation of 277 phages

In this study, strains of different genera within the *Roseobacteraceae* family were used as bait bacteria for phage isolation (Table 10). In the large-scale isolation experiment, phages were isolated by direct plating. A collection of 388 Roseobacter Group strains (M#), which were until then not further classified, and two *Lentibacter* strains were challenged with four different water samples (P1, P2, P4, and NHS). In addition, the same two *Lentibacter* strains and one *Octadecabacter* strain were used for direct phage isolation from seawater sample HE504-33. Furthermore, enrichment cultures containing *Lentibacter* sp. SH36 and different water samples (HE396-6 and HE440-S) were used for phage isolation. All potential host strains originated from different sampling sites in the southern North Sea, either directly at the East Frisian coast, in the open sea or at the channel of Helgoland, and provided by different collaborators (see Fig. 18 and chapter 4.2.2.).

Table 10: Pairing of the phage water samples with the bacterial hosts during the phage isolation procedures.

Phage water source	Hosts	Isolation procedure
HE396-6 HE440-S	*Lentibacter* sp. SH36	enrichment
HE504-33	*Lentibacter* sp. SH36, MPI-62, MM282	direct-plating
P1, P2, P4, NHS	*Lentibacter* sp. SH36, MPI-62, Roseobacter Group strains M#	

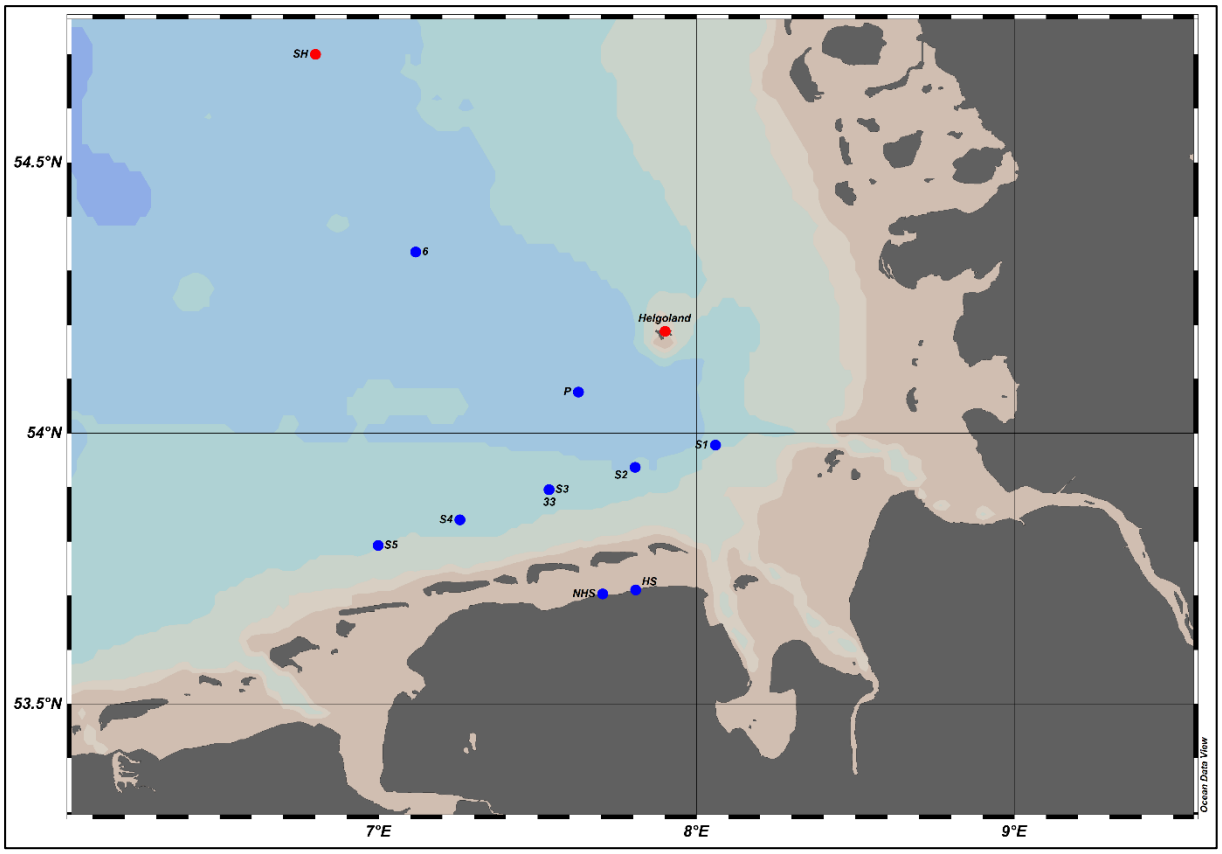


Fig. 18: Map of the german bight showing sampling sites for the isolation of bacterial strains (red) and phages (blue). SH = origin of *Lentibacter* sp. SH36 (Hahnke et al. 2013), Helgoland = channel of Helgoland (time series station), P = origin of mesocosm inoculum (Mori et al. 2021), 33 = origin of sample HE504-33, 6 = origin of sample HE396-6, S1-5 = origins of pooled samples HE504-S, NHS = Neuharlingersiel, HS = Harlesiel.

A total of 277 phage isolates were obtained (Table 11). Three of them were isolated from enrichment cultures, while the majority originated from direct plating. They have been named as follows: the isolation host genus name, followed by "phage", followed by "ICBM" (from the Institute of Chemistry and Biology of the Marine Environment) and a number (see Table 13). The last part (ICBM#) also served as a short name, and it will be used throughout the text.

The majority of phage isolates in this study (115 unique phage genomes) originated from infection of 64 strains from the Roseobacter Group collection M#, which had been isolated from Helgoland Roads (Fig. 18). Almost all of these strains formed a cohesive cluster based on 16S rRNA gene sequence similarity with their closest relative being *Sulfitobacter marinus* (Fig. 19). Only strain M315, host of phage ICBM153, clustered differently within the *Sulfitobacter* genus. The tree also includes *Sulfitobacter* sp. SH24-1b, which is the host strain of "Ascunsovirus oldenburgi" ICBM5, a novel ssDNA microvirus, which is described in chapter 6. This strain was most closely related to *Sulfitobacter dubius* and was isolated from the same sampling site as *Lentibacter* sp. SH36 (Hahnke et al. 2013).

Lentibacter sp. SH36, host of the above-described cobaviruses, was again successfully used for phage isolation from enrichment cultures (phages ICBM4, ICBM6, ICBM7) as well as from direct plating (phages ICBM8, ICBM161 - ICBM166). Another strain of this genus, *Lentibacter* sp. MPI-62, isolated from the channel of Helgoland (North Sea) in 2017, was used for direct phage isolation and was infected by four phages (phages ICBM157 - ICBM160). *Octadecabacter* sp. MM282 was isolated in 2017 at the North Sea coast in Harlesiel, Germany, and was successfully used for direct isolation of phage ICBM156. In conclusion, phages have been isolated from three different genera, which are *Lentibacter*, *Sulfitobacter* and *Octadecabacter*.

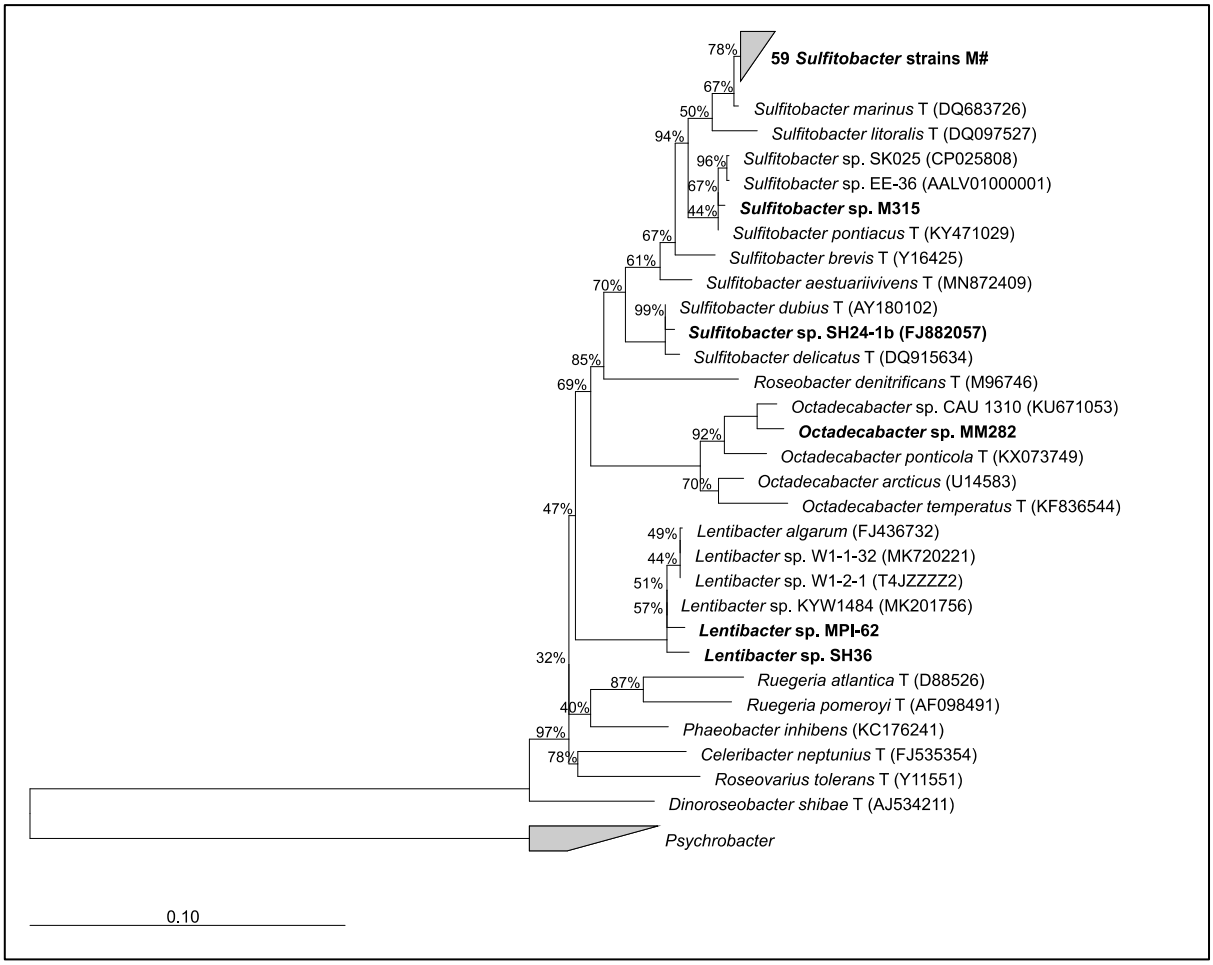


Fig. 19: Neighbor-joining tree based on 16S rRNA gene sequence similarity showing the phylogenetic affiliation of the bacterial strains used for phage isolation in this study (in bold). The tree was calculated with Jukes-Cantor correction and the termini filter. Bootstrap values were derived from 1000 replicates. The bar represents 0.10 substitutions per nucleotide position. Four strains of the genus *Psychrobacter* (Gammaproteobacteria) (FJ039851, AJ309940, JN411455, and U46139) served as outgroup. Genbank accession numbers are given in parentheses. T type strain of species. Strains M71 and M172 are missing from the tree. However, pairwise sequence comparison revealed that the partial 16S rRNA gene as well as the internal transcribed spacer (ITS) were 99.9% identical to *Sulfitobacter* sp. M290, which belongs to the cluster of 59 *Sulfitobacter* strains (see chapter 5.3.1, Fig. S15)

4.3.2. Quick screening for unique phages by RAPD-PCR

For the large number of phage isolates, a screening method was needed in order to prevent from expensive genome sequencing of duplicate phages. For this purpose, a randomly amplified polymorphic DNA (RAPD) PCR with subsequent gel electrophoresis was performed, which allowed for differentiation of banding patterns and a quick typing of the isolates. Phage isolates originating from the same host strain often had identical banding patterns and were thus regarded as duplicates (Fig. 20 and S13, SI file S4-2). Only one representative, for example phage ICBM76, was chosen to be genome sequenced. However, in other cases phages from the same host strain had different banding patterns and were both genome sequenced (e.g., phages ICBM71 and ICBM104).

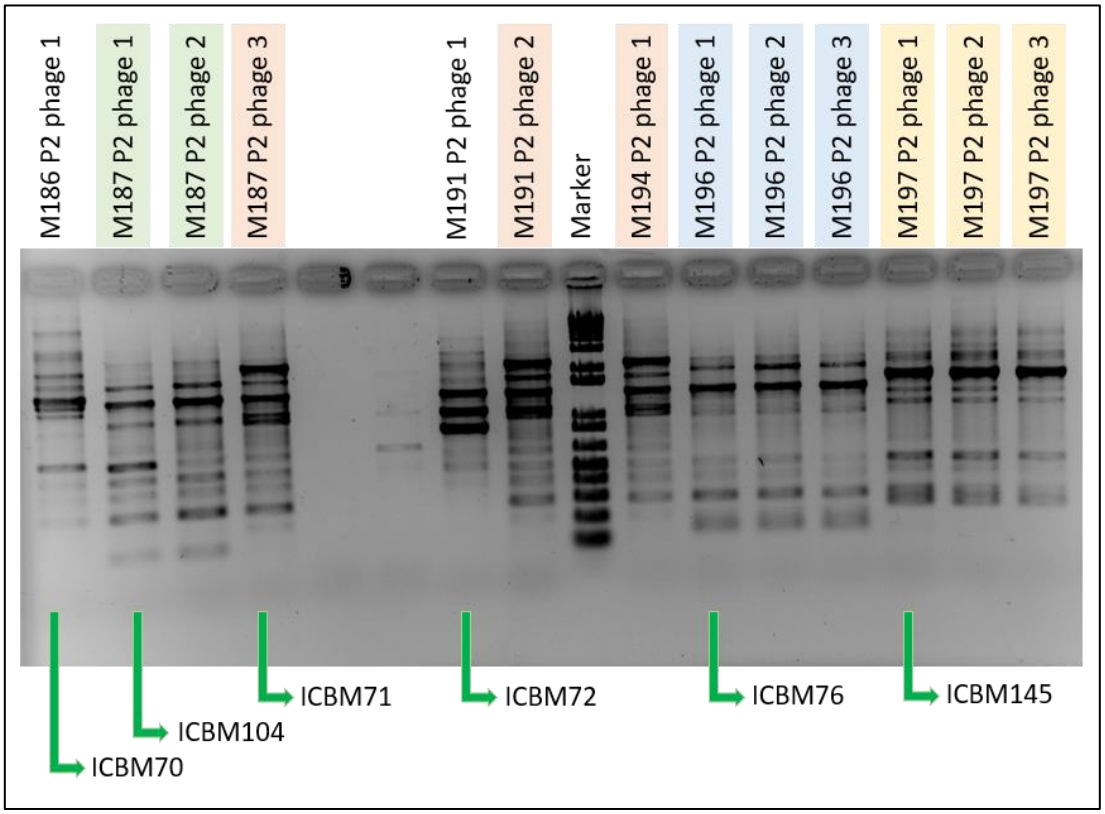


Fig. 20: Gel electrophoresis of RAPD-PCR products for 13 phage isolates. Phages with identical band patterns are colored accordingly. For each unique pattern, one isolate was chosen to be genome sequenced (green arrows). Marker 1kb Plus DNA ladder (Invitrogen™).

Out of 261 phage isolates that had been isolated by direct plating and submitted to RAPD-PCR, 138 showed unique banding patterns and were chosen to be genome sequenced (Table 11, SI file S4-3). From 242 phages that had been isolated from the *Sulfitobacter* sp. M# strains, more than half had unique patterns (127 phages). Six phages had been isolated from *Lentibacter* sp. MPI-62 by direct plating, three from seawater sample P1 and three from sample P2, respectively. From each water sample, two isolates showed the same banding pattern,

respectively. Thus, in the end four phages from *Lentibacter* sp. MPI-62 were selected for sequencing. Out of the 12 phages isolated from *Lentibacter* sp. SH36 by direct plating, six were selected for sequencing, including two phages from each seawater sample (P1, P2, and HE504-33), respectively. Phage ICBM156 isolated from *Octadecabacter* sp. MM282 also showed a banding pattern different from all other isolates and was chosen for genome sequencing.

4.3.3. Genome sequencing and assembling of selected phage isolates

Apart from the 145 phage isolates that were chosen for genome sequencing based on their unique RAPD-PCR banding patterns, 23 additional phages were genome sequenced. They included phages ICBM4, ICBM6 and ICBM7 isolated from enrichment cultures with *Lentibacter* sp. SH36 as well as phage ICBM8, isolated by direct plating with the same host and seawater sample HE504-33. The remaining additional phages were isolated by direct plating from different *Sulfitobacter* sp. M# strains.

Out of the 161 phages sequenced, a total number of 143 phage genomes were successfully assembled (Table 11), 130 of them infecting *Sulfitobacter* strains, twelve infecting *Lentibacter* strains and one phage infecting *Octadecabacter* sp. MM282. From the *Sulfitobacter* phage gen omes, 36 could only be assembled partially (Table 11). For the remaining 94 *Sulfitobacter* pha ges, the *Octadecabacter* phage and all *Lentibacter* phages, the obtained genome contigs were c ircularly closed as observed upon the quality check using Bandage (Wick et al. 2015) and ther efore could be considered complete. PhageTerm (Garneau et al. 2017) was used to predict the genome ends and to cut the circular genomes into linear contigs accordingly. Furthermor e, the genome sequence was reordered in a way that genes of the DNA replication module wer e encoded first and the morphology genes were located towards the genome end. We used this genome arrangement as a convention in order to simplify genome comparisons, still having in mind that the chosen genome termini and directions were artificial. Especially for circularly p ermuted phage genomes, genome ends cannot be universally determined,

because they differ from virion to virion (see introduction chapter 1.3.4.).

The 107 complete genomes ranged in size from 33.35 to 80.76 kb (Table 12). A comparison of the sequenced genome lengths to those predicted through the PFGE measurements (perfomed only for 56 phages), showed differences of less than 4 kb (Table 12). Only in two cases (*Sulfitobacter* phages ICBM76 and ICBM94), the genome size measured by

PFGE was more than 5 kb smaller than the actual genome size. Therefore, PFGE proved to be a decent tool for genome size estimation. All phage genomes obtained were of dsDNA nature, as indicated by the fact that they had been amenable to RAPD-PCR and to genome sequencing using a library preparation protocol for dsDNA genomes.

Table 11: Counts of phage isolates, sequenced and final unique phage genomes. *additional phages were sequenced, that had not been analyzed with RAPD-PCR **nucleotide-based intergenomic identity <100%.

	All phages	***Sulfitobacter* phages**	***Octadecabacter* phages**	***Lentibacter* phages**		
				Total	From direct plating	From enrichment
Total phage isolates	**277**	**254**	**1**	**22**	19	3
Isolates used for RAPD-PCR	261	242	1	18	18	-
Isolates chosen after RAPD-PCR	138	127	1	10	10	-
Isolates sent for sequencing*	161	146	1	14	11	3
Assembled genomes	**143**	**130**	**1**	**12**	9	3
- complete	107	94	1	12	9	3
- partial	36	36	-	-	-	-
Unique genomes**	**128**	**115**	**1**	**12**	9	3
- complete	94	81	1	12	9	3
- partial	34	34	-	-	-	-

Table 12: Sizes of complete phage genomes comparing PFGE and sequencing. Size differences >5kb are marked in red. *n.d. = not determined.

	Genome size [bp]				**Genome size [bp]**		
Phage name	**Sequenced**	**PFGE**	**Difference (PFGE - Seq.)**	**Phage name**	**Sequenced**	**PFGE**	**Difference (PFGE - Seq.)**
Lentibacter phage ICBM4	43.101	n.d.	-	Sulfitobacter phage ICBM95	54.142	50.54	-3.60
Lentibacter phage ICBM6	40.273	n.d.	-	Sulfitobacter phage ICBM97	51.373	51.56	0.19
Lentibacter phage ICBM7	45.55	n.d.	-	Sulfitobacter phage ICBM99	52.071	51.56	-0.51
Lentibacter phage ICBM8	38.666	n.d.	-	Sulfitobacter phage ICBM100	54.118	54.63	0.51
Sulfitobacter phage ICBM12	54.842	55.65	0.80	Sulfitobacter phage ICBM101	53.772	53.60	-0.17
Sulfitobacter phage ICBM13	53.694	54.63	0.93	Sulfitobacter phage ICBM102	50.475	48.50	-1.98
Sulfitobacter phage ICBM16	53.115	52.58	-0.53	Sulfitobacter phage ICBM103	50.611	50.38	-0.23
Sulfitobacter phage ICBM18	52.071	51.56	-0.51	Sulfitobacter phage ICBM105	51.379	49.52	-1.86
Sulfitobacter phage ICBM21	52.607	52.58	-0.02	Sulfitobacter phage ICBM107	52.423	n.d.	-
Sulfitobacter phage ICBM22	54.842	54.63	-0.22	Sulfitobacter phage ICBM109	51.998	n.d.	-

Sulfitobacter phage ICBM23	53.772	55.65	1.87	Sulfitobacter phage ICBM110	55.587	n.d.	-
Sulfitobacter phage ICBM24	53.772	52.58	-1.19	Sulfitobacter phage ICBM111	33.588	n.d.	-
Sulfitobacter phage ICBM25	55.591	58.71	3.12	Sulfitobacter phage ICBM113	52.367	n.d.	-
Sulfitobacter phage ICBM38	51.389	53.21	1.82	Sulfitobacter phage ICBM117	34.081	n.d.	-
Sulfitobacter phage ICBM39	53.256	49.44	-3.81	Sulfitobacter phage ICBM118	34.114	n.d.	-
Sulfitobacter phage ICBM40	53.772	50.38	-3.39	Sulfitobacter phage ICBM119	53.787	n.d.	-
Sulfitobacter phage ICBM41	53.787	50.38	-3.40	Sulfitobacter phage ICBM121	36.856	n.d.	-
Sulfitobacter phage ICBM42	54.434	51.33	-3.11	Sulfitobacter phage ICBM122	36.856	n.d.	-
Sulfitobacter phage ICBM43	54.119	52.27	-1.85	Sulfitobacter phage ICBM123	33.588	n.d.	-
Sulfitobacter phage ICBM45	53.766	52.27	-1.50	Sulfitobacter phage ICBM124	54.296	n.d.	-
Sulfitobacter phage ICBM47	52.388	52.27	-0.12	Sulfitobacter phage ICBM125	34.08	n.d.	-
Sulfitobacter phage ICBM48	53.624	53.21	-0.41	Sulfitobacter phage ICBM126	53.881	n.d.	-
Sulfitobacter phage ICBM49	54.336	54.15	-0.18	Sulfitobacter phage ICBM127	52.781	n.d.	-
Sulfitobacter phage ICBM50	52.071	51.33	-0.74	Sulfitobacter phage ICBM128	54.612	n.d.	-
Sulfitobacter phage ICBM51	55.142	55.36	0.22	Sulfitobacter phage ICBM129	70.738	n.d.	-
Sulfitobacter phage ICBM52	54.694	55.36	0.67	Sulfitobacter phage ICBM130	33.346	n.d.	-
Sulfitobacter phage ICBM53	53.465	51.66	-1.80	Sulfitobacter phage ICBM131	52.646	n.d.	-
Sulfitobacter phage ICBM54	53.931	52.72	-1.21	Sulfitobacter phage ICBM133	54.967	n.d.	-
Sulfitobacter phage ICBM55	34.08	n.d.	-	Sulfitobacter phage ICBM134	54.461	n.d.	-
Sulfitobacter phage ICBM57	54.46	54.83	0.37	Sulfitobacter phage ICBM137	53.554	n.d.	-
Sulfitobacter phage ICBM58	54.373	53.77	-0.60	Sulfitobacter phage ICBM138	52.782	n.d.	-
Sulfitobacter phage ICBM59	54.967	54.83	-0.14	Sulfitobacter phage ICBM139	53.722	n.d.	-
Sulfitobacter phage ICBM60	53.548	53.77	0.22	Sulfitobacter phage ICBM143	52.012	n.d.	-
Sulfitobacter phage ICBM61	52.642	52.72	0.08	Sulfitobacter phage ICBM144	50.195	n.d.	-
Sulfitobacter phage ICBM62	50.122	48.50	-1.62	Sulfitobacter phage ICBM145	53.772	n.d.	-
Sulfitobacter phage ICBM64	54.336	53.77	-0.56	Sulfitobacter phage ICBM146	52.265	n.d.	-
Sulfitobacter phage ICBM65	54.288	53.77	-0.52	Sulfitobacter phage ICBM147	51.69	n.d.	-
Sulfitobacter phage ICBM67	51.759	n.d.	-	Sulfitobacter phage ICBM148	52.071	n.d.	-
Sulfitobacter phage ICBM68	52.277	52.72	0.44	Sulfitobacter phage ICBM150	51.373	n.d.	-
Sulfitobacter phage ICBM69	51.379	51.66	0.28	Sulfitobacter phage ICBM151	53.772	n.d.	-
Sulfitobacter phage ICBM70	53.694	54.83	1.13	Sulfitobacter phage ICBM152	53.772	n.d.	-
Sulfitobacter phage ICBM76	50.122	35.70	-14.42	Sulfitobacter phage ICBM153	80.756	n.d.	-
Sulfitobacter phage ICBM77	50.195	47.48	-2.72	Sulfitobacter phage ICBM154	52.265	n.d.	-
Sulfitobacter phage ICBM78	53.429	51.56	-1.87	Sulfitobacter phage ICBM155	52.071	n.d.	-
Sulfitobacter phage ICBM79	53.781	51.56	-2.22	Octadecabacter phage ICBM156	60.762	n.d.	-
Sulfitobacter phage ICBM82	53.759	51.56	-2.20	Lentibacter phage ICBM157	42.72	n.d.	-
Sulfitobacter phage ICBM86	52.921	49.52	-3.40	Lentibacter phage ICBM158	42.906	n.d.	-
Sulfitobacter phage ICBM87	53.362	50.54	-2.82	Lentibacter phage ICBM159	42.936	n.d.	-
Sulfitobacter phage ICBM88	53.044	49.52	-3.52	Lentibacter phage ICBM160	43.138	n.d.	-

Sulfitobacter phage ICBM89	53.957	50.54	-3.42	Lentibacter phage ICBM163	37.2	n.d.	-
Sulfitobacter phage ICBM90	51.373	47.48	-3.89	Lentibacter phage ICBM164	43.051	n.d.	-
Sulfitobacter phage ICBM91	51.389	48.50	-2.89	Lentibacter phage ICBM165	37.385	n.d.	-
Sulfitobacter phage ICBM93	53.694	50.54	-3.15	Lentibacter phage ICBM166	44.947	n.d.	-
Sulfitobacter phage ICBM94	54.72	43.70	-11.02				

Table 13: Total unique phages isolated and genome sequenced in this study. *different North Sea water samples, from the shore (NHS), a mesocosm experiment (P1, P2 and P4), and from the open sea (HE504-33, HE396-6 and HE440-S).

Isolation host	Phage name (full)	Phage name (short)	Isolation source*	Isolation procedure	Genome status	Genome size [kb]
Lentibacter sp. SH36	Lentibacter phage ICBM4	ICBM4	HE440-S	Enrichment	complete	43.101
Lentibacter sp. SH36	Lentibacter phage ICBM6	ICBM6	HE396-6	Enrichment	complete	40.273
Lentibacter sp. SH36	Lentibacter phage ICBM7	ICBM7	HE440-S	Enrichment	complete	45.55
Lentibacter sp. SH36	Lentibacter phage ICBM8	ICBM8	HE504-33	Direct plating	complete	38.666
Sulfitobacter sp. M43	Sulfitobacter phage ICBM9	ICBM9	NHS	Direct plating	partial	52.743
Sulfitobacter sp. M43	Sulfitobacter phage ICBM10	ICBM10	NHS	Direct plating	partial	51.641
Sulfitobacter sp. M43	Sulfitobacter phage ICBM11	ICBM11	NHS	Direct plating	partial	53.354
Sulfitobacter sp. M45	Sulfitobacter phage ICBM12	ICBM12	P2	Direct plating	complete	54.842
Sulfitobacter sp. M45	Sulfitobacter phage ICBM13	ICBM13	P2	Direct plating	complete	53.694
Sulfitobacter sp. M47	Sulfitobacter phage ICBM15	ICBM15	P2	Direct plating	partial	51.401
Sulfitobacter sp. M53	Sulfitobacter phage ICBM16	ICBM16	P1	Direct plating	complete	53.115
Sulfitobacter sp. M53	Sulfitobacter phage ICBM17	ICBM17	P2	Direct plating	partial	50.382
Sulfitobacter sp. M53	Sulfitobacter phage ICBM18	ICBM18	P2	Direct plating	complete	52.071
Sulfitobacter sp. M53	Sulfitobacter phage ICBM19	ICBM19	NHS	Direct plating	partial	32.449
Sulfitobacter sp. M53	Sulfitobacter phage ICBM21	ICBM21	NHS	Direct plating	complete	52.607
Sulfitobacter sp. M54	Sulfitobacter phage ICBM22	ICBM22	P2	Direct plating	complete	54.842
Sulfitobacter sp. M55	Sulfitobacter phage ICBM23	ICBM23	P2	Direct plating	complete	53.772
Sulfitobacter sp. M55	Sulfitobacter phage ICBM24	ICBM24	NHS	Direct plating	complete	53.772
Sulfitobacter sp. M63	Sulfitobacter phage ICBM25	ICBM25	NHS	Direct plating	complete	55.591
Sulfitobacter sp. M66	Sulfitobacter phage ICBM26	ICBM26	P2	Direct plating	partial	51.329
Sulfitobacter sp. M66	Sulfitobacter phage ICBM27	ICBM27	P2	Direct plating	partial	53.89
Sulfitobacter sp. M66	Sulfitobacter phage ICBM28	ICBM28	P2	Direct plating	partial	50.382
Sulfitobacter sp. M66	Sulfitobacter phage ICBM29	ICBM29	P2	Direct plating	partial	53.899
Sulfitobacter sp. M66	Sulfitobacter phage ICBM31	ICBM31	NHS	Direct plating	partial	26.874
Sulfitobacter sp. M66	Sulfitobacter phage ICBM32	ICBM32	NHS	Direct plating	partial	49.909
Sulfitobacter sp. M67	Sulfitobacter phage ICBM33	ICBM33	P2	Direct plating	partial	50.382
Sulfitobacter sp. M67	Sulfitobacter phage ICBM34	ICBM34	P2	Direct plating	partial	50.383
Sulfitobacter sp. M67	Sulfitobacter phage ICBM35	ICBM35	NHS	Direct plating	partial	52.693
Sulfitobacter sp. M67	Sulfitobacter phage ICBM36	ICBM36	NHS	Direct plating	partial	50.678

Sulfitobacter sp. M68	Sulfitobacter phage ICBM38	ICBM38	P4	Direct plating	complete	51.389
Sulfitobacter sp. M69	Sulfitobacter phage ICBM39	ICBM39	NHS	Direct plating	complete	53.256
Sulfitobacter sp. M71	Sulfitobacter phage ICBM40	ICBM40	P2	Direct plating	complete	53.772
Sulfitobacter sp. M72	Sulfitobacter phage ICBM41	ICBM41	NHS	Direct plating	complete	53.787
Sulfitobacter sp. M81	Sulfitobacter phage ICBM42	ICBM42	NHS	Direct plating	complete	54.434
Sulfitobacter sp. M83	Sulfitobacter phage ICBM43	ICBM43	P2	Direct plating	complete	54.119
Sulfitobacter sp. M85	Sulfitobacter phage ICBM45	ICBM45	P4	Direct plating	complete	53.766
Sulfitobacter sp. M86	Sulfitobacter phage ICBM46	ICBM46	NHS	Direct plating	partial	45.405
Sulfitobacter sp. M86	Sulfitobacter phage ICBM47	ICBM47	P4	Direct plating	complete	52.388
Sulfitobacter sp. M86	Sulfitobacter phage ICBM48	ICBM48	P4	Direct plating	complete	53.624
Sulfitobacter sp. M90	Sulfitobacter phage ICBM49	ICBM49	P2	Direct plating	complete	54.336
Sulfitobacter sp. M92	Sulfitobacter phage ICBM51	ICBM51	P2	Direct plating	complete	55.142
Sulfitobacter sp. M92	Sulfitobacter phage ICBM52	ICBM52	P4	Direct plating	complete	54.694
Sulfitobacter sp. M105	Sulfitobacter phage ICBM53	ICBM53	P2	Direct plating	complete	53.465
Sulfitobacter sp. M105	Sulfitobacter phage ICBM54	ICBM54	NHS	Direct plating	complete	53.931
Sulfitobacter sp. M157	Sulfitobacter phage ICBM55	ICBM55	P2	Direct plating	complete	34.08
Sulfitobacter sp. M157	Sulfitobacter phage ICBM56	ICBM56	NHS	Direct plating	partial	54.001
Sulfitobacter sp. M165	Sulfitobacter phage ICBM57	ICBM57	P2	Direct plating	complete	54.46
Sulfitobacter sp. M165	Sulfitobacter phage ICBM58	ICBM58	P2	Direct plating	complete	54.373
Sulfitobacter sp. M165	Sulfitobacter phage ICBM59	ICBM59	NHS	Direct plating	complete	54.967
Sulfitobacter sp. M170	Sulfitobacter phage ICBM60	ICBM60	NHS	Direct plating	complete	53.548
Sulfitobacter sp. M170	Sulfitobacter phage ICBM61	ICBM61	NHS	Direct plating	complete	52.642
Sulfitobacter sp. M171	Sulfitobacter phage ICBM62	ICBM62	P2	Direct plating	complete	50.122
Sulfitobacter sp. M173	Sulfitobacter phage ICBM63	ICBM63	NHS	Direct plating	partial	53.474
Sulfitobacter sp. M176	Sulfitobacter phage ICBM65	ICBM65	NHS	Direct plating	complete	54.288
Sulfitobacter sp. M180	Sulfitobacter phage ICBM67	ICBM67	NHS	Direct plating	complete	51.759
Sulfitobacter sp. M180	Sulfitobacter phage ICBM68	ICBM68	NHS	Direct plating	complete	52.277
Sulfitobacter sp. M183	Sulfitobacter phage ICBM69	ICBM69	P2	Direct plating	complete	51.379
Sulfitobacter sp. M186	Sulfitobacter phage ICBM70	ICBM70	P2	Direct plating	complete	53.694
Sulfitobacter sp. M187	Sulfitobacter phage ICBM71	ICBM71	P2	Direct plating	partial	52.603
Sulfitobacter sp. M191	Sulfitobacter phage ICBM72	ICBM72	P2	Direct plating	partial	51.034
Sulfitobacter sp. M192	Sulfitobacter phage ICBM73	ICBM73	NHS	Direct plating	partial	52.973
Sulfitobacter sp. M192	Sulfitobacter phage ICBM74	ICBM74	NHS	Direct plating	partial	54.605

Sulfitobacter sp. M196	Sulfitobacter phage ICBM76	ICBM76	P2	Direct plating	complete	50.122
Sulfitobacter sp. M196	Sulfitobacter phage ICBM77	ICBM77	NHS	Direct plating	complete	50.195
Sulfitobacter sp. M197	Sulfitobacter phage ICBM78	ICBM78	NHS	Direct plating	complete	53.429
Sulfitobacter sp. M197	Sulfitobacter phage ICBM79	ICBM79	NHS	Direct plating	complete	53.781
Sulfitobacter sp. M199	Sulfitobacter phage ICBM80	ICBM80	NHS	Direct plating	partial	52.003
Sulfitobacter sp. M200	Sulfitobacter phage ICBM81	ICBM81	NHS	Direct plating	partial	53.792
Sulfitobacter sp. M200	Sulfitobacter phage ICBM82	ICBM82	NHS	Direct plating	complete	53.759
Sulfitobacter sp. M201	Sulfitobacter phage ICBM83	ICBM83	NHS	Direct plating	partial	51.497
Sulfitobacter sp. M201	Sulfitobacter phage ICBM84	ICBM84	NHS	Direct plating	partial	52.69
Sulfitobacter sp. M207	Sulfitobacter phage ICBM86	ICBM86	NHS	Direct plating	complete	52.921
Sulfitobacter sp. M207	Sulfitobacter phage ICBM87	ICBM87	NHS	Direct plating	complete	53.362
Sulfitobacter sp. M207	Sulfitobacter phage ICBM88	ICBM88	NHS	Direct plating	complete	53.044
Sulfitobacter sp. M244	Sulfitobacter phage ICBM89	ICBM89	P2	Direct plating	complete	53.957
Sulfitobacter sp. M260	Sulfitobacter phage ICBM90	ICBM90	P2	Direct plating	complete	51.373
Sulfitobacter sp. M260	Sulfitobacter phage ICBM91	ICBM91	P4	Direct plating	complete	51.389
Sulfitobacter sp. M261	Sulfitobacter phage ICBM92	ICBM92	P2	Direct plating	partial	52.169
Sulfitobacter sp. M271	Sulfitobacter phage ICBM94	ICBM94	P2	Direct plating	complete	54.72
Sulfitobacter sp. M283	Sulfitobacter phage ICBM95	ICBM95	P2	Direct plating	complete	54.142
Sulfitobacter sp. M300	Sulfitobacter phage ICBM99	ICBM99	P2	Direct plating	complete	52.071
Sulfitobacter sp. M303	Sulfitobacter phage ICBM100	ICBM100	P2	Direct plating	complete	54.118
Sulfitobacter sp. M68	Sulfitobacter phage ICBM102	ICBM102	P2	Direct plating	complete	50.475
Sulfitobacter sp. M86	Sulfitobacter phage ICBM103	ICBM103	P2	Direct plating	complete	50.611
Sulfitobacter sp. M187	Sulfitobacter phage ICBM104	ICBM104	P2	Direct plating	partial	52.169
Sulfitobacter sp. M276	Sulfitobacter phage ICBM105	ICBM105	P2	Direct plating	complete	51.379
Sulfitobacter sp. M47	Sulfitobacter phage ICBM106	ICBM106	P2	Direct plating	partial	51.972
Sulfitobacter sp. M53	Sulfitobacter phage ICBM107	ICBM107	P2	Direct plating	complete	52.423
Sulfitobacter sp. M55	Sulfitobacter phage ICBM109	ICBM109	NHS	Direct plating	complete	51.998
Sulfitobacter sp. M63	Sulfitobacter phage ICBM110	ICBM110	NHS	Direct plating	complete	55.587
Sulfitobacter sp. M63	Sulfitobacter phage ICBM111	ICBM111	P2	Direct plating	complete	33.588
Sulfitobacter sp. M68	Sulfitobacter phage ICBM113	ICBM113	P4	Direct plating	complete	52.367
Sulfitobacter sp. M69	Sulfitobacter phage ICBM114	ICBM114	P4	Direct plating	partial	52.636
Sulfitobacter sp. M70	Sulfitobacter phage ICBM117	ICBM117	P2	Direct plating	complete	34.081
Sulfitobacter sp. M70	Sulfitobacter phage ICBM118	ICBM118	P2	Direct plating	complete	34.114

Sulfitobacter sp. M72	Sulfitobacter phage ICBM119	ICBM119	NHS	Direct plating	complete	53.787
Sulfitobacter sp. M72	Sulfitobacter phage ICBM120	ICBM120	P4	Direct plating	partial	53.159
Sulfitobacter sp. M73	Sulfitobacter phage ICBM121	ICBM121	P2	Direct plating	complete	36.856
Sulfitobacter sp. M73	Sulfitobacter phage ICBM122	ICBM122	P2	Direct plating	complete	36.856
Sulfitobacter sp. M85	Sulfitobacter phage ICBM124	ICBM124	P4	Direct plating	complete	54.296
Sulfitobacter sp. M86	Sulfitobacter phage ICBM126	ICBM126	NHS	Direct plating	complete	53.881
Sulfitobacter sp. M86	Sulfitobacter phage ICBM127	ICBM127	NHS	Direct plating	complete	52.781
Sulfitobacter sp. M91	Sulfitobacter phage ICBM128	ICBM128	NHS	Direct plating	complete	54.612
Sulfitobacter sp. M92	Sulfitobacter phage ICBM129	ICBM129	NHS	Direct plating	complete	70.738
Sulfitobacter sp. M92	Sulfitobacter phage ICBM130	ICBM130	NHS	Direct plating	complete	33.346
Sulfitobacter sp. M105	Sulfitobacter phage ICBM131	ICBM131	NHS	Direct plating	complete	52.646
Sulfitobacter sp. M165	Sulfitobacter phage ICBM133	ICBM133	NHS	Direct plating	complete	54.967
Sulfitobacter sp. M170	Sulfitobacter phage ICBM134	ICBM134	NHS	Direct plating	complete	54.461
Sulfitobacter sp. M172	Sulfitobacter phage ICBM135	ICBM135	NHS	Direct plating	partial	51.877
Sulfitobacter sp. M176	Sulfitobacter phage ICBM137	ICBM137	NHS	Direct plating	complete	53.554
Sulfitobacter sp. M180	Sulfitobacter phage ICBM138	ICBM138	NHS	Direct plating	complete	52.782
Sulfitobacter sp. M186	Sulfitobacter phage ICBM139	ICBM139	NHS	Direct plating	complete	53.722
Sulfitobacter sp. M194	Sulfitobacter phage ICBM142	ICBM142	NHS	Direct plating	partial	52.186
Sulfitobacter sp. M196	Sulfitobacter phage ICBM143	ICBM143	NHS	Direct plating	complete	52.012
Sulfitobacter sp. M197	Sulfitobacter phage ICBM145	ICBM145	P2	Direct plating	complete	53.772
Sulfitobacter sp. M242	Sulfitobacter phage ICBM146	ICBM146	P2	Direct plating	complete	52.265
Sulfitobacter sp. M260	Sulfitobacter phage ICBM147	ICBM147	P2	Direct plating	complete	51.69
Sulfitobacter sp. M315	Sulfitobacter phage ICBM153	ICBM153	P2	Direct plating	complete	80.756
Sulfitobacter sp. M351	Sulfitobacter phage ICBM154	ICBM154	P2	Direct plating	complete	52.265
Octadecabacter sp. MM282	Octadecabacter phage ICBM156	ICBM156	33	Direct plating	complete	60.762
Lentibacter sp. MPI-62	Lentibacter phage ICBM157	ICBM157	P1	Direct plating	complete	42.72
Lentibacter sp. MPI-62	Lentibacter phage ICBM158	ICBM158	P1	Direct plating	complete	42.906
Lentibacter sp. MPI-62	Lentibacter phage ICBM159	ICBM159	P2	Direct plating	complete	42.936
Lentibacter sp. MPI-62	Lentibacter phage ICBM160	ICBM160	P2	Direct plating	complete	43.138
Lentibacter sp. SH36	Lentibacter phage ICBM163	ICBM163	P2	Direct plating	complete	37.2
Lentibacter sp. SH36	Lentibacter phage ICBM164	ICBM164	P2	Direct plating	complete	43.051
Lentibacter sp. SH36	Lentibacter phage ICBM165	ICBM165	HE504-33	Direct plating	complete	37.385
Lentibacter sp. SH36	Lentibacter phage ICBM166	ICBM166	HE504-33	Direct plating	complete	44.947

4.3.4. The new roseophage isolates belong to 12 genus-level genomic clusters

Calculation of the nucleotide-based intergenomic identities between the 143 assembled phage genomes using VIRIDIC (Moraru et al. 2020) (SI files S4-4 and S4-5) revealed that 15 sequences were 100% identical to others (all *Sulfitobacter* phages; 13 complete, two partial). T hese duplicate phages were removed from the dataset, leaving behind 128 unique novel r oseophage genomes (Table 11). From 146 *Sulfitobacter* phage isolates, 130 were successfull y genome sequenced (resulting in either complete or partial genomes) and 115 of these geno mes turned out to be unique (Table 11). Therefore, one can assume that most duplicate phage i solates had been sorted out successfully. RAPD-PCR proved to be an appropriate method for q uick typing of closely related phages circumventing laborious DNA purification and sequencin g, as it was described before (Comeau et al. 2004; Gutiérrez et al. 2011).

Out of the 128 unique phage genomes, 94 were complete and used for further classification. Clustering based on nucleotide-based intergenomic identity (NBII) using VIRIDIC (Moraru et al. 2020) grouped them into 14 genus-level and 49 species-level clusters, based on the thresholds recommended by ICTV for species (95%) and for genera (70%) (Fig. 21, Table 14). The great majority of the *Sulfitobacter* phage genomes (72 phages) fell into a large cluster comprising the genus-level clusters 9, 10 and 11 (Fig. 21). Within this large cluster, most of the nucleotide-based intergenomic identities are higher the than the genus threshold of 70%, with very few being as low as 68.4%. Thus, we merged genus-level clusters 9, 10 and 11 into one genus cluster (labeled 9-11). The members of this genus cluster were provisionally named here “sulfiviruses”. In this chapter, we will characterize only six sulfiviruses (ICBM13, ICBM16, ICBM21, ICBM23, ICBM25, and ICBM47). A separate chapter of this thesis has been dedicated to the sulfiviruses, in which also the rest of them is further characterized (see chapter 5). The remaining 22 viruses (nine *Sulfitobacter* phages, the twelve *Lentibacter* phages and the *Octadecabacter* phage) were divided into eleven genus-level and 16 species-level clusters (Fig. 21, Table 14). Phages isolated from different host genera clustered together, respectively. Only genus cluster 2 contained two *Sulfitobacter* phages (ICBM121 and ICBM122) and one *Lentibacter* phage (ICBM163). All of these phages were selected for further taxonomic and genomic analysis, while from the sulfivirus cluster only representatives were chosen, as mentioned above (Table 14).

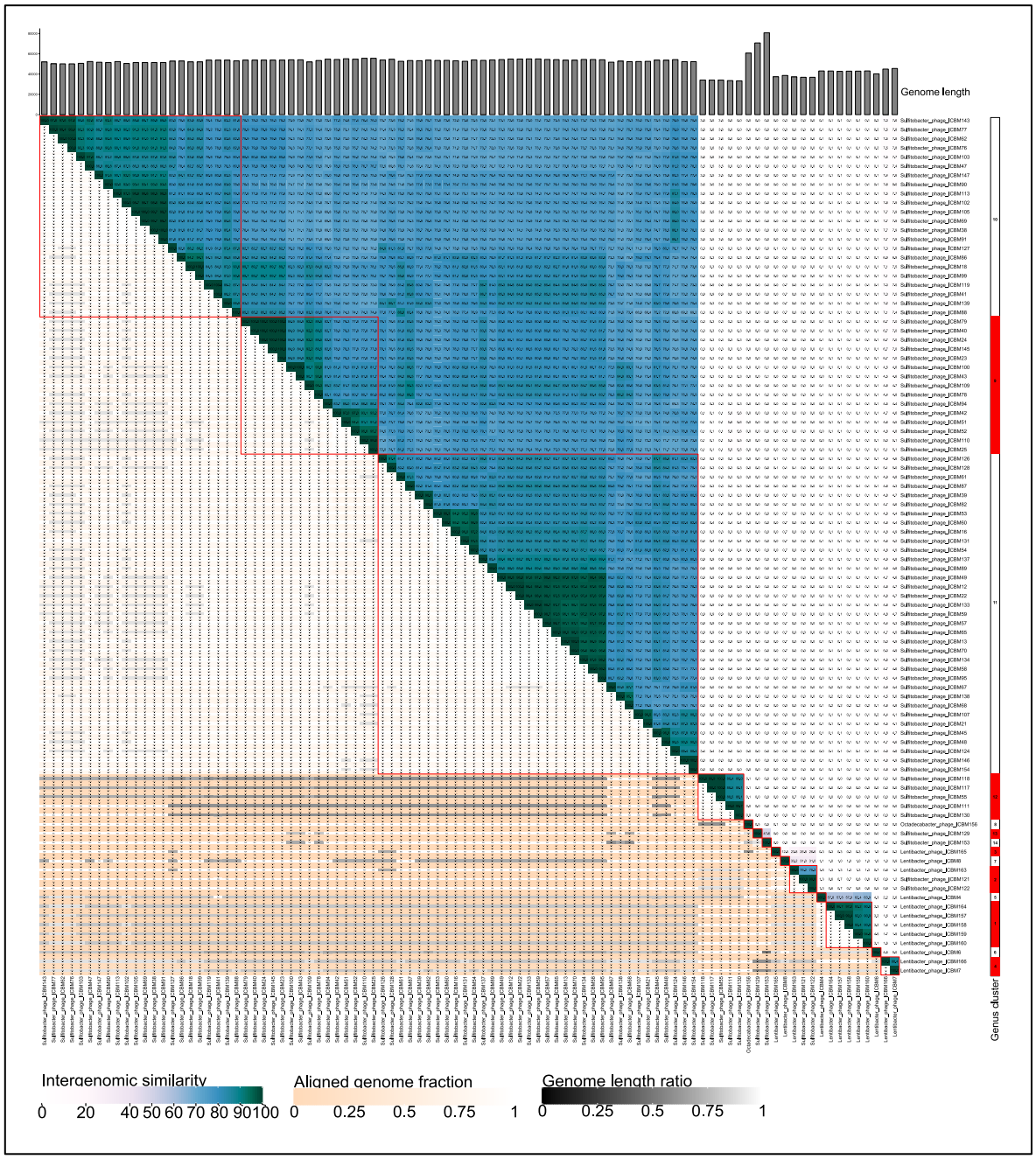


Fig. 21: Nucleotide-based intergenomic identities of 94 unique and complete roseophage genomes, calculated with VIRIDIC. Genus-level clusters suggested by VIRIDIC are indicated by red rectangles and labeled on the right side. (Figure available with higher resolution as SI file S4-6).

Table 14: Clustering of 94 unique and complete roseophage genomes into genera and species clusters. Phages chosen for further taxonomic and genomic analysis are written in bold.

Genus cluster	Species cluster	Newly proposed binomial name	Phage genomes	Host genus
1	1	„Annekevirus eins“	**ICBM157, ICBM158, ICBM164**	*Lentibacter*
	2	„Annekevirus zwei“	**ICBM159**	*Lentibacter*
	3	„Annekevirus drei“	**ICBM160**	*Lentibacter*
2	4	„Benvirus unu“	**ICBM163**	*Lentibacter*
	22	„Benvirus doi“	**ICBM121, ICBM122**	*Sulfitobacter*
3	5	„Martinvirus patru“	**ICBM165**	*Lentibacter*
4	6	„Falkvirus eni“	**ICBM166**	*Lentibacter*

	9	„Falkvirus tva“	**ICBM7**	*Lentibacter*
5	7	„Heinckevirus cuatro“	**ICBM4**	*Lentibacter*
6	8	„Siovirus sase“	**ICBM6**	*Lentibacter*
7	10	„Maryvirus opt“	**ICBM8**	*Lentibacter*
8	11	„Octadecavirus uma“	**ICBM156**	*Octadecabacter*
9-11	12		ICBM43, ICBM100	*Sulfitobacter*
	16		ICBM78, ICBM109	*Sulfitobacter*
	17	„Sulfivirus twee“	**ICBM25**, ICBM52, ICBM110	*Sulfitobacter*
	33	„Sulfivirus een“	**ICBM23**, ICBM24, ICBM40, ICBM79, ICBM145	*Sulfitobacter*
	39		ICBM42, ICBM51	*Sulfitobacter*
	49		ICBM94	*Sulfitobacter*
	13		ICBM38, ICBM69, ICBM91, ICBM102, ICBM105, ICBM113	*Sulfitobacter*
	14	„Sulfivirus drie“	**ICBM47**, ICBM103	*Sulfitobacter*
	20		ICBM41, ICBM119	*Sulfitobacter*
	25		ICBM127	*Sulfitobacter*
	31		ICBM139	*Sulfitobacter*
	32		ICBM62, ICBM76, ICBM77, ICBM143	*Sulfitobacter*
	35		ICBM147	*Sulfitobacter*
	37		ICBM18, ICBM99	*Sulfitobacter*
	44		ICBM86	*Sulfitobacter*
	46		ICBM88	*Sulfitobacter*
	48		ICBM90	*Sulfitobacter*
	15	„Sulfivirus zes“	**ICBM21**, ICBM107	*Sulfitobacter*
	21	„Sulfivirus vif“	ICBM12, **ICBM13**, ICBM22, ICBM49, ICBM57, ICBM58, ICBM59, ICBM65, ICBM70, ICBM95, ICBM133, ICBM134	*Sulfitobacter*
	23		ICBM124	*Sulfitobacter*
	24		ICBM126	*Sulfitobacter*
	26		ICBM128	*Sulfitobacter*
	28	„Sulfivirus vier“	**ICBM16**, ICBM54, ICBM131	*Sulfitobacter*
	29		ICBM137	*Sulfitobacter*
	30		ICBM68, ICBM138	*Sulfitobacter*
	34		ICBM146, ICBM154	*Sulfitobacter*
	38		ICBM39, ICBM82	*Sulfitobacter*
	40		ICBM45, ICBM48	*Sulfitobacter*
	41		ICBM53, ICBM60	*Sulfitobacter*
	42		ICBM61	*Sulfitobacter*
	43		ICBM67	*Sulfitobacter*
	45		ICBM87	*Sulfitobacter*
	47		ICBM89	*Sulfitobacter*
12	18	„Viktorvirus dva“	**ICBM111**, **ICBM130**	*Sulfitobacter*
	19	„Viktorvirus adin“	**ICBM55**, **ICBM117**, **ICBM118**	*Sulfitobacter*
13	27	„Carlotavirus una“	**ICBM129**	*Sulfitobacter*
14	36	„Annevirus trei“	**ICBM153**	*Sulfitobacter*

4.3.5. Taxonomic classification and genomic characterization of selected new phage isolates

Phage classification is a fast evolving discipline, which has undergone many recent changes, and its exact rules are still being considered by the scientific community. Further on, we aimed to place the new isolated roseophages into the taxonomic context of their related phages and all other roseophages from previous studies.

For this whole-genome-based classification approach, in a first step phage genomes related to the new roseophage isolates were collected from the GenBank database using a BLASTP search. Only cultivated phages and only one phage from each genus were kept. Out of the new roseophage isolates, 28 were selected representing all of the detected clusters (Table 14). In addition, the genomes of eleven phages infecting *Sulfitobacter* sp. SW_H+_2_149 (phages Ebeline 1-11) were included into this project, provided by Nina Bartlau and Rudolf Amman (MPI, Bremen). The query dataset was completed by almost all publicly available genomes of dsDNA roseophages that were described in literature at the time of data collection (August 2022, 94 genomes). A first hierarchical clustering using VirClust (Moraru 2023) enabled the reduction of the dataset to those viral genome clusters (VGCs) that contained roseophages. An alignment with the ICTV-recognized taxonomy at that time was made in order to be able to use the classification of existing families as a guide for the following subdivision into VGCs.

The complete dataset comprising 965 phage genomes was further analysed with VirClust, for the purpose of taxonomic classification and protein annotation (Moraru 2023). Initially, proteins were predicted and grouped into protein clusters (PCs) based on their sequence similarity (BLASTP-based). The presence/absence of PCs in the viral genomes was used to determine pairwise intergenomic distances, which were further used for the calculation of a genome tree (Fig. 22, for a more detailed view see Fig. S14, and SI files S4-8 and S4-9). This tree was further split into viral genome clusters (VGC) of the family-level, using a threshold at which most of the ICTV-recognized families were in one VGC.

The roseophages (previous and from this study) were spread all over the tree. Partly, they clustered within already existing and ICTV-recognized viral families. The novel roseophages fell into eight VGCs. Four of these clusters contained already recognized families: *Mesyanzhinoviridae* (VGC_11), *Casjensviridae* (VGC_9), *Zobellviridae* (VGC_8) and *Autographiviridae* (VGC_4). The eleven Ebeline phages infecting *Sulfitobacter* sp.

SW_H+_2_149 clustered closely together in a separate VGC. They are not described in more detail here, as they were not part of our isolation campaign.

In the following subchapters of this thesis, the new roseophages clustering in each of the VGCs will be described in terms of their genomic compositions and characteristics as well as their newly proposed taxonomic classification. Predicted genome termini and DNA replication strategies (Table 15) will be discussed as well as the predicted morphologies and the lifestyle suspected based on gene annotations. Detailed genome content is only shown from representative phages. The complete gene annotations of all ICBM phages can be found in SI file S4-10. Table 16 provides an overview of the newly described and classified phages, summarizing their most important features and the proposed taxonomic classification. A breakdown of the isolation sites of the new roseophages within the assigned families is given in figure 23. In all families that comprise more than one of the new roseophages, these isolates originate from different water samples. In general, most phages were obtained from water samples NHS and P2 (Fig. 23).

To assign the new roseophages to species, genera and (sub-) families, the following tools and thresholds were used. In accordance to the recently compiled guidelines for virus taxonomy (Simmonds et al. 2023), which suggest that genetic relationships of very similar genomes and assignment of the lower rank taxa genus and species should be based on nucleotide or amino acid sequence alignments, we consulted the nucleotide-based pairwise intergenomic identity (NBII) calculated with VIRIDIC (Moraru et al. 2020) for the classification of the new phages on genus and species level. The ICTV-recommended thresholds of 70% for the genus level and 95% for the species level were applied. Simmonds et al. (2023) further suggested determining relationships at the intermediate ranks of family, order and class using protein profile comparisons. In this study, we delineated new potential phage families based on the protein cluster (PC)-based intergenomic similarity calculated with VirClust (Moraru 2023). A threshold of 30% was used, meaning that phages within a family must share more than 30% PC-based intergenomic similarity. New families were only proposed outside already existing families, i.e. if no phages of an ICTV-recognized family were found in the same VGC as our novel roseophages. When the new roseophages clustered together with an existing family, but within a subgroup with PC-based intergenomic similarities higher than 30%, a new subfamily was proposed.

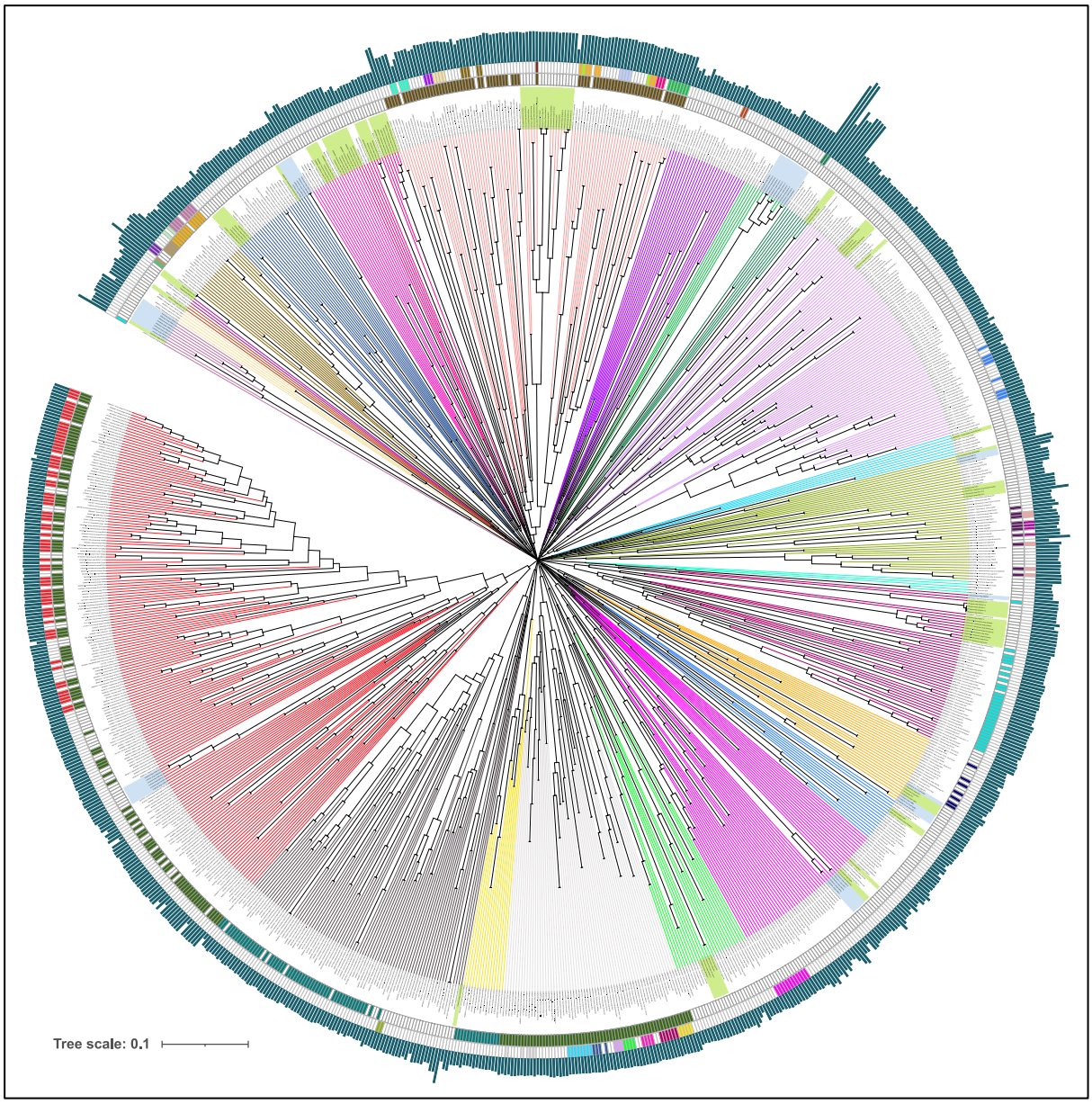


Fig. 22: Whole-genome based proteomic tree of 965 dsDNA phages. Names of cultivated roseophages are marked in green (previous) and blue (this study). Color-strips indicate affiliation to ICTV-recognized families and subfamilies, respectively. Genome lengths are displayed as bar chart. Colored branches indicate viral genome clusters (VGCs). (Figure available with higher resolution as SI file S4-8).

Table 15: PhageTerm results of genome end determination for the representative ICBM phages.

VGC	Phage	Ends	Left	pvalue_left	Right	pvalue_right	Permuted	Orientation	Class	Type
25	Sulfitobacter phage ICBM21	Redundant	32674	1.7E-79	33001	1.55E-115	No	NA	DTR (short)	T7
25	Sulfitobacter phage ICBM13	Redundant	11460	1.45E-86	11788	6.81E-145	No	NA	DTR (short)	T7
25	Sulfitobacter phage ICBM16	Redundant	14465	1.6E-103	14792	1.97E-123	No	NA	DTR (short)	T7
25	Sulfitobacter phage ICBM47	Redundant	36702	1.21E-82	37007	5.13E-68	No	NA	DTR (short)	T7
25	Sulfitobacter phage ICBM23	Redundant	35965	2.39E-22	36294	6.67E-68	No	NA	DTR (short)	T7
25	Sulfitobacter phage ICBM25	Redundant	18266	2.44E-103	18593	1.79E-64	No	NA	DTR (short)	T7
25	Lentibacter phage ICBM7	Redundant	32113	8.29E-116	32221	9.81E-85	No	NA	DTR (short)	T7
25	Lentibacter phage ICBM166	Redundant	40649	2.12E-23	40757	1.18E-10	No	NA	DTR (short)	T7
19	Lentibacter phage ICBM165	Redundant	36112	4.24E-56	36418	1.61E-43	No	NA	DTR (short)	T7
19	Lentibacter phage ICBM8	Redundant	34120	3.15E-38	34440	7.6E-59	No	NA	DTR (short)	T7
19	Lentibacter phage ICBM163	Redundant	27248	1.13E-52	27568	3.89E-10	No	NA	DTR (short)	T7
19	Sulfitobacter phage ICBM121	Redundant	18537	4.99E-55	Distributed		Yes	Forward	Headful (pac)	P1
19	Sulfitobacter phage ICBM122	Redundant	24435	3.34E-57	Distributed		Yes	Forward	Headful (pac)	P1
9	Octadecabacter phage ICBM156	Redundant	Random		Random		Yes		unknown	
11	Sulfitobacter phage ICBM129	Redundant	Random		Random		Yes		unknown	
11	Sulfitobacter phage ICBM153	Redundant	Random		Random		Yes		unknown	
6	Sulfitobacter phage ICBM118	Redundant	20479	9.14E-15	Distributed		Yes	Forward	Headful (pac)	P1
6	Sulfitobacter phageICBM117	Redundant	Distributed		4295	1.05E-24	Yes	Reverse	Headful (pac)	P1
6	Sulfitobacter phageICBM55	Redundant	Random		Random		Yes		unknown	
6	Sulfitobacter phageICBM111	Redundant	22348	1.73E-11	Distributed		Yes	Forward	Headful (pac)	P1
6	Sulfitobacter phageICBM130	Redundant	Distributed		28208	0.000000102	Yes	Reverse	Headful (pac)	P1
8	Lentibacter phage ICBM6	Redundant	369	2.18E-71	538	1.58E-58	No	NA	DTR (short)	T7
4	Lentibacter phage ICBM4	Redundant	Distributed		2744	1.69E-10	Yes	Reverse	Headful (pac)	P1
4	Lentibacter phage ICBM160	Redundant	32449	9.44E-42	32726	5.01E-43	No	NA	DTR (short)	T7
4	Lentibacter phage ICBM164	Redundant	18408	2.04E-08	18685	5.78E-38	No	NA	DTR (short)	T7
4	Lentibacter phage ICBM159	Redundant	17128	6.17E-94	17405	1.04E-50	No	NA	DTR (short)	T7
4	Lentibacter phage ICBM158	Redundant	16342	9.74E-38	16620	2.06E-13	No	NA	DTR (short)	T7
4	Lentibacter phage ICBM157	Redundant	7417	5.65E-21	7695	6.11E-35	No	NA	DTR (short)	T7

Table 16: Overview of the newly described and classified ICBM phages. *predicted by Virfam. **predicted by PhageTerm. ***Lysogenic potential indicated by possession of genes of category integration/excision or lytic/lysogenic decision. RM = restriction-modification system. Abi = abortive infection system.

Viral genome cluster	Phage	Isolation host	Isolation source	Updated (newly proposed + existing) taxonomic classification				Genome length [bp]	G+C content [%]	Predicted morphology*	Predicted genome replication and packaging strategy**	Lysogenic potential***	tRNAs	Special genes
				Family	Subfamily	Genus	Species							
	ICBM4	*Lentibacter* sp. *SH36*	HE440-S			„Heinckevirus“	„cuatro“				Headful (pac) (circularly permuted genome), P1-type rolling circle			
VGC_4	ICBM157	*Lentibacter* sp. MPI-62	P1	*Autographi-viridae*	„Incetivirinae“			42720 - 43138	47.7 - 47.9	podoviral		yes	-	Regulators of chromosome condensation (RCC1)
	ICBM158						„eins“							
	ICBM164	*Lentibacter* sp. *SH36*	P2			„Annekevirus“					DTR (short), T7-type bidirectional			
	ICBM159	*Lentibacter* sp. MPI-62					„zwei“							
	ICBM160						„drei“							
VGC_9	ICBM156	*Octadecabacter* sp. MM282	HE504-33	*Casjensviridae*		„Octadecavirus“	„uma“	60762	61.7	siphoviral	unknown	yes	-	
VGC_11	ICBM129	*Sulfitobacter* sp. M92	NHS	*Mesyanzhino-viridae*	„Maresulfi-virinae“	„Carlotavirus“	„una“	70738 - 80756	57.7 - 61.5	siphoviral	unknown	yes	-	Que biosynthesis genes (anti restriction)
	ICBM153	*Sulfitobacter* sp. M315	P2			„Annevirus“	„trei“						7	
	vB_DshS_R5C	*Dinoroseobacter shibae* DFL12T				*Nanhaivirus*	*D5C*							

VGC_25	ICBM23	*Sulfitobacter* sp. M55	P2	"Hayaniviridae"	"Sulfivirus"	"een"	52607-55591	44.6 - 45.0	siphoviral	DTR (short), T7-type bidirectional	no	-	Methylases (anti RM)
	ICBM25	*Sulfitobacter* sp. M63	NHS			"twee"						1	
	ICBM47	*Sulfitobacter* sp. M86	P4			"drie"						-	
	ICBM16	*Sulfitobacter* sp. M53	P1			"vier"						1	
	ICBM13	*Sulfitobacter* sp. M45	P2			"vif"							
	ICBM21	*Sulfitobacter* sp. M53	NHS			"zes"							
	ICBM166	*Lentibacter* sp. SH36	HE504-33	"Schlingloff-viridae"	"Falkvirus"	"eni"	44927 - 45550	47.0 - 47.4				-	SaV-like protein (anti Abi), methylase (anti RM)
	ICBM7		HE440-S			"tva"						-	
	CRP-6	*Roseobacter* strain FZCC0042											
VGC_19	ICBM163	*Lentibacter* sp. SH36	P2	"Diferite-viridae"	"Benvirus"	"unu"	36856 - 38666	55.9 - 57.3	podoviral	DTR (short), T7-type bidirectional	yes	-	Methylases (anti RM), PAPS reductase (AMG) (only in "Benvirus" genus)
	ICBM121	*Sulfitobacter* sp. M73				"doi"				Headful (pac) (circularly permuted genome), P1-type rolling circle			
	ICBM122												
	ICBM165	*Lentibacter* sp. SH36	HE504-33		"Martinvirus"	"patru"				DTR (short), T7-type bidirectional	no		
	ICBM8				"Maryvirus"	"opt"					yes		

	ICBM55	*Sulfitobacter* sp. M157										unknown				
	ICBM117	*Sulfitobacter* sp. M70	P2				„adin“									Methylase (anti RM)
	ICBM118											Headful (pac) (circularly permuted genome), P1-type rolling circle	yes	-		
VGC_6	ICBM111	*Sulfitobacter* sp. M63		„Woolley-viridae“	„Viktorvirus“			„dva“	33346 - 34114	60.5 - 60.7	siphoviral					
	ICBM130	*Sulfitobacter* sp. M92	NHS													
VGC_8	ICBM6	*Lentibacter* sp. SH36	HE396-6	*Zobellviridae*	*Cobavirinae*	*Siovirus*	„sase“		40273	46.3	podoviral	DTR (short), T7-type bidirectional	no	-	Methylase (anti RM)	

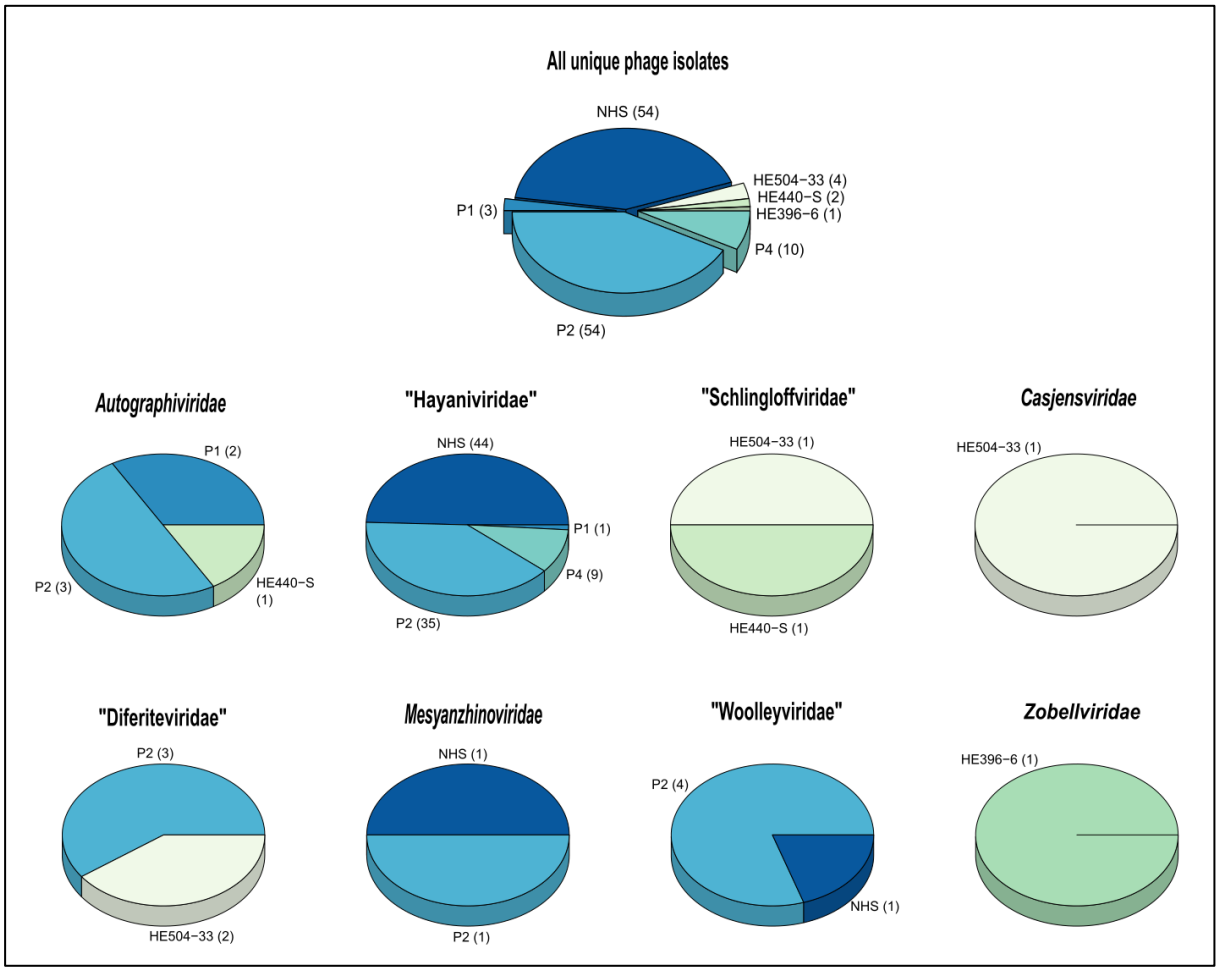


Fig. 23: Origins of the new phage isolates from different North Sea water samples, from the shore (NHS, July 2018), a mesocosm experiment (P1, P2 and P4, March 2018), and from the open sea (HE504-33 (2018), HE396-6 (2013) and HE440-S (2015)). Pie chart on the top shows origins of all unique phage isolates. Below, origins of the new ICBM phages within each newly proposed or existing family are shown. Total numbers of phage isolates are written in parentheses.

4.3.5.1. First roseophage isolates to cluster within the *Autographiviridae* family

In the whole-genome based proteomic tree, one of the largest VGCs containing roseophages corresponded to the *Autographiviridae* family (Fig. 24, SI files S4-9 and S4-11). Most of the previously characterized *Autographiviridae* from this VGC_4 infect bacteria from *Cyanophyceae*, *Gammaproteobacteria*, *Betaproteobacteria*, *Desulfovibrionia*, and *Alphaproteobacteria (Hyphomicrobiales, Pelagibacterales, Hyphomicrobiales, Caulobacterales*). Viruses in this VGC with known morphology are podoviruses (Table 17).

The new roseophage isolates grouping into *Autographiviridae* formed a separate cluster, with a maximum PC-based intergenomic similarity with other *Autographiviridae* of 18% (Fig. 26). Within this roseophage cluster, the minimum PC-based intergenomic similarity was of 78% (Fig. 26). Therefore, we propose here that this cluster forms a new subfamily within the *Autographiviridae*, which we tentatively named here "Incetivirinae", from the Romanian word "incet" meaning "slow", in accordance to the derivation of the host genus name *Lentibacter* from the latin word "lentus" meaning "slow". We expect that this subfamily will form a family on its own, once the *Autographiviridae* classification will be overhauled.

The six novel roseophages within this family were phages ICBM157, ICBM158, ICBM159, and ICBM160 infecting *Lentibacter* sp. MPI-62 and phages ICBM164 and ICBM4 infecting *Lentibacter* sp. SH36 (Table 17). They have been isolated from three different water samples (HE440-S, P2 and P1). According to their nucleotide-based intergenomic identity, they formed two new genera. While phage ICBM4 formed its own genus, the other five viruses showed high intergenomic identity to each other and clustered into three species within one genus (Fig. 25). We tentatively named these two genera "Heinckevirus" (ICBM4), from the research vessel Heincke, and "Annekevirus" (ICBM157, ICBM158, ICBM159, ICBM160, and ICBM164), after Anneke Heins (MPI, Bremen), who provided the host strain *Lentibacter* sp. MPI-62.

The new roseophages had genomes of 42.7 - 43.1 kb in size, 47.7 - 47.9% G+C content and 45 - 49 predicted ORFs. Functional annotation was possible for 27 of these ORFs. Phages of the "Annekevirus" genus all had short DTRs of 278 - 279 bp at the genome ends, identicating a T7-type DNA packaging technique (Table 15). In contrast, the PhageTerm results for Lentibacter phage ICBM4 were not as conclusive. A circularly permuted genome and a headful packaging strategy were suggested. However, the typical coverage pattern was missing (Garneau et al. 2017). The six phages showed an identical genome architecture and all but one annotated genes were found in all genomes (Fig. 27, SI file S4-10). Genes were arranged in

subsequent functional modules (Fig. 27, Table 18). In the DNA replication module, nine genes could be annotated: a ssDNA binding protein (PC_123), a RusA-like Holliday junction resolvase (PC_124), a bifunctional DNA primase/helicase (PC_88), a HNH endonuclease (PC_126), a DNA polymerase I (PC_89), an endo/exo-ribonuclease (PC_127), a ribonucleotide reductase (PC_91) and a thymidylate synthase ThyX (PC_1). In addition, a DNA directed RNA-polymerase was encoded (PC_122), which is a hallmark of the *Autographiviriae* family. The second half of the genome harboured genes for virion morphology. Those responsible for capsid structure and packaging included a capsid assembly and scaffolding protein (PC_132), two internal virion proteins (PC_138 and PC_136), the major capsid protein (PC_133) and the terminase large subunit (PC_145). Furthermore, a head-tail connecting protein (PC_131), a tail fiber protein (PC_139), tail tubular protein A (PC_134) and tail tubular protein B (PC_135) were annotated. Three lysis genes coded for an endolysin (PC_95), a holin (PC_142) and a protein with a lysozyme-domain (PC_202) (Fig. 27, Table 18). The latter gene (PC_202) was missing in the genome of phage ICBM4. In addition, a nucleotide pyrophosphohydrolase (PC_128) and an acetyl transferase (GNAT) (PC_111) have been annotated. Two genes annotated as "regulator of chromosome condensation (RCC1)" (PC_140 and PC_141) were found in the genomes. Such proteins have been frequently described in eukaryotes, regulating DNA condensation (Ohtsubo et al. 1989). So far, they have been found only in the two related Jumbo phages, Chronobacter phage vB_CsaM_GAP32 and Escherichia coli virus phAPEC6 (Abbasifar et al. 2014; Wagemans et al. 2020). In neither of the genomes were any tRNAs predicted.

The detection of DTRs and the annotation of a T7-type DNA polymerase I suggested that the new roseophages use a T7-like bidirectional replication and DNA packaging technique (Tables 15 and 18). Furthermore, Virfam prediction and gene annotation (absence of major tail protein, tail completion protein and sheath), indicated a podoviral morphology with a short tail (Tables 17 and 18). A Lambda repressor-like gene was annotated in the genomes. In bacteriophage Lambda, this protein regulates viral gene expression as part of the lytic/lysogenic decision (Maniatis and Ptashne 1973). Together with the annotation of an integrase gene, this indicated the potential for a lysogenic lifestyle of these phages (Table 18). The phages of the "Incetivirinae" subfamily shared 37 core proteins, comprising 26 out of 27 annotated proteins and 11 hypothetical proteins (Table 19).

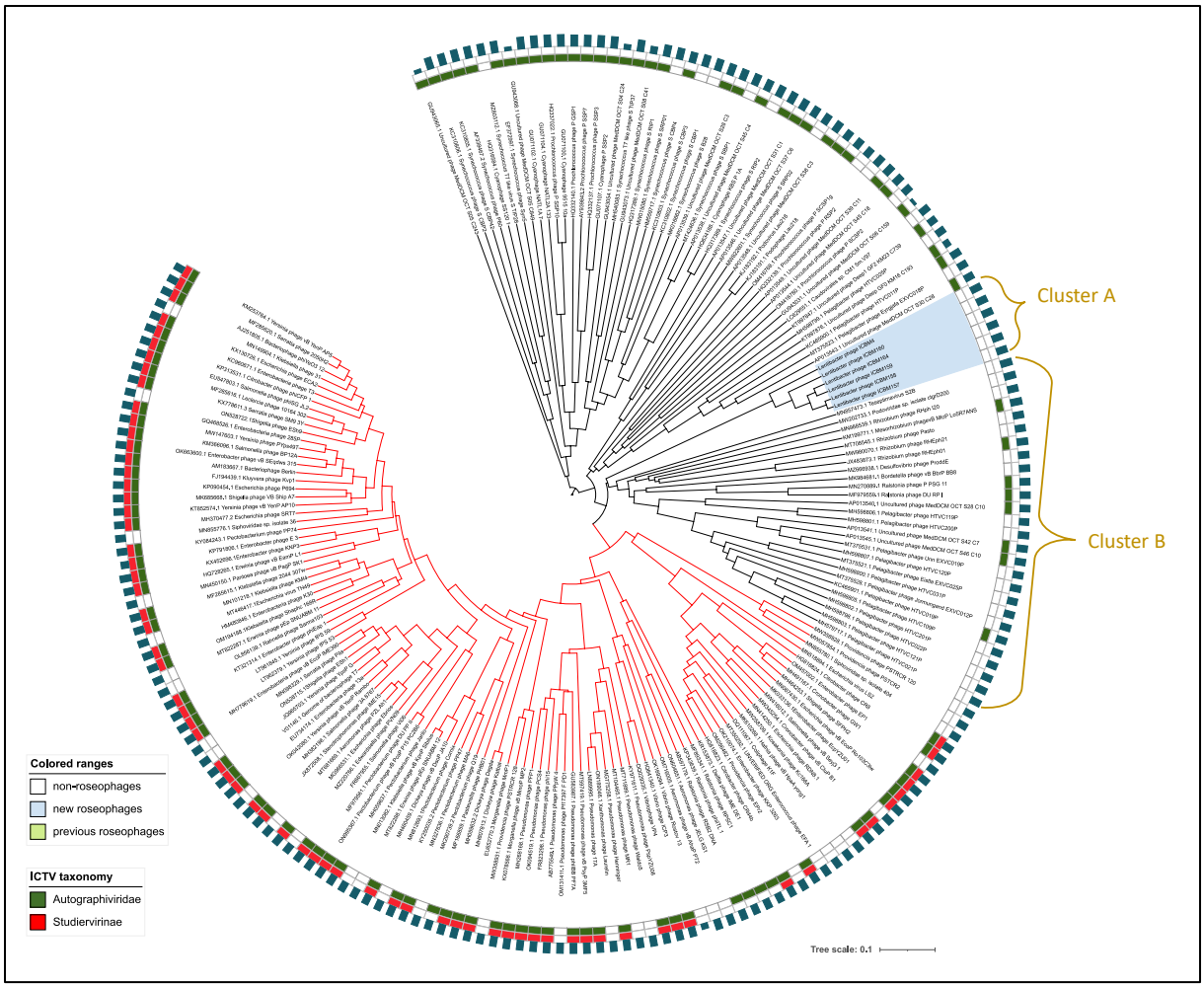


Fig. 24: Section of the whole-genome based proteomic tree showing the *Autographiviridae* family. Names of cultivated roseophages are marked in green (previous) and blue (this study). Color-strips indicate affiliation to ICTV-recognized families and subfamilies, respectively. Genome lengths are displayed as bar chart. Branches of the *Studiervirinae* subfamily are marked in red.

Table 17: Phages of cluster A and B in the *Autographiviridae* family. Roseophages are marked in blue. Bp = base pairs. n.a. = not available. *Morphology predicted by Virfam.

Cluster	Phage	Host	Genome size [bp]	GC content [%]	Accession	Previous classification (reference)	ICTV or newly proposed taxonomy			Morphology	Isolation source	Reference
							Family	Genus	Species			
A	Lentibacter phage ICBM4	*Lentibacter* sp. SH36	43101	47.80	this study		*Autographiviridae*, „Incetivirinae“	„Heinckevirus“	„cuatro“	podoviral*	seawater (HE440-S)	this study
	Lentibacter phage ICBM160	*Lentibacter* sp. MPI-62	43138	47.70	this study			„Annekevirus“	„drei“	podoviral*	seawater (P2)	this study
	Lentibacter phage ICBM164	*Lentibacter* sp. SH36	43051	47.90	this study				„eins“	podoviral*	seawater (P2)	this study
	Lentibacter phage ICBM159	*Lentibacter* sp. MPI-62	42936	47.80	this study				„zwei“	podoviral*	seawater (P2)	this study
	Lentibacter phage ICBM158	*Lentibacter* sp. MPI-62	42906	47.80	this study				„eins“	podoviral*	seawater (P1)	this study
	Lentibacter phage ICBM157	*Lentibacter* sp. MPI-62	42720	47.80	this study					podoviral*	seawater (P1)	this study
B	Teseptimavirus S2B	*Caulobacter vibrioides* CB15	45682	66.80	MN857473					podoviral	freshwater river	Ely et al. (2022)
	Podoviridae sp. isolate ctgrD200	n.a.	45384	65.30	MW202733						freshwater	unpublished
	Rhizobium phage RHph I20	*Rhizobium* sp. N324	41785	60.10	MN988539						agricultural lands	unpublished
	Mesorhizobium phagevB MloP Lo5R7ANS	*Mesorhizobium loti* R7ANS	45718	61.10	KM199771		*Autographiviridae*	*Pairvirus*			n.a.	unpublished
	Rhizobium phage Pasto	*Bradyrhizobium japonicum*	42407	58.60	MT708545					podoviral	rhizosphere, agriculture	Manuel et al. (2021)
	Rhizobium phage RHEph21	*Rhizobium gallicum* sv. phaseoli R72	40454	58.60	MW980070						rhizosphere, agriculture	unpublished
	Rhizobium phage RHEph01	*Rhizobium etli* Brasil5	43444	59.20	JX483873		*Autographiviridae*	*Paadamvirus*		podoviral	rhizosphere, agriculture	Santamaría et al. (2014)
	Desulfovibrio phage ProddE	*Desulfovibrio desulfuricans* strain Edelweiss	42637	51.40	MZ666938						freshwater river sediment	Boeckman et al. (2022)
	Bordetella phage vB BbrP BB8	*Bordetella bronchiseptica*	41593	58.80	MK984681					podoviral	wastewater	Szymczak et al. (2020)
	Ralstonia phage P PSG 11	*Ralstonia solanacearum*	40313	58.80	MN270889						freshwater river	Wei et al. (2017)

	Ralstonia phage DU RP I	*Ralstonia solanacearum*	40657	58.80	MF979559		*Autographiviridae*	*Gyeongsanvirus*		n.a.	unpublished
	Uncultured phage MedDCM OCT S28 C10	n.a.	40866	32.70	AP013540		*Autographiviridae*	*Votkovvirus*		seawater	Mizuno et al. (2013)
	Pelagibacter phage HTVC119P	Candidatus *Pelagibacter* sp. HTCC7211	38357	32.00	MH598806	HTVC019P-related podoviridae (Zhao et al. 2018)				seawater	Zhao et al. (2018)
	Pelagibacter phage HTVC200P	Candidatus *Pelagibacter* sp. FZCC0015	42221	33.20	MH598801					seawater	Zhao et al. (2018)
	Uncultured phage MedDCM OCT S42 C7	n.a.	42399	33.90	AP013541		*Autographiviridae*	*Sieqvirus*		seawater	Mizuno et al. (2013)
	Uncultured phage MedDCM OCT S46 C10	n.a.	41803	33.80	AP013545		*Autographiviridae*	*Foussvirus*		seawater	
	Pelagibacter phage Unn EXVC019P	Candidatus *Pelagibacter* sp H2P3alpha	41069	33.50	MT375531	podoviridae				seawater	Buchholz et al. (2021)
B	Pelagibacter phage HTVC120P	Candidatus *Pelagibacter* sp. HTCC7211	42622	33.30	MH598807	HTVC019P-related podoviridae (Zhao et al. 2018)				seawater	Zhao et al. (2018)
	Pelagibacter phage Eistla EXVC025P	Candidatus *Pelagibacter ubique* HTCC1062	39638	32.70	MT375521	podoviridae				seawater	Buchholz et al. (2021)
	Pelagibacter phage HTVC031P	Candidatus *Pelagibacter ubique* HTCC1062	41046	33.40	MH598800	HTVC019P-related podoviridae (Zhao et al. 2018)				seawater	Zhao et al. (2018)
	Pelagibacter phage Jormungand EXVC012P	Candidatus *Pelagibacter* sp H2P3alpha	41529	34.10	MT375528	podoviridae				seawater	Buchholz et al. (2021)
	Pelagibacter phage HTVC019P	Candidatus *Pelagibacter ubique* HTCC1062	42084	34.00	KC465901		*Autographiviridae*	*Pelagivirus*	podoviral	seawater	Zhao et al. (2013)

	Pelagibacter phage HTVC109P	Candidatus *Pelagibacter* sp. HTCC7211	41323	35.50	MH598805		seawater	
	Pelagibacter phage HTVC201P	Candidatus *Pelagibacter* sp. FZCC0015	41415	33.10	MH598802		seawater	
B	Pelagibacter phage HTVC022P	Candidatus *Pelagibacter ubique* HTCC1062	42010	34.20	MH598798	HTVC019P-related podoviridae (Zhao et al. 2018)	seawater	Zhao et al. (2018)
	Pelagibacter phage HTVC121P	Candidatus *Pelagibacter* sp. HTCC7211	42600	33.50	MH598803		seawater	
	Pelagibacter phage HTVC021P	Candidatus *Pelagibacter ubique* HTCC1062	42809	33.60	MH579717		seawater	

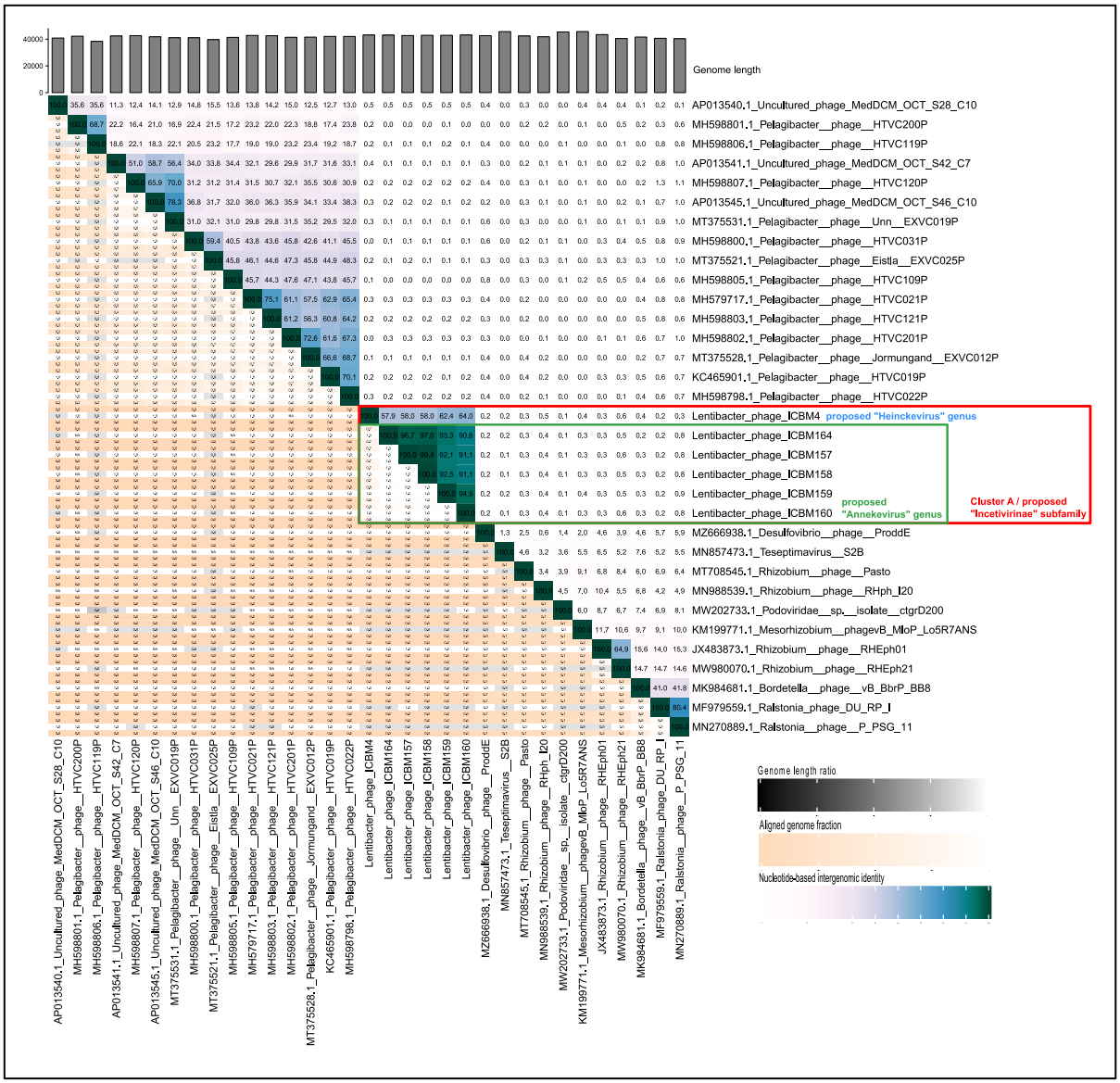


Fig. 25: Nucleotide-based intergenomic identities of phages in clusters A and B of the *Autographiviridae* family (VGC_4), calculated with VIRIDIC. Members of the newly proposed genera "Heinkevirus" and "Annekevirus" and the newly proposed subfamily "Incetivirinae" are annotated with boxes and colored labels. (The VIRIDIC heatmap and cluster table for the complete VGC_4 can be found in SI files S4-11 and S4-12).

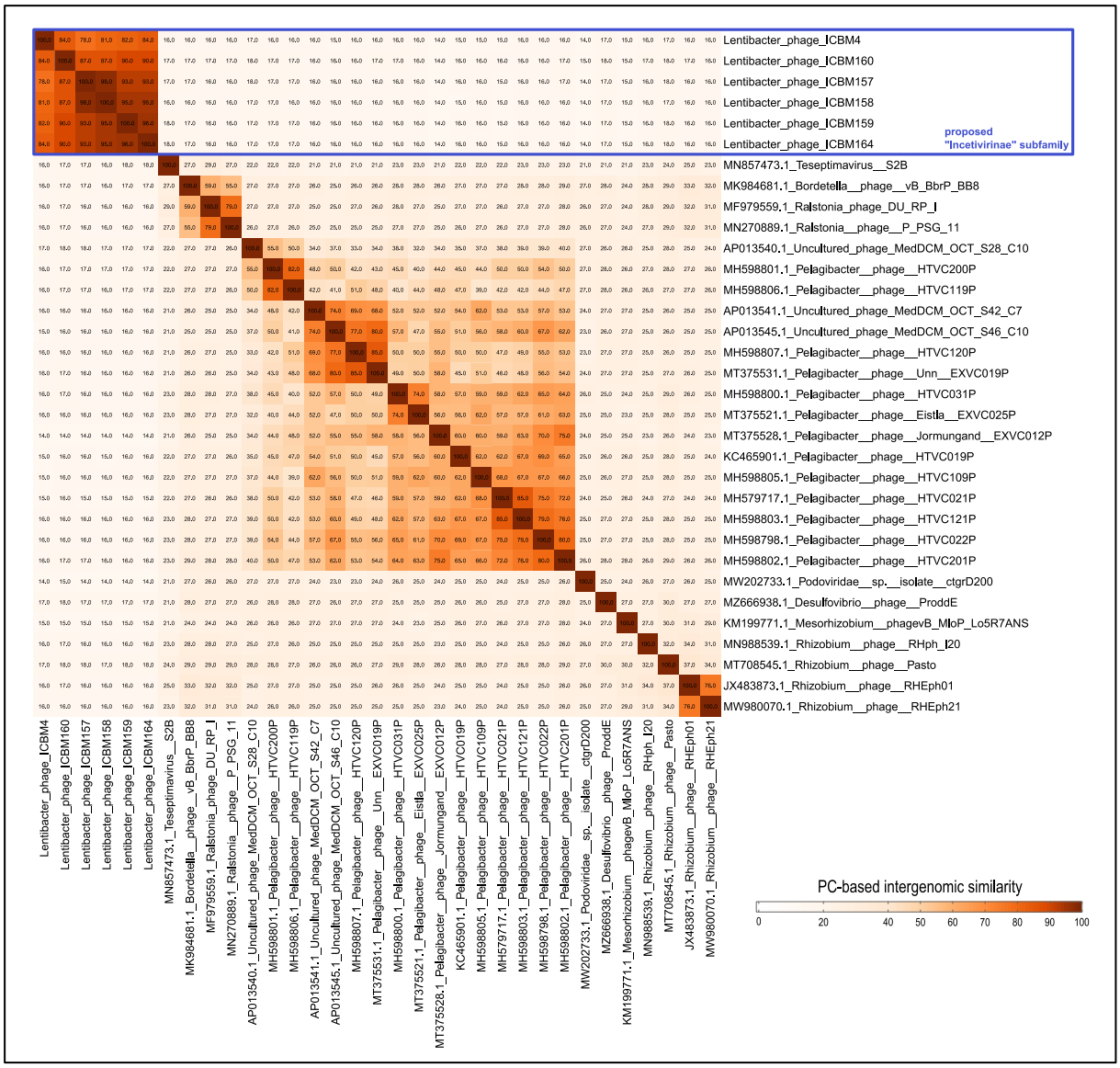


Fig. 26: PC-based intergenomic similarities of phages in clusters A and B of the *Autographiviridae* family (VGC_4), calculated with VirClust (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%). Members of the newly proposed genera “Heinkevirus” and “Annekevirus” and the newly proposed subfamily “Incetivirinae” are annotated with boxes and colored labels.

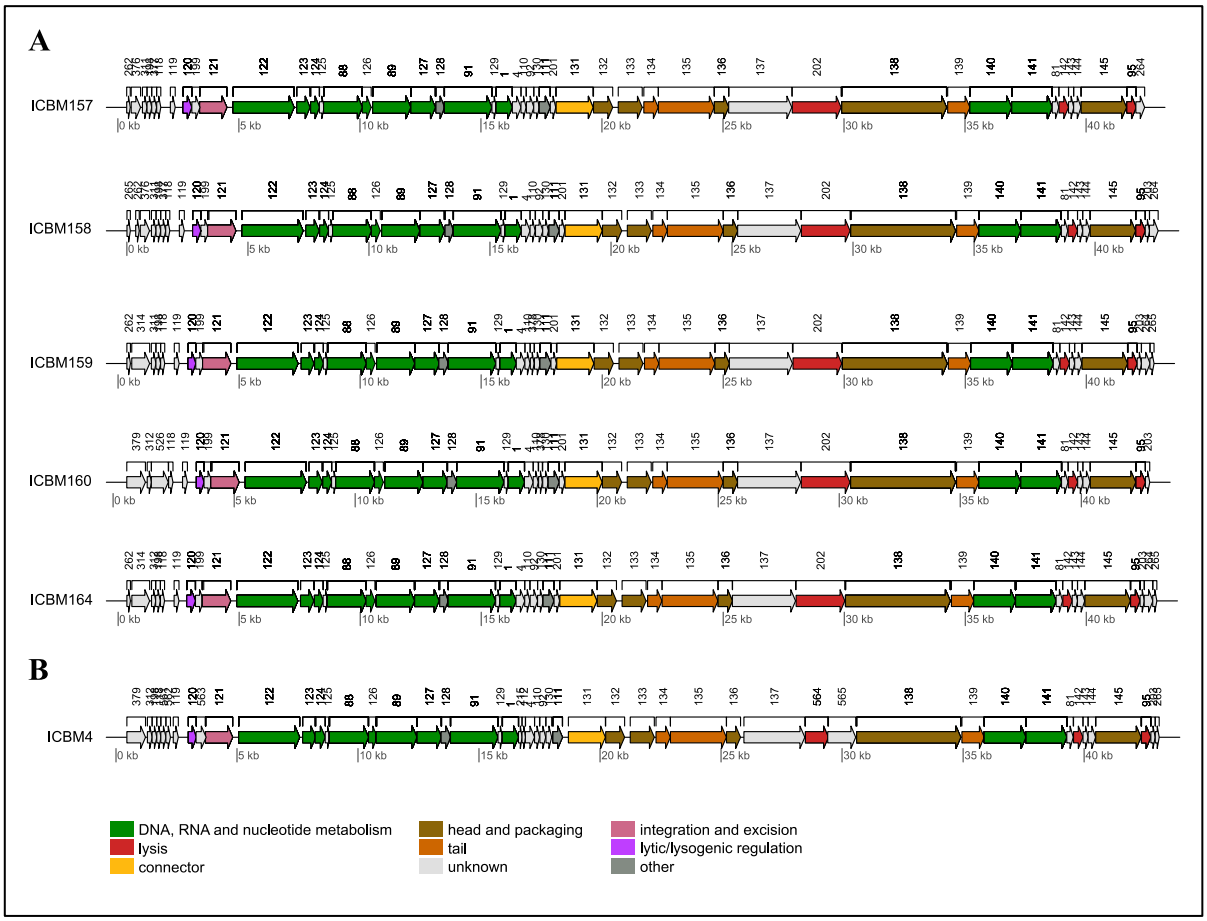


Fig. 27: Genome map of members of the *Autographiviridae* family. **A** Members of the "Annekevirus" genus (*Lentibacter* phages ICBM157, ICBM158, ICBM159, ICBM160 and ICBM164). **B** Lentibacter phage ICBM4 (genus "Heinckevirus").

Table 18: Gene annotations of Lentibacter phage ICBM157. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core proteins of phages in the "Incetiviridae" family.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	363	509	147	1	262	hp	unknown
gene_2	550	936	387	1	376	hp	unknown
gene_3	1014	1205	192	1	311	hp	unknown
gene_4	1202	1396	195	1	198	hp	unknown
gene_5	1397	1600	204	1	377	hp	unknown
gene_6	1602	1772	171	1	118*	hp*	unknown
gene_7	2167	2370	204	1	119*	hp*	unknown
gene_8	2694	3053	360	1	120*	Lambda repressor-like*	lytic/lysogenic regulation
gene_9	3053	3391	339	1	199	hp	unknown
gene_10	3381	4502	1122	1	121*	Integrase*	integration and excision
gene_11	4761	7277	2517	1	122*	DNA directed RNA-polymerase*	DNA, RNA and nucleotide metabolism
gene_12	7393	7938	546	1	123*	ssDNA binding protein*	DNA, RNA and nucleotide metabolism
gene_13	7964	8329	366	1	124*	RusA-like Holliday junction resolvase*	DNA, RNA and nucleotide metabolism
gene_14	8326	8523	198	1	125*	hp*	unknown

gene_15	8498	10099	1602	1	88*	DNA primase/helicase*	DNA, RNA and nucleotide metabolism
gene_16	10099	10467	369	1	126*	HNH endonuclease*	DNA, RNA and nucleotide metabolism
gene_17	10526	12100	1575	1	89*	DNA polymerase I*	DNA, RNA and nucleotide metabolism
gene_18	12109	13116	1008	1	127*	endo/exo-ribonuclease*	DNA, RNA and nucleotide metabolism
gene_19	13142	13492	351	1	128*	nucleotide pyrophosphohydrolase*	other
gene_20	13489	15447	1959	1	91*	ribonucleotide reductase*	DNA, RNA and nucleotide metabolism
gene_21	15452	15640	189	1	129*	hp*	unknown
gene_22	15627	16298	672	1	1*	thymidylate synthase ThyX*	DNA, RNA and nucleotide metabolism
gene_23	16295	16657	363	1	4*	hp*	unknown
gene_24	16654	16863	210	1	110*	hp*	unknown
gene_25	16879	17187	309	1	92	hp	unknown
gene_26	17180	17416	237	1	130*	hp*	unknown
gene_27	17409	17861	453	1	111*	acetyl transferase (GNAT)*	other
gene_28	17864	18091	228	1	201	hp	unknown
gene_29	18104	19636	1533	1	131*	head-tail connecting protein*	connector
gene_30	19649	20443	795	1	132*	capsid assembly and scaffolding protein*	head and packaging
gene_31	20673	21665	993	1	133*	major capsid protein*	head and packaging
gene_32	21728	22324	597	1	134*	tail tubular protein A*	tail
gene_33	22328	24637	2310	1	135*	tail tubular protein B*	tail
gene_34	24640	25236	597	1	136*	internal virion protein*	head and packaging
gene_35	25233	27842	2610	1	137*	hp*	unknown
gene_36	27864	29879	2016	1	202	Lysozym-domain	lysis/tail
gene_37	29906	34234	4329	1	138*	internal virion protein D*	head and packaging
gene_38	34281	35204	924	1	139*	tail fiber protein*	tail
gene_39	35201	36922	1722	1	140*	Regulator of chromosome condensation, RCC1*	DNA, RNA and nucleotide metabolism
gene_40	36932	38605	1674	1	141*	Regulator of chromosome condensation, RCC1*	DNA, RNA and nucleotide metabolism
gene_41	38605	38874	270	1	81*	hp*	unknown
gene_42	38884	39276	393	1	142*	Holin*	lysis
gene_43	39273	39491	219	1	143*	hp*	unknown
gene_44	39488	39787	300	1	144*	hp*	unknown
gene_45	39797	41674	1878	1	145*	terminase large subunit*	head and packaging
gene_46	41685	42074	390	1	95*	endolysin (L-alanyl-D-glutamate peptidase)*	lysis
gene_47	42058	42420	363	1	264	hp	unknown

Table 19: Core proteins of the newly proposed subfamily "Incetivirinae" in the family *Autographiviridae*. *maximum number of predicted ORFs divided by number of core proteins.

predicted ORFs#	core proteins #	Annotated core proteins #	Annotated core proteins (sorted by functional category)	Percentage core proteins / total ORFs*
45-49	37	26	DNA, RNA and nucleotide metabolism: DNA directed RNA-polymerase, (PC_122), ssDNA binding protein (PC_123), RusA-like Holliday junction resolvase (PC_124), DNA primase/helicase (PC_88), HNH endonuclease (PC_126), DNA polymerase I (PC_89), endo/exo-ribonuclease (PC_127), ribonucleotide reductase (PC_91), thymidylate synthase ThyX (PC_1), Regulator of chromosome condensation, RCC1 (PC_140 and PC_141), Head and packaging: capsid assembly and scaffolding protein (PC_132), major capsid protein (PC_133), internal virion protein (PC_136 and PC_138), terminase large subunit (PC_145), Connector: head-tail connecting protein (PC_131), Tail: tail tubular protein A (PC_134), tail tubular protein B (PC_135), tail fiber protein (PC_139), Lysis: endolysin (L-alanyl-D-glutamate peptidase) (PC_95), holin (PC_142) Lytic/lysogenic regulation: Lambda repressor-like protein (PC_120), Integration and excision: Integrase (PC_121), Other: nucleotide pyrophosphohydrolase (PC_128), acetyl transferase (GNAT) (PC_111)	75.51%

4.3.5.2. A novel roseophage infecting *Octadecabacter* falls within the *Casjensviridae* family

In the whole-genome based proteomic tree, the new roseophage ICBM156 clustered in a VGC together with phages of the ICTV-recognized *Casjensviridae* family (VGC_9) (Fig. 28). The previously characterized *Casjensviridae* in this VGC infect bacteria from *Gammaproteobacteria*, *Betaproteobacteria*, *Cyanophyceae* and *Alphaproteobacteria* (*Loktanella*). The have been isolated from diverse habitats, such as seawater, freshwater lakes, sewage, soil and agricultural samples. All viruses with confirmed morphology are siphoviruses (Table 20).

Phage ICBM156 distantly grouped into a separate cluster, together with other marine (roseo-) phages (Fig. 28, Table 20). It comprised eight *Rhodobacter* phages, five *Ruegeria* phages, Loktanella phage pCB2051_A, Synechococcus virus S_ESS1, Rhizobium phage RHph_X2_30 and our phage ICBM156. The only previously assigned *Casjensviridae* were Loktanella phage pCB2051_A and Synechococcus virus S_ESS1. These two phages shared more than 30% PC-based intergenomic similarity with phage ICBM156 (Fig. 30), but phage pCB2051_A clustered further apart in the proteomic tree (Fig. 28). Nevertheless, ICBM156, S_ESS1 and pCB2051_A could potentially belong to one subfamily. Phage ICBM156 was

isolated from seawater sample HE504-33, infecting *Octadecabacter* sp. MM282. With the maximum nucleotide-based intergenomic identity to other phages accounting for 29.3%, Octadecabacter phage ICBM156 formed its own genus (Fig. 29). We tentatively named it here the "Octadecavirus" genus. We propose here Octadecabacter phage ICBM156 as member of the *Casjensviridae* family, in the new "Octadecavirus" genus. However, the PC-based intergenomic similarities (Fig. 30) indicate that this family might be split into smaller families in future classification procedures.

Octadecabacter phage ICBM156 had a genome of ~60.8 kb and a G+C content of 61.7%. It encoded 76 ORFs, 34 of which could be annotated. Genome termini detection using PhageTerm (Garneau et al. 2017) revealed that the genome of Octadecabacter phage ICBM156 had random ends (Table 15). The genome was organized in different modules, with the "DNA, RNA and nucleotide metabolism" module being split by the chosen genome ends (Fig. 31). The genes of this module comprised two DNA helicases (PC_629 and PC_687), a DNA polymerase (PC_626), a DNA topoisomerase (PC_669), an exonuclease (PC_691), a ribonucleotide reductase (PC_355), a dihydrofolate reductase (PC_664), a ssDNA binding protein (PC_624), a thymidylate synthase ThyX (PC_1) and a tRNA endonuclease (PC_628) (Table 21). "Head and packaging" genes included the capsid decoration protein (PC_635), the capsid maturation protease (PC_634), the major capsid protein (PC_636), the portal protein (PC_633) and the terminase small and large subunit (PC_630 and PC_631). Three "connectors" were encoded in the genome: head-tail adaptor protein Ad1 (PC_632), neck protein Ne1 (PC_639) and tail completion protein Tc1 (PC_640). Moreover, five tail proteins (PC_641, PC_642, PC_643, PC_644, PC_648 and PC_649), two tail assembly proteins (PC_353 and PC_645), two tail assembly chaperons (PC_646 and PC_647) and the tail length tape measure protein (PC_644) were annotated. One lysis gene, an endolysin was found (PC_354). In addition, the genome encoded a cysteine dioxygenase (PC_662) and a protease (PC_650). No tRNAs could be predicted. No genome replication and packaging strategy could be predicted by PhageTerm (Table 15). Furthermore, the annotated DNA polymerase was of no specific type. Virfam predicted a siphoviral morphology for ICBM156, which was also reflected by the annotation of the tail tape measure protein and the absence of a tail sheath (Tables 20 and 21). The possession of a gene coding for a Lambda repressor-like protein indicated that phage ICBM156 might have a lysogenic potential.

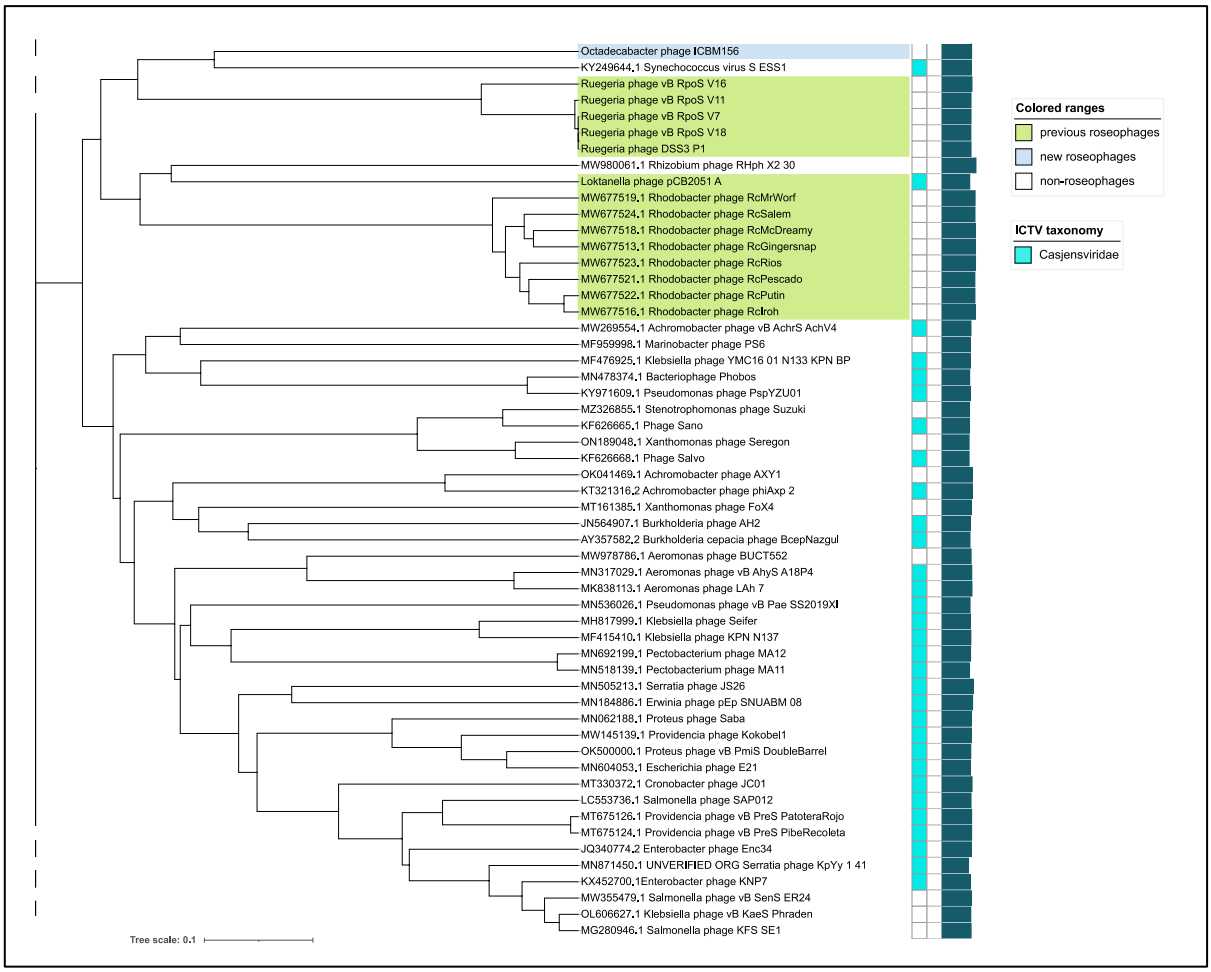


Fig. 28: Section of the whole-genome based proteomic tree showing VGC_9. Names of cultivated roseophages are marked in green (previous) and blue (this study). Color-strips indicate affiliation to ICTV-recognized families and subfamilies, respectively. Genome lengths are displayed as bar chart.

Table 20: Phages of VGC_9. Roseophages are marked in blue. Bp = base pairs. N.a. = not available. *Morphology predicted by Virfam.

Phage	Host	Genome size [bp]	GC content [%]	Accession	Previous classification (reference)	ICTV or newly proposed taxonomy			Morphology	Isolation source	Reference
						Family	Genus	Species			
Octadecabacter phage ICBM156	*Octadecabacter* sp. MM282	60762	61.70%	this study			„Octadecavirus“	„uma“	siphoviral*	Seawater (HE504-33)	this study
Synechococcus virus S_ESS1	*Synechococcus* sp.	60362	60.90%	KY249644		*Casjensviridae*	*Sessunavirus*		siphoviral	coastal seawater	Han et al. (2017)
Ruegeria phage vB_RpoS_V16	*Ruegeria pomeroyi* DSS-3	61382	63.60%	MH015258	Chi-like (Siphoviridae, cluster 3) (Zhan and Chen 2019a)				siphoviral	coastal seawater	Zhan et al. (2018)
Ruegeria phage vB_RpoS_V11	*Ruegeria pomeroyi* DSS-3	59549	64.00%	MH015254					siphoviral	coastal seawater	
Ruegeria phage vB_RpoS_V7	*Ruegeria pomeroyi* DSS-3	59573	64.10%	MH015249					siphoviral	coastal seawater	
Ruegeria phage vB_RpoS_V18	*Ruegeria pomeroyi* DSS-3	59111	64.00%	MH015252					siphoviral	coastal seawater	
Ruegeria phage DSS3_P1	*Ruegeria pomeroyi* DSS-3	59601	64.10%	KM581061					siphoviral	coastal seawater	
Rhizobium phage RHph_X2_30	*Sinorhizobium americanum* X2_30	69699	57.00%	MW980061						agricultural lands	unpublished
Loktanella phage pCB2051_A	*Loktanella* sp. CB2051	56958	55.00%	HQ632859	*Caudoviricetes, Casjensviridae, Broinstvirus (Tolstoy et al. 2021)*	*Casjensviridae*	*Broinstvirus*		siphoviral	seawater	unpublished
Rhodobacter phage RcMrWorf	*Rhodobacter capsulatus* YW1	67929	60.10%	MW677519	RcD (Rapala et al. 2021)				siphoviral	freshwater	Rapala et al. (2021)
Rhodobacter phage RcSalem	*Rhodobacter capsulatus* YW1	67718	60.00%	MW677524					siphoviral	treated sewage water	
Rhodobacter phage RcMcDreamy	*Rhodobacter capsulatus* YW1	68244	60.00%	MW677518					siphoviral	freshwater	
Rhodobacter phage RcGingersnap	*Rhodobacter capsulatus* YW1	68225	60.20%	MW677513					siphoviral	freshwater	
Rhodobacter phage RcRios	*Rhodobacter capsulatus* YW1	68774	60.30%	MW677523					siphoviral	treated sewage water	
Rhodobacter phage RcPescado	*Rhodobacter capsulatus* YW1	67494	60.40%	MW677521					siphoviral	freshwater	
Rhodobacter phage RcPutin	*Rhodobacter capsulatus* YW1	67605	60.30%	MW677522					siphoviral	freshwater	
Rhodobacter phage RcIroh	*Rhodobacter capsulatus* YW1	68475	60.20%	MW677516					siphoviral	freshwater	
Achromobacter phage vB_AchrS_ AchV4	*Achromobacter* sp.	59489	62.80%	MW269554		*Casjensviridae*	*Gediminasvirus*		siphoviral	garden grapes	Kaliniene et al. (2021)

Marinobacter phage PS6	*Marinobacter* sp.	58226	51.90%	MF959998				n.a.	unpublished
Klebsiella phage YMC16_01_ N133_KPN_BP	*Klebsiella pneumoniae* YMC16/01/N133	58387	58.90%	MF476925	*Casjensviridae*	*Seodaemunguvirus*	siphoviral	n.a.	unpublished
Bacteriophage Phobos	*Pseudomonas syringae*	56734	63.30%	MN478374	*Casjensviridae*	*Phobosvirus*	siphoviral	freshwater	Amarillas et al. (2020)
Pseudomonas phage PspYZU01	*Pseudomonas* sp.	58279	63.10%	KY971609			siphoviral	sewage	unpublished
Stenotrophomonas phage Suzuki	*Stenotrophomonas maltophilia*	56042	62.60%	MZ326855			siphoviral	freshwater	Sullivan et al. (2022)
Phage Sano	*Xylella fastidiosa*	56147	62.40%	KF626665	*Casjensviridae*	*Sanovirus*	siphoviral	sewage	Ahern et al. (2014)
Xanthomonas phage Seregon	*Xanthomonas campestris*	55527	63.20%	ON189048			siphoviral	n.a.	Erdrich et al. (2022)
Phage Salvo	*Xylella fastidiosa*	55601	63.00%	KF626668	*Casjensviridae*	*Salvovirus*	siphoviral	sewage	Ahern et al. (2014)
Achromobacter phage AXY1	*Achromobacter xylosoxidans*	61950	60.30%	OK041469				n.a.	unpublished
Achromobacter phage phiAxp_2	*Achromobacter xylosoxidans*	62220	60.10%	KT321316	*Casjensviridae*	*Fengtaivirus*	siphoviral	hospital sewage	Li et al. (2016b)
Xanthomonas phage FoX4	*Xanthomonas campestris*	60418	61.80%	MT161385		*Foxquatrovirus*	siphoviral	soil	unpublished
Burkholderia phage AH2	*Burkholderia cenocepacia* C6433	58065	61.30%	JN564907	*Casjensviridae*	*Ahduovirus*	siphoviral	soil, rhizosphere	Lynch et al. (2012)
Burkholderia cepacia phage BcepNazgul	*Burkholderia cepacia*	57455	60.60%	AY357582	*Casjensviridae*	*Nazgulvirus*	siphoviral	soil	unpublished
Aeromonas phage BUCT552	*Aeromonas hydrophila*	59685	60.00%	MW978786			siphoviral	sewage	Chen et al. (2022)
Aeromonas phage vB_AhyS_ A18P4	Aeromonas hydrophila A18	60975	62.00%	MN317029	*Casjensviridae*	*Sharonstreetvirus*	siphoviral	n.a.	unpublished
Aeromonas phage Lah_7	*Aeromonas hydrophila*	61426	62.20%	MK838113			siphoviral	wastewater	Kabwe et al. (2020)
Pseudomonas phage vB_Pae_SS2019XI	*Pseudomonas aeruginosa*	57567	60.10%	MN536026	*Casjensviridae*	*Maxdohrnvirus*		wastewater	unpublished
Klebsiella phage Seifer	*Klebsiella pneumoniae*	58197	56.10%	MH817999	*Casjensviridae*	*Yonseivirus*	siphoviral	wastewater	Salazar et al. (2019)
Klebsiella phage KPN_N137	*Klebsiella pneumoniae* YMC15/11/N137	59100	56.30%	MF415410			siphoviral	n.a.	unpublished
Pectobacterium phage MA12	*Pectobacterium carotovorum*	58573	54.50%	MN692199	*Casjensviridae*	*Newforgelanevirus*	siphoviral	agricultural wastewater	Zaczek-Moczydłowska et al. (2020)
Pectobacterium phage MA11	*Pectobacterium carotovorum*	55830	54.50%	MN518139			siphoviral	agricultural wastewater	
Serratia phage JS26	*Serratia* sp. ATCC 39006	63971	57.00%	MN505213	*Casjensviridae*	*Dunedinvirus*	siphoviral	n.a.	unpublished
Erwinia phage pEp_SNUABM_08	*Erwinia pyrifoliae*	62716	57.20%	MN184886	*Casjensviridae*	*Gwanakrovirus*	siphoviral	soil	Kim et al. (2021)

Proteus phage Saba	*Proteus mirabilis*	60056	48.80%	MN062188		*Casjensviridae*	*Cenphatecvirus*	siphoviral	wastewater	Nguyen et al. (2019)
Providencia phage Kokobel1	*Providencia stuartii*	59837	48.90%	MW145139		*Casjensviridae*	*Kokobelvirus*		sewage	unpublished
Proteus phage vB_PmiS_DoubleBarrel	*Proteus mirabiliis*	59089	46.80%	OK500000	Chi-like (Cobbley et al. 2022)				sewage	Cobbley et al. (2022)
Escherichia phage E21	*Escherichia coli*	58536	46.90%	MN604053		*Casjensviridae*	*Lavrentievaviru s*		agricultural wastewater	unpublished
Cronobacter phage JC01	*Cronobacter sakazakii*	61736	58.50%	MT330372		*Casjensviridae*	*Jacunavirus*	siphoviral	sewage	Jiang et al. (2022)
Salmonella phage SAP012	*Salmonella* sp.	59618	54.10%	LC553736		*Casjensviridae*	*Zhonglingvirus*		sewage	Shahin et al. (2022)
Providencia phage vB_PreS_PatoteraRojo	*Providencia rettgeri* B0142	60728	49.40%	MT675126					worm farm effluent	unpublished
Providencia phage vB_PreS_PibeRecoleta	*Providencia rettgeri* B0142	60727	49.30%	MT675124		*Casjensviridae*	*Redjacvirus*		worm farm effluent	unpublished
Enterobacter phage Enc34	*Enterobacter cancerogenus*	60496	51.10%	JQ340774		*Casjensviridae*	*Enchivirus*	siphoviral	n.a.	Kazaks et al. (2012)
Serratia phage KpYy_1_41	*Klebsiella pneumoniae*	54417	56.90%	MN871450		*Casjensviridae*	*Chivirus*	siphoviral	seawater	unpublished
Enterobacter phage KNP7	*Enterobacter* sp.	58058	56.50%	KX452700		*Casjensviridae*	*Chivirus*		sewage	unpublished
Salmonella phage vB_SenS_ER24	*Salmonella enterica*	60438	56.60%	MW355479		*Casjensviridae*	*Chivirus*	siphoviral	sewage	Rodwell et al. (2021)
Klebsiella phage vB_KaeS_Phraden	*Klebsiella aerogenes*	59052	56.60%	OL606627	Chi-like (Cobbley et al. 2022)				sewage	Cobbley et al. (2022)
Salmonella phage KFS_SE1	*Salmonella enterica*	59715	56.70%	MG280946		*Casjensviridae*	*Chivirus*	siphoviral	poultry processing plant	Choi et al. (2017)

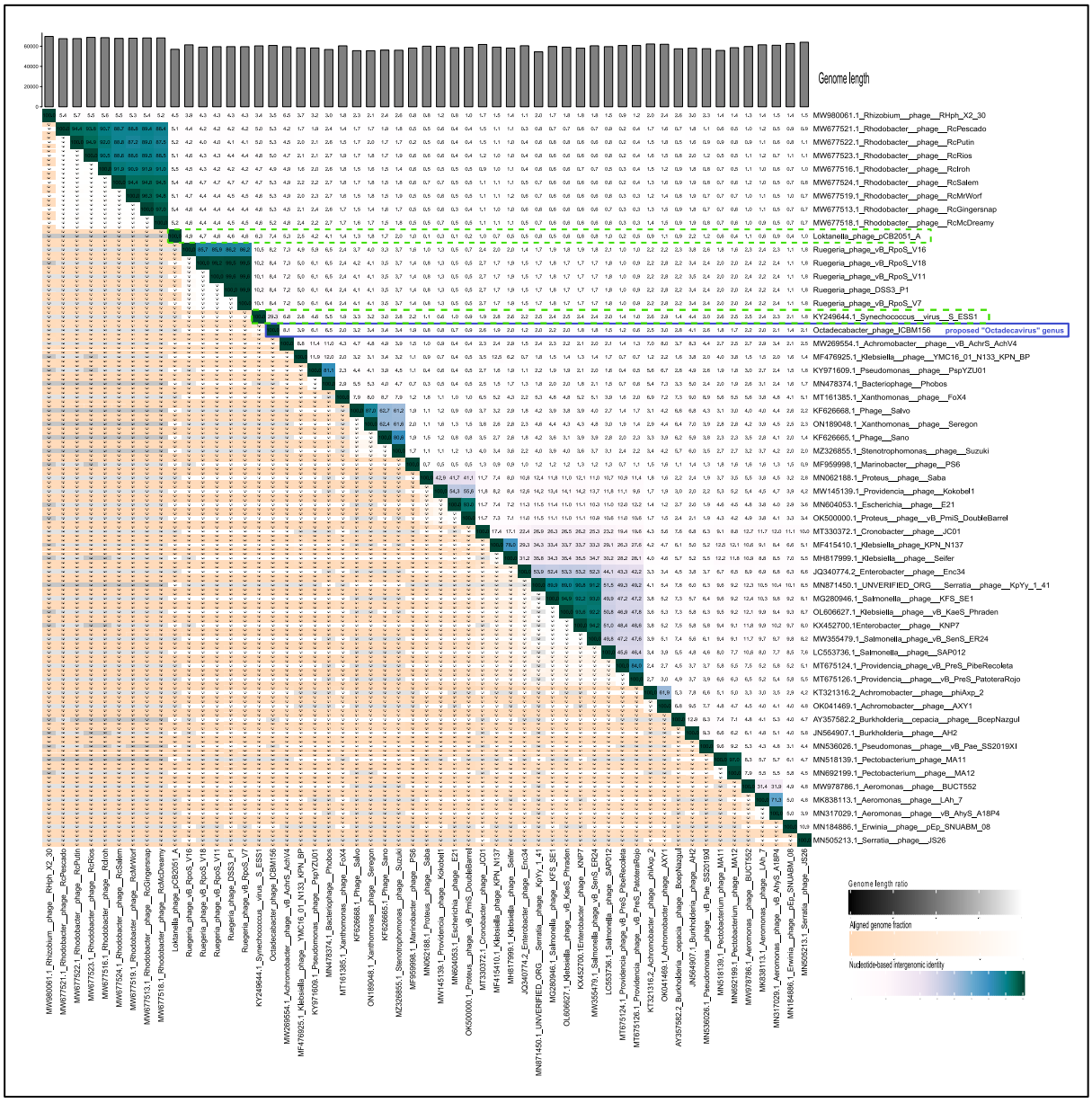


Fig. 29: Nucleotide-based intergenomic identities of phages in the *Casjensviridae* family (VGC_9), calculated with VIRIDIC. Octadecabacter phage ICBM156 is marked in blue. Two phages that potentially form a subfamily with phage ICBM156 are marked with dashed-lined green boxes.

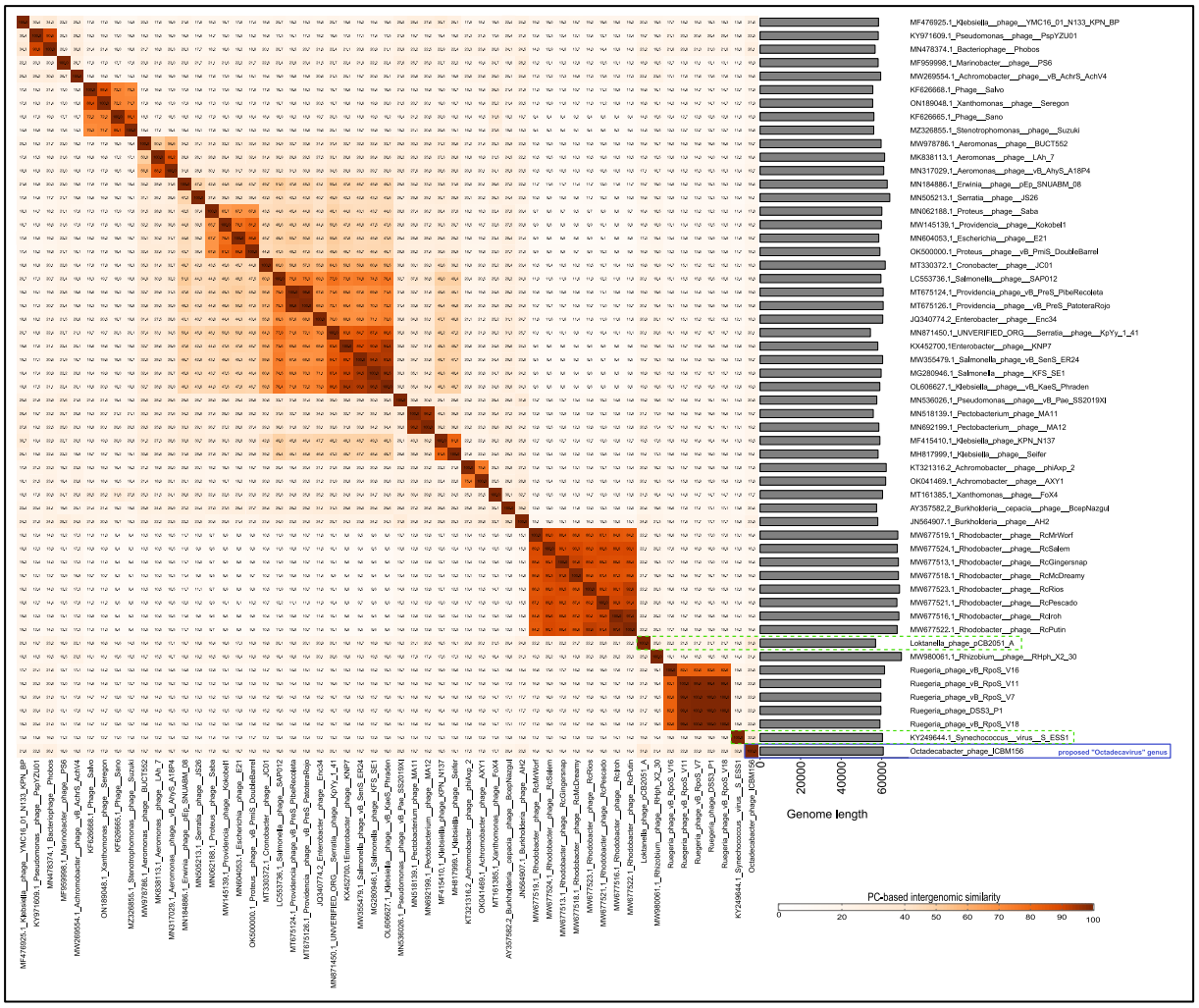


Fig. 30: PC-based intergenomic similarities of phages in the *Casjensviridae* family (VGC_9), calculated with VirClust (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%). Octadecabacter phage ICBM156 is marked in blue. Two phages that potentially form a subfamily with phage ICBM156 are marked with dashed-lined green boxes.

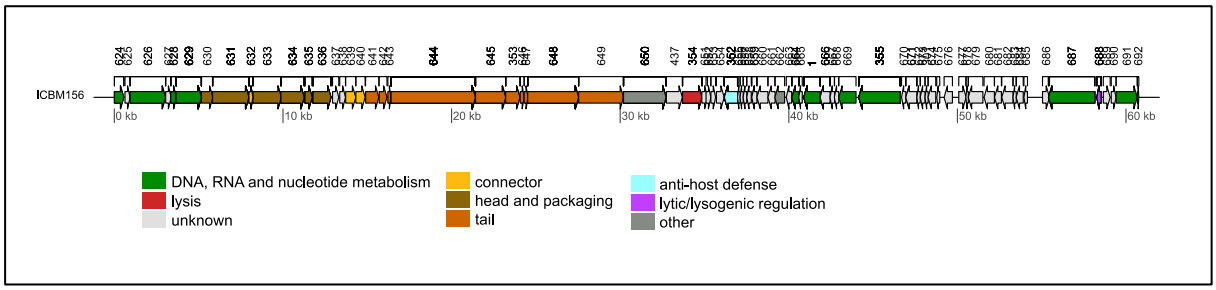


Fig. 31: Genome map Octadecabacter phage ICBM156, member of VGC_9.

Table 21: Gene annotations of Octadecabacter phage ICBM156. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	1	559	559	1	624	ssDNA binding protein	DNA, RNA and nucleotide metabolism
gene_2	630	929	300	1	625	hp	unknown
gene_3	931	3039	2109	1	626	DNA polymerase	DNA, RNA and nucleotide metabolism
gene_4	3044	3367	324	1	627	hp	unknown
gene_5	3348	3668	321	1	628	tRNA endonuclease	DNA, RNA and nucleotide metabolism
gene_6	3665	5167	1503	1	629	DNA helicase	DNA, RNA and nucleotide metabolism
gene_7	5164	5847	684	1	630	terminase small subunit	head and packaging
gene_8	5822	7981	2160	1	631	terminase large subunit	head and packaging
gene_9	7981	8214	234	1	632	head-tail adaptor Ad1	connector
gene_10	8215	9906	1692	1	633	portal protein	head and packaging
gene_11	9875	11260	1386	1	634	capsid maturation protease	head and packaging
gene_12	11305	11736	432	1	635	capsid decoration protein	head and packaging
gene_13	11781	12848	1068	1	636	major capsid protein	head and packaging
gene_14	12938	13309	372	1	637	hp	unknown
gene_15	13367	13705	339	1	638	hp	unknown
gene_16	13720	14319	600	1	639	neck protein Ne1	connector
gene_17	14321	14875	555	1	640	tail completion Tc1	connector
gene_18	14902	15684	783	1	641	tail protein	tail
gene_19	15704	16168	465	1	642	tail protein	tail
gene_20	16219	16398	180	1	643	tail protein	tail
gene_21	16408	21417	5010	1	644	tail length tape measure protein	tail
gene_22	21421	23211	1791	1	645	tail assembly protein	tail
gene_23	23208	24056	849	1	353	tail assembly protein	tail
gene_24	24070	24303	234	1	646	tail assembly chaperone	tail
gene_25	24300	24530	231	1	647	tail assembly chaperone	tail
gene_26	24514	27531	3018	1	648	tail protein	tail
gene_27	27535	30198	2664	1	649	tail protein	tail
gene_28	30195	32723	2529	1	650	protease	other
gene_29	32738	33688	951	1	437	hp	unknown
gene_30	33699	34853	1155	1	354	endolysin (N-acetylmuramidase)	lysis
gene_31	34840	35175	336	1	651	hp	unknown
gene_32	35090	35377	288	1	652	hp	unknown
gene_33	35378	35659	282	1	653	hp	unknown
gene_34	35704	36171	468	1	654	hp	unknown
gene_35	36174	36962	789	-1	352	DNA methylase N-4/N-6	anti-host defense
gene_36	36959	37114	156	-1	655	hp	unknown
gene_37	37107	37295	189	-1	656	hp	unknown
gene_38	37292	37507	216	-1	657	hp	unknown
gene_39	37504	37773	270	-1	658	hp	unknown
gene_40	37805	38107	303	-1	659	hp	unknown
gene_41	38107	38763	657	-1	660	hp	unknown
gene_42	38764	39189	426	-1	661	hp	unknown
gene_43	39173	39751	579	-1	662	cystein dioxygenase	other

gene_44	39850	40170	321	-1	663	hp	unknown
gene_45	40182	40604	423	-1	664	dihydrofolate reductase	DNA, RNA and nucleotide metabolism
gene_46	40591	40797	207	-1	665	hp	unknown
gene_47	40887	41858	972	-1	1	thymidylate synthase ThyX	DNA, RNA and nucleotide metabolism
gene_48	41855	42448	594	-1	666	hp	unknown
gene_49	42445	42711	267	-1	667	hp	unknown
gene_50	42704	42961	258	-1	668	hp	unknown
gene_51	42961	43986	1026	-1	669	DNA topoisomerase	DNA, RNA and nucleotide metabolism
gene_52	44154	46607	2454	-1	355	ribonucleotide reductase	DNA, RNA and nucleotide metabolism
gene_53	46684	46941	258	-1	670	hp	unknown
gene_54	46938	47594	657	-1	671	hp	unknown
gene_55	47594	47794	201	-1	672	hp	unknown
gene_56	47794	48078	285	-1	673	hp	unknown
gene_57	48080	48268	189	-1	301	hp	unknown
gene_58	48253	48717	465	-1	674	hp	unknown
gene_59	48772	48957	186	-1	675	hp	unknown
gene_60	49209	49703	495	-1	676	hp	unknown
gene_61	50079	50516	438	1	677	hp	unknown
gene_62	50499	50648	150	-1	678	hp	unknown
gene_63	50645	51511	867	-1	679	hp	unknown
gene_64	51580	52194	615	-1	680	hp	unknown
gene_65	52199	52609	411	-1	681	hp	unknown
gene_66	52631	53299	669	-1	682	hp	unknown
gene_67	53296	53493	198	-1	683	hp	unknown
gene_68	53484	53906	423	-1	684	hp	unknown
gene_69	53945	54157	213	-1	685	hp	unknown
gene_70	55050	55463	414	1	686	hp	unknown
gene_71	55403	58150	2748	-1	687	DNA helicase	DNA, RNA and nucleotide metabolism
gene_72	58272	58523	252	-1	688	Lambda repressor-like	lytic/lysogenic regulation
gene_73	58650	59060	411	1	689	hp	unknown
gene_74	59122	59409	288	1	690	hp	unknown
gene_75	59414	60664	1251	1	691	Exonuclease	DNA, RNA and nucleotide metabolism
gene_76	60674	60762	89	1	692	hp	unknown

4.3.5.3. Two new *Sulfitobacter* infecting roseophages form a new subfamily within the *Mesyanzhinoviridae* family

Two new roseophages grouped into a VGC with the family *Mesyanzhinoviridae* (Fig. 32). The VGC_11 comprised 46 phages, which had been isolated from diverse aquatic (seawater, freshwater) and terrestrial (sewage, soil) habitats (Table 22). The previously described phages of the *Mesyanzhinoviridae* infect *Gammaproteobacteria (Pseudomonadales* and *Xanthomonadales)* and *Betaproteobacteria (Burkholderiales).* Those family members with a confirmed morphology are siphoviruses (Table 22).

The new roseophage isolates grouping into the *Mesaynzhinoviridae* formed a separate cluster together with Dinoroseobacter phage vB_DshS_R5C, a lytic siphovirus that had been isolated from seawater infecting *Dinoroseobacter shibae* DFL12^T (Yang et al. 2017). The cluster showed a maximum PC-based intergenomic similarity with other *Mesyanzhinoviridae* of 12.0% (Fig. 34). Within this cluster, the minimum PC-based intergenomic similarity accounted for 49.3% (Fig. 34). Thus, we propose that this cluster represents a new subfamily of the *Mesyanzhinoviridae*, which we provisionally named here “Maresulfivirinae”, from the romanian word “mare” for “large”, as the family comprised largest *Sulfitobacter* phage genomes described so far. We expect that it will become a family on its own, when the *Mesyanzhinoviridae* family is divided into smaller families by future classification procedures. Four other roseophages clustered into this VGC (Ruegeria phage Tedan, Rhodobacter phage RcSimone_Hastad and Roseobacter phages RDJL_Phi_1 and RDJL_Phi_2). However, they belonged to a different subcluster than the “Maresulfivirinae” (Fig. 32).

The two new roseophages in this family were ICBM129 and ICBM153 infecting *Sulfitobacter* strains M92 and M315, respectively. They originated from seawater samples NHS and P2, respectively (Table 22). According to their nucleotide-based intergenomic identity (NBII = 47.6%), both form a genus of their own (Fig. 33). We provisionally termed these genera here “Carlotavirus” (ICBM129) and “Annevirus” (ICBM153).

The *Sulfitobacter* phages ICBM129 and ICBM153 had large genomes of 70.7 - 80.8 kb, 57.7 - 57.8% G+C content and 98 - 116 predicted ORFs. From these genes, 45 - 47 could be functionally annotated. For both viruses, phageTerm analysis determined random genome ends (Table 15). This was reflected by the fact that with the chosen termini, the functional gene module “DNA, RNA and nucleotide metabolism” was being split in the depicted genome map (Fig. 35). This module comprised genes coding for the following proteins: a DNA polymerase I (PC_510), another DNA/RNA polymerase (PC_710, only in ICBM129), a DNA polymerase

processivity factor (PC_453), a DNA polymerase exonuclease subunit (PC_454), a clamp loader of the DNA polymerase (PC_508), an exonuclease (PC_502), a thymidylate synthase ThyX (PC_470), a DNA primase (PC_465), a 5'-3' deoxyribonucleotidase (PC_463) and a cobalamin-dependent ribonucleotide reductase (PC_355) (Table 24). In addition, five DNA helicases were found (PC_456, PC_504, PC_464, PC_469, and PC_501). "Head and packaging" proteins comprised a capsid morphogenesis protein (PC_490), capsid scaffolding protein (PC_489), major capsid protein (PC_488), portal protein (PC_492), two structural proteins (PC_477 and PC_478) and the terminase large subunit (PC_82). Further morphogenesis genes were the neck protein Ne1 (PC_483), the tail completion protein Tc1 (PC_482), two tail proteins (PC_475 and PC_481), a tail assembly chaperone (PC_480), two tail assembly proteins (PC_353 and PC_476) and the tail length tape measure protein (PC_479). In addition, the connectors head closure protein Hc1 (PC_484) and head-tail adaptor protein Ad1 (PC_485) were annotated. Two lysis genes were encoded in the genomes: a spanin (PC_472) and an endolysin (PC_354). Furthermore, a GTP cyclohydrolase (PC_459), a DNA transfer protein (PC_486), a dATP/dGTP pyrophosphohydrolase (PC_455) were encoded in both genomes. The genome of phage ICBM129 additionally contained a hemolysin gene (PC_702). The genome of phage ICBM153 possessed an additional peptide chain release factor (PC_760), a dCMP deaminase (PC_782) and a PhoH-like phosphate starvation-inducible protein (PC_113) (SI file S4-10). Moreover, genes coding for queuosine (Que) biosynthesis proteins were present in the genomes of *Sulfitobacter* phages ICBM129 and ICBM153 (Table 24). Que is a hypermodified guanosine analogue in tRNAs (Asp, Asn, His, or Tyr) that leads to improved translation efficiency (El Yacoubi et al. 2012). Que or $preQ_0$ (precursor 7-cyano-7-deazaguanine) biosynthesis genes have been detected in phages and viral metagenomes before (Sabri et al. 2011; Holmfeldt et al. 2013; Kulikov et al. 2014), also in the genome of Dinoroseobacter phage vB_DshS_R5C (Yang et al. 2017). Recently, it was shown that they serve for the insertion of 7-deazaguanine derivatives into phage DNA, which is then protected against host restriction enzymes (Hutinet et al. 2017; Hutinet et al. 2019). All four required Que biosynthesis genes *queCDEF* (QueC-like queuosine biosynthesis gene, QueD-like 6-pyruvoyl-tetrahydropterin synthase, QueE-like radical SAM domain and QueF-like queuosine biosynthesis gene) as well as a GTP cyclohydrolase and a queuine tRNA-ribosyltransferase could be found in the genomes of *Sulfitobacter* phages ICBM129 and ICBM153. While in the genome of Sulfitobacter phage ICBM129 no tRNAs could be predicted, in the genome of Sulfitobacter phage ICBM153, seven tRNAs of seven different types were found (Table 23).

PhageTerm could not predict any replication and packaging strategy for these two phages (Table 15). The presence of a T7-like DNA polymerase I could point towards a T7-like bidirectional replication and DNA packaging technique. Gene annotations resembled the predicted siphoviral morphology, as tail completion protein Tc1 and tail length tape measure protein, but no tail sheath were annotated (Tables 22 and 24). Both genomes contained a Lambda-repressor like gene, indicating a lysogenic potential of the phages. The "Maresulfivirinae" phages shared 52 core proteins, with 36 of them being annotated (Table 25). The core proteins included the five queuosine biosynthesis proteins involved in anti-host defense.

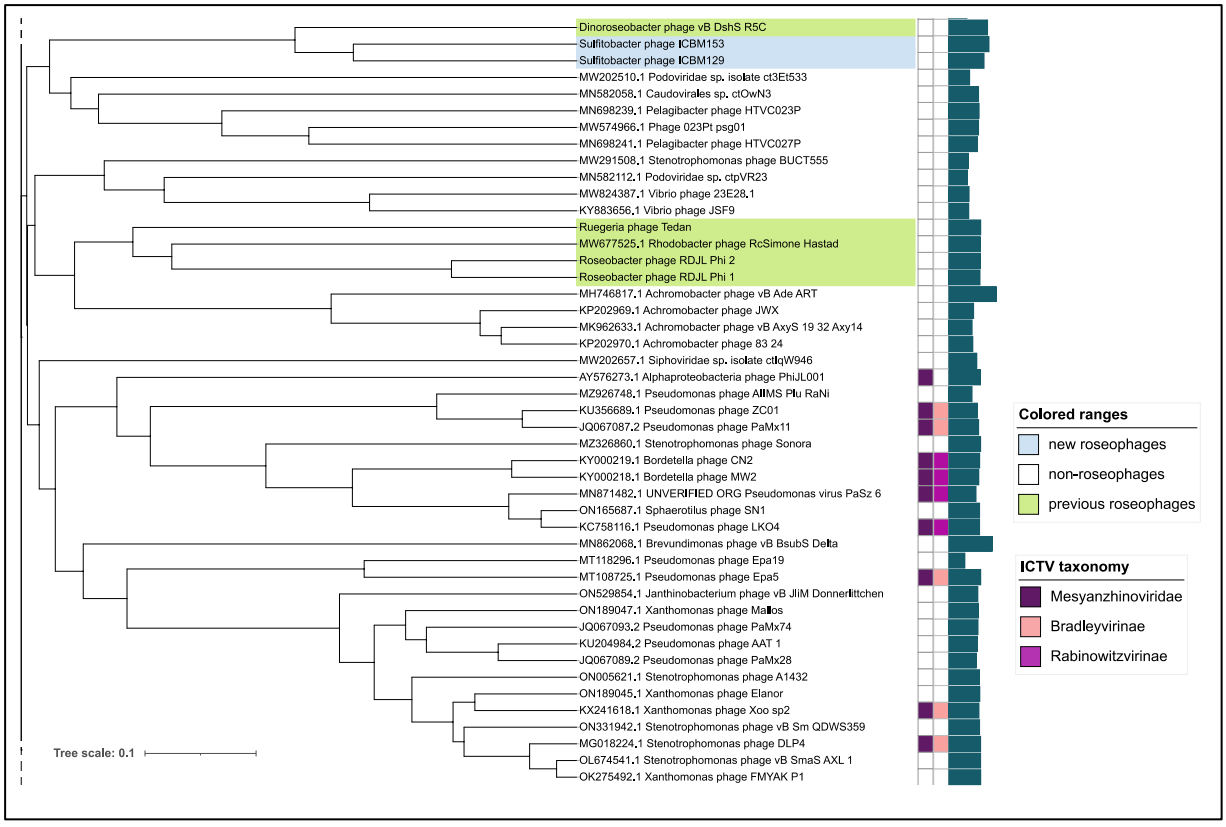


Fig. 32: Section of the whole-genome based proteomic tree showing VGC_11. Names of cultivated roseophages are marked in green (previous) and blue (this study). Color-strips indicate affiliation to ICTV-recognized families and subfamilies, respectively. Genome lengths are displayed as bar chart.

Table 22: Phages of VGC_11. Roseophages are marked in blue. Bp = base pairs. N.a. = not available. *Morphology predicted by Virfam.

Phage	Host	Genome size [bp]	GC content [%]	Accession	Previous classification (reference)	ICTV or newly proposed taxonomy			Morphology	Isolation source	Reference
						Family/Subfamily	Genus	Species			
Dinoroseobacter phage vB_DshS_R5C	*Dinoroseobacter shibae* DFL12T	77874	61.50%	KY606587	*Caudoviricetes*, *Nanhaivirus* (Kropinski et al. 2018; Turner et al. 2020)	*Mesyanzhinoviridae*, "Maresulfivirinae"	*Nanhaivirus*	*D5C*	siphoviral	seawater	Yang et al. (2017)
Sulfitobacter phage ICBM153	*Sulfitobacter* sp. M315	80756	57.70%	this study			"Annevirus"	"trei"	siphoviral*	Seawater (P2)	this study
Sulfitobacter phage ICBM129	*Sulfitobacter* sp. M92	70738	57.80%	this study			"Carlotavirus"	"una"	siphoviral*	Seawater (NHS)	this study
Podoviridae sp. isolate ct3Et533	n.a.	41714	45.10%	MW202510						freshwater river	unpublished
Caudovirales sp. ctOwN3	n.a.	59743	43.80%	MN582058						freshwater river	unpublished
Pelagibacter phage HTVC023P	Candidatus *Pelagibacter* sp. HTCC1062	60878	35.00%	MN698239					podoviral	seawater	Zhang et al. (2021)
Phage 023Pt_psg01	n.a.	59769	36.30%	MW574966						seawater	unpublished
Pelagibacter phage HTVC027P	Candidatus *Pelagibacter* sp. HTCC1062	57595	34.80%	MN698241	HTVC023P-type (Zhang et al. 2021)				podoviral	seawater	Zhang et al. (2021)
Stenotrophomonas phage BUCT555	*Stenotrophomonas maltophilia* 1207	39440	61.40%	MW291508					podoviral	hospital sewage	Han et al. (2021)
Podoviridae sp. ctpVR23	n.a.	37417	49.40%	MN582112						freshwater river	unpublished
Vibrio phage 23E28.1	*Vibrio crassostreae*	40729	49.90%	MW824387						seawater	unpublished
Vibrio phage JSF9	*Vibrio cholerae*	40007	50.60%	KY883656						freshwater river	Naser et al. (2017)
Ruegeria phage Tedan	*Ruegeria* sp. AU67	64084	54.40%	MT764845	*Xiamenvirus* (Baum et al. 2021)		*Xiamenvirus*		siphoviral	sponge	Baum et al. (2021)
Rhodobacter phage RcSimone-Håstad	*Rhodobacter capsulatus* SB1003	63102	60.70%	MW677525	Caudovirales, Siphoviridae (Rapala et al. 2021)				siphoviral	freshwater river	Rapala et al. (2021)
Roseobacter phage RDJL_Phi_2	*Roseobacter denitrificans* OCh114	63513	57.30%	KT266805	*Caudoviricetes*, *Xiamenvirus* (Kropinski et al. 2016; Turner et al. 2020)		*Xiamenvirus*		siphoviral	seawater	Liang et al. (2016)
Roseobacter phage RDJL_Phi_1	*Roseobacter denitrificans* OCh114	62668	57.90%	HM151342					siphoviral	seawater	Zhang and Jiao (2009)

Achromobacter phage vB_Ade_ART	*Achromobacter denitrificans* PR1	95343	55.00%	MH746817				wastewater	unpublished
Achromobacter phage JWX	*Achromobacter xylosoxidans*	49714	55.40%	KP202969		*Steinhofvirus*	siphoviral	sewage	Dreiseikelmann et al. (2017)
Achromobacter phage vB_AxyS_19_32_Axy14	*Achromobacter xylosoxidans* I2BC	46703	55.00%	MK962633			siphoviral	sewage	Essoh et al. (2020)
Achromobacter phage 83_24	*Achromobacter xylosoxidans*	48216	54.90%	KP202970		*Steinhofvirus*	siphoviral	sewage	Dreiseikelmann et al. (2017)
Siphoviridae sp. isolate ctIqW946	n.a.	56278	59.00%	MW202657				freshwater river	unpublished
Alphaproteobacteria phage PhiJL001	n.a.	63649	62.10%	AY576273	*Mesyanzhinovviridae*	*Keylargovirus*	siphoviral	seawater, sponge associated	Lohr et al. (2005)
Pseudomonas phage AIIMS_Plu_RaNi	*Pseudomonas luteola*	46647	64.50%	MZ926748			siphoviral	sewage	Rathor et al. (2022)
Pseudomonas phage ZC01	*Pseudomonas aeruginosa*	57061	63.50%	KU356689	*Mesyanzhinovviridae/ Bradleyvirinae*	*Abidjanvirus*	siphoviral	composting	Amgarten et al. (2017)
Pseudomonas phage PaMx11	*Pseudomonas aeruginosa*	59878	64.50%	JQ067087			siphoviral	freshwater	Sepúlveda-Robles et al. (2012)
Stenotrophomonas phage Sonora	*Stenotrophomonas maltophilia*	63825	63.00%	MZ326860			siphoviral	soil	Teve et al. (2022)
Bordetella phage CN2	*Bordetella bronchiseptica* ATCC 10580	62030	64.10%	KY000219	*Mesyanzhinovviridae/ Rabinowitzvirinae*	*Vojvodinavirus*	siphoviral	canal	Petrovic et al. (2017)
Bordetella phage MW2	*Bordetella bronchiseptica* ATCC 10580	60160	64.10%	KY000218			siphoviral	wastewater	
Pseudomonas virus PaSz_6	*Pseudomonas aeruginosa* PAO1	54656	64.70%	MN871482				Seawater	unpublished
Sphaerotilus phage SN1	*Sphaerotilus natans*	61858	64.40%	ON165687				wastewater	Gunathilake et al. (2022)
Pseudomonas phage LKO4	*Pseudomonas aeruginosa*	61818	64.40%	KC758116	*Mesyanzhinovviridae/ Rabinowitzvirinae*	*Yuavirus*		n.a.	unpublished
Brevundimonas phage vB_BsubS_Delta	*Brevundimonas subvibrioides* ATCC 15264	87225	62.50%	MN862068			siphoviral	pond water	Mascolo et al. (2022)
Pseudomonas phage Epa19	*Pseudomonas* sp.	32090	61.40%	MT118296				n.a.	unpublished
Pseudomonas phage Epa5	*Pseudomonas* sp.	64096	62.20%	MT108725	*Mesyanzhinovviridae/ Bradleyvirinae*	*Epaquintavirus*		sewage	Farlow et al. (2020)
Janthinobacterium phage vB_JliM_Donnerlittchen	*Janthinobacterium lividum* EIF1	58220	67.80%	ON529854				sewage	unpublished
Xanthomonas phage Mallos	*Xanthomonas translucens*	59242	61.80%	ON189047			siphoviral	sewage	Erdrich et al. (2022)

Pseudomonas phage PaMx74	*Pseudomonas aeruginosa*	58637	68.40%	JQ067093			siphoviral	sewage	Sepúlveda-Robles et al. (2012)
Pseudomonas phage AAT_1	*Pseudomonas aeruginosa* PA14	57599	65.90%	KU204984				fetal bovine serum	Andrade-Domínguez and Kolter (2016)
Pseudomonas phage PaMx28	*Pseudomonas aeruginosa*	55108	66.50%	JQ067089			siphoviral	sewage	Sepúlveda-Robles et al. (2012)
Stenotrophomonas phage A1432	*Stenotrophomonas maltophilia* YCR3A-1	61660	61.90%	ON005621				n.a.	unpublished
Xanthomonas phage Elanor	*Xanthomonas translucens*	62341	64.50%	ON189045			siphoviral	soil	Erdrich et al. (2022)
Xanthomonas phage Xoo_sp2	*Xanthomonas oryzae* pv. *oryzae*	60497	66.50%	KX241618	*Mesyanzhinovviridae/ Bradleyvirinae*	*Xooduovirus*	siphoviral	soil	Dong et al. (2018)
Stenotrophomonas phage vB_Sm_QDWS359	*Stenotrophomonas maltophilia* SM102	61862	67.50%	ON331942				n.a.	unpublished
Stenotrophomonas phage DLP4	*Stenotrophomonas maltophilia* D1585	63945	65.10%	MG018224	*Mesyanzhinovviridae/ Bradleyvirinae*	*Bosavirus*	siphoviral	soil	Peters et al. (2019)
Stenotrophomonas phage vB_SmaS_AXL_1	*Stenotrophomonas maltophilia* D1585	63962	67.30%	OL674541			siphoviral	soil	McCutcheon et al. (2022)
Xanthomonas phage FMYAK_P1	*Xanthomonas* sp.	64088	67.80%	OK275492				freshwater	unpublished

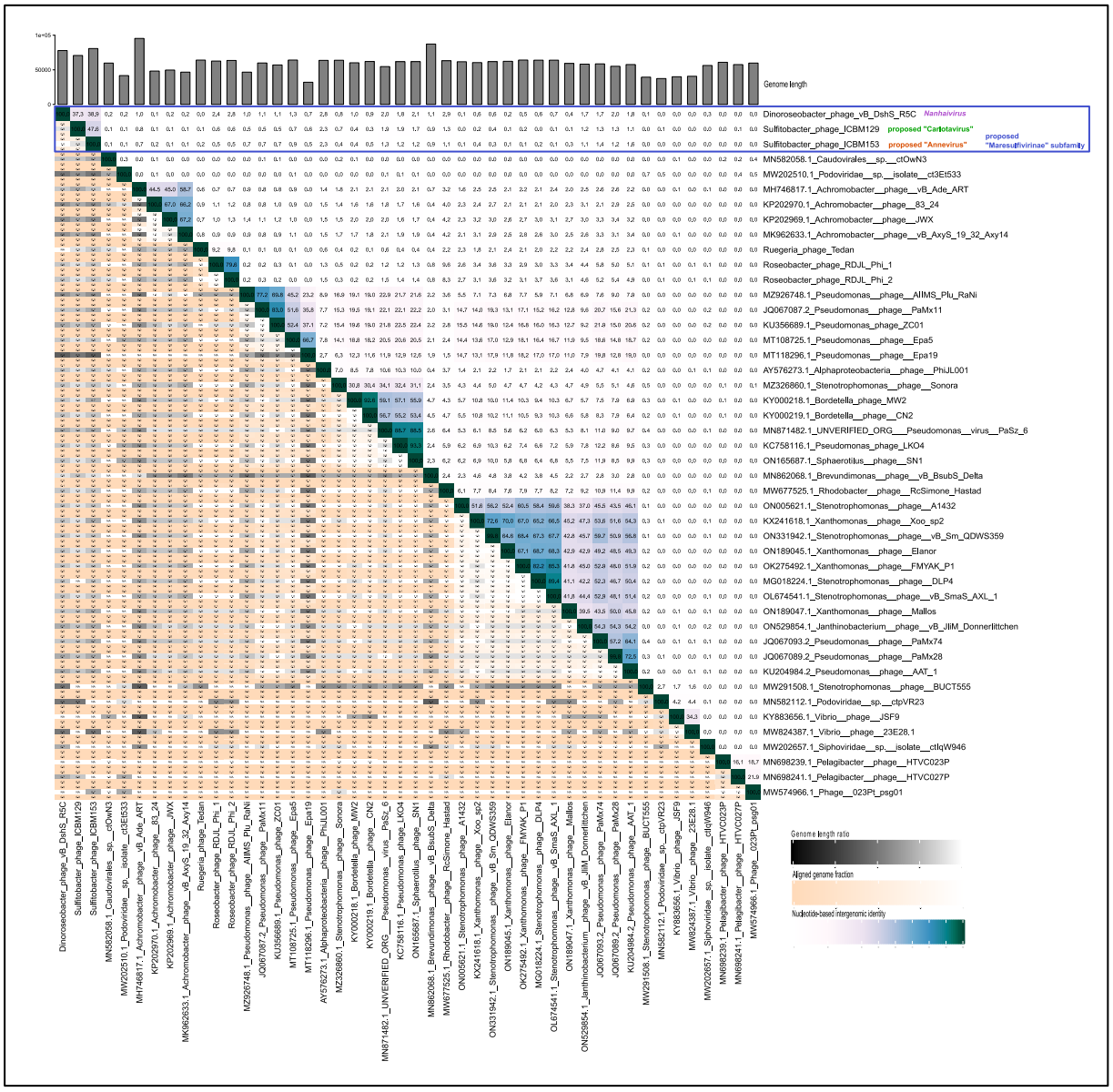


Fig. 33: Nucleotide-based intergenomic identities of phages in VGC_11, calculated with VIRIDIC. Members of the newly proposed genera “Carlotavirus” and “Annevirus”, the existing genus Nanhaivirus, and the newly proposed subfamily “Maresulfivirinae” of the family *Mesyanzhinoviridae* are annotated with a blue box and colored labels.

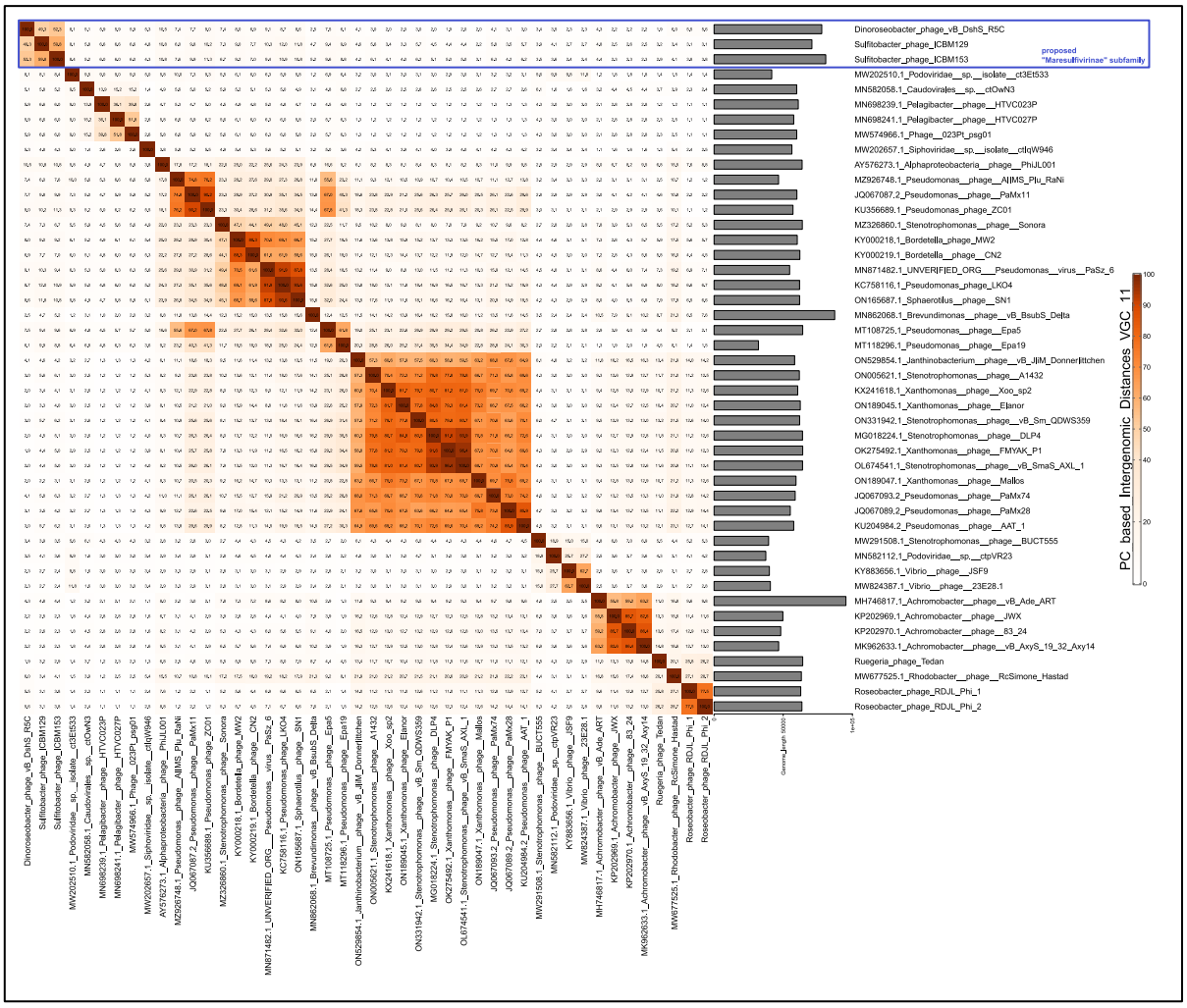


Fig. 34: PC-based intergenomic similarities of phages in VGC_11, calculated with VirClust (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%). Members of the newly proposed subfamily "Maresulfivirinae" of the family *Mesyanzhinoviridae* are annotated with a blue box.

Table 23: tRNAs found in the genome of Sulfitobacter phage ICBM153.

Phage	tRNA_#	tRNA_Begin	tRNA_End	tRNA_Type
Sulfitobacter phage ICBM153	1	51850	51776	Gln
Sulfitobacter phage ICBM153	2	51518	51441	Thr
Sulfitobacter phage ICBM153	3	51222	51149	Gly
Sulfitobacter phage ICBM153	4	50993	50915	Glu
Sulfitobacter phage ICBM153	5	50798	50723	Asn
Sulfitobacter phage ICBM153	6	50717	50641	Met
Sulfitobacter phage ICBM153	7	50386	50311	Phe

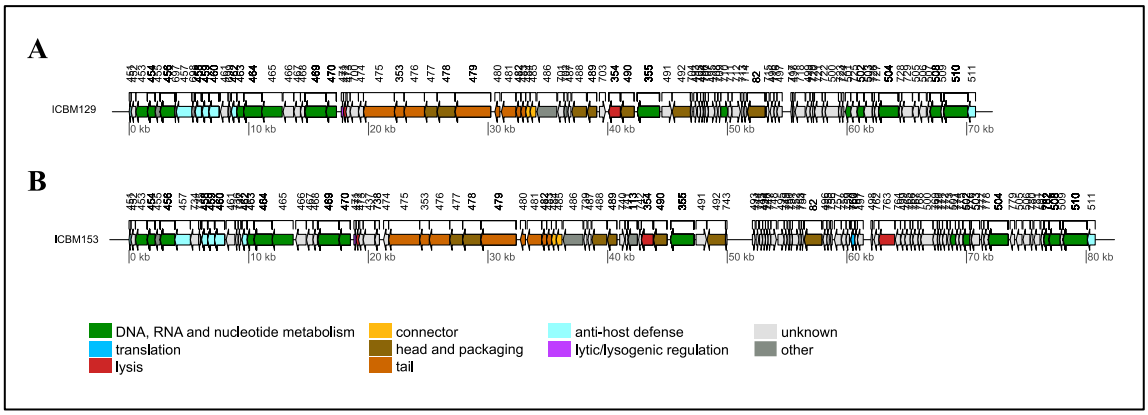


Fig. 35: Genome map of *Sulfitobacter* phages ICBM129 and ICBM153.

Table 24: Gene annotations of Sulfitobacter phage ICBM129. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core protein of the phages in the "Maresulfivirinae".

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	1	159	159	-1	451	QueE-like radical SAM domain	Anti-host defense
gene_2	159	545	387	-1	452*	hp*	unknown
gene_3	557	1519	963	-1	453*	DNA polymerase processivity factor*	DNA, RNA and nucleotide metabolism
gene_4	1524	2174	651	-1	454*	DNA polymerase exonuclease subunit*	DNA, RNA and nucleotide metabolism
gene_5	2167	2604	438	-1	455*	dATP/dGTP pyrophosphohydrolase*	other
gene_6	2597	3808	1212	-1	456*	Dda-like helicase*	DNA, RNA and nucleotide metabolism
gene_7	3808	3930	123	-1	697	hp	unknown
gene_8	3930	5231	1302	-1	457*	queuine tRNA-ribosyltransferase*	Anti-host defense
gene_9	5231	5425	195	-1	698	hp	unknown
gene_10	5506	6030	525	-1	458*	QueF-like queuosine biosynthesis gene*	Anti-host defense
gene_11	6090	6650	561	-1	459*	GTP cyclohydrolase*	other
gene_12	6659	7495	837	-1	460*	QueC-like queuosine biosynthesis*	Anti-host defense
gene_13	7626	8252	627	-1	461*	hp*	unknown
gene_14	8275	8445	171	-1	699	hp	unknown
gene_15	8531	8980	450	-1	462*	QueD-like 6-pyruvoyl-tetrahydropterin synthase*	Anti-host defense
gene_16	8982	9563	582	-1	463*	5'-3' deoxyribonucleotidase*	DNA, RNA and nucleotide metabolism
gene_17	9560	11080	1521	-1	464*	DNA helicase*	DNA, RNA and nucleotide metabolism
gene_18	11067	12785	1719	-1	465*	DNA primase*	DNA, RNA and nucleotide metabolism
gene_19	12860	13741	882	-1	466*	hp*	unknown
gene_20	13769	14311	543	-1	467*	hp*	unknown
gene_21	14313	14657	345	-1	468*	hp*	unknown
gene_22	14693	16462	1770	-1	469*	DNA helicase*	DNA, RNA and nucleotide metabolism
gene_23	16462	17319	858	-1	470*	thymidylate synthase ThyX*	DNA, RNA and nucleotide metabolism
gene_24	17724	17918	195	1	471*	Lambda repressor-like*	lytic/lysogenic regulation
gene_25	17905	18213	309	-1	472*	Spanin*	lysis

gene_26	18107	18466	360	-1	473*	hp*	unknown
gene_27	18463	19158	696	-1	700	hp	unknown
gene_28	19155	19577	423	-1	474	hp	unknown
gene_29	19577	22165	2589	-1	475	tail protein	tail
gene_30	22162	22971	810	-1	353	tail assembly protein	tail
gene_31	22972	24672	1701	-1	476	tail assembly protein	tail
gene_32	24672	25736	1065	-1	477	structural protein	head and packaging
gene_33	25738	27243	1506	-1	478	structural protein	head and packaging
gene_34	27243	30203	2961	-1	479*	tail length tape measure protein*	tail
gene_35	30561	30986	426	-1	480*	tail assembly chaperone*	tail
gene_36	31112	32311	1200	-1	481*	tail protein*	tail
gene_37	32330	32737	408	-1	482*	tail completion Tc1*	tail
gene_38	32734	33144	411	-1	483*	neck protein Ne1*	tail
gene_39	33137	33511	375	-1	484*	head closure Hc1*	connector
gene_40	33508	33990	483	-1	485*	head-tail adaptor Ad1*	connector
gene_41	34083	35741	1659	-1	486*	DNA transfer protein*	other
gene_42	35810	36343	534	-1	701	hp	unknown
gene_43	36330	36599	270	-1	702	hemolysin	other
gene_44	36671	36922	252	-1	487*	hp*	unknown
gene_45	36998	38218	1221	-1	488*	major capsid protein*	head and packaging
gene_46	38296	39075	780	-1	489*	capsid scaffolding protein*	head and packaging
gene_47	39302	39802	501	1	703	hp	unknown
gene_48	40085	41047	963	-1	354	endolysin (N-acetylmuramidase)	lysis
gene_49	41087	42199	1113	-1	490*	capsid morphogenesis protein*	head and packaging
gene_50	42471	44324	1854	-1	355*	ribonucleotide reductase*	DNA, RNA and nucleotide metabolism
gene_51	44526	45350	825	1	491*	hp*	unknown
gene_52	45386	46906	1521	-1	492*	portal protein*	head and packaging
gene_53	46906	47049	144	-1	704	hp	unknown
gene_54	47188	47427	240	1	493	hp	unknown
gene_55	47496	47756	261	1	705	hp	unknown
gene_56	47756	47992	237	1	494	hp	unknown
gene_57	47989	48138	150	1	706	hp	unknown
gene_58	48149	48412	264	1	707	hp	unknown
gene_59	48409	48951	543	1	495	hp	unknown
gene_60	48948	49172	225	1	708	hp	unknown
gene_61	49224	49448	225	1	709	hp	unknown
gene_62	49445	50056	612	1	710	DNA/RNA pol	DNA, RNA and nucleotide metabolism
gene_63	50053	50388	336	1	711	hp	unknown
gene_64	50385	51080	696	1	712	hp	unknown
gene_65	51133	51318	186	-1	713	hp	unknown
gene_66	51374	51580	207	-1	714	hp	unknown
gene_67	51671	53143	1473	-1	82*	terminase large subunit*	head and packaging
gene_68	53258	53488	231	1	715	hp	unknown
gene_69	53571	53894	324	-1	496*	hp*	unknown
gene_70	53887	54141	255	-1	716*	hp*	unknown
gene_71	54138	54578	441	-1	497*	hp*	unknown
gene_72	55306	55443	138	-1	717	hp	unknown

gene_73	55443	55568	126	-1	498	hp	unknown
gene_74	55658	56497	840	-1	718	hp	unknown
gene_75	56548	56907	360	-1	499	hp	unknown
gene_76	56904	57176	273	-1	719	hp	unknown
gene_77	57119	57280	162	-1	720	hp	unknown
gene_78	57277	57918	642	-1	721	hp	unknown
gene_79	57918	58181	264	-1	722	hp	unknown
gene_80	58165	59283	1119	-1	500	hp	unknown
gene_81	59286	59507	222	-1	723	hp	unknown
gene_82	59553	59732	180	-1	724	hp	unknown
gene_83	59862	60317	456	-1	501*	DNA helicase*	DNA, RNA and nucleotide metabolism
gene_84	60299	60838	540	-1	725	hp	unknown
gene_85	60835	61413	579	-1	502*	Exonuclease*	DNA, RNA and nucleotide metabolism
gene_86	61410	62093	684	-1	503*	hp*	unknown
gene_87	62109	62282	174	-1	726	hp	unknown
gene_88	62282	62632	351	-1	727	hp	unknown
gene_89	62632	64287	1656	-1	504*	DNA helicase	DNA, RNA and nucleotide metabolism
gene_90	64348	64572	225	-1	728	hp	unknown
gene_91	64572	65456	885	-1	729	hp	unknown
gene_92	65479	65970	492	-1	505*	hp*	unknown
gene_93	66036	66623	588	-1	506*	hp*	unknown
gene_94	66697	66945	249	-1	507*	hp*	unknown
gene_95	66945	67868	924	-1	508*	clamp loader of DNA polymerase*	DNA, RNA and nucleotide metabolism
gene_96	67843	68082	240	-1	509*	hp*	unknown
gene_97	68079	70076	1998	-1	510*	DNA polymerase I*	DNA, RNA and nucleotide metabolism
gene_98	70079	70738	660	-1	511*	QueE-like radical SAM domain*	Anti-host defense

Table 25: Core proteins of the newly proposed subfamily "Maresulfivirinae" in the family *Mesyanzhinoviridae*. *maximum number of predicted ORFs divided by number of core proteins.

predicted ORFs#	core proteins #	Annotated core proteins #	Annotated core proteins (sorted by functional category)	Percentage core proteins / total ORFs*
98-116	52	36	DNA, RNA and nucleotide metabolism: DNA polymerase (PC_510), DNA polymerase processivity factor (PC_453), DNA polymerase exonuclease subunit (PC_454), clamp loader of DNA polymerase (PC_508), exonuclease (PC_502), Dda-like helicase (PC_456), 5'-3' deoxyribonucleotidase (PC_463), DNA helicases (PC_464, PC_501 and PC_504), DNA primase (PC_465), DNA helicase (PC_469), thymidylate synthase ThyX (PC_470), and ribonucleotide reductase (PC_355) Head and packaging: major capsid protein (PC_488), capsid scaffolding protein (PC_489), capsid morphogenesis protein (PC_490), portal protein (PC_492), terminase large subunit (PC_82) Connector: head closure protein Hc1 (PC_484) and the head-tail adaptor Ad1 (PC_485) Tail: tail length tape measure protein (PC_479), tail assembly chaperone (PC_480), tail protein (PC_481), tail completion Tc1 (PC_482), neck protein Ne1 (PC_483) Lysis: spanin (PC_472) Lytic/lysogenic regulation: Lambda repressor-like protein (PC_471), Anti-host defense: queuine tRNA-ribosyltransferase (PC_457), QueF-like queuosine biosynthesis gene (PC_458), QueC-like queuosine biosynthesis (PC_460), QueD-like 6-pyruvoyl-tetrahydropterin synthase (PC_462), QueE-like radical SAM domain (PC_511) Other: dATP/dGTP pyrophosphohydrolase (PC_455), GTP cyclohydrolase (PC_459), DNA transfer protein (PC_486)	44.83%

4.3.5.4. "Hayaniviridae" and "Schlingloffviridae" - two new families of lytic roseophages

One viral genome cluster consisted only of roseophages (VGC_25). It comprised eight of our new isolates and Roseophage CRP-6 (Zhang et al. 2019b) (Fig. 36). Caulobacter phage Sansa (KT001913.1) distantly clustered within this VGC, too. However, it was earlier classified as member of the *Casjensviridae* family and both, the intergenomic identities calculated with VIRIDIC (Moraru et al. 2020) (Fig. 37a) and the hierarchical clustering based on shared protein content determined with VirClust (Moraru 2023) (SI file S4-13), suggested a misclustering due to long branch attraction. The phage was not included in the further classification.

The phages in this VGC formed two distinct subclusters, sharing less than 20% PC-based intergenomic similarity with each other (Fig. 36 and 37b). Within the two clusters, the minimum PC-based intergenomic similarity was 74.1 and 75.0%, respectively. Therefore, we suggest referring to them as two potential phage families, which we named here "Hayaniviridae" family and "Schlingloffviridae" family.

The "Hayaniviridae" family contained six phages (ICBM21, ICBM13, ICBM16, ICBM47, ICBM23 and ICBM25) infecting closely related *Sulfitobacter* strains (Table 26). They were representatives of the large "sulfivirus" cluster, which will be analyzed in more detail in chapter 5. The six phages have been isolated from four different water samples (P1, P2, P4 and NHS) (Table 26). Based on their nucleotide-based intergenomic identities (NBII), all six phages formed one genus (Fig. 37a). We provisionally termed this genus here "Sulfivirus" genus. Each of the phages represented its own species ($70.3\% \leq NBII \leq 86.3\%$) (Fig. 37a). Members of the "Schlingloffviridae" family comprised two phages infecting *Lentibacter* sp. SH36 (ICBM7 and ICBM166) as well as Roseophage CRP-6 (Zhang et al. 2019b) (Table 26). Both *Lentibacter* phages have been isolated from North Sea water samples, but from different years (HE440-S, HE504-33). Phage CRP-6 has been isolated from the subtropical Pingtan coast in China, infecting *Roseobacter* strain FZCC0042, which belongs to the RCA cluster (Table 26). All three phages belonged to one genus, based on their nucleotide-based intergenomic identity ($NBII \geq 76.8$) (Fig. 37a). We provisionally called this genus here "Falkvirus" genus. Each of the phages represented its own species ($76.8\% \leq NBII \leq 88.4\%$).

Genome composition of the "Hayaniviridae" family

Phages of the "Hayaniviridae" family had genome lengths of 52.4 - 55.6 kb and a G+C content of 44.6 - 45.0, with 86 - 93 predicted ORFs (Table 26). Out of these genes, 28 - 30 could be annotated (Table 28). PhageTerm analysis indicated the presence of short direct

terminal repeats (DTRs) of 306 to 330 bp at the genome ends (Table 15). The genomes had a modular architecture with two major arms, the left arm encoding genes on the forward strand, the right arm on the reverse strand. In addition, there were a few genes (5-10) on the reverse strand at the very beginning of the genome, which could only be annotated as hypothetical (Fig. 38). The DNA replication and nucleotide metabolism module located on the left genomic arm comprised a DNA primase/helicase (PC_200 or PC_306), a DNA polymerase I (PC_94), a thymidylate synthase ThyX (PC_1), an exonuclease (PC_90) and a cobalamin-dependent ribonucleotide reductase (PC_300, PC_5223 or PC_525) (Table 28). Phages ICBM13, ICBM23, ICBM25 and ICBM47 had a second DNA polymerase (PC_375 or PC_731) encoded and in the genome of ICBM25, an RNA ligase (PC_805) and a polynucleotide kinase (PC_806) were additionally annotated. The genomes had varying numbers of HNH endonucleases endoded in their genomes, which differed also in their positions (Fig. 38, SI file S4-10). On the right genomic arm, there were the morphology and the lysis module encoded. An endolysin (PC_3) was encoded close to the center of the genome. The structural genes comprised the major capsid protein (PC_104), the portal protein (PC_105), a capsid maturation protease (PC_196) and another capsid protein (PC_151) as well as the terminase large subunit (PC_82) (Table 28). Furthermore, the head closure protein Hc1 (PC_101), tail completion protein Tc1 (PC_100), a tail fiber protein (PC_2), a minor tail protein (PC_99), a tail length tape measure protein (PC_310 or PC_522) were annotated. Another protein in the tail module was annotated either as a concanavalin A-like lectin/glucanase (PC_368 in ICBM13, ICBM16, and ICBM23), as hydrolase/lipase (PC_359 in ICBM25) or as tail fiber protein (PC_790 in ICBM21). In addition, a heat-shock protein DnaJ (PC_168), a metallo-phosphoesterase (PC_96) and a nucleotide pyrophosphohydrolase (PC_181) were found (Table 28). Different numbers of DNA methylases were encoded in the genomes, which are likely involved in strategies to circumvent host defense. Searching the genomes of *Sulfitobacter* phages ICBM21 and ICBM47 with tRNAscan-SE (Lowe and Chan 2016), one asparagine - tRNA was predicted, respectively (Table 27). No tRNAs were found in the other phage genomes of this family.

The presence of DTRs at the genome ends and a T7-like polymerase A indicated that the phages of the “Hayaniviridae” family use a T7-like bidirectional genome replication and packaging strategy (Tables 15 and 28). A siphoviral morphology with an icosahedral capsid and a long, non-contractile tail was predicted by Virfam. It was also resembled by the gene annotations (tail length tape measure protein and tail completion protein Tc1 present, no tail sheath) (Table 28). The fact that no lysogeny-related proteins have been annotated suggests that these phages are strictly lytic. The phages of the “Hayaniviridae” family shared 63 core

proteins, including 20 with an annotated function that belonged to all functional categories (Table 30).

Genome composition of the "Schlingloffviridae" family

The two *Lentibacter* phages of the "Schlingloffviridae" family (ICBM7 and ICBM166) had genome lengths of 44.9 - 45.6 kb and a G+C content of 47.1 - 47.4% with 72 - 73 predicted ORFs (Table 26). In both genomes, 31 of the genes could be annotated (Table 29). Short DTRs (109 bp) were predicted at the genome ends (Table 15). The genomes were organized in two arms (Fig. 38). On the left arm (forward strand), mainly genes of the DNA, RNA and nucleotide metabolism were annotated, comprising a thymidylate synthase (PC_1), an RNA polymerase sigma factor (PC_398), a DNA primase/helicase (PC_200), a DNA polymerase I (PC_94), an exonuclease (PC_90), an endonuclease (PC_148), a ribonucleotide reductase (PC_5) and a glutaredoxin (PC_150) (Table 29). The right arm of the genomes (reverse strand) encoded the structural module containing the portal protein (PC_105), the major capsid protein (PC_104), a capsid scaffolding protein (PC_414) and the terminase large subunit (PC_82) as well as two connectors (tail completion Tc1 (PC_100), head closure Hc1 (PC_101)). In addition, three tail proteins (PC_410, PC_411, and PC_423)), two tail fiber proteins (PC_408 and PC_422)), a tail length tape measure protein (PC_412), and a minor tail protein (PC_99) were annotated (Table 29). Two endolysin-encoding genes were present in both genomes (PC_95 and PC_146). In addition, genes coding for a bifunctional heptose 7-phosphate kinase / heptose 1-phosphate adenyltransferase (PC_389), a transcriptional regulator (Trp repressor; PC_334), a metallo-phosphoesterase (PC_96), as well as a MazG-like pyrophosphatase (PC_149) and PhoH-like phosphate starvation-inducible protein (PC_113) were encoded. The genomes of *Lentibacter* phages ICBM7 and ICBM166 contained two proteins involved in anti-host defense: a SaV-like protein (PC_397), which is involved in the sensitivity to the host abortive infection mechanism AbiV (Haaber et al. 2009), and a S-adenosyl-L-methionine-dependent methyltransferase (PC_402), which can serve for protection against restriction-modification systems (Murphy et al. 2013) (Table 29). In addition, an AMG was found coding for a 2OG-Fe(II) oxygenase (PC_419). In cyanophages, this enzyme is thought to affect the nitrogen metabolism of the host during infection (Sullivan et al. 2010; Wang et al. 2022). No tRNAs were predicted in these phage genomes.

As phage ICBM7 and ICBM166 possessed DTRs at the genome termini and encoded a T7-like DNA polymerase I, a T7-like bidirectional genome replication and packaging technique can be suspected (Tables 15, 28, and 29). For Phage CRP-6, genome ends and packaging strategy are not described (Zhang et al. 2019b). The annotation of a tail length tape measure

protein and the tail completion protein Tc1 and the absence of a tail sheath matched the prediction of a siphoviral morphology by Virfam (Lopes et al. 2014) (Tables 28 and 29). Phage CRP-6 was described as having a podoviral morphology with an icosahedral capsid and a short tail (Zhang et al. 2019b). This would mean there were two different tail morphologies within one genus. Virfam analysis of the CRP-6 genome could now specify a gene that the authors annotated as" hypothetical protein" as the tail completion protein Tc1, which indicates that phage CRP-6 actually also has a long tail morphology. A siphoviral structure was predicted (Table 26). The podoviral morphology observed by transmission electron microscopy could potentially be due to loss of the phage tails in the course of sample preparation. As no lysogeny-related proteins have been annotated in these genomes, a strictly lytic lifestyle of the phages can be suspected. The phages of the "Schlingloffviridae" family shared 54 core proteins, including 31 with an annotated function (Table 30).

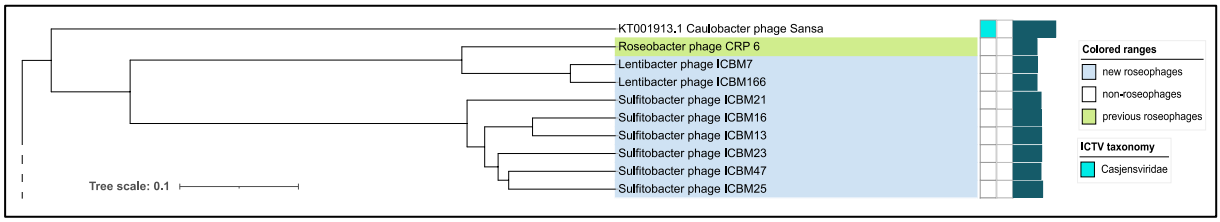


Fig. 36: Section of the whole-genome based proteomic tree showing VGC_25. Names of cultivated roseophages are marked in green (previous) and blue (this study). Genome lengths are displayed as bar chart.

Table 26: Phages of VGC_25. *Morphology predicted by Virfam.

Phage	Host	Genome size [bp]	GC content [%]	Previous classification (reference)	ICTV or newly proposed taxonomy			Morphology	Isolation source	Reference / Accession
					Family	Genus	Species			
Sulfitobacter phage ICBM21	*Sulfitobacter* sp. M53	52,607	44.6				"zes"	siphoviral*	Seawater (NHS)	This study
Sulfitobacter phage ICBM13	*Sulfitobacter* sp. M45	53,694	44.8				"vif"	siphoviral*	Seawater (P2)	This study
Sulfitobacter phage ICBM16	*Sulfitobacter* sp. M53	53,115	45.0				"vier"	siphoviral*	Seawater (P1)	This study
Sulfitobacter phage ICBM47	*Sulfitobacter* sp. M86	52,388	44.8		"Hayaniviridae"	"Sulfivirus"	"drie"	siphoviral*	Seawater (P4)	This study
Sulfitobacter phage ICBM23	*Sulfitobacter* sp. M55	53,772	44.9				"een"	siphoviral*	Seawater (P2)	This study
Sulfitobacter phage ICBM25	*Sulfitobacter* sp. M63	55,591	45.0				"twee"	siphoviral*	Seawater (NHS)	This study
Lentibacter phage ICBM7	*Lentibacter* sp. SH36	45,550	47.1				"tva"	siphoviral*	Seawater (HE440-S)	This study
Lentibacter phage ICBM166	*Lentibacter* sp. SH36	44,947	47.4		"Schlingloffviridae"	"Falkvirus"	"eni"	siphoviral*	Seawater (HE504-33)	This study
Roseobacter phage CRP-6	*Roseobacter* strain FZCC0042	44,927	47.0	CRP-6-type (Zhang et al. 2019a)				siphoviral* (podoviral)	Seawater (Pingtan coast, Taiwan Strait)	Zhang et al. (2019a), MK613348

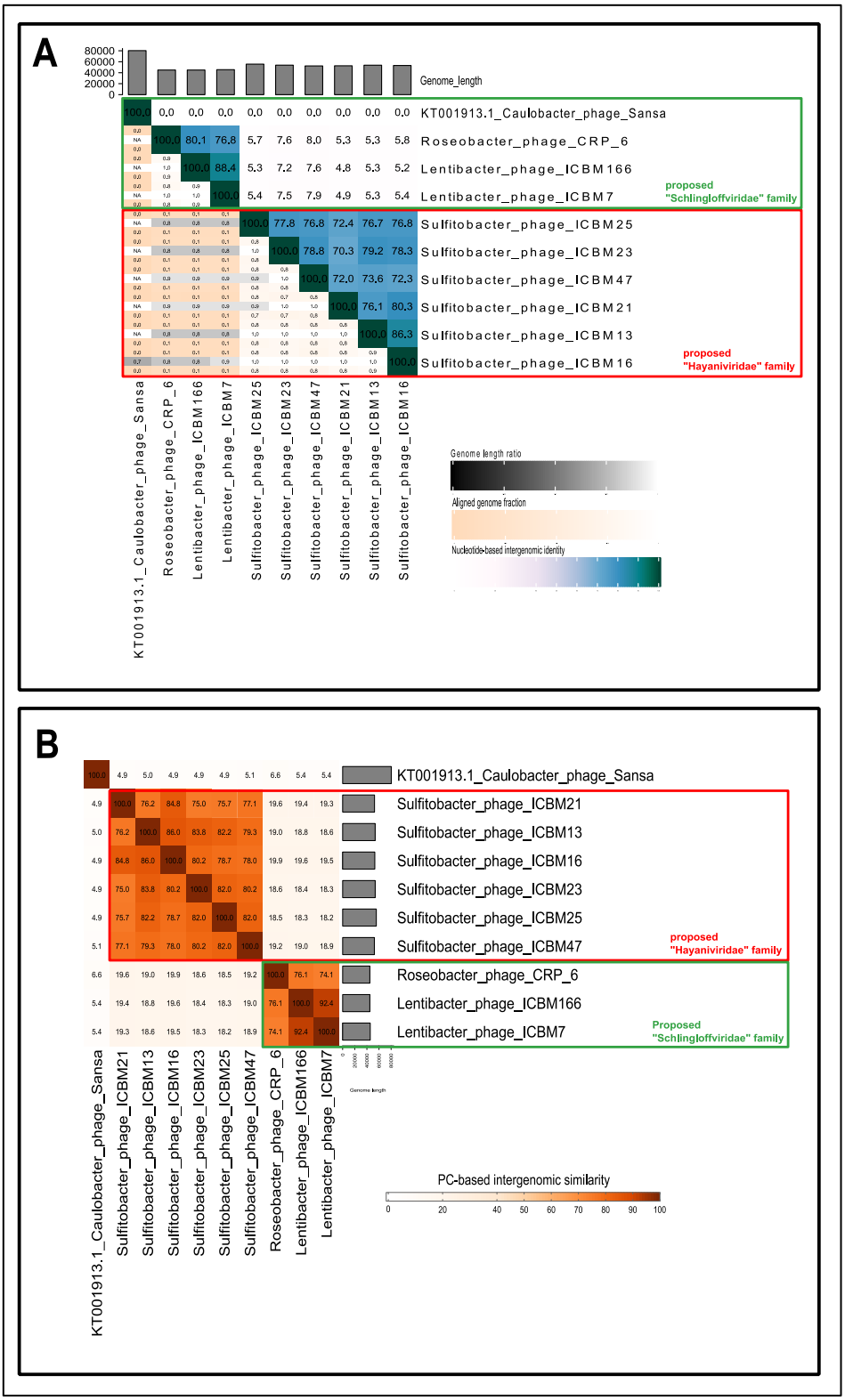


Fig. 37: Internal clustering of VGC_25. **A** Nucleotide-based intergenomic identities calculated with VIRIDIC. **B** PC-based intergenomic similarities calculated with VirClust (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%). The newly proposed families “Hayaniviridae” and “Schlingloffviridae” are annotated with colored boxes.

Table 27: tRNAs found in the genomes of "Sulfivirus" phages.

Phage	tRNA_#	tRNA_Begin	tRNA_End	tRNA_Type
Sulfitobacter phage ICBM21	1	50164	50088	Asn
Sulfitobacter phage ICBM47	1	50941	50865	Asn

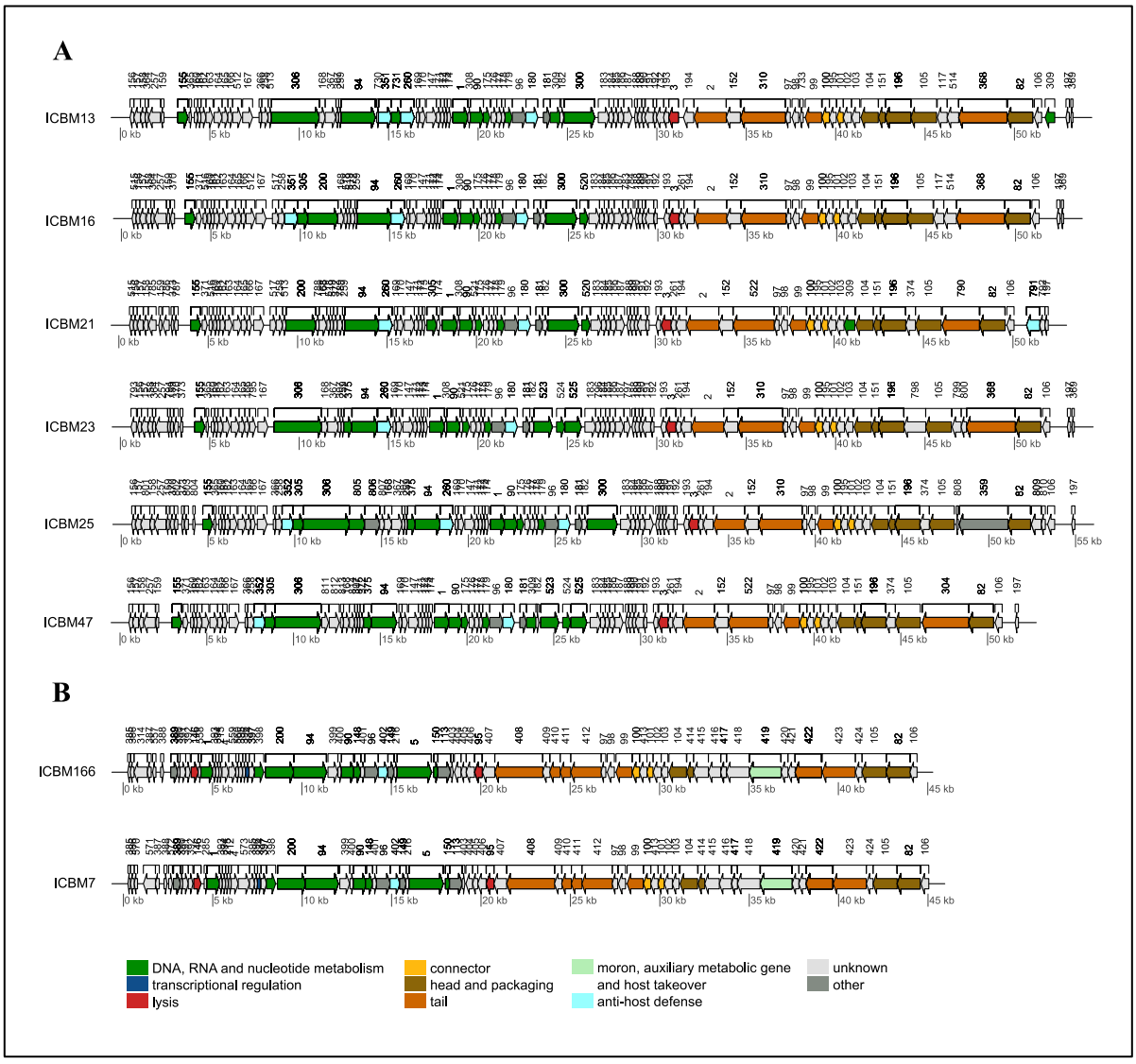


Fig. 38: Genome maps of VGC_25. **A** Phages of the "Hayaniviridae" family. **B** Phages of the "Schlingloffviridae" family.

Table 28: Gene annotations of Lentibacter phage ICBM16. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core proteins of phages in the "Hayaniviridae" family.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	595	810	216	-1	515	hp	unknown
gene_2	810	1028	219	-1	156*	hp*	unknown

gene_3	1025	1357	333	-1	157*	hp*	unknown
gene_4	1381	1641	261	-1	158*	hp*	unknown
gene_5	1641	1823	183	-1	364	hp	unknown
gene_6	1839	2501	663	-1	257	hp	unknown
gene_7	2533	2703	171	-1	159*	hp*	unknown
gene_8	2746	2970	225	-1	370	hp	unknown
gene_9	3553	4104	552	1	155*	HNH endonuclease*	DNA, RNA and nucleotide metabolism
gene_10	4190	4474	285	1	371	hp	unknown
gene_11	4565	4780	216	1	516	hp	unknown
gene_12	4822	5016	195	1	160*	hp*	unknown
gene_13	5019	5204	186	1	161*	hp*	unknown
gene_14	5194	5475	282	1	162*	hp*	unknown
gene_15	5477	5872	396	1	163*	hp*	unknown
gene_16	5933	6379	447	1	164*	hp*	unknown
gene_17	6376	6705	330	1	165*	hp*	unknown
gene_18	6702	6953	252	1	166*	hp*	unknown
gene_19	6968	7459	492	1	512	hp	unknown
gene_20	7526	8083	558	1	167*	hp*	unknown
gene_21	8444	8785	342	1	517	hp	unknown
gene_22	8782	9141	360	1	258	hp	unknown
gene_23	9138	9857	720	1	351	DNA methylase N-4/N-6	anti-host defense
gene_24	9854	10438	585	1	305	HNH endonuclease	DNA, RNA and nucleotide metabolism
gene_25	10435	12105	1671	1	200	DNA primase/helicase	DNA, RNA and nucleotide metabolism
gene_26	12105	12431	327	1	168*	heat-shock protein DnaJ*	unknown
gene_27	12443	12640	198	1	518	hp	unknown
gene_28	12640	12813	174	1	519	hp	unknown
gene_29	12825	13070	246	1	372	hp	unknown
gene_30	13054	13209	156	1	259	hp	unknown
gene_31	13206	15095	1890	1	94*	DNA polymerase I*	DNA, RNA and nucleotide metabolism
gene_32	15079	15834	756	1	260	C-5 cytosine methyltransferase	anti-host defense
gene_33	15949	16122	174	1	169*	hp*	unknown
gene_34	16137	16472	336	1	170*	hp*	unknown
gene_35	16509	17072	564	1	147*	hp*	unknown
gene_36	17072	17311	240	1	171*	hp*	unknown
gene_37	17304	17522	219	1	172*	hp*	unknown
gene_38	17519	17677	159	1	173*	hp*	unknown
gene_39	17677	17880	204	1	174*	hp*	unknown
gene_40	17996	18835	840	1	1*	thymidylate synthase ThyX*	DNA, RNA and nucleotide metabolism
gene_41	18832	18975	144	1	308	hp	unknown
gene_42	18972	19703	732	1	90*	Exonuclease*	DNA, RNA and nucleotide metabolism
gene_43	19700	20071	372	1	175*	HNH endonuclease*	DNA, RNA and nucleotide metabolism
gene_44	20132	20449	318	1	176*	hp*	unknown
gene_45	20446	20721	276	1	177*	hp*	unknown
gene_46	20711	20911	201	1	178*	hp*	unknown
gene_47	20918	21316	399	1	179*	HNH endonuclease*	DNA, RNA and nucleotide metabolism

gene_48	21316	22083	768	1	96*	metallo-phosphoesterase*	other
gene_49	22080	22742	663	1	180*	DNA methyltransferase*	anti-host defense
gene_50	23062	23439	378	1	181*	nucleotide pyrophosphohydrolase*	other
gene_51	23436	23723	288	1	182*	hp*	unknown
gene_52	23749	25452	1704	1	300	Cobalamin-dependent ribonucleotide reductase	DNA, RNA and nucleotide metabolism
gene_53	25636	26094	459	1	520	HNH endonuclease	DNA, RNA and nucleotide metabolism
gene_54	26135	26737	603	1	183*	hp*	unknown
gene_55	26734	27003	270	1	184*	hp*	unknown
gene_56	27006	27236	231	1	185*	hp*	unknown
gene_57	27260	27562	303	1	186*	hp*	unknown
gene_58	27559	28023	465	1	187*	hp*	unknown
gene_59	28030	28344	315	1	783	hp	unknown
gene_60	28341	28619	279	1	784	hp	unknown
gene_61	28719	28922	204	-1	188*	hp*	unknown
gene_62	28885	29124	240	-1	189*	hp*	unknown
gene_63	29121	29330	210	-1	190*	hp*	unknown
gene_64	29290	29766	477	-1	191*	hp*	unknown
gene_65	29763	29957	195	-1	192*	hp*	unknown
gene_66	30336	30638	303	-1	193*	hp*	unknown
gene_67	30631	31176	546	-1	3*	endolysin (lysozyme-peptidase)*	lysis
gene_68	31173	31454	282	-1	261	hp	unknown
gene_69	31454	32032	579	-1	194*	hp*	unknown
gene_70	32054	33853	1800	-1	2*	tail fiber protein*	tail
gene_71	33846	34637	792	-1	152*	hp*	unknown
gene_72	34659	37157	2499	-1	310	tail length tape measure protein	tail
gene_73	37169	37435	267	-1	97*	hp*	unknown
gene_74	37531	37905	375	-1	98*	hp*	unknown
gene_75	38039	38959	921	-1	99*	minor tail protein*	tail
gene_76	38984	39391	408	-1	100*	tail completion Tc1*	connector
gene_77	39388	39777	390	-1	195*	hp*	unknown
gene_78	39810	40169	360	-1	101*	head closure Hc1*	connector
gene_79	40169	40618	450	-1	102*	hp*	unknown
gene_80	40638	41090	453	-1	103*	hp*	unknown
gene_81	41156	42124	969	-1	104*	major capsid protein*	head and packaging
gene_82	42139	42519	381	-1	151*	capsid protein*	head and packaging
gene_83	42527	43942	1416	-1	196*	capsid maturation protease*	head and packaging
gene_84	43952	45388	1437	-1	105*	portal protein*	head and packaging
gene_85	45425	45988	564	-1	117	hp	unknown
gene_86	45985	46698	714	-1	514	hp	unknown
gene_87	46702	49383	2682	-1	368	Concanavalin A-like lectins/glucanases	tail
gene_88	49399	50820	1422	-1	82*	terminase large subunit*	head and packaging
gene_89	50913	51323	411	-1	106*	hp*	unknown
gene_90	52303	52491	189	-1	197*	hp*	unknown
gene_91	52565	52690	126	-1	369	hp	unknown

Table 29: Gene annotations of Lentibacter phage ICBM7. Gene annotations for Lentibacter phage ICBM166 are almost identical, there is just one hypothetical gene less prior to the phosphate kinase. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core proteins of phages in the "Schlingloffviridae" family.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	254	394	141	-1	385	hp	unknown
gene_2	399	611	213	-1	386	hp	unknown
gene_3	608	790	183	-1	570	hp	unknown
gene_4	1116	1793	678	-1	571	hp	unknown
gene_5	1793	2011	219	-1	387	hp	unknown
gene_6	2270	2440	171	1	388	hp	unknown
gene_7	2533	2742	210	1	572	hp	unknown
gene_8	2799	3194	396	1	389*	bifunctional heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase*	other
gene_9	3184	3324	141	1	390*	hp*	unknown
gene_10	3308	3631	324	1	391*	hp*	unknown
gene_11	3631	3888	258	1	392*	hp*	unknown
gene_12	3941	4345	405	1	146*	endolysin (cell wall hydrolase)*	lysis
gene_13	4515	4685	171	1	285	hp	unknown
gene_14	4685	5377	693	1	1*	thymidylate synthase ThyX*	DNA, RNA and nucleotide metabolism
gene_15	5367	5537	171	1	393	hp	unknown
gene_16	5534	5719	186	1	394	hp	unknown
gene_17	5716	5865	150	1	215	hp	unknown
gene_18	5887	6036	150	1	112	hp	unknown
gene_19	6029	6424	396	1	4	hp	unknown
gene_20	6424	7053	630	1	573	hp	unknown
gene_21	7055	7342	288	1	395	hp	unknown
gene_22	7335	7505	171	1	396*	hp*	unknown
gene_23	7498	7704	207	1	334*	Trp repressor*	transcriptional regulation
gene_24	7697	8005	309	1	397*	SaV-like*	unknown
gene_25	8002	8517	516	1	398*	RNA polymerase sigma factor*	DNA, RNA and nucleotide metabolism
gene_26	8617	10188	1572	1	200*	DNA primase/helicase*	DNA, RNA and nucleotide metabolism
gene_27	10185	12023	1839	1	94*	DNA polymerase I*	DNA, RNA and nucleotide metabolism
gene_28	12104	12655	552	1	399*	hp*	unknown
gene_29	12667	12867	201	1	400*	hp*	unknown
gene_30	12868	13587	720	1	90*	Exonuclease*	DNA, RNA and nucleotide metabolism
gene_31	13548	13946	399	1	148*	Endonuclease*	DNA, RNA and nucleotide metabolism
gene_32	13934	14167	234	1	401*	hp*	unknown
gene_33	14160	14924	765	1	96*	metallo-phosphoesterase*	other
gene_34	14917	15462	546	1	402*	S-adenosyl-L-methionine-dependent methyltransferase*	anti-host defense
gene_35	15459	15806	348	1	149*	MazG-like pyrophosphatase*	other
gene_36	15803	15985	183	1	216*	hp*	unknown
gene_37	15982	17889	1908	1	5*	ribonucleotide reductase*	DNA, RNA and nucleotide metabolism
gene_38	18011	18247	237	1	150*	Glutaredoxin*	DNA, RNA and nucleotide metabolism

gene_39	18247	18951	705	1	113*	PhoH-like phosphate starvation-inducible protein*	other
gene_40	18948	19100	153	1	403*	hp*	unknown
gene_41	19159	19527	369	-1	404	hp	unknown
gene_42	19531	19824	294	-1	405*	hp*	unknown
gene_43	19814	20287	474	-1	406*	hp*	unknown
gene_44	20293	20718	426	-1	95*	endolysin (L-alanyl-D-glutamate peptidase)*	lysis
gene_45	20722	21432	711	-1	407	hp	unknown
gene_46	21437	24121	2685	-1	408*	tail fiber protein*	tail
gene_47	24108	24527	420	-1	409*	hp*	unknown
gene_48	24506	25057	552	-1	410*	tail protein*	tail
gene_49	25054	25644	591	-1	411*	tail protein*	tail
gene_50	25644	27347	1704	-1	412*	tail length tape measure protein*	tail
gene_51	27340	27645	306	-1	97*	hp*	unknown
gene_52	27684	28082	399	-1	98*	hp*	unknown
gene_53	28171	29070	900	-1	99*	minor tail protein*	tail
gene_54	29088	29495	408	-1	100*	tail completion Tc1*	connector
gene_55	29488	29901	414	-1	413*	hp*	unknown
gene_56	29901	30242	342	-1	101*	head closure Hc1*	connector
gene_57	30254	30700	447	-1	102*	hp*	unknown
gene_58	30700	31077	378	-1	103*	hp*	unknown
gene_59	31133	32113	981	-1	104*	major capsid protein*	head and packaging
gene_60	32138	32509	372	-1	414*	capsid scaffolding protein*	head and packaging
gene_61	32512	33327	816	-1	415*	hp*	unknown
gene_62	33327	34016	690	-1	416*	hp*	unknown
gene_63	34016	34339	324	-1	417*	hp*	unknown
gene_64	34339	35607	1269	-1	418*	hp*	unknown
gene_65	35617	37392	1776	-1	419*	2OG-Fe(II) oxygenase*	moron, auxiliary metabolic gene and host takeover
gene_66	37397	37777	381	-1	420	hp	unknown
gene_67	37774	38184	411	-1	421	hp	unknown
gene_68	38184	39659	1476	-1	422*	tail fiber protein*	tail
gene_69	39696	41543	1848	-1	423*	tail protein*	tail
gene_70	41554	41925	372	-1	424*	hp*	unknown
gene_71	41925	43271	1347	-1	105*	portal protein*	head and packaging
gene_72	43273	44559	1287	-1	82*	terminase large subunit*	head and packaging
gene_73	44603	45043	441	-1	106*	hp*	unknown

Table 30: Core proteins of the newly proposed families "Hayaniviridae" and "Schlingloffviridae". *maximum number of predicted ORFs divided by number of core proteins.

Proposed family	predicted ORFs#	core proteins #	Annotated core proteins #	Annotated core proteins (sorted by functional category)	Percentage core proteins / total ORFs*
„Hayaniviridae"	86-93	63	20	DNA, RNA and nucleotide metabolism: HNH endonuclease (PC_155), DNA polymerase I (PC_94), thymidylate synthase ThyX (PC_1), Exonuclease (PC_90), HNH endonuclease (PC_175), HNH endonuclease (PC_179), Head and packaging: major capsid protein (PC_104), capsid protein (PC_151), capsid maturation protease (PC_196), portal protein (PC_105), terminase large subunit (PC_82) Connector: head closure Hc1 (PC_101), tail completion Tc1 (PC_100), Tail: tail fiber protein (PC_2), minor tail protein (PC_99), Lysis: endolysin (lysozyme-peptidase) (PC_3), Anti-host defense: DNA methyltransferase (PC_180), Other: heat-shock protein DnaJ (PC_168), metallo-phosphoesterase (PC_96), nucleotide pyrophosphohydrolase (PC_181)	67.74%
„Schlingloffviridae"	69-73	54	31	DNA, RNA and nucleotide metabolism: thymidylate synthase ThyX (PC_1), RNA polymerase sigma factor (PC_398), DNA primase/helicase (PC_200), DNA polymerase I (PC_94), Exonuclease (PC_90), Endonuclease (PC_148), ribonucleotide reductase (PC_5), Glutaredoxin (PC_150), Head and packaging: major capsid protein (PC_104), capsid scaffolding protein (PC_414), portal protein (PC_105), terminase large subunit (PC_82), Connector: tail completion Tc1 (PC_100), head closure Hc1 (PC_101), Tail: tail fiber protein (PC_408), tail protein (PC_410), tail protein (PC_411), tail length tape measure protein (PC_412), minor tail protein (PC_99), tail fiber protein (PC_422), tail protein (PC_423), Lysis: endolysin (cell wall hydrolase) (PC_146), endolysin (L-alanyl-D-glutamate peptidase) (PC_95), Anti-host defense: S-adenosyl-L-methionine-dependent methyltransferase (PC_402), Transcriptional regulation: Trp repressor (PC_334), Moron, AMG and host takeover: 2OG-Fe(II) oxygenase (PC_419), Other: bifunctional heptose 7-phosphate kinase/heptose 1-phosphate adenyltransferase (PC_389), metallo-phosphoesterase (PC_96), MazG-like pyrophosphatase (PC_149), PhoH-like phosphate starvation-inducible protein (PC_113),	73.97%

4.3.5.5. The "Diferiteviridae" - a new family of lytic and potentially temperate roseophages infecting *Lentibacter* and *Sulfitobacter*

Five of the new roseophages clustered into one VGC together with known roseophages infecting the genera *Sulfitobacter* (phiGT1, NYA 2014a, pCB2047-A, pCB2047-C), *Octadecabacter* (Antarctic DB virus 2) and *Roseobacter* (CRP 7) (Ankrah et al. 2014b; Luhtanen et al. 2018; Zhang et al. 2019a; Hwang et al. 2020). In addition, a number of *Vibrio* and *Pseudomonas* phages belonged to this VGC_19, which comprised 34 phages in total (Fig. 39, Table 31). The members originated from marine, freshwater and terrestrial habitats such as soil and sewage. Those phages with confirmed morphology information had a podoviral shape (Table 31).

The new roseophages formed a separate cluster within this VGC_19 (Fig. 39). The maximum PC-based intergenomic similarity with other phages in this VGC was 17.9% (Fig. 41). Within their cluster, the minimum PC-based intergenomic similarity was 30.9% (with one exception of 20.4 % between two phages). Additionally taking into consideration that ICBM8 had similar ecological features predicted as the phage ICBM165, i.e. the fact that the phages infected the same host strain, had the same podoviral morphology and the same DNA packaging strategy predicted, phage ICBM8 was also included into this family. Thus, we propose here that the five new roseophages form a new family, which we tentatively named here "Diferiteviridae", from the Romanian word "diferite" meaning "diverse", because the members infected different host genera.

The five new roseophages in this family were ICBM8, ICBM163 and ICBM165 infecting *Lentibacter* sp. SH36 and phages ICBM121 and ICBM122 infecting *Sulfitobacter* sp. M73 (Table 31). Based on their nucleotide-based intergenomic identity (NBII) they clustered into three genera (Fig. 40). Lentibacter phage ICBM163 belonged to one genus together with *Sulfitobacter* phages ICBM121 and ICBM122 (NBII ≥ 75.2), which we provisionally called here "Benvirus". The *Sulfitobacter* phages were almost identical to each other and belonged to one species. They have been isolated from the same water sample (P2) and with the same isolation host *Sulfitobacter* sp. M73 (Table 31). Phage ICBM163 infecting *Lentibacter* sp. SH36 was isolated from the same water sample (P2). The other two *Lentibacter* viruses ICBM8 and ICBM165 belonged to two separate genera, which we tentatively named "Martinvirus" (ICBM165) and "Maryvirus" (ICBM8) (Fig. 40). They had the same isolation host, *Lentibacter* sp. SH36, but originated from a different water sample (HE504-33) (Table 31).

The new roseophages had genomes of 36.9 - 38.7 kb in size, 55.9 - 57.3 % G+C content and 54 - 57 predicted ORFs. Out of these genes, 22 - 28 could be functionally annotated. Determination of the genome termini with PhageTerm (Garneau et al. 2017), showed that *Lentibacter* phages ICBM8 and ICBM165 had short, T7-type direct terminal repeats (DTRs) (Table 15). Also for Lentibacter phage ICBM163, DTRs were predicted, even though they were less supported. For the two *Sulfitobacter* phages ICBM121 and ICBM122, although being so closely related to phage ICBM163, circularly permuted genomes were predicted. The genome termini were chosen in correspondence to ICBM163 to facilitate genome comparisons.

The genome composition within this family was less conserved, thus the annotated DNA replication, lysis and morphology genes are described separately for each of the three genera. In the genomes of the "Benvirus" genus (ICBM121, ICBM122, and ICBM163), genes were arranged in three modules in terms of transcriptional direction (Fig. 42). Functional categories were not that clearly separated. However, genes for DNA, RNA and nucleotide metabolism rather grouped in the first half of the genome and morphology genes grouped in the second half of the genome. DNA replication genes comprised a DNA primase/helicase (PC_316), a nuclease, a DNA repair exonuclease (PC_323), a HNH endonuclease (only in ICBM121 and ICBM122) and a Holliday junction resolvase (Table 32). No DNA polymerase could be annotated. The structural genes comprised the portal protein (PC_209), a capsid protein (PC_278), a capsid scaffolding protein (PC_208), two virion structural proteins (PC_212 and PC_213), a tail protein (PC_381 or PC_447), a tail fiber protein (PC_448, only in ICBM121 and ICBM122) and a phage tail assembly chaperone (PC_281 or PC_328). The major capsid protein was encoded twice (PC_207 and PC_325). Furthermore, the terminase large and small subunits could be annotated (PC_214 and PC_280). Five lysis genes were annotated: two spanins (PC_271 and PC_272), an amidase (PC_273) and two endolysins (PC_146 and PC_326).

For Lentibacter phage ICBM165, the genomic organization looked a bit different (Fig. 42). The majority of the genes were encoded on the reverse strand, thus the genome was not as clearly separated into modules. Still, the overall order of the genes was similar to the genomes of the "Benvirus" genus. Genes of the functional category "DNA, RNA and nucleotide metabolism" comprised a DNA primase/helicase (PC_537), an exonuclease (PC_539), a Holliday junction resolvase (PC_267) and a tRNA endonuclease (PC_543) (Table 33). Also in this genome, no DNA polymerase could be annotated. The structural and lysis genes were more similar to the "Benvirus" phages. Morphology genes comprised the major capsid protein (PC_207), the capsid scaffolding protein (PC_208), the portal protein (PC_209), a capsid

protein (PC_278), two virion structural proteins (PC_212 and PC_213), and the terminase small and large subunit (PC_535 and PC_214). Further, one tail protein (PC_381) and three phage tail assembly chaperones (PC_281, PC_328, and PC_548). Four lysis genes were annotated: an amidase (PC_273), two spanins (PC_271 and 272) and an endolysin (PC_556).

In corresponce with the lower PC-based intergenomic similarity, the genomic content of Lentibacter phage ICBM8 differed the most from the other genomes in this family. The genes were organized in two genomic arms with opposing transcriptional directions (Fig. 42). The DNA replication module on the left genomic arm contained a DNA helicase (PC_580), a nuclease (PC_578), a Sak4-like ssDNA annealing protein (PC_6), a ssDNA binding protein (PC_577), and a tRNA endonuclease (PC_581) (Table 34). Furthermore, a bifunctional DNA primase/polymerase (PC_574) was found. The “head and packaging” genes comprised the capsid scaffolding protein (PC_208), the major capsid protein (PC_207), the portal protein (PC_209), two virion structural proteins (PC_212 and PC_213) the terminase large and small subunit (PC_214 and PC_280). Moreover, a tail collar-fiber protein (PC_596), one tail protein (PC_599) and one tail assembly chaperone (PC_597) were annotated. Four lysis genes were spread over the genome, including a spanin (PC_588) and three endolysin (PC_146, PC_589 and PC_594).

In all genomes of the “Diferiteviridae” family, a DNA injection protein (PC_210) and a phosphofructokinase (PC_204) were found. DNA methylases were encoded in some of the genomes (two in ICBM121, ICBM122 and ICBM163, and one in ICBM8), which can be part of the viral anti-host defense protection (Fig. 42). Furthermore, a phosphoadenosine phosphosulfate (PAPS) reductase was annotated in the “Benvirus” genomes (Table 32), which has been found in other phages (Summer et al. 2006) and could function as an AMG giving advantage to the host under sulfur limited conditions (Summer et al. 2007b). No tRNAs were predicted in these genomes.

The genome replication strategy of the phages in this family was not easily determined. No characteristic DNA polymerase was annotated in either of the genomes. PhageTerm provided varying predictions for the members of this family (Table 15). While the detection of DTRs in the genomes of phage ICBM8, ICBM163 and ICBM165 indicated a T7-like bidirectional genome replication and packaging, the genomes of phages ICBM121 and ICBM122 seemed to be circularly permuted using P1-type rolling circle replication and the headful (pac) DNA packaging strategy. However, the latter prediction was less supported (Table 15). Considering that ICBM121, ICBM122 and ICBM165 belonged to the same genus

and should thus have similar packaging strategies, the phageTerm results should be interpreted with caution. According to Virfam prediction and the gene annotation (absence of major tail protein, tail completion protein and sheath), the phages in the “Diferiteviridae” family had a podoviral morphology with a short tail (Table 31). The genomes of the “Benvirus” genus (ICBM121, ICBM122, and ICBM163) had a gene encoded annotated as “excisionase and transcriptional regulator” (PC_317). In phage ICBM163, a lambda repressor-like protein (PC_528) could be annotated next to the excisionase. In phage ICBM8, only the lambda repressor-like protein was found (PC_575). Therefore, these phages could have lysogenic potential. The five phages of this family shared 12 core proteins (Table 35). Eight of them had an annotated function, with most of them being morphology proteins.

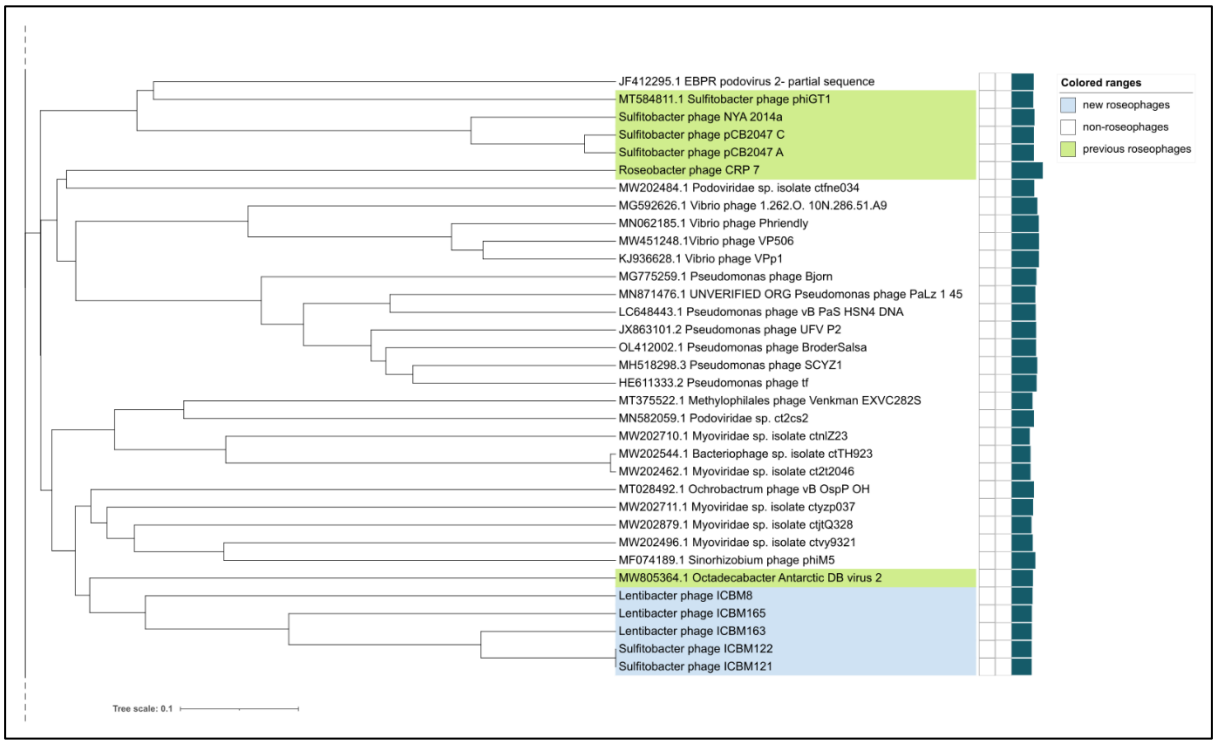


Fig. 39: Section of the whole-genome based proteomic tree showing VGC_19. Names of cultivated roseophages are marked in green (previous) and blue (this study). Genome lengths are displayed as bar chart.

Table 31: Phages of VGC_19. Roseophages are marked in blue. Bp = base pairs. N.a. = not available. *Morphology predicted by Virfam.

Phage	Host	Genome size [bp]	GC content [%]	Accession	Previous classification (reference)	ICTV or newly proposed taxonomy			Morphology	Isolation source	Reference
						Family	Genus	Species			
EBPR podovirus 2	n.a.	40628	63.9	JF412295	podoviridae					bioreactor	Skennerton et al. (2011)
Sulfitobacter phage phiGT1	*Sulfitobacter* sp. HGT1	40019	56.4	MT584811	unclassified				podoviral	coastal sediment	Hwang et al. (2020)
Sulfitobacter phage NYA 2014a	*Sulfitobacter* sp. 2047	42092	58.5	KM233261	unclassified				podoviral	n.a.	unpublished
Sulfitobacter phage pCB2047_A	*Sulfitobacter* sp. 2047	40929	58.8	HQ332142	Podoviridae, cluster 3 (Zhan and Chen 2019a)				podoviral	algal bloom mesocosm (seawater)	Ankrah et al. (2014b)
Sulfitobacter phage pCB2047_C	*Sulfitobacter* sp. 2047	40931	59	HQ317384					podoviral	algal bloom mesocosm (seawater)	
Roseobacter phage CRP-7	*Roseobacter* RCA FZCC0042	58106	40.3	MK613349	CRP-7-type (Zhang et al. 2019a)				podoviral	seawater	Zhang et al. (2019a)
Podoviridae sp. isolate ctfne034	n.a.	41762	45	MW202484						freshwater river	unpublished
Vibrio phage 1.262.O. 10N.286.51.A9	*Vibrio breoganii*	47635	39.1	MG592626					podoviral	seawater	unpublished
Vibrio phage Phriendly	*Vibrio natriegens* ATCC 14048	50218	41.4	MN062185					podoviral	seawater	Clark et al. (2019)
Vibrio phage VP506	*Vibrio parahaemolyticus* VP506	50577	41.6	MW451248						n.a.	unpublished
Vibrio phage VPp1	*Vibrio parahaemolyticus* VP 17802	50431	41.3	KJ936628						n.a.	Li et al. (2018)
Pseudomonas phage Bjorn	*Pseudomonas* sp. CT12	45936	53.1	MG775259			*Bjornvirus*			plant compost	unpublished
Pseudomonas phage PaLz 1_45	*Pseudomonas aeruginosa* PAO1	43890	52.1	MN871476						seawater	unpublished
Pseudomonas phage vB_PaS_HSN4	*Pseudomonas aeruginosa*	44534	53.5	LC648443						n.a.	unpublished
Pseudomonas phage UFV P2	*Pseudomonas fluorescens*	45517	51.5	JX863101			*Vicosavirus*			wastewater	Eller et al. (2013)
Pseudomonas phage BroderSalsa	*Pseudomonas* sp. ERG2	45307	51.9	OL412002						wastewater	unpublished
Pseudomonas phage SCYZ1	*Pseudomonas fluorescens*	47475	52.7	MH518298					podoviral	soil	unpublished
Pseudomonas phage tf	*Pseudomonas putida* PpG1	46271	53.2	HE611333			*Krylovvirus*			n.a.	Glukhov et al. (2012)

Methylophilales phage Venkman EXVC282S	*Methylophilales bacterium* H5P1	38624	34.4	MT375522					podoviral	seawater	Buchholz et al. (2021)
Podoviridae sp. ct2cs2	n.a.	41219	35.7	MN582059						freshwater river	unpublished
Myoviridae sp. isolate ctnlZ23	n.a.	33284	48.9	MW202710						freshwater river	unpublished
Bacteriophage sp. isolate ctTH923	n.a.	34712	46.9	MW202544						freshwater river	unpublished
Myoviridae sp. isolate ct2t2046	n.a.	34712	46.9	MW202462						freshwater river	unpublished
Ochrobactrum phage vB_OspP_OH	*Ochrobactrum* sp. POC9	41227	55.2	MT028492			*Wolominvirus*		podoviral	sewage	Decewicz et al. (2020)
Myoviridae sp. isolate ctyzp037	n.a.	39701	51.3	MW202711						freshwater river	unpublished
Myoviridae sp. isolate ctjtQ328	n.a.	36472	55.8	MW202879						freshwater river	unpublished
Myoviridae sp. isolate ctvy9321	n.a.	38949	57.5	MW202496						freshwater river	unpublished
Sinorhizobium phage phiM5	*Sinorhizobium meliloti*	44005	61	MF074189					podoviral	induced prophage	Johnson et al. (2017)
Octadecabacter Antarctic DB virus 2	*Octadecabacter* sp. IceBac 430	39241	53.3	MW805364	*Caudoviricetes (Demina et al. 2021)*				podoviral	antarctic sea ice	Luhtanen et al. (2018)
Lentibacter phage ICBM8	*Lentibacter* sp. SH36	38666	55.9	This study		“Diferiteviridae”	“Maryvirus”	“opt”	podoviral*	Seawater (HE504-33)	this study
Lentibacter phage ICBM165	*Lentibacter* sp. SH36	37385	57.3	This study			“Martinvirus”	“patru”	podoviral*	Seawater (HE504-33)	this study
Lentibacter phage ICBM163	*Lentibacter* sp. SH36	37200	57	This study				“unu”	podoviral*	Seawater (P2)	this study
Sulfitobacter phage ICBM122	*Sulfitobacter* sp. M73	36856	57.1	This study			“Benvirus”	“doi”	podoviral*	Seawater (P2)	this study
Sulfitobacter phage ICBM121	*Sulfitobacter* sp. M73	36856	57.1	This study					podoviral*	Seawater (P2)	this study

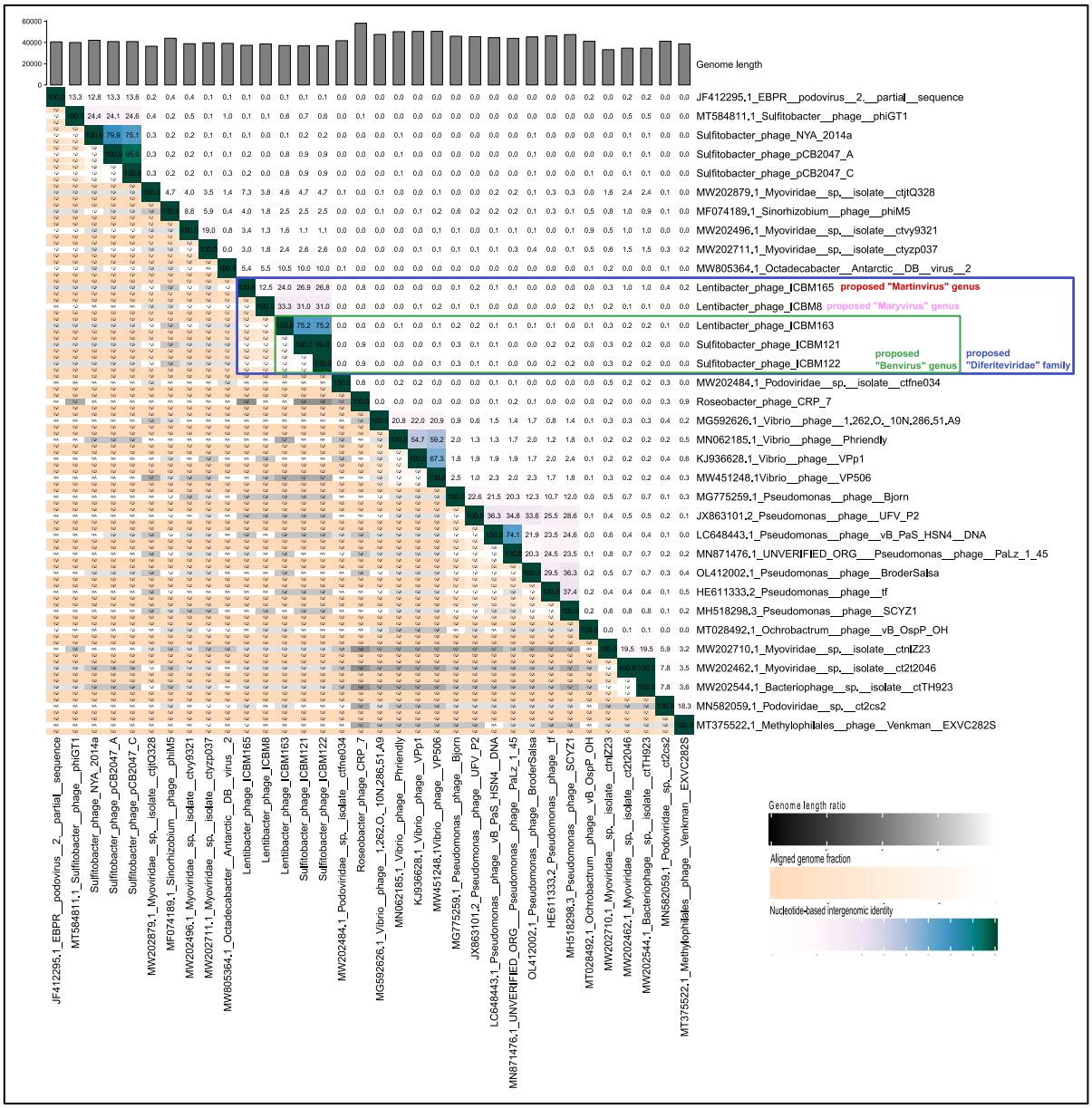


Fig. 40: Nucleotide-based intergenomic identities of the phages in VGC_19 calculated with VIRIDIC. Members of the newly proposed genera “Martinvirus”, “Maryvirus” and “Benvirus” and the new proposed “Diferiteviridae” family are annotated with boxes and colored labels.

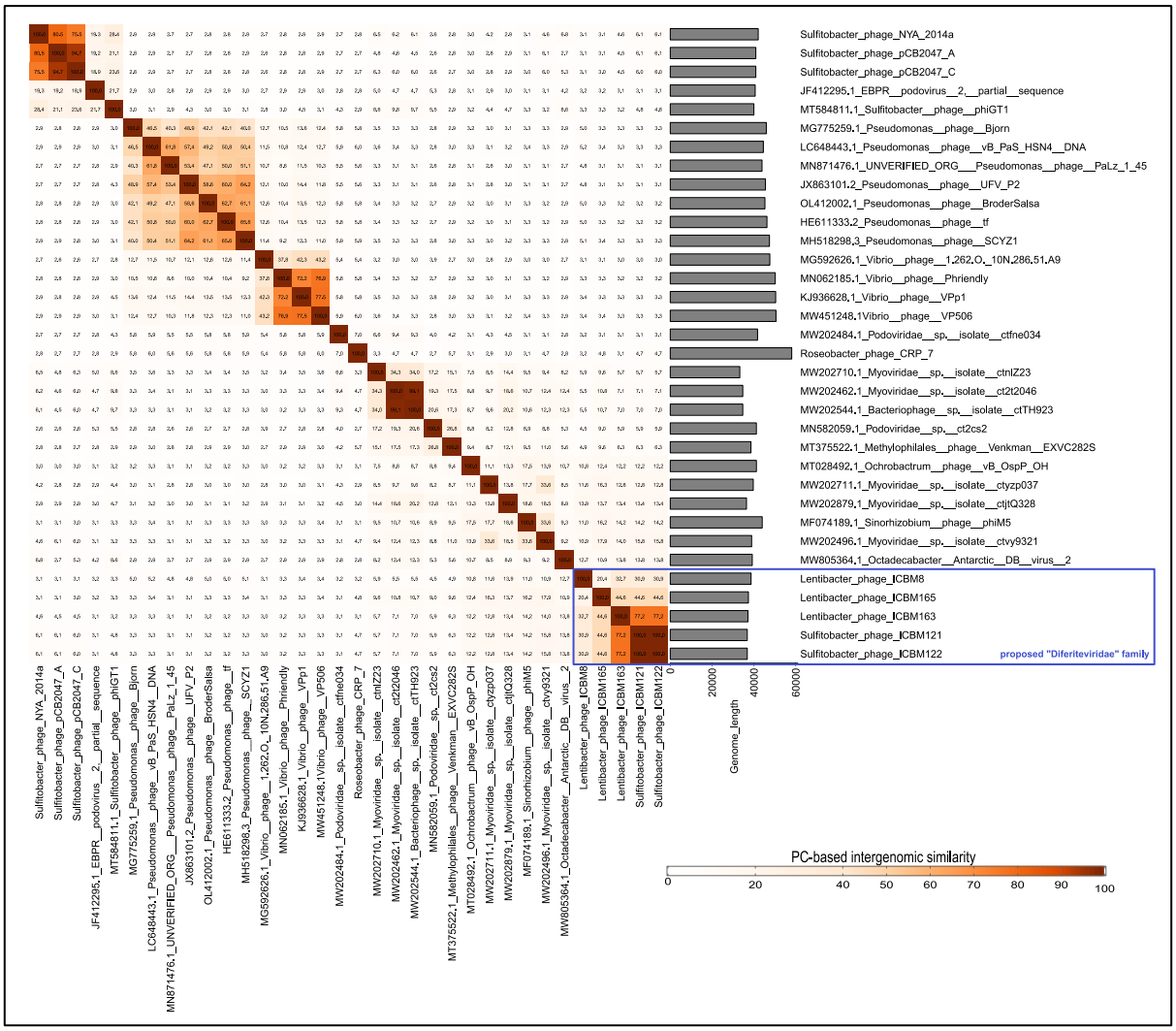


Fig. 41: PC-based intergenomic similarities of the phages in VGC_19, calculated with VirClust (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%). Members of the newly proposed “Diferiteviridae” family are annotated with a blue box.

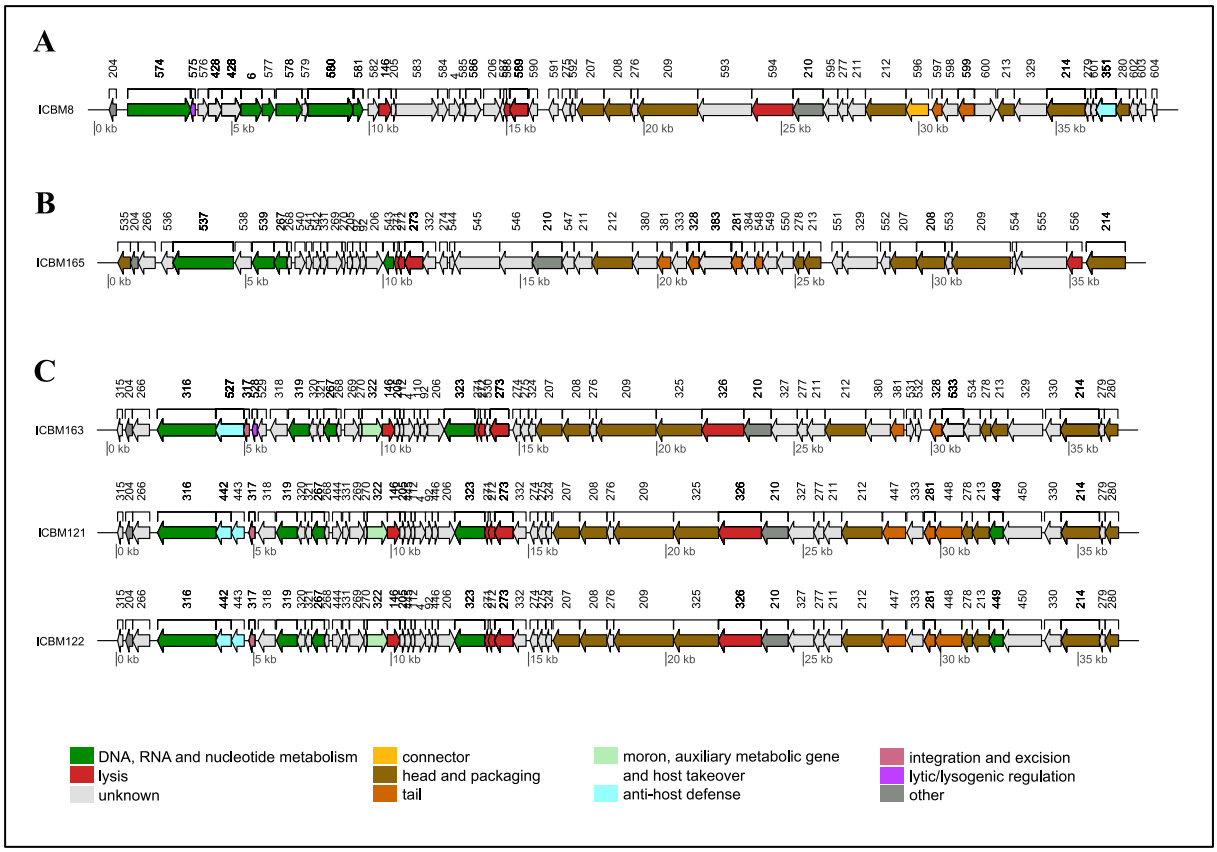


Fig. 42: Genome maps of VGC_19. A Lentibacter phage ICBM8. B Lentibacter phage ICBM165. C Phages of the "Benvirus" genus (Lentibacter phage ICBM163 and *Sulfitobacter* phages ICBM121 and ICBM122).

Table 32: Gene annotations of Lentibacter phage ICBM121. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core proteins of phages of the "Diferiteviridae" family.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	47	238	192	-1	315	hp	unknown
gene_2	338	595	258	-1	204*	Phosphofructokinase*	other
gene_3	592	1215	624	-1	266	hp	unknown
gene_4	1502	3625	2124	-1	316	DNA primase/helicase	DNA, RNA and nucleotide metabolism
gene_5	3622	4179	558	-1	442	DNA (citosine) methyltransferase	anti-host defense
gene_6	4188	4655	468	-1	443	DNA methyltransferase	anti-host defense
gene_7	4827	5042	216	-1	317	excisionase and transcriptional regulator	integration and excision
gene_8	5166	5783	618	-1	318	hp	unknown
gene_9	5828	6580	753	-1	319	nuclease	DNA, RNA and nucleotide metabolism
gene_10	6577	6867	291	-1	320	hp	unknown
gene_11	6882	7106	225	-1	321	hp	unknown
gene_12	7106	7573	468	-1	267	Holliday junction resolvase	DNA, RNA and nucleotide metabolism
gene_13	7573	7740	168	-1	268	hp	unknown
gene_14	7869	8189	321	1	444	hp	unknown
gene_15	8237	8470	234	1	331	hp	unknown
gene_16	8494	9015	522	1	269	hp	unknown

gene_17	9012	9134	123	1	270	hp	unknown
gene_18	9131	9868	738	1	322	phosphoadenosine phosphosulfate (PAPS)reductase	moron, auxiliary metabolic gene and host takeover
gene_19	9865	10323	459	1	146	endolysin (cell wall hydrolase)	lysis
gene_20	10320	10508	189	1	205*	hp*	unknown
gene_21	10505	10729	225	1	445	hp	unknown
gene_22	10726	10875	150	1	112	hp	unknown
gene_23	10868	11251	384	1	4	hp	unknown
gene_24	11251	11502	252	1	92	hp	unknown
gene_25	11495	11713	219	1	446	hp	unknown
gene_26	11713	12324	612	1	206*	hp*	unknown
gene_27	12308	13405	1098	-1	323	DNA repair exonuclease	DNA, RNA and nucleotide metabolism
gene_28	13402	13608	207	-1	271	Spanin	lysis
gene_29	13532	13786	255	-1	272	Spanin	lysis
gene_30	13765	14436	672	-1	273	Amidase	lysis
gene_31	14436	14912	477	-1	332	hp	unknown
gene_32	15058	15336	279	-1	274	hp	unknown
gene_33	15336	15644	309	-1	275	hp	unknown
gene_34	15619	15840	222	-1	324	hp	unknown
gene_35	15895	16851	957	-1	207*	major capsid protein*	head and packaging
gene_36	16873	17853	981	-1	208*	capsid scaffolding protein*	head and packaging
gene_37	17858	18100	243	-1	276	hp	unknown
gene_38	18097	20274	2178	-1	209*	portal protein*	head and packaging
gene_39	20285	21913	1629	-1	325	major capsid protein	head and packaging
gene_40	21916	23475	1560	-1	326	endolysin (D-alanyl-D-alanine carboxypeptidase)	lysis
gene_41	23475	24449	975	-1	210*	DNA injection protein*	other
gene_42	24451	25386	936	-1	327	hp	unknown
gene_43	25386	25748	363	-1	277*	hp*	unknown
gene_44	25741	26397	657	-1	211*	hp*	unknown
gene_45	26394	27869	1476	-1	212*	virion structural protein*	head and packaging
gene_46	27893	28723	831	-1	447	tail protein	tail
gene_47	28768	29358	591	-1	333	hp	unknown
gene_48	29400	29798	399	-1	281	Phage tail assembly chaperone	tail
gene_49	29800	30768	969	-1	448	tail fiber protein	tail
gene_50	30772	31152	381	-1	278	capsid protein	head and packaging
gene_51	31152	31772	621	-1	213*	virion structural protein*	head and packaging
gene_52	31763	32272	510	-1	449	HNH endonuclease	DNA, RNA and nucleotide metabolism
gene_53	32272	33672	1401	-1	450	hp	unknown
gene_54	33778	34386	609	-1	330	hp	unknown
gene_55	34386	35780	1395	-1	214*	terminase large subunit*	head and packaging
gene_56	35777	35989	213	-1	279	hp	unknown
gene_57	35982	36476	495	-1	280	terminase small subunit	head and packaging

Table 33: Gene annotations of Lentibacter phage ICBM165. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core proteins of phages of the "Diferiteviridae" family.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	345	806	462	-1	535	terminase small subunit	head and packaging
gene_2	806	1087	282	-1	204*	Phosphofructokinase*	other
gene_3	1090	1707	618	-1	266	hp	unknown
gene_4	1932	2360	429	-1	536	hp	unknown
gene_5	2357	4552	2196	-1	537	DNA primase/helicase	DNA, RNA and nucleotide metabolism
gene_6	4613	5206	594	-1	538	hp	unknown
gene_7	5242	6033	792	-1	539	exonuclease	DNA, RNA and nucleotide metabolism
gene_8	6030	6500	471	-1	267	Holliday junction resolvase	DNA, RNA and nucleotide metabolism
gene_9	6500	6661	162	-1	268	hp	unknown
gene_10	6796	7158	363	1	540	hp	unknown
gene_11	7204	7407	204	1	541	hp	unknown
gene_12	7456	7728	273	1	542	hp	unknown
gene_13	7728	7961	234	1	331	hp	unknown
gene_14	7985	8506	522	1	269	hp	unknown
gene_15	8503	8625	123	1	270	hp	unknown
gene_16	8702	8893	192	1	205*	hp*	unknown
gene_17	8890	9171	282	1	92	hp	unknown
gene_18	9164	9382	219	1	92	hp	unknown
gene_19	9384	9980	597	1	206*	hp*	unknown
gene_20	10011	10379	369	-1	543	tRNA endonuclease	DNA, RNA and nucleotide metabolism
gene_21	10407	10592	186	-1	271	Spanin	lysis
gene_22	10540	10782	243	-1	272	Spanin	lysis
gene_23	10782	11438	657	-1	273	Amidase	lysis
gene_24	11438	11914	477	-1	332	hp	unknown
gene_25	12060	12338	279	-1	274	hp	unknown
gene_26	12423	12605	183	1	544	hp	unknown
gene_27	12602	14242	1641	-1	545	hp	unknown
gene_28	14242	15429	1188	-1	546	hp	unknown
gene_29	15431	16510	1080	-1	210*	DNA injection protein*	other
gene_30	16510	16956	447	-1	547*	hp*	unknown
gene_31	16949	17605	657	-1	211*	hp*	unknown
gene_32	17602	19074	1473	-1	212*	virion structural protein*	head and packaging
gene_33	19086	19976	891	-1	380	hp	unknown
gene_34	19976	20470	495	-1	381	tail protein	tail
gene_35	20515	21045	531	-1	333	hp	unknown
gene_36	21087	21503	417	-1	328	Phage tail assembly chaperone	tail
gene_37	21503	22684	1182	-1	383	hp	unknown
gene_38	22660	23061	402	-1	281	Phage tail assembly chaperone	tail
gene_39	23061	23525	465	-1	384	hp	unknown
gene_40	23522	23815	294	-1	548	Phage tail assembly chaperone	tail
gene_41	23815	24336	522	-1	549	hp	unknown
gene_42	24338	24928	591	-1	550	hp	unknown

gene_43	24932	25312	381	-1	278	capsid protein	head and packaging
gene_44	25312	25932	621	-1	213*	virion structural protein*	head and packaging
gene_45	26333	26716	384	-1	551	hp	unknown
gene_46	26716	27984	1269	-1	329	hp	unknown
gene_47	28103	28447	345	-1	552	hp	unknown
gene_48	28450	29406	957	-1	207*	major capsid protein*	head and packaging
gene_49	29418	30437	1020	-1	208*	capsid scaffolding protein*	head and packaging
gene_50	30459	30713	255	-1	553	hp	unknown
gene_51	30694	32829	2136	-1	209*	portal protein*	head and packaging
gene_52	32898	33053	156	1	554	hp	unknown
gene_53	33050	34879	1830	-1	555	hp	unknown
gene_54	34879	35436	558	-1	556	endolysin	lysis
gene_55	35595	37016	1422	-1	214*	terminase large subunit*	head and packaging

Table 34: Gene annotations of Lentibacter phage ICBM8. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Proteins shared between ICBM8 and the phages of the “Diferiteviridae” family.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	535	798	264	-1	204*	Phosphofructokinase*	other
gene_2	1204	3510	2307	1	574	DNA primase/polymerase	DNA, RNA and nucleotide metabolism
gene_3	3488	3670	183	-1	575	Lambda repressor-like	lytic/lysogenic regulation
gene_4	3769	4149	381	1	576	hp	unknown
gene_5	4149	4628	480	1	428	hp	unknown
gene_6	4628	5332	705	1	428	hp	unknown
gene_7	5332	6087	756	1	6	Sak4-like ssDNA annealing protein	DNA, RNA and nucleotide metabolism
gene_8	6099	6551	453	1	577	ssDNA binding protein	DNA, RNA and nucleotide metabolism
gene_9	6610	7554	945	1	578	nuclease	DNA, RNA and nucleotide metabolism
gene_10	7551	7775	225	1	579	hp	unknown
gene_11	7772	9433	1662	1	580	DNA helicase	DNA, RNA and nucleotide metabolism
gene_12	9411	9773	363	1	581	tRNA endonuclease	DNA, RNA and nucleotide metabolism
gene_13	9960	10349	390	1	582	hp	unknown
gene_14	10346	10804	459	1	146	endolysin (cell wall hydrolase)	lysis
gene_15	10801	10977	177	1	205*	hp*	unknown
gene_16	10974	12485	1512	1	583	hp	unknown
gene_17	12499	12846	348	1	584	hp	unknown
gene_18	12907	13296	390	1	4	hp	unknown
gene_19	13296	13511	216	1	585	hp	unknown
gene_20	13508	14056	549	1	586	hp	unknown
gene_21	14164	14775	612	1	206*	hp*	unknown
gene_22	14759	14995	237	-1	587	hp	unknown
gene_23	14895	15125	231	-1	588	spanin	lysis
gene_24	15104	15769	666	-1	589	endolysin (lysozyme)	lysis
gene_25	15821	16126	306	-1	590	hp	unknown
gene_26	16536	16877	342	-1	591	hp	unknown
gene_27	17021	17308	288	-1	275	hp	unknown

gene_28	17309	17497	189	-1	592	hp	unknown
gene_29	17559	18533	975	-1	207*	major capsid protein*	head and packaging
gene_30	18547	19527	981	-1	208*	capsid scaffolding protein*	head and packaging
gene_31	19532	19774	243	-1	276	hp	unknown
gene_32	19771	21948	2178	-1	209*	portal protein*	head and packaging
gene_33	21959	23920	1962	-1	593	hp	unknown
gene_34	23922	25424	1503	-1	594	endolysin (lysozyme)	lysis
gene_35	25424	26515	1092	-1	210*	DNA injection protein*	other
gene_36	26517	27050	534	-1	595	hp	unknown
gene_37	27050	27412	363	-1	277*	hp*	unknown
gene_38	27405	28061	657	-1	211*	hp*	unknown
gene_39	28058	29533	1476	-1	212*	virion structural protein*	head and packaging
gene_40	29537	30355	819	-1	596	tail collar fiber protein	connector
gene_41	30482	30847	366	-1	597	tail assembly chaperone	tail
gene_42	30840	31424	585	-1	598	hp	unknown
gene_43	31435	32019	585	-1	599	tail protein	tail
gene_44	32036	32803	768	1	600	hp	unknown
gene_45	32871	33476	606	-1	213*	virion structural protein*	head and packaging
gene_46	33477	34655	1179	-1	329	hp	unknown
gene_47	34655	36052	1398	-1	214*	terminase large subunit*	head and packaging
gene_48	36046	36258	213	-1	279	hp	unknown
gene_49	36251	36469	219	-1	601	hp	unknown
gene_50	36453	37175	723	-1	351	DNA methylase N-4/N-6	anti-host defense
gene_51	37168	37662	495	-1	280	terminase small subunit	head and packaging
gene_52	37675	37956	282	-1	602	hp	unknown
gene_53	37946	38257	312	-1	603	hp	unknown
gene_54	38486	38666	181	-1	604	hp	unknown

Table 35: Core proteins of the newly proposed family "Diferiteviridae". *maximum number of predicted ORFs divided by number of core proteins.

predicted ORFs#	**core proteins #**	**Annotated core proteins #**	**Annotated core proteins (sorted by functional category)**	**Percentage core proteins / total ORFs***
55-57	12	8	Head and packaging: major capsid protein (PC_207), capsid scaffolding protein (PC_208), portal protein (PC_209), virion structural protein (PC_212), virion structural protein (PC_213), terminase large subunit (PC_214), Other: Phosphofructokinase (PC_204), DNA injection protein (PC_210)	21.05%

4.3.5.6. The “Woolleyviridae” - a new family of potentially temperate roseophage infecting *Sulfitobacter*

Five new roseophages formed a separate cluster within a VGC containing 58 viruses, many of them being temperate phages (Fig. 43, Table 36). The described viruses in this VGC_6 mainly infect *Gammaproteobacteria* and *Alphaproteobacteria (Rhodobacterales, Hyphomicrobiales* and *Caulobacterales)*, with many of the hosts being human pathogens. Accordingly, the majority of the viruses was isolated from terrestrial, anthropogenic environments such as sewage, clinics, dairy products or agricultural lands, but there were also members originating from freshwater rivers, a salt mine and even antarctic soil. All described phages of this VGC apart from one (Marinobacter phage AS1) were reported with a siphoviral morphology (Table 36).

The new roseophages grouping in this VGC formed a separate cluster, with a maximum PC-based intergenomic similarity to other phages of 22.5% (Fig. 45). Within the cluster, the minium PC-based intergenomic similarity accounted for 87.1%. Therefore, we propose that this cluster forms a new family, which we tentatively named here “Woolleyviridae”. Three other roseophages clustered into this VGC: Paracoccus phage Shpa, Rhodobacter phage RcapNL and Paracoccus phage vB_PthS_Pthi. They grouped in a larger cluster with our novel roseophages, but outside the “Woolleyviridae” family (Fig. 43 and 45).

The five novel roseophages in the “Woolleyviridae” family were ICBM55, ICBM111, ICBM117, ICBM118 and ICBM130 (Table 36). Apart from phage ICBM130 (from seawater sample NHS), they all originated from the same seawater sample (P2). They infected the four closely related *Sulfitobacter* strains M63, M70, M92 and M157 (Table 36). According to their nucleotide-based intergenomic identity, all five phages belonged to the same genus (NBII ≥ 88.0), forming two species (Fig. 44). We provisionally named this genus here “Viktorvirus”.

The genomes of the “Woolleyviridae” phages were of 33.4 - 34.1 kb in size, with 60.5 - 60.7% GC content and 41 - 45 predicted ORFs (Table 36). Out of these ORFs, 22 could be functionally annotated. PhageTerm predicted circularly permuted genomes with random termini at both genomes ends (ICBM55) or with one defined end (other phages) (Table 15). The genomic architecture was highly conserved amongst the members of this family and the genes that could be annotated were almost identical (Fig. 46, Table 37). The genes were arranged in functional modules. The “DNA, RNA and nucleotide metabolism” module comprised a ssDNA binding protein (PC_235) and two DNA primases (PC_224 and PC_227). No DNA polymerase was annotated. “Head and packaging” genes included the capsid

maturation protease (PC_253), the major capsid protein (PC_252), the portal protein (PC_254) and the terminase small and large subunit (PC_211 and PC_255). Further morphogenesis genes comprised the head-tail adaptor Ad1 (PC_250), the head-tail adaptor (PC_249), the neck protein Ne1 (PC_248), the tail completion protein Tc1 (PC_247), a major tail protein (PC_246), and the tail length tape measure protein (PC_243) (Table 37). One large gene in the morphology module was either annotated as a tail fiber protein (PC_304 in ICBM55, ICBM117, and ICBM118) or as an esterase/lipase (PC_359 in ICBM111 and ICBM130) (Fig. 46). Two lysis genes were encoded in the genomes: a spanin (PC_236) and an endolysin (PC_239). In addition, the genomes contained a deoxyribonucleoside 5' monophosphate phosphatase (PC_225) and a transcriptional regulator (PC_228). Phages ICBM55, ICBM117 and ICBM118 encoded a DNA methylase in their genomes, which can be part of the resistance against restriction-modification systems. No tRNAs could be predicted in the genomes of these phages.

Genome end determination with PhageTerm indicated that these phages use a headful (pac) packaging strategy and are circularly permuted (Table 15). Only for Sulfitobacter phage ICBM55, no packaging type could be assigned. However, it had random termini at both ends and was circularly permuted. Thus, it might use the headful packaging without a specific pac site (see chapter 1.3.4.). A siphoviral morphology was predicted by Virfam and reflected by the annotation of the major tail protein and the tail completion protein Tc1 (Table 37). A Lambda repressor like protein and an integrase were annotated in all genomes of the “Woolleyviridae” family, which indicates that these phages are capable to follow the lysogenic pathway. The phages of the “Woolleyviridae” family shared 37 core proteins, with 20 of them being functionally annotated (Table 38).

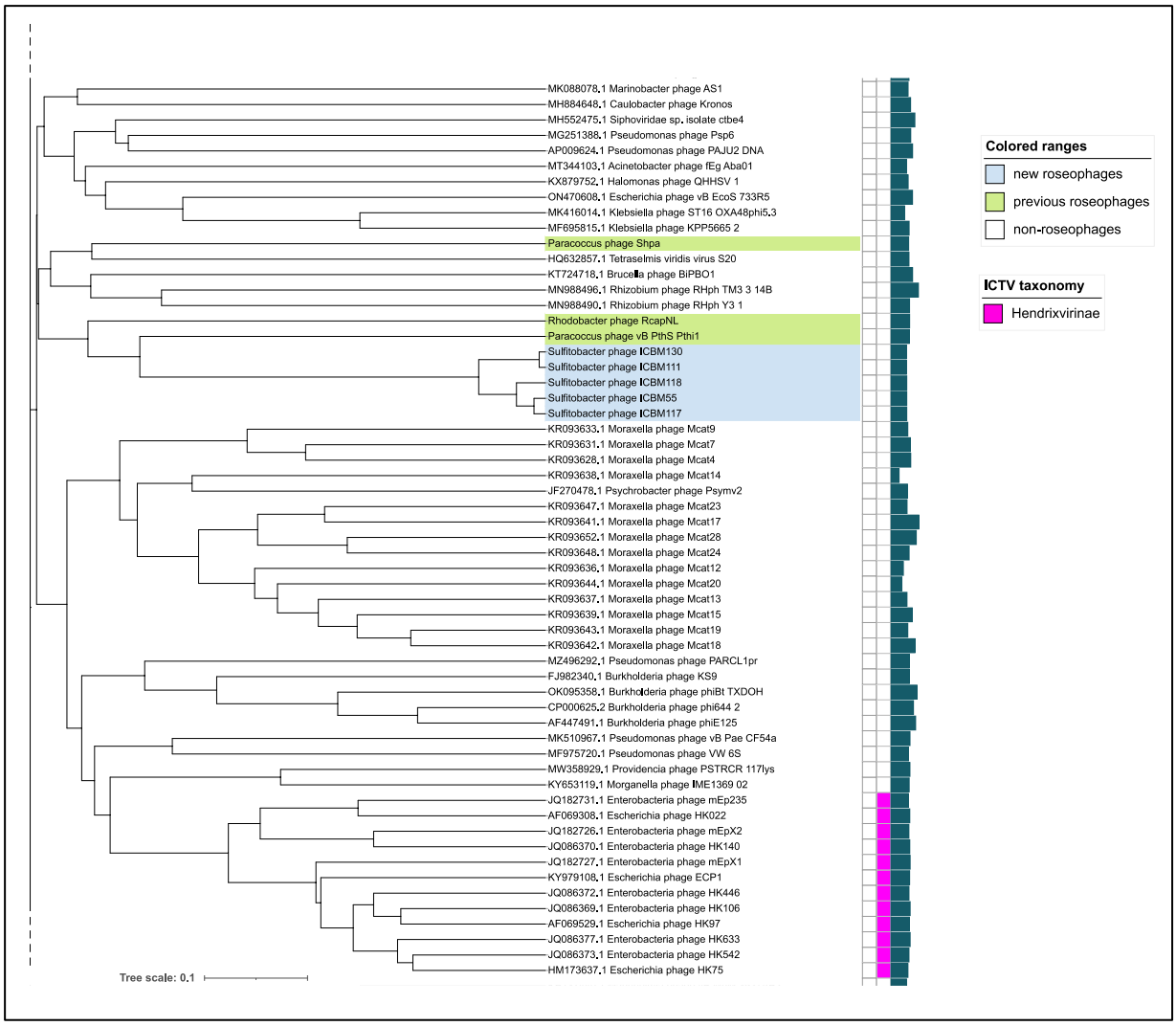


Fig. 43: Section of the whole-genome based proteomic tree showing VGC_6. Names of cultivated roseophages are marked in green (previous) and blue (this study). Color-strips indicate affiliation to ICTV-recognized families and subfamilies, respectively. Genome lengths are displayed as bar chart.

Table 36: Phages of VGC_6. Roseophages are marked in blue. Bp = base pairs. N.a. = not available. *Morphology predicted by Virfam.

Phage	Host	Genome size [bp]	GC content [%]	Accession	Previous classification	ICTV or newly proposed taxonomy			Morphology	Isolation source	Reference
						Family/Subfamily	Genus	Species			
Marinobacter phage AS1	*Marinobacter* sp. D1S9	36994	57.00%	MK088078					podoviral	seawater	Aparna et al. (2019)
Caulobacter phage Kronos	*Caulobacter* sp. CBR1	42424	66.30%	MH884648					siphoviral	rhizosphere	Berrios and Ely (2019)
Siphoviridae sp. isolate ctbe4	n.a.	51816	55.80%	MH552475						nematode tissue metagenome	unpublished
Pseudomonas phage Psp6	*Pseudomonas putida*	42789	57.10%	MG251388						freshwater river	unpublished
Pseudomonas phage PAJU2	*Pseudomonas aeruginosa*	46872	56.30%	AP009624					siphoviral	freshwater river	Uchiyama et al. (2009)
Acinetobacter phage fEg_Aba01	*Acinetobacter baumannii* DSM 106838	33779	40.10%	MT344103					siphoviral	prophage, sewage	Badawy et al. (2020)
Halomonas phage QHHSV_1	*Halomonas ventosae*	37270	66.80%	KX879752						salt mine	Fu et al. (2017)
Escherichia phage vB_EcoS_733R5	*Escherichia coli*	46651	50.20%	ON470608					siphoviral	prophage	Lopez-Diaz et al. (2022)
Klebsiella phage ST16_OXA48phi5.3	*Klebsiella pneumoniae* ST16-OXA48	29301	50.20%	MK416014					siphoviral	prophage, clinical strain	Bleriot et al. (2020)
Klebsiella phage KPP5665_2	*Klebsiella pneumoniae*	39241	51.60%	MF695815					siphoviral	prophage, milk	Carl et al. (2017)
Paracoccus phage Shpa	*Paracoccus* sp. HS3	38261	64.70%	KR072689	*Caudoviricetes*, *Vhulanivirus* (van Zyl et al. 2018)		*Vhulanivirus*		siphoviral	soda lake	van Zyl et al. (2016)
Tetraselmis viridis virus S20	n.a.	38987	61.20%	HQ632857						seawater	Stepanova et al. (2013)
Brucella phage BiPBO1	*Brucella inopinata* BO1	46877	53.30%	KT724718					siphoviral	prophage	Hammerl et al. (2016)
Rhizobium phage RHph_TM3_3_14B	*Rhizobium etli* T3_14	59457	58.90%	MN988496						agricultural lands	unpublished
Rhizobium phage RHph_Y3_1	*Rhizobium phaseoli* Y3_1	40173	58.50%	MN988490						agricultural lands	unpublished
Rhodobacter phage RcapNL	*Rhodobacter capsulatus SB1003*	40489	65.10%	JQ066768	unclassified		*Capnelvirus*			n.a.	unpublished
Paracoccus phage vB_PthS_Pthi1	*Paracoccus thiocyanatus* JCM 20756	39547	63.80%	MK291444	Siphoviridae (Decewicz et al. 2019)				siphoviral	prophage	Decewicz et al. (2019)
Sulfitobacter phage ICBM130	*Sulfitobacter* sp. M92	33346	60.70%	this study		"Woolleyviridae"	"Viktorvirus"	"dva"	siphoviral*	Seawater (NHS)	this study

Sulfitobacter phage ICBM111	*Sulfitobacter* sp. M63	33588	60.70%	this study				siphoviral *	Seawater (P2)	
Sulfitobacter phage ICBM118	*Sulfitobacter* sp. M70	34114	60.50%	this study				siphoviral *	Seawater (P2)	
Sulfitobacter phage ICBM55	*Sulfitobacter* sp. M157	34080	60.50%	this study			"adin"	siphoviral *	Seawater (P2)	
Sulfitobacter phage ICBM117	*Sulfitobacter* sp. M70	34081	60.50%	this study				siphoviral *	Seawater (P2)	
Moraxella phage Mcat9	*Moraxella catarrhalis*	36406	43.90%	KR093633						
Moraxella phage Mcat7	*Moraxella catarrhalis*	42027	43.60%	KR093631	siphoviridae				prophage, clininal strain	Ariff et al. (2015)
Moraxella phage Mcat4	*Moraxella catarrhalis*	42819	43.20%	KR093628						
Moraxella phage Mcat14	*Moraxella catarrhalis*	17422	43.40%	KR093638						
Psychrobacter phage Psymv2	*Psychrobacter* sp. MV2	35725	44.50%	JF270478				siphoviral	antarctic soil	Meiring et al. (2012)
Moraxella phage Mcat23	*Moraxella catarrhalis*	34758	42.60%	KR093647						
Moraxella phage Mcat17	*Moraxella catarrhalis*	60303	42.50%	KR093641						
Moraxella phage Mcat28	*Moraxella catarrhalis*	54840	42.20%	KR093652						
Moraxella phage Mcat24	*Moraxella catarrhalis*	39290	42.80%	KR093648						
Moraxella phage Mcat12	*Moraxella catarrhalis*	26845	43.60%	KR093636						
Moraxella phage Mcat20	*Moraxella catarrhalis*	23629	42.70%	KR093644	siphoviridae				prophage, clininal strain	Ariff et al. (2015)
Moraxella phage Mcat13	*Moraxella catarrhalis*	34664	42.80%	KR093637						
Moraxella phage Mcat15	*Moraxella catarrhalis*	46083	42.70%	KR093639						
Moraxella phage Mcat19	*Moraxella catarrhalis*	36471	42.40%	KR093643						
Moraxella phage Mcat18	*Moraxella catarrhalis*	51889	42.20%	KR093642						
Pseudomonas phage PARCL1pr	*Pseudomonas* sp. PARCI1	39565	63.00%	MZ496292					prophage, marine sponge	Oliveira et al. (2022)
Burkholderia phage KS9	*Burkholderia pyrrocinia* LMG 21824	39896	60.70%	FJ982340				siphoviral	prophage	Seed and Dennis (2005)
Burkholderia phage phiBt_TXDOH	*Burkholderia pseudomallei* BP82	56453	60.30%	OK095358					prophage	unpublished
Burkholderia phage phi644_2	*Burkholderia pseudomallei*	48674	60.40%	CP000625		*Stanholtvirus*			n.a.	unpublished
Burkholderia phage phiE125	*Burkholderia* sp.	53373	61.20%	AF447491				siphoviral	prophage	Woods et al. (2002)
Pseudomonas phage vB_Pae_CF54a	*Pseudomonas aeruginosa*	41186	61.20%	MK510967					Prophage, lung of cystic	Tariq et al. (2019)

								fibrosis patients	
Pseudomonas phage VW_6S	*Pseudomonas fluorescens* W-6	37917	56.90%	MF975720				freshwater river	unpublished
Providencia phage PSTRCR_117lys	n.a.	41059	49.40%	MW358929				sewage	unpublished
Morganella phage IME1369_02	*Morganella morganii* IME1369	39387	50.00%	KY653119				prophage	unpublished
Enterobacteria phage mEp235	*Escherichia coli*	37595	50.00%	JQ182731		*Nochtlivirus*		n.a.	unpublished
Escherichia phage HK022	*Escherichia coli*	40751	49.50%	AF069308		*Shamshuipovirus*	siphoviral	Prophage, sewage	Juhala et al. (2000), Dhillon et al. (1980b)
Enterobacteria phage mEpX2	*Escherichia coli*	38759	50.10%	JQ182726				n.a.	unpublished
Enterobacteria phage HK140	*Escherichia coli*	40710	49.90%	JQ086370		*Yautsimvirus*		n.a.	unpublished
Enterobacteria phage mEpX1	*Escherichia coli*	41567	49.30%	JQ182727		*Cuauhtlivirus*		n.a.	unpublished
Escherichia phage ECP1	*Escherichia coli* MG1655	40469	51.00%	KY979108		*Wongtaivirus*		Prophage, chicken feces	unpublished
Enterobacteria phage HK446	*Escherichia coli*	39026	50.10%	JQ086372	*Hendrixvirinae*	*Kwaitsingvirus*		n.a.	unpublished
Enterobacteria phage HK106	*Escherichia coli*	41468	49.30%	JQ086369		*Wanchaivirus*		Prophage, sewage	Dhillon et al. (1980b)
Escherichia phage HK97	*Escherichia coli*	39732	49.80%	AF069529		*Byrnievirus*	siphoviral	Prophage, pig dung	Juhala et al. (2000), Dhillon et al. (1980a)
Enterobacteria phage HK633	*Escherichia coli*	41528	49.70%	JQ086377		*Saikungvirus*		n.a.	unpublished
Enterobacteria phage HK542	*Escherichia coli*	38964	50.90%	JQ086373		*Wongtaivirus*		Prophage, sewage	Dhillon et al. (1980b)
Escherichia phage HK75	*Escherichia coli* K-12	36661	50.20%	HM173637		*Saikungvirus*		Prophage, mammal feces	Dhillon et al. (1976)

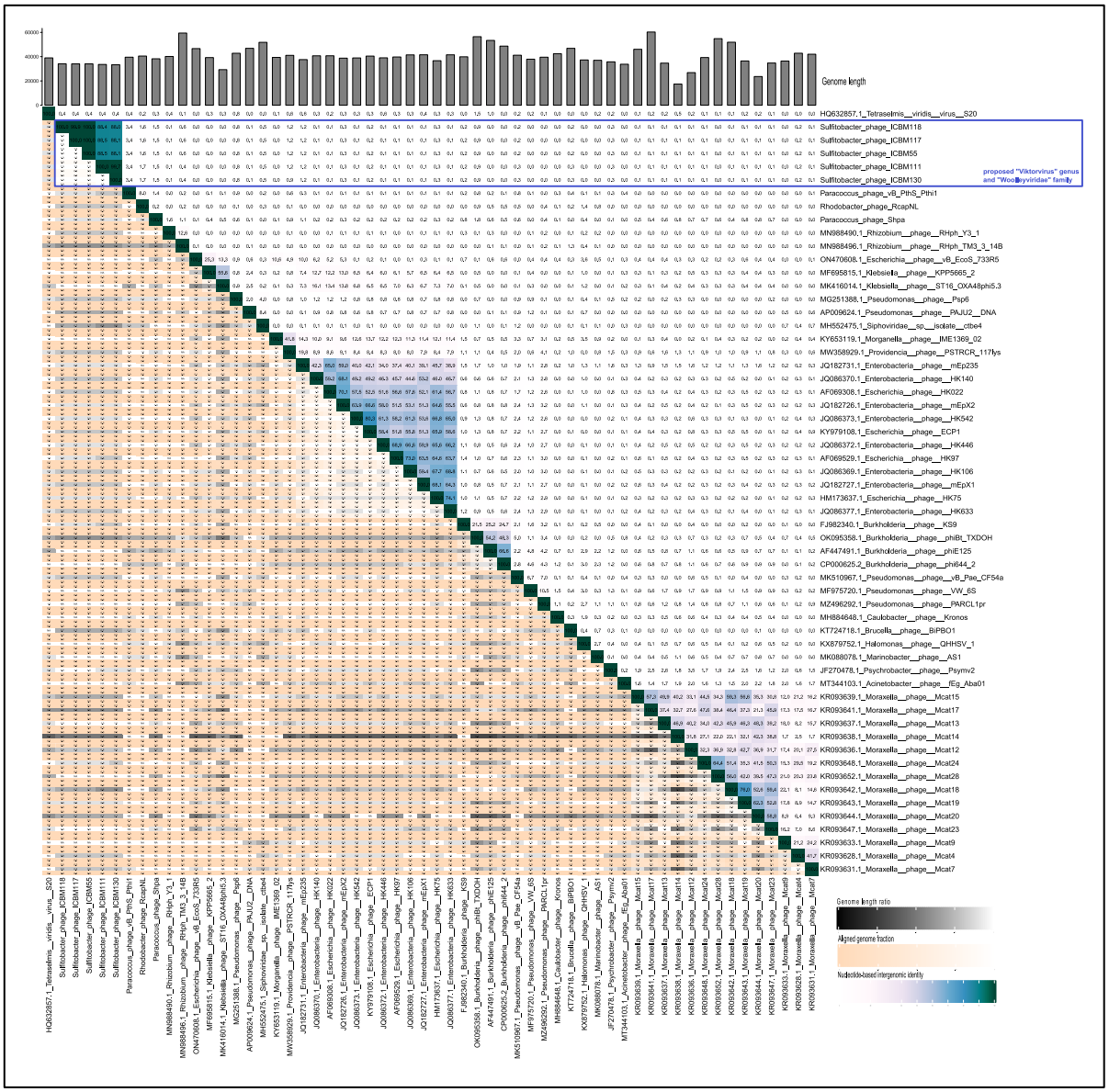


Fig. 44: Nucleotide-based intergenomic identities of the phages in VGC_6, calculated with VIRIDIC. Members of the newly proposed "Viktorvirus" genus and "Woolleyviridae" family are annotated with a blue box. Intergenomic identity of *Sulfitobacter* phages ICBM117 and ICBM55 is 99.997 %, rounded up to 100 % in the heatmap.

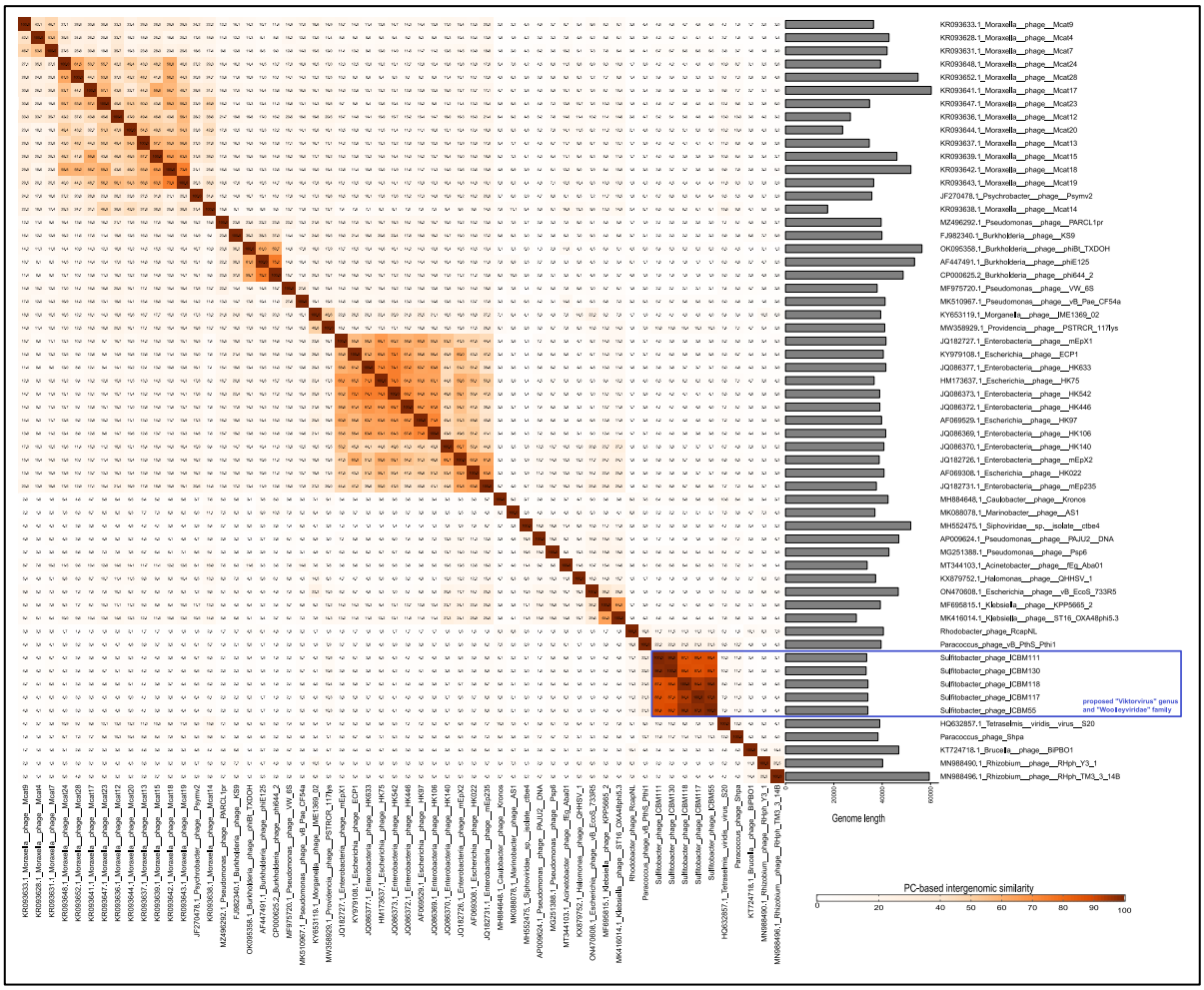


Fig. 45: PC-based intergenomic similarities of the phages in VGC_6, calculated with VirClust (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%). Members of the newly proposed "Viktorvirus" genus and "Woolleyviridae" family are annotated with a blue box.

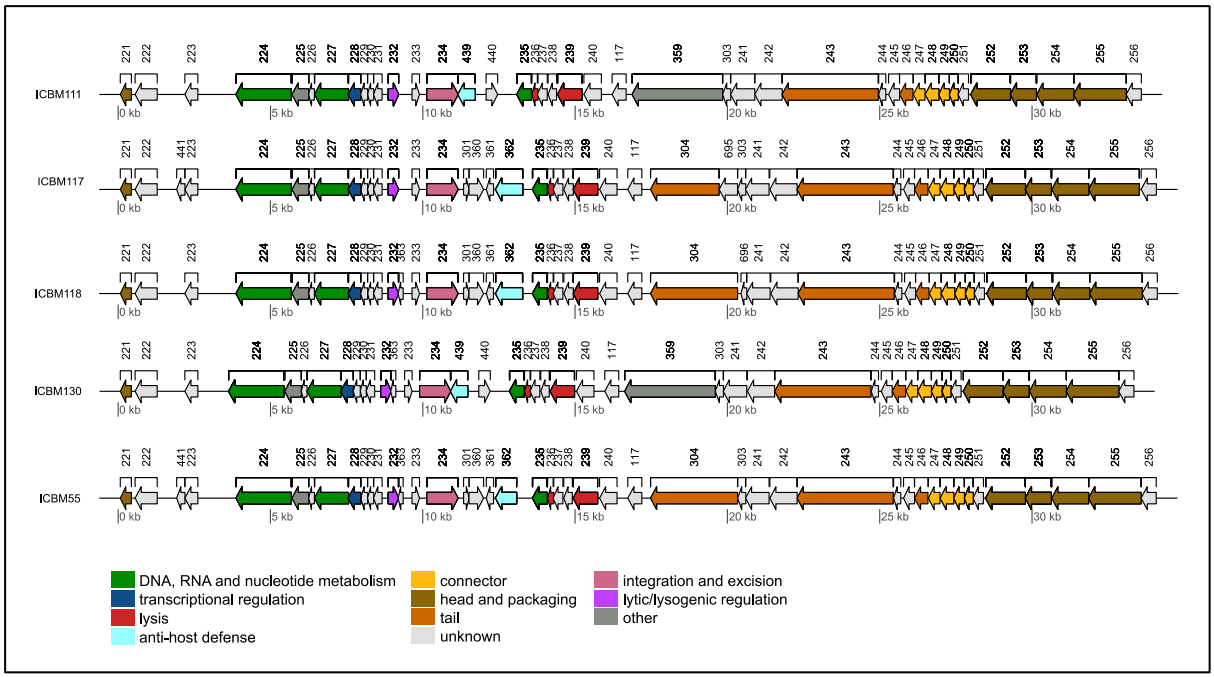


Fig. 46: Genome map of *Sulfitobacter* phages ICBM111, ICBM117, ICBM118, ICBM130 and ICBM55, members of the "Viktorvirus" genus.

Table 37: Gene annotations of Sulfitobacter phage ICBM55. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core protein of the phages in "Woolleyviridae" family.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	69	434	366	-1	221*	terminase small subunit*	head and packaging
gene_2	557	1282	726	-1	222*	hp*	unknown
gene_3	1923	2192	270	-1	441	hp	unknown
gene_4	2189	2617	429	-1	223*	hp*	unknown
gene_5	3870	5684	1815	-1	224*	DNA primase*	DNA, RNA and nucleotide metabolism
gene_6	5710	6258	549	-1	225*	deoxyribonucleoside 5' monophosphate phosphatase*	other
gene_7	6255	6443	189	-1	226*	hp*	unknown
gene_8	6448	7563	1116	-1	227*	DNA primase*	DNA, RNA and nucleotide metabolism
gene_9	7553	7963	411	-1	228*	transcription regulator*	transcriptional regulation
gene_10	7960	8190	231	-1	229*	hp*	unknown
gene_11	8187	8390	204	-1	230*	hp*	unknown
gene_12	8383	8664	282	-1	231*	hp*	unknown
gene_13	8868	9206	339	1	232*	Lambda repressor-like*	lytic/lysogenic regulation
gene_14	9191	9358	168	-1	363	hp	unknown
gene_15	9647	9892	246	1	233*	hp*	unknown
gene_16	10137	11159	1023	1	234*	Integrase*	integration and excision
gene_17	11339	11521	183	1	301	hp	unknown
gene_18	11518	11997	480	1	360	hp	unknown
gene_19	12084	12323	240	-1	361	hp	unknown
gene_20	12390	13079	690	-1	362	DNA methylase (N-4/N-6 adenine)	anti-host defense
gene_21	13604	14083	480	-1	235*	ssDNA binding protein*	DNA, RNA and nucleotide metabolism

gene_22	14097	14300	204	-1	236*	Spanin*	lysis
gene_23	14278	14601	324	-1	237*	hp*	unknown
gene_24	14630	14908	279	-1	238*	hp*	unknown
gene_25	14941	15738	798	-1	239*	endolysin (lysozyme)*	lysis
gene_26	15803	16378	576	-1	240*	hp*	unknown
gene_27	16731	17192	462	-1	117*	hp*	unknown
gene_28	17478	20333	2856	-1	304	tail fiber protein	tail
gene_29	20335	20604	270	-1	303*	hp*	unknown
gene_30	20595	21368	774	-1	241	hp	unknown
gene_31	21392	22291	900	-1	242*	hp*	unknown
gene_32	22288	25443	3156	-1	243*	tail length tape measure protein*	tail
gene_33	25456	25692	237	-1	244*	hp*	unknown
gene_34	25782	26144	363	-1	245*	hp*	unknown
gene_35	26141	26584	444	-1	246*	major tail protein*	tail
gene_36	26596	26979	384	-1	247*	tail completion Tc1*	connector
gene_37	26984	27433	450	-1	248*	neck protein Ne1*	connector
gene_38	27430	27780	351	-1	249*	head-tail adaptor*	connector
gene_39	27780	28070	291	-1	250*	head-tail adaptor Ad1*	connector
gene_40	28067	28405	339	-1	251*	hp*	unknown
gene_41	28480	29784	1305	-1	252*	major capsid protein*	head and packaging
gene_42	29781	30635	855	-1	253*	capsid maturation protease*	head and packaging
gene_43	30649	31863	1215	-1	254*	portal protein*	head and packaging
gene_44	31860	33584	1725	-1	255*	terminase large subunit*	head and packaging
gene_45	33581	34080	500	-1	256*	hp*	unknown

Table 38: Core proteins of the newly proposed family "Woolleyviridae". *maximum number of predicted ORFs divided by number of core proteins.

predicted ORFs#	**core proteins #**	**Annotated core proteins #**	**Annotated core proteins (sorted by functional category)**	**Percentage core proteins / total ORFs***
41-45	37	20	DNA, RNA and nucleotide metabolism: DNA primase (PC_224 and PC_227), ssDNA binding protein (PC_235), Head and packaging: terminase small subunit (PC_221), major capsid protein (PC_252), capsid maturation protease (PC_253), portal protein (PC_254), terminase large subunit (PC_255), Connector: tail completion Tc1 (PC_247), neck protein Ne1 (PC_248), head-tail adaptor (PC_249), head-tail adaptor Ad1 (PC_250), Tail: tail length tape measure protein (PC_243), major tail protein (PC_246), Lysis: Spanin (PC_236), endolysin (lysozyme) (PC_239), Lytic/lysogenic regulation: Lambda repressor-like (PC_232), Integration and excision: Integrase (PC_234), Transcriptional regulation: transcription regulator (PC_228), Other: deoxyribonucleoside 5' monophosphate phosphatase (PC_225)	82.22%

4.3.5.7. The *Zobellviridae* family is extended by a new roseophage species

One of the new roseophage isolates clustered into a VGC corresponding to the previously characterized *Zobellviridae* family (VGC_8) (Fig. 47). The isolation of the cobaviruses (*Lentibacter* phages ICBM1, ICBM2, and ICBM3) and the creation of the family *Zobellviridae* have been described in chapters 2 and 3. Meanwhile, the order *Caudovirales* has been abolished and the *Zobellviridae* now belongs to the newly created class *Caudoviricetes* (Adriaenssens et al. 2021). Moreover, the binomial naming format for bacteriophages has been introduced, which also changed the species designations for the *Zobellviridae* members (Table 39) (Turner et al. 2020). The described members of the *Zobellviridae* infect *Alphaproteobacteria* (*Lentibacter, Roseobacter,* and *Celeribacter*) and *Gammaproteobacteria* (Table 39). They originate from diverse habitats, ranging from seawater to rhizosphere soil, saline lakes and hospital associated samples. For all described members, a podoviral morphology was attested.

The novel roseophage isolate grouped into the *Cobavirinae* subfamily, which had a maximum PC-based intergenomic similarity with other viruses in the family of 16.0% (Fig. 49). Within the subfamily, the minimum PC-based intergenomic similarity was 33.0%. Therefore, it could become a family of its own, once future classification procedure split the *Zobellviridae* into several families. The new roseophage within the *Cobavirinae* was phage ICBM6, which was isolated from seawater sample HE396-6 infecting *Lentibacter* sp. SH36. Sharing 82.3 - 82.8% nucleotide-based intergenomic identity with the *Lentibacter* phages vB_LenP_ICBM1 and vB_LenP_ICBM 3, it formed a new species within the *Siovirus* genus (Fig. 48 and 49).

Similar to the already described genomes of the *Cobavirinae* subfamily (see chapter 2), Lentibacter phage ICBM6 had a linear genome of 40.3 kb, a G+C content of 46.3% and 55 predicted ORFs, out of which 21 could be annotated. PhageTerm determined short direct terminal repeats (DTRs) of 170 bp at both ends of the genome (Table 15). The genomic organization and annotated gene content was almost identical to that of Lentibacter phage ICBM1 (Fig. 50, Table 40). The genome was organized in two arms, with the left arm encoding genes involved in replication and nucleotide metabolism and the right arm encoding lysis and morphogenesis genes (compare chapter 2.3.4.). The “DNA, RNA and nucleotide metabolism” genes comprised a DNA polymerase I (PC_89), a DNA primase/helicase (PC_88), an endonuclease (PC_148), a glutaredoxin (PC_150), a nuclease (PC_345), a cobalamin-dependent ribonucleotide reductase (PC_91) and a thymidylate synthase ThyX (PC_1) (Table 40). Morphogenesis genes included an internal virion protein D (PC_288), the major

capsid protein (PC_294), the terminase large subunit (PC_298), the head-tail adaptor protein (PC_297), a phage tail assembly chaperone (PC_382), a tail appendage protein (PC_290), a tail protein (PC_292) and a tail fiber protein (PC_569). Two lysis genes were encoded in the genome: a lysozyme (PC_348) and a spanin (PC_287). In addition, an acetyl transferase (GNAT) (PC_313), a MazG-like pyrophosphatase (PC_149) and a PhoH-like phosphate starvation-inducible protein (PC_113) were found. The only difference to the genome of phage ICBM1 was the absence of two additional tail genes (a second phage tail assembly chaperone and a second tail fiber protein). A methyltransferase (PC_336) was encoded in the genome, which can serve for the protection against restriction-modification systems. No tRNAs could be predicted in the genome of phage ICBM6.

In accordance with the other cobaviruses, phage ICBM6 had DTRs and a T7-like DNA polymerase I (Tables 15 and 40). This indicated the use of a T7-like bidirectional replication and DNA packaging strategy. Moreover, a podoviral morphology was predicted for phage ICBM6, reflected by the annotation of the head-tail adaptor protein and the absence of a tail completion protein, a major tail protein or a tail sheath (Table 40). As no lysogeny-related genes were annotated in the genome, a strictly lytic lifestyle can be assumed. The members of the *Cobavirinae* subfamily including the novel phage ICBM6 shared 17 core proteins, with 10 having an annotated function (Table 41).

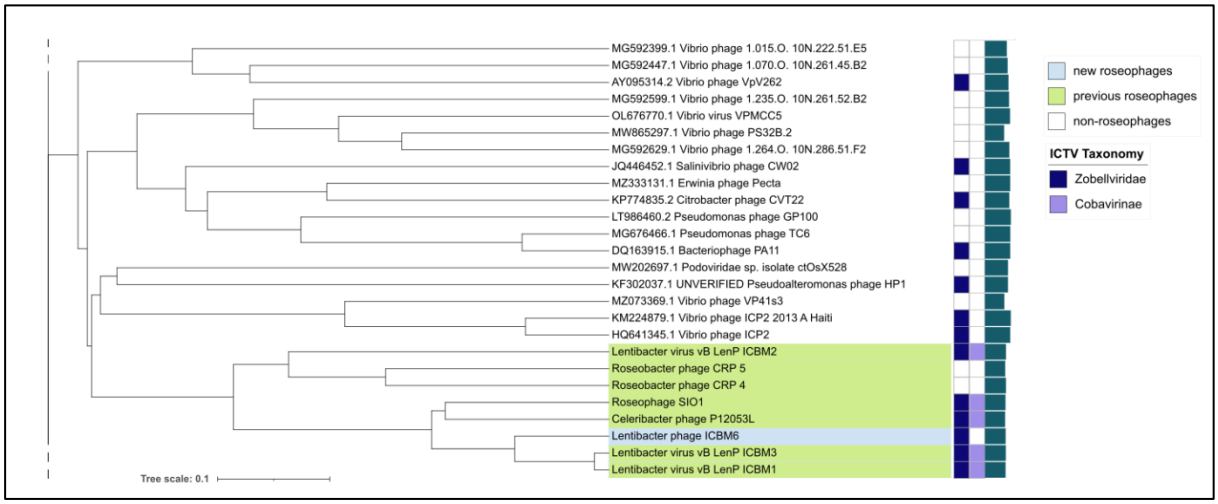


Fig. 47: Section of the whole-genome based proteomic tree showing the *Zobellviridae* family. Names of cultivated roseophages are marked in green (previous) and blue (this study). Color-strips indicate affiliation to ICTV-recognized families and subfamilies, respectively. Genome lengths are displayed as bar chart.

Table 39: Phages of the *Zobellviridae* family. Roseophages are marked in blue. Bp = base pairs. N.a. = not available. *Morphology predicted by Virfam.

Phage	Host	Genome size [bp]	GC content [%]	Accession	Previous classification (reference)	ICTV or newly proposed taxonomy			Morphology	Isolation source	Reference
						Family/Subfamily	Genus	Species			
Vibrio phage 1.015.O.10N.222.51.E5	*Vibrio breoganii*	42586	47.60%	MG592399						seawater	Kauffman et al. (2018)
Vibrio phage 1.070.O.10N.261.45.B2	*Enterovibrio norvegicus*	44087	49.40%	MG592447						seawater	
Vibrio phage VpV262	*Vibrio parahaemolyticus*	46012	49.50%	AY095314		*Zobellviridae*	*Vipivirus*			n.a.	unpublished
Vibrio phage 1.235.O.10N.261.52.B2	*Vibrio lentus*	47017	43.00%	MG592599						seawater	Kauffman et al. (2018)
Vibrio virus VPMCC5	*Vibrio harveyi*	48938	40.70%	OL676770	*Zobellviridae* (Kar et al. 2022)				podoviral	n.a.	Kar et al. (2022)
Vibrio phage PS32B.2	*Vibrio crassostreae*	37147	43.80%	MW865297					podoviral	seawater	Piel et al. (2022)
Vibrio phage 1.264.O.10N.286.51.F2	*Vibrio splendidus*	47739	42.80%	MG592629						seawater	Kauffman et al. (2018)
Salinivibrio phage CW02	*Salinivibrio costicola*	49390	47.70%	JQ446452		*Zobellviridae*	*Salinovirus*		podoviral	Great Salt Lake, USA	Shen et al. (2012)
Erwinia phage Pecta	*Erwinia* sp.	49573	47.90%	MZ333131					podoviral	organic household waste	Alanin et al. (2022)
Citrobacter phage CVT22	*Citrobacter* sp. TM1552	47636	41.60%	KP774835		*Zobellviridae*	*Citrovirus*		podoviral	Termite gut	Tikhe et al. (2015)

Pseudomonas phage GP100	*Pseudomonas protegens* CHA0	50547	50.90%	LT986460	Podoviridae (Vacheron et al. 2018)					rhizosphere soil	Vacheron et al. (2018)
Pseudomonas phage TC6	*Pseudomonas aeruginosa* PA1	49796	45.10%	MG676466	"Pa11virus" (Tang et al. 2018)				podoviral	sewage of Southwest Hospital (Chongqing, China)	Tang et al. (2018)
Bacteriophage PA11	n.a.	49639	44.80%	DQ163915	"Pa11virus" (Tang et al. 2018)	*Zobellviridae*	*Paundecim virus*			n.a.	Kwan et al. (2006)
Podoviridae sp. Isolate ctOsX528	n.a.	45102	36.80%	MW202697						freshwater river	unpublished
UNVERIFIED Pseudoalteromonas phage HP1	*Pseudoalteromonas* sp. str. H100	45035	44.70%	KF302037		*Zobellviridae*	*Melvirus*		podoviral	North Sea	Duhaime et al. (2017)
Vibrio phage VP41s3	*Vibrio parahaemolyticus*	37751	40.90%	MZ073369						supermarket seafood, China, Wuhan	unpublished
Vibrio phage ICP2_2013_A_Haiti	*Vibrio cholerae*	50440	42.80%	KM224879						Haitian cholera patient stool samples	Seed et al. (2014)
Vibrio phage ICP2	*Vibrio cholerae*	49675	42.70%	HQ641345		*Zobellviridae*	*Icepovirus*		podoviral	cholera patient stool samples, Bangladesh	Seed et al. (2011)
Lentibacter virus vB_LenP_ICBM2	*Lentibacter* sp. SH36	40907	47.80%	MF431616	(Bischoff et al. 2020; Adriaenssens et al. 2021)	*Zobellviridae/ Cobavirinae*	*Veravirus*	*septentrionalis*	podoviral	seawater	Bischoff et al. (2019)
Roseobacter phage CRP 5	*Roseobacter* RCA strain FZCC0040	39600	45.60%	MK613347	Cobavirus-like (Zhang et al. 2019a)				podoviral	seawater	Zhang et al. (2019a)
Roseobacter phage CRP 4	*Roseobacter* RCA strain FZCC0023	40768	45.20%	MK613346					podoviral	seawater	
Roseophage SIO1	*Roseobacter* sp. SIO67	40072	46.20%	AF189021	(Bischoff et al. 2020;	*Zobellviridae/ Cobavirinae*	*Siovirus*	*americense*	podoviral	seawater	Rohwer et al. (2000)

Celeribacter phage P12053L	*Celeribacter* sp. strain IMCC12053	39061	46.10%	JQ809650	Adrianenssens et al. 2021)	*Zobellviridae*/ *Cobavirinae*	*Siovirus*	*coreense*	podoviral *	seawater	Kang et al. (2012)
Lentibacter phage ICBM6	*Lentibacter* sp. SH36	40273	46.30%	this study				"sase"		Seawater (HE396-6)	this study
Lentibacter virus vB_LenP_ICBM3	*Lentibacter* sp. SH36	40497	47.30%	MF431615	(Bischoff et al. 2020; Adrianenssens et al. 2021)	*Zobellviridae*/ *Cobavirinae*	*Siovirus*	*germanense*	podoviral	seawater	Bischoff et al. (2019)
Lentibacter virus vB_LenP_ICBM1	*Lentibacter* sp. SH36	40163	47.00%	MF431617					podoviral	seawater	

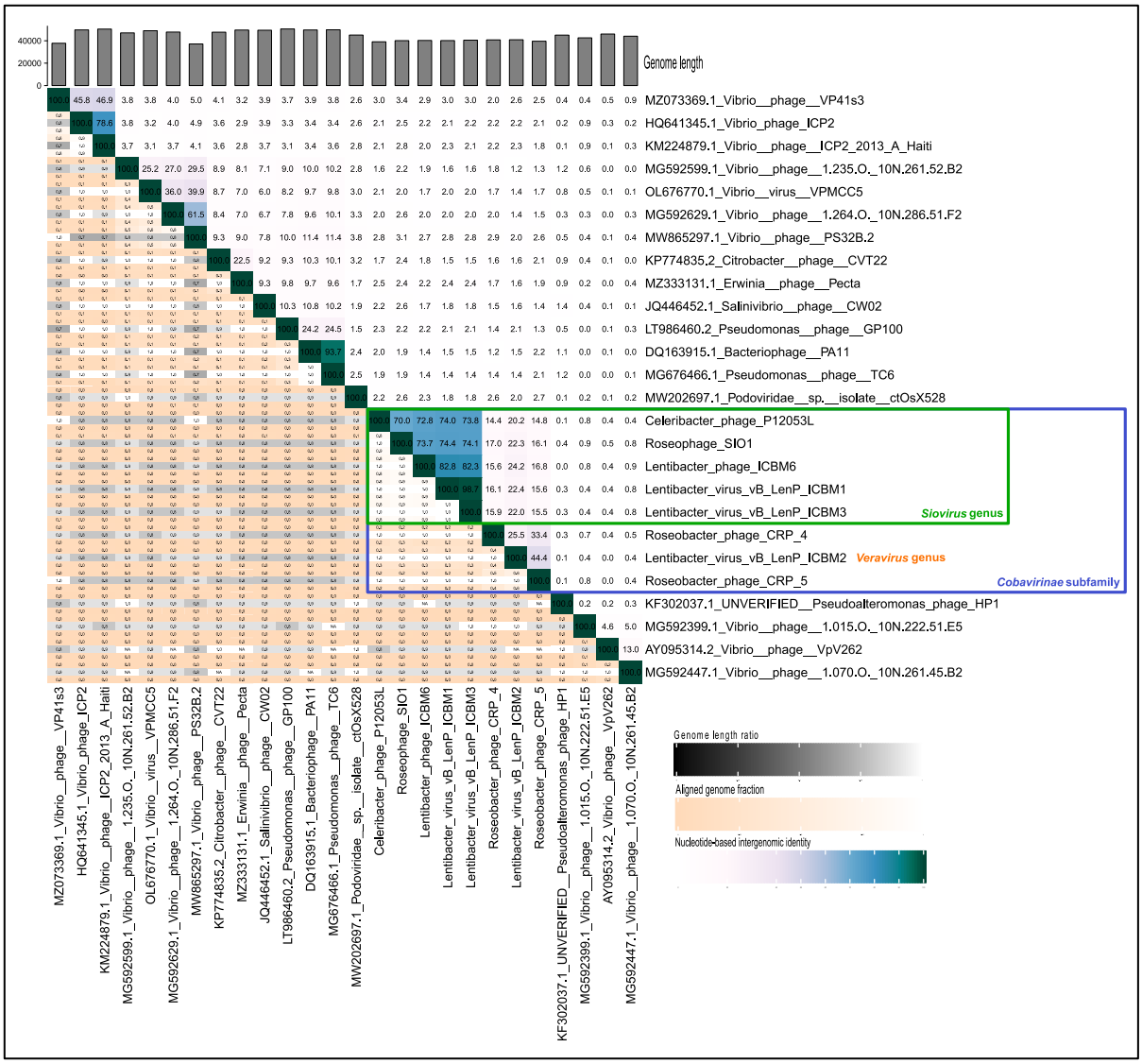


Fig. 48: Nucleotide-based intergenomic identities of the phages in the *Zobellviridae* family, calculated with VIRIDIC. Members of the subfamily *Cobavirinae* and the genera *Siovirus* and *Veravirus* are annotated with boxes and colored labels.

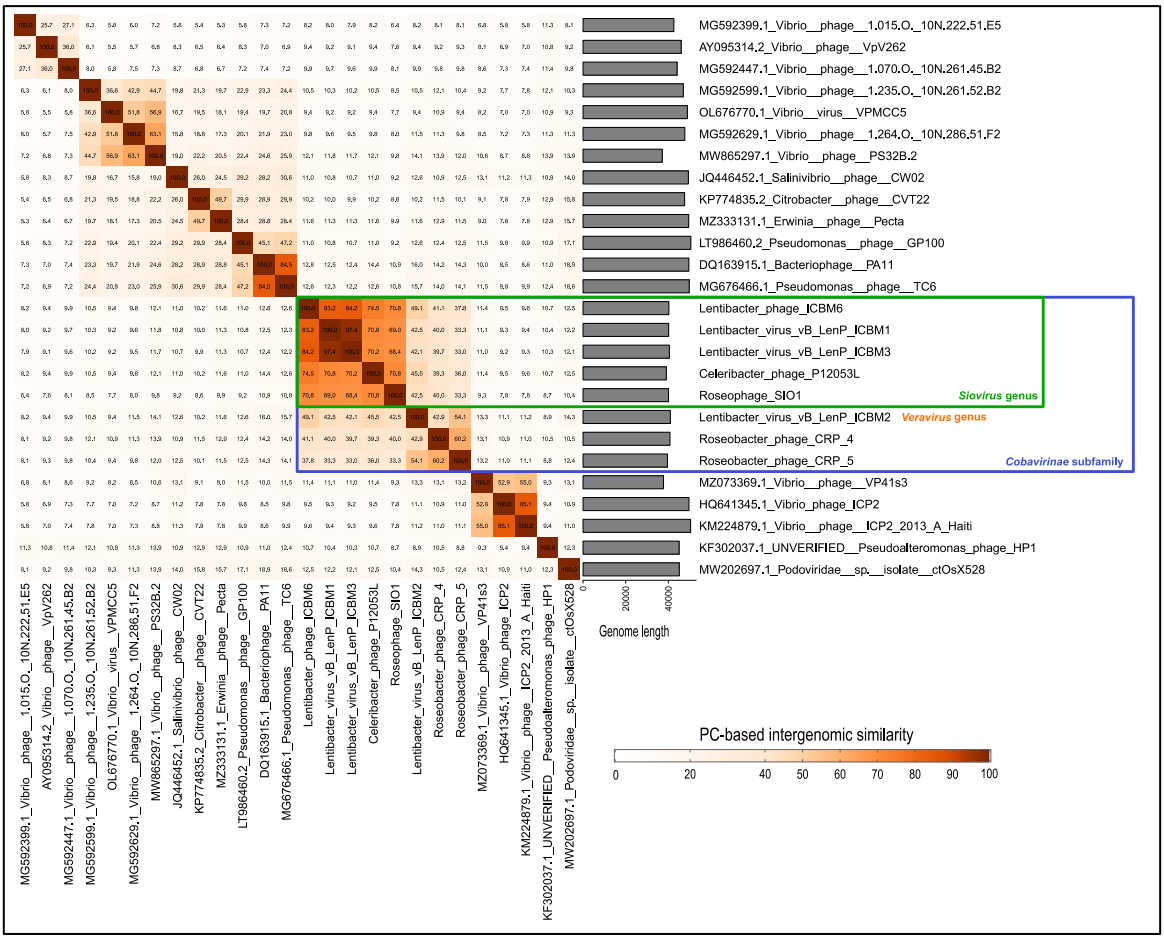


Fig. 49: PC-based intergenomic similarities of the phages in the *Zobellviridae* family, calculated with VirClust (log e-value clustering, matches kept if bitscore >= 30, coverage >= 70, evalue < 0.00001, identity >= 0%). Members of the subfamily *Cobavirinae* and the genera *Siovirus* and *Veravirus* are annotated with boxes and colored labels.

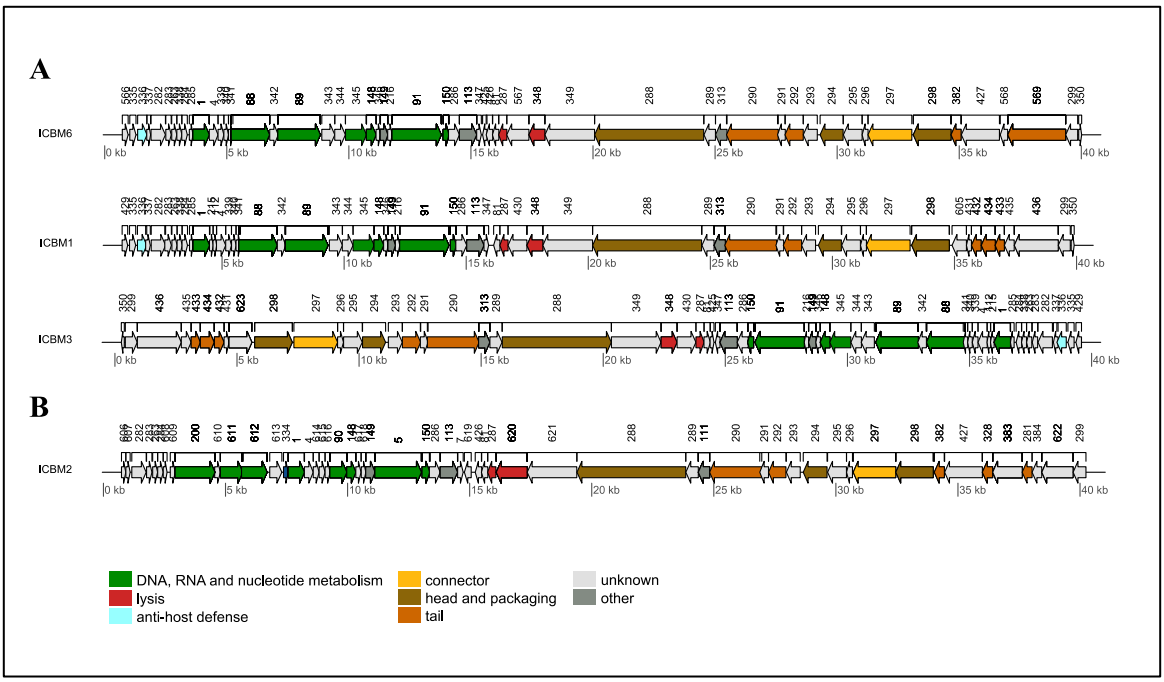


Fig. 50: Genome map of the *Lentibacter* phages belonging to the *Zobellviridae* family. **A** Cobaviruses ICBM1, ICBM3 and the new phage ICBM6. **B** Cobavirus ICBM2.

Table 40: Gene annotations of Lentibacter phage ICBM6. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core proteins of the phages in the *Cobavirinae* subfamily.

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	722	961	240	1	566	hp	unknown
gene_2	1028	1282	255	1	335	hp	unknown
gene_3	1350	1733	384	1	336	S-adenosyl-L-methionine-dependent methyltransferase	anti-host defense
gene_4	1723	1878	156	1	337	hp	unknown
gene_5	1898	2488	591	1	282	hp	unknown
gene_6	2503	2724	222	1	283	hp	unknown
gene_7	2744	2956	213	1	263	hp	unknown
gene_8	2953	3165	213	1	338	hp	unknown
gene_9	3158	3385	228	1	284	hp	unknown
gene_10	3468	3617	150	1	285	hp	unknown
gene_11	3617	4285	669	1	1*	thymidylate synthase ThyX*	DNA, RNA and nucleotide metabolism
gene_12	4282	4644	363	1	4	hp	unknown
gene_13	4644	4880	237	1	339	hp	unknown
gene_14	4882	5070	189	1	340	hp	unknown
gene_15	5067	5249	183	1	341	hp	unknown
gene_16	5185	6753	1569	1	88*	DNA primase/helicase*	DNA, RNA and nucleotide metabolism
gene_17	6759	7106	348	1	342	hp	unknown
gene_18	7090	8832	1743	1	89	DNA polymerase I	DNA, RNA and nucleotide metabolism
gene_19	8891	9427	537	1	343	hp	unknown
gene_20	9420	9860	441	1	344	hp	unknown
gene_21	9862	10725	864	1	345	nuclease	DNA, RNA and nucleotide metabolism
gene_22	10718	11119	402	1	148*	Endonuclease*	DNA, RNA and nucleotide metabolism
gene_23	11112	11297	186	1	346	hp	unknown
gene_24	11284	11595	312	1	149	MazG-like pyrophosphatase	other
gene_25	11592	11774	183	1	216	hp	unknown
gene_26	11771	13786	2016	1	91*	ribonucleotide reductase*	DNA, RNA and nucleotide metabolism
gene_27	13862	14098	237	1	150*	Glutaredoxin*	DNA, RNA and nucleotide metabolism
gene_28	14082	14531	450	1	286*	hp*	unknown
gene_29	14528	15235	708	1	113	PhoH-like phosphate starvation-inducible	other
gene_30	15236	15424	189	1	347	hp	unknown
gene_31	15479	15658	180	-1	425	hp	unknown
gene_32	15624	15893	270	-1	426	hp	unknown
gene_33	15883	16143	261	-1	81*	hp*	unknown
gene_34	16136	16471	336	-1	287	spanin	lysis
gene_35	16468	17376	909	-1	567	hp	unknown
gene_36	17376	18029	654	-1	348	lysozyme	lysis
gene_37	18031	20064	2034	-1	349	hp	unknown
gene_38	20068	24531	4464	-1	288*	internal virion protein D*	head and packaging
gene_39	24545	25015	471	-1	289	hp	unknown
gene_40	25017	25487	471	-1	313	acetyl transferase (GNAT)	other
gene_41	25465	27588	2124	-1	290*	tail appendage*	tail
gene_42	27557	27874	318	-1	291*	hp*	unknown

gene_43	27861	28610	750	-1	292*	tail protein*	tail
gene_44	28612	29184	573	-1	293*	hp*	unknown
gene_45	29300	30241	942	-1	294*	major capsid protein*	head and packaging
gene_46	30254	31021	768	-1	295*	hp*	unknown
gene_47	31024	31266	243	-1	296*	hp*	unknown
gene_48	31263	33056	1794	-1	297	head-tail adaptor	connector
gene_49	33118	34662	1545	-1	298*	terminase large subunit*	head and packaging
gene_50	34662	35075	414	-1	382	Phage tail assembly chaperone	tail
gene_51	35085	36647	1563	-1	427	hp	unknown
gene_52	36681	36983	303	-1	568	hp	unknown
gene_53	36988	39357	2370	-1	569	tail fiber protein	tail
gene_54	39359	39850	492	-1	299*	hp*	unknown
gene_55	39889	40014	126	1	350	hp	unknown

Table 41: Core proteins of the subfamily *Cobavirinae* in the family *Zobellviridae*. *maximum number of predicted ORFs divided by number of core proteins.

predicted ORFs#	**core proteins #**	**Annotated core proteins #**	**Annotated core proteins (sorted by functional category)**	**Percentage core proteins / total ORFs***
55-58	17	10	DNA, RNA and nucleotide metabolism: thymidylate synthase ThyX (PC_1), DNA primase/helicase (PC_88), Endonuclease (PC_148), ribonucleotide reductase (PC_91), Glutaredoxin (PC_150), Head and packaging: internal virion protein D (PC_288), major capsid protein (PC_294), terminase large subunit (PC_298), Tail: tail appendage (PC_290), tail protein (PC_292)	29.31%

4.4. Discussion

When we started the large-scale isolation campaign, only four described roseophage isolates had been recovered from the southern North Sea, the above described cobaviruses (ICBM1, ICBM2, and ICBM3) and Roseobacter phage CRP-235 (Qin et al. 2022). With this study, we were able to increase the knowledge on the diversity of dsDNA roseophages in this habitat significantly. We obtained more than a hundred new lytic roseophages, which we then classified taxonomically and genome characterized. We sequenced a total of 128 unique dsDNA roseophage genomes, belonging to twelve different genera. From these, 28 representative genomes were further investigated and could be assigned to eight families. Four of them represent existing, ICTV-recognized families: *Autographiviridae* (six phages), *Casjensviridae* (one phage), *Mesyanzhinoviridae* (two phages) and *Zobellviridae* (one phage). The remaining four families are here newly proposed: “Hayaniviridae” (six phages), “Schlingloffviridae” (two phages), “Diferiteviridae” (five phages), and “Woolleyviridae” (five phages) (Table 16). Furthermore, we propose here two new subfamilies: the “Incetivirinae” subfamily within the

Autographiviridae family, and the "Maresulfivirinae" within the *Mesyanzhinoviridae* family. The novel roseophages infect hosts of the genera *Sulfitobacter*, *Lentibacter* and *Octadecabacter*. They have dsDNA genomes ranging in size from 33.3 kb (Sulfitobacter phage ICBM130, "Woolleyviridae") to 80.8 kb (Sulfitobacter phage ICBM153, *Mesyanzhinoviridae*). About two thirds of the genomically analysed new roseophages show evidence of a lysogenic lifestyle.

4.4.1. Thresholds for phage taxonomic classification

In our phage classification approach, we performed a hierarchical clustering of the viral genomes based on the intergenomic distances of their shared protein content (proteins were clustered with 70% coverage and 100% identity). Subsequently, we split this proteomic tree into viral genome clusters (VGCs) using a distance threshold that fit best the existing ICTV families (99.5% distance). However, looking at the individual VGCs, we observed a high genomic diversity, underlined by the absence or a low number of core protein clusters or, in some cases, by the presense of different virion morphologies in the same VGC (e.g., siphoviruses and podoviruses in VGC_11, see chapter 4.3.5.3.). In our opinion, viral families should cluster viruses that share a strong core of protein clusters (especially belonging to the virion structure and morphogenesis module), have a similar morphology, DNA-replication and life-style. These criteria are in line with the new recommendations for viral classification (Simmonds et al. 2023). Within our data set, mostly viral groups smaller than the complete VGCs showed such common characteristics that they can be classified as families. Accordingly, we used a threshold of 30% PC-based intergenomic similarities to delineate new phage families (or subfamilies in the case of subclusters within already ICTV recognized families). Furthermore, virion morphology, genome replication strategy and host taxonomy further strengthened our classification approach. The predicted morphologies of the new ICBM phages were consistent within each family, being either of podoviral type (*Autographiviridae*, *Zobellviridae* and "Diferiteviridae") or siphoviral type ("Hayaniviridae", "Schlingloffviridae", *Casjensviridae*, *Mesyanzhinoviridae* and "Woolleyviridae") (Table 16). The predicted mode of genome replication was not as uniform within each family, but overall was limited to two types, either T7-type bidirectional replication and direct terminal repeats (DTRs), or P1-type rolling circle replication and headful (pac) packaging with circularly permuted genomes (Table 16). As a future task, phylogenetic analysis of viral hallmark genes, e.g., major capsid protein, which was detected in the core proteome of all families, should be performed in order to confirm the proposed family affiliation and intra-family taxonomy. For the classification into levels higher

than family rank, our PC-based tree is not sufficient. To determine order-level ranks, VirClust could be used again to calculate a hierarchical clustering of the viral genomes based on protein super clusters (PSCs) (Moraru 2023). For viral classification into higher ranks such as phyla, phylogenetic analysis of highly conserved hallmark genes and protein structure comparisons should be applied (Simmonds et al. 2023).

4.4.2. Diversity of the discovered phage families

Using the direct-plating approach for phage isolation, and partly also the enrichment approach, we were able to capture a high roseophage diversity. Although they are all dsDNA phages, and were all isolated using only three host genera from *Roseobacteraceae*, the taxonomic classification revealed quite diverse family compositions. Here, we provide a small window into the true roseophage diversity in the North Sea and the ecology behind. The phage families with more than one new roseophage comprise isolates from two to four different water samples, indicating the different degrees of diversity of the families as well as their relevance at different locations and years in the North Sea. Some families include phages infecting closely related hosts ("Hayaniviridae", "Woolleyviridae", infecting *Sulfitobacter*), while other families include viruses infecting two different host genera (e.g., "Diferiteviridae", infecting *Sulfitobacter* and *Lentibacter*). It would be an exacting task in the future to test the capability of all these phages to infect different genera on the one hand, and highly similar strains of the same species on the other hand. Furthermore, the proposed phage (sub-) families display different degrees of conservation regarding protein content. The number of core proteins ranges from eight ("Diferiteviridae") to 31 on family level ("Schlingloffviridae"), and 36 on subfamily level ("Maresulfivirinae"). In accordance with the prerequisite of having a similar DNA replication and virion morphology to be a family, the majority of the core proteins in all families are involved in those critical functions belonging to the functional categories of "head and packaging" and "DNA replication". In all families, the major capsid protein and the terminase large subunit belong to the annotated core proteins. Furthermore, the portal protein, the capsid scaffolding protein as well as the DNA polymerase, DNA primase/helicase and the thyX thymidylate synthase are core proteins in many families.

4.4.3. Genetic particularities of the new roseophages

The newly described roseophages possess varying interesting features as indicated by different auxiliary metabolic genes (AMGs) and genes involved in anti host-defense mechanisms encoded in their genomes. In most families, methylases are encoded, which are

well known as part of the viral response to the host defensive restriction-modification systems (Murphy et al. 2013). Additionally, the genomes of the "Schlingloffviridae" family encode a SaV-like protein, which is involved in the recognition of abortive infection systems (Haaber et al. 2009). Members of the "Maresulfivirinae" subfamily have the genes for queuosine biosynthesis encoded in their genomes, which was shown more recently to serve for protection of the viral DNA against bacterial restriction enzymes (Hutinet et al. 2019) (Table 16). Another interesting gene was found in the genomes of the "Benvirus" genus of the "Diferiteviridae" family. The phosphoadenosine phosphosulfate (PAPS) reductase can function as an AMG which benefits the host in low-sulfur conditions (Summer et al. 2006). Furthermore, the new roseophage isolates in the *Autographiviridae* family possess genes for the regulation of host chromosome condensation (RCC1). These regulators are common in eukaroytic genomes, but have so far only been described in two other phages (Abbasifar et al. 2014; Wagemans et al. 2020).

Strinkingly, we found no DNA polymerases in the phage genomes from the "Diferiteviridae" and "Woolleyviridae". We can only hypothesize that these phages use their host´s DNA polymerase for replication, as it has been shown for example for *Bacillus subtilis* phage SPP1 (Seco et al. 2013). However, it is theoretically also possible that they possess a polymerase gene that is so different from all known genes that it could not be annotated. Cai et al. (2023) recently detected a completely new clade of cyanopodophages without a DNA polymerase encoded in their genomes and showed that they are very abundant in surface oceans worldwide.

Phages are known to have their own set of tRNAs to enhance translation efficiency during infection and thus facilitate phage production (Enav et al. 2012). Accordingly, Holmfeldt et al. (2013) detected a correlation of the amount of phage encoded tRNAs in *Cellulophaga* phages and the number of infected host strains. *Sulfitobacter* phages ICBM129 and ICBM153 belong to the same newly proposed subfamily "Maresulfivirinae", but possess different numbers of tRNA genes (none and seven, respectively). It would be an interesting task for future research to determine their host range in order to test if this correlation holds true here as well.

4.4.4. Conclusions

Once more, we could significantly extend the knowledge of dsDNA roseophages by isolation, genome sequencing, and taxonomic classification of phages infecting three *Roseobacteraceae* genera. Taxonomic assignment of the new viruses to eight different families, including four

completely new ones, indicates again the vast diversity of marine phages still awaiting to be uncovered. The obtained collection of genome-sequenced phage isolates provides a wide range of opportunities for further research, both at genomic and cultivation level.

5. Microdiversity of the sulfiviruses and their hosts (preliminary results)

5.1. Chapter summary

Many of the new dsDNA roseophage isolates obtained in the isolation campaign described in the previous chapter grouped into a large genus-level cluster – the "Sulfivirus" genus – infecting closely related *Sulfitobacter* strains. From the complete genus, only six representatives were considered for further classification in the previous chapter, and they were placed in the newly proposed family "Hayaniviridae". Having this large collection of phages from one single genus at hand, together with the collection of highly related, but different host strains, it was a great opportunity to investigate the genomic microdiversity of both phages and hosts. Furthermore, the influence of this microdiversity on phage-host interactions, more precisely the host range, could be examined.

The results described in this chapter represent a preliminary work. Thus, there is no separate subchapter for discussion. Instead, initial discussion approaches are presented together with the results.

Contributions to this work:

Cristina Moraru designed the research and contributed to data analysis. I performed most of the laboratory work and the data analysis and wrote this chapter. We had great help in the laboratory work by Andrea Schlingloff (16S rRNA gene and ITS sequencing), Aaron Woolley and Martin Sackmann (phage purification and host range assay) and Ismail Hayani (TEM). Whole genome sequencing of the bacterial strains was performed by our collaborators Anja Poehlein and Mechthild Bömeke at the Göttingen Genomics Laboratory. Anne Kupczok (Wageningen University and Research, NL) helped us with the investigation of recombination events in the viral genomes.

5.2. Materials and methods

5.2.1. 16S / ITS phylogenetic analysis of sulfivirus host strains

Sequencing of the 16S rRNA gene and the internal transcribed spacer (ITS) region of the bacterial hosts are described in chapter 4.2.3. The phylogenetic tree based on the 16S rRNA gene showed, that all sulfivirus host strains were identical or almost identical at the 16S rRNA level (chapter 4.3.1). Thus, we performed further phylogenetic analysis including the ITS region. Sequence alignment and tree calculation were performed using Geneious Prime® (version 2021.2.2). The neighbor-joining tree was calculated with Jukes-Cantor correction and 1000 bootstrap replicates. The closest relative of the host strains was retrieved by a BLASTN (Camacho et al. 2009) search against the nucleotide collection (nr/nt) and included into the phylogenetic tree.

In addition, phylogenetic analysis based only on the 16S rRNA gene was performed with the ARB software package (Ludwig et al. 2004). Using the reference data set SSU Ref NR 138.1, a neighbor-joining tree was calculated with Jukes-Cantor correction, 1000 bootstrap replicates and the termini filter. Members of the genus *Psychrobacter* served as an outgroup.

5.2.2. Whole genome sequencing of sulfivirus host strains

Thirty representative host strains were cultivated in liquid MB50 medium for 1 - 2 days at 20 °C and 100 rpm. DNA extraction for the two sequencing techniques Illumina and Nanopore was performed using different protocols. For Illumina sequencing it was performed in our laboratory as described further. From a culture with an OD_{600} higher than 0.5, 500 µl were transferred to a PCR clean reaction tube and centrifuged for 15 min at 4000 x g and 20 °C. The supernatant was discarded and the pellet resuspended in 100 µl of the ChargeSwitch gDNA Mini Bacteria Kit (ThermoFisher Scientific). The genomic DNA was extracted following the instruction manual of the kit with the incubation step for cell lysis performed for 1 h at 80 °C. DNA concentrations were determined using the Qubit 2.0 fluorometer and the Qubit dsDNA HS Assay kit (ThermoFisher Scientific). Extracted DNA was stored at 4 °C. For Nanopore sequencing, frozen cell pellets were sent to our collaborators for DNA extraction and sequencing. For the preparation of these cell pellets 2 x 35 ml culture were grown to an OD_{600} higher than 0.5. Culture vessels were cooled on ice. After centrifugation for 15 min at 4000 x g and 4 °C, the supernatant was removed leaving about 2 ml of culture medium for pellet resuspension. The resuspended cells were split into 500 - 1000 µl aliquots in smaller reaction tubes and

centrifuged again for 15 min at 4000 x g and 4 °C. The supernatant was removed and centrifugation repeated with the same parameters. Finally, the pellet was frozen in liquid nitrogen and stored at -80 °C. Further sample preparation, i.e. DNA extraction, was performed by our collaborators Anja Poehlein and Mechthild Bömeke at the Göttingen Genomics Laboratory (University of Göttingen). Whole genome sequencing by Illumina and Nanopore technologies was carried out by our collaborators in Göttingen, who provided us with the readily assembled and annotated genomes.

5.2.3. Phylogenomic analysis of the sulfivirus host strains

The most closely related strains were retrieved by running 20 of the host genomes in the Type strain Genome Server (TYGS) (Meier-Kolthoff and Göker 2019). The genomes of the 17 related strains in the output tree were downloaded from the NCBI Nucleotide database and together with all 30 sulfivirus host genomes were submitted to the Genome-to-Genome Distance Calculator 3.0 (GGDC) (Meier-Kolthoff et al. 2021), running each genome against all other genomes in a separate project, respectively. The GGDC output for all genomes was combined in a d6 intergenomic distance matrix and a hierarchical clustering tree was calculated using the hclust package in the R programming environment (R Core Team (2020), https://www.r-project.org/).

5.2.4. Initial comparative analysis of host genomes

First, the plasmids of all 30 strains were clustered with VirClust (Moraru 2023), using BLASTp for protein clustering (bitscore threshold of 100, coverage threshold of 80, evalue threshold of 0.0001, clustering based on log evalues), the complete agglomeration method for plasmid clustering and a 0.7 clustering distance for tree splitting. After the plasmids were assigned to different clusters, the *Sulfitobacter* strains were clustered based on absence/presence of these plasmid clusters (distance matrix computation method “binary”, agglomeration method “ward.D2”) and the result was visualized as a heatmap including the infection patterns of the sulviruses using the R programming environment (R Core Team (2020), https://www.r-project.org/).

A search for prophages in the chromosomes and plasmids was performed using Prophage Hunter (Song et al. 2019) and PHASTER (Arndt et al. 2016). Genomes of the detected active prophages were compared determining their nucleotide-based intergenomic sequence identity using VIRIDIC with default settings (Moraru et al. 2020).

5.2.5. Phage genomic analysis

5.2.5.1. Clustering of sulfivirus genomes on species and genus level

Nucleotide-based intergenomic identities of all sulfivirus pairs was calculated with VIRIDIC (Moraru et al. 2020), using the default options ("-word_size 7 -reward 2 -penalty -3 -gapopen 5 -gapextend 2"). A whole genome-based phylogenetic tree was created by pairwise comparisons of the nucleotide sequences using the Virus Classification and Tree Building Online Resource (VICTOR, available at https://victor.dsmz.de), and the underlying Genome-BLAST Distance Phylogeny (GBDP) method under settings recommended for prokaryotic viruses (Meier-Kolthoff et al. 2013; Meier-Kolthoff and Göker 2017). Taxon boundaries at the species, genus and family level were estimated with the OPTSIL program (Göker et al. 2009; Meier-Kolthoff et al. 2013) and the clustering thresholds suggested by Meier-Kolthoff et al. (2013).

5.2.5.2. Protein-based hierarchical clustering of the sulfivirus genomes and gene annotations using VirClust

A hierarchical clustering of all sulfivirus genomes was performed with VirClust (Moraru 2023) with the following parameters: BLASTp for the creation of protein clusters (clustering based on log evalues, thresholds for matches being removed: evalue >0.00001, bitscore <50, coverage <0, identity <0%), the "complete" agglomeration method for genome tree calculations and a distance of 0.9 to split the genomes into clusters. Functional annotations of predicted genes and their proteins was performed with VirClust, by searching against the databases InterPro (Finn et al. 2017), Prokaryotic Virus Orthologous Groups (pVOGs, Grazziotin et al. (2017)) and Virus Orthologous Groups (VOGDB, https://vogdb.org/help). A final, consensus annotation was then assigned manually. Core proteins were determined for each viral genome cluster (VGC), also using VirClust. Genome maps were generated using the genoPloR package (Guy et al. 2010) from the R programming environment (https://www.rproject.org/). tRNAs were identified with the online tool tRNAscan-SE v. 2.0 (http://lowelab.ucsc.edu/tRNAscan-SE/index.html) using the option "bacterial" as sequence source (Lowe and Chan 2016). Further, a second VirClust analysis was performed with settings able to capture fine differences between proteins, to enable investigation of the sulfivirus microdiversity. In a first step, proteins were clustered only if they were identical (thresholds used for filtering the BLASTp hits: evalue >0.00001, bitscore <50, coverage <100, identity <100%). Then, intergenomic distances were calculated based on these protein clusters, and used further for hierarchical clustering of the sulfivirus genomes (the "complete" agglomeration method for tree building, and an 0.9 distance for tree spitting).

5.2.5.3. Detection of recombination events

The ProgressiveMauve alignment (Darling et al. 2010) implemented in Geneious Prime® (version 2021.2.2) was used with default options to create a whole genome alignment of all sulfiviruses. After extraction of collinear blocks, gap removal and concatenation, an alignment of the core regions was obtained with MAFFT (Katoh and Standley 2013). Maximum likelihood (ML) phylogenies were estimated using IQ-TREE (v1.6.12) (Nguyen et al. 2015) and based on them, recombination events were identified with ClonalFrameML (v1.12) (Didelot and Wilson 2015). Apart from the visual display of recombination and mutation sites in the genomes, the results from ClonalFrameML were used to calculate the relative effect of recombination to mutation (r/m).

5.2.6. Host range determination

Cross-infectivity of each sulfivirus was tested against 59 original *Sulfitobacter* host strains by spot assay and streak assay. For initial screening, a spot assay was performed in which dilutions 10^0, 10^{-3} and 10^{-5} of a fresh viral lysate were spotted on an agar plate with a soft agar layer containing the respective bacterial strain. Dilutions of the viral lysate (prepared from a liquid infection culture as described in chapter 4.2.5.2) were created by mixing with ASWbase medium. An aliquot of exponentially growing host culture (final OD_{600} in 3 ml = 0.0233) was pipetted in the middle of an MB50% agar plate (1.8% agar) and 3 ml of MB50%-soft agar (0.6% low melting Biozym Plaque GeneticPure agarose, Biozym, kept warm at 40 °C) were added into the middle of the culture droplet. The plate was shaken for mixing and even distribution of the soft agar layer. After solidification of the soft agar layer, 10 µl droplets of the phage dilutions were pipetted on top. Plates were incubated for up to one week at 20 °C and regularly checked for plaque formation.

To confirm the results, all sulfiviruses were again tested against all *Sulfitobacter* host strains in a streak assay, an adaptation of the method described by Kauffman and Polz (2018). Here, a droplet of undiluted phage lysate was also spotted on the soft agar layer containing the bacterial strain. Immediately afterwards, while the soft agar was still liquid, a dilution streaking was performed (Fig. 51a). For this, a sterile toothpick was swiped through the phage droplet and the soft agar creating three parallel streaks. The plate was turned a bit counterclockwise and a fresh toothpick was taken to make another three streaks crossing the end of the previous streaks. The plate was turned again and a final toothpick was taken to make one serpentine streak starting at the end of the secondary streaks. This way, single virus plaques could be observed in case of successful infection and a false positive result due to spontaneous cell lysis

could be excluded. Plates were incubated at 20 °C and checked or plaque formation after three days (Fig. 51b).

For those phage-host combinations that gave contradictory results in the spot assay and streak assay, another round of streak assay was conducted. Clustering of the sulfiviruses according to their infectivity patterns (distance matrix computation method "binary", agglomeration method "ward.D2") and heatmap visualization was performed using the R programming environment (R Core Team (2020), https://www.r-project.org/). Likewise, further heatmaps with the phages clustered by isolation hosts or based on the core region-phylogenetic tree were prepared.

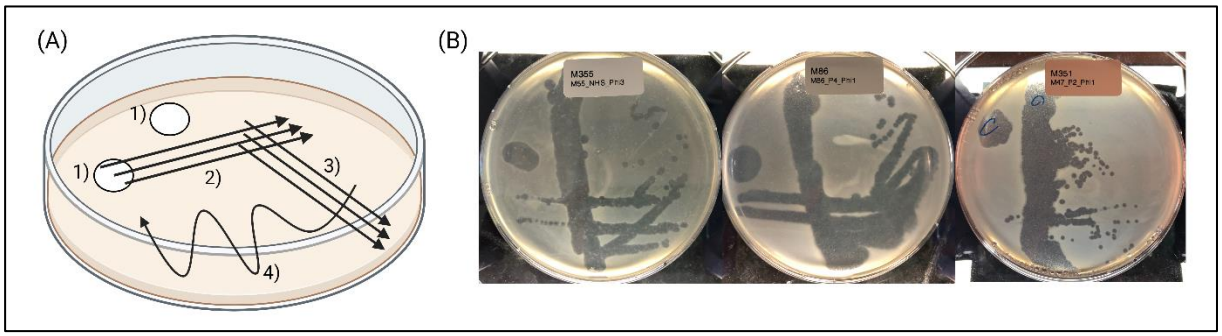


Fig. 51: Host range determination. **(A)** Scheme of the streak assay for host range determination. Steps 1 and 2 were performed with the pipette tip, for steps 3 and 4 a fresh sterile tooth pick was taken, respectively. **(B)** Examples for plates positive for infection and with single plaques visible. (Image created with Biorender.com).

5.2.7. TEM of sulfiviruses ICBM16 and ICBM18

In order to determine the morphology of the sulfiviruses, fresh lysates of two representatives (phages ICBM16 and ICBM18) were concentrated by polyethylene glycol (PEG) precipitation and examined under the transmission electron microscope (TEM). The phage lysates were prepared from two subsequent infection cultures, similar to the method used for the high titer lysates of the cobaviruses (see chapter 2.2.4). The first culture was set up inoculating 100 ml MB medium with *Sulfitobacter* sp. M53 to a final OD_{600} of 0.06 and 1 ml of the respective phage stock. In parallel, two cultures without phage were incubated, one as preculture, one as control. After incubation for 12 h at 20 °C and 100 rpm, lysis was indicated by cell debris and a decrease in OD_{600} as compared to the non-infected cultures. The phage fraction was harvested by centrifugation (15 min at 5752 rcf, 20 °C) and keeping the supernatant. The preculture was also centrifuged, the supernatant discarded and the pellet resuspended in 100 ml 2x MB, to which the phage fraction was added. This way, a highly concentrated phage-host mixture was established for the second infection round. The new infection culture and a control (the control culture from the first round of infection diluted with 100 ml 2x MB) were incubated for 10 h at

20 °C and 100 rpm. The OD_{600} of the cultures was monitored hourly. When a decrease of the OD_{600} in the infection culture below 0.25 was observed, the phage lysate was harvested by centrifugation (15 min, 5752 rcf, 20 °C) and stored at 4 °C. For concentration by PEG precipitation, the phage lysates were incubated for 12 h at 4 °C with PEG (final concentration 10%) and NaCl (final concentration 0.06 M). The supernatant was discarded after centrifugation for 2 h at 7197 x g and 4 °C and the pellet resuspended in 500 µl SM buffer, pooling all pellets of the same phage.

For transmission electron microscopy, 30 µl of the phage lysate were placed on the sterile side of a piece of Parafilm® and a carbon coated grid (Formvar 162, 200 mesh) was placed on top with the shiny side facing the droplet. After 3 min of absorption, the grid was transferred to a droplet of 30 µl 2% uranyl acetate solution for 20 s. Immediately afterwards, the grid was picked up and the remaining liquid was carefully removed using filter paper, leaving a thin film behind. After air drying for 15 min, the grids were examined with the transmission electron microscope Zeiss EM902A. Images were documented with the Proscan High Speed SSCCD camera and analyzed using the software ImageSP viewer (Version 1.2.10.36, SYSPROG). Phages negatively stained were used for size measurements.

5.3. Results and discussion

5.3.1. A micro-diverse bacterial collection served for isolation of Sulfivirus phages

The closely related sulfiviruses analyzed in this project have been isolated from 48 *Sulfitobacter* host strains. Analysis of the 16S rRNA gene and the 16S-23S rRNA internal transcribed spacer (ITS) region revealed that most of the host strains had an identical 16S rRNA gene, and even an identical IST (Fig. 52, SI File S5-1), indicating that they belong to the same species. In the 16S rRNA gene based phylogenetic analysis (Fig. 53) these strains formed a highly supported clade together with *Sulfitobacter marinus* DSM 23422 (99.31 - 99.42 % sequence identity, see SI File S5-2), suggesting that they belonged to a new species within the genus *Sulfitobacter*. Furthermore, whole genome analysis of a selection of 30 host strains confirmed this (Fig. 54). The very low intergenomic distances (max 0.0885, see SI file S5-3) showed that these host strains were highly related and belonged to the same microdiverse species level population.

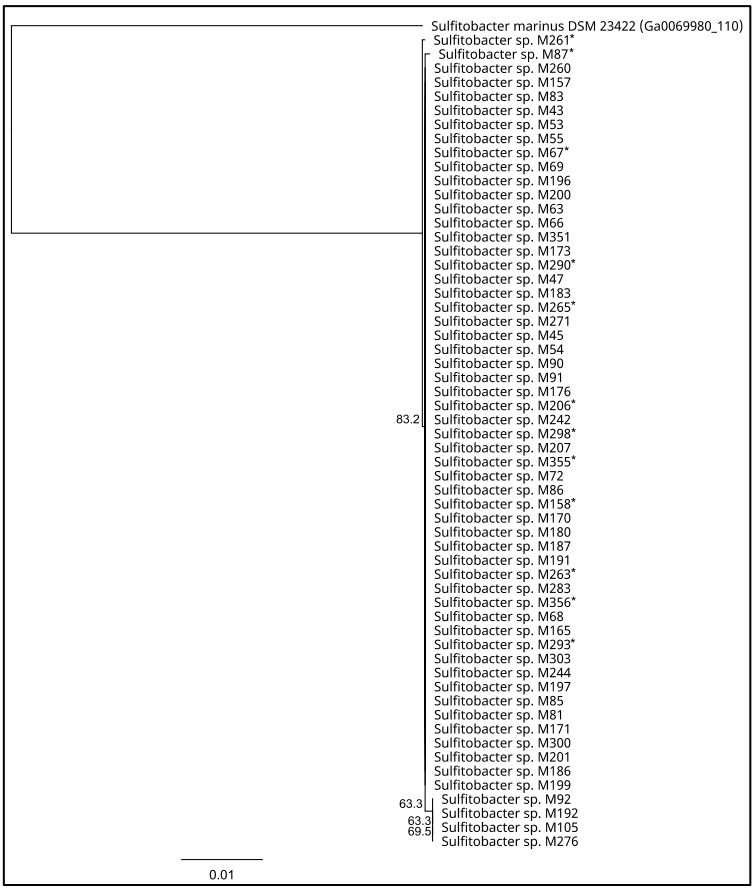


Fig. 52: Neighbor-joining tree based on the 16S rRNA gene and ITS region of 46 sulfivirus host strains plus their closest relative *Sulfitobacter marinus* DSM23422. Strains M71 and M172 are missing from the tree, but their sequences were 99.9% identical to that of strain M290 (see Fig. S15). *additional host strains of phages that were either not successfully sequenced or turned out to be duplicates.

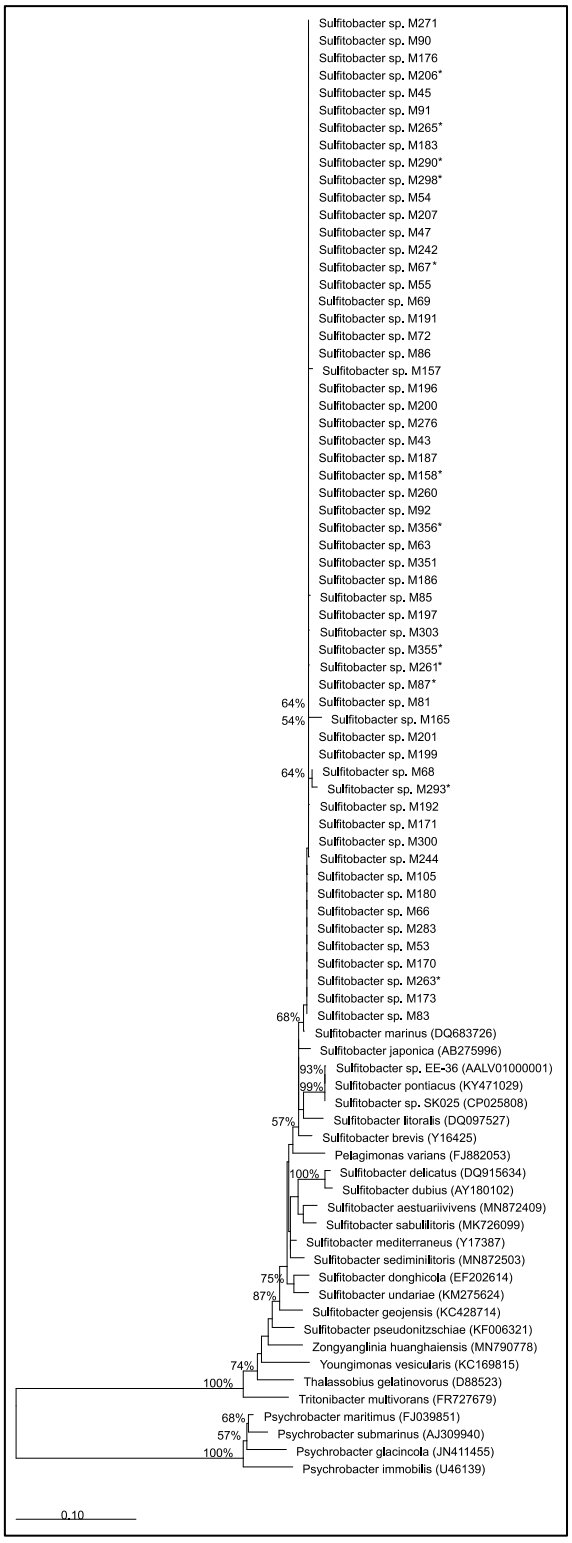


Fig. 53: **N**eighbor-joining tree based on the 16S rRNA gene of 46 sulfivirus host strains and other *Sulfitobacter* strains, plus reference genomes. *Psychrobacter* strains served as outgroup. Strains M71 and M172 are missing from the tree, but their 16S rRNA gene sequences were 99.9% identical to that of strain M290 (see Fig. S15). *additional host strains of phages that were either not successfully sequenced or turned out to be duplicates.

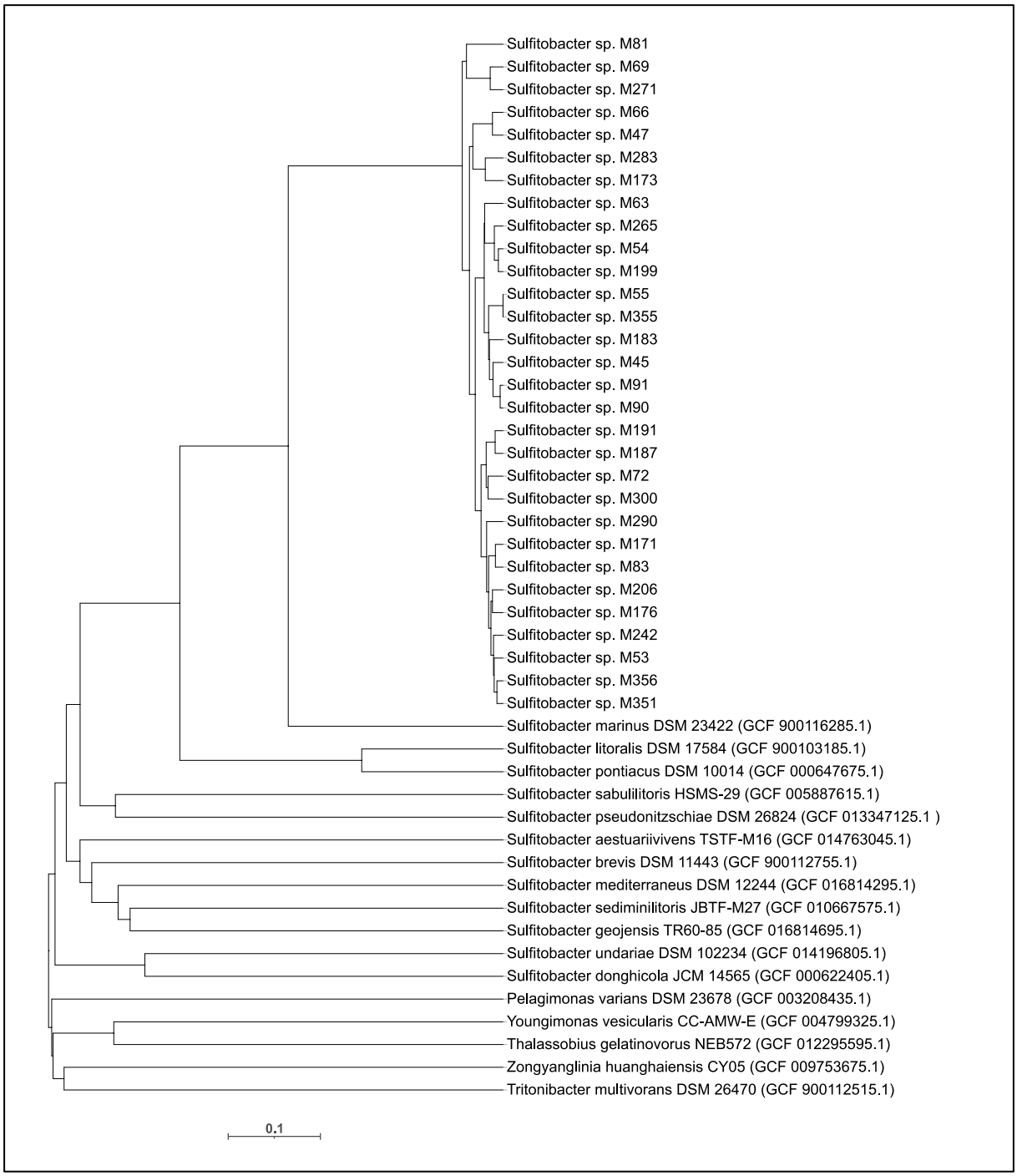


Fig. 54: Hierachical clustering tree based on whole genome distances of 30 *Sulfitobacter* host strains plus 17 related strains.

5.3.2. 89 phage isolates of one genus

As mentioned before, the 89 phages (72 complete + 17 partial genomes) isolated from the above-described host strains were highly similar. Sharing more than 68.4% nucleotide-based intergenomic identity, they belonged to a single phage genus, provisionally called the "Sulfivirus" genus (Fig. 56). Within this genus, the phages could be provionally assigned to 48 species clusters (Table 42, note: the cluster IDs are different from those assigned in chapter 4.3.4.). The sulfiviruses were isolated by direct plating from four different seawater samples, either directly from the shore (NHS) or from the mesocosm experiment (P1, P2, and P4). Examination of two sulfiviruses with transmission electron microscopy (TEM) revealed a siphoviral morphology with a long, non-contractile tail (Fig. 55). Sulfitobacter phage ICBM16 had a capsid size of 63 ± 3 nm and a tail of 141 ± 8 nm in length (46 virions measured). Sulfitobacter phage ICBM18 had a capsid size of 63 ± 4 nm and a tail of 142 ± 7 nm in length (50 virions measured). The siphoviral morphology was also reflected by gene annotations (see chapter 5.3.2.1.). Whole genome-based classification of six representatives with VirClust (Moraru 2023) and analysis of shared protein clusters showed that they clustered together with three other roseophages (*Lentibacter* phages ICBM7 and ICBM166 and Roseobacter phage CRP-6), but constituted their own phage family (see chapter 4.3.5.4). In addition, the whole genome-based phylogeny obtained with the VICTOR web service (Meier-Kolthoff and Göker 2017) and the taxon boundaries estimated with OPTSIL (Göker et al. 2009) assigned all phages to one genus and one family (Fig. 57).

The size of complete sulfivirus genomes ranged from 50.122 - 55.591 kb and the G+C content from 44.5 - 45.0% (Table 42). In these genomes, 80 to 95 ORFs were predicted (see SI file S5-4). The genomes of 17 sulfiviruses could be sequenced only partially, but they were almost complete having genome sizes of 45.405 - 54.605 kb. Genome ends had short direct terminal repeats (DTRs) of 306 - 361 bp, indicating a T7-like DNA packaging strategy (see Table S13). With protein clustering using relaxed thresholds (evalue >0.00001, bitscore <50, coverage <0, identity <0%), more than two thirds (65) of all protein clusters (PCs) were present in all sulfivirus genomes (Fig. 58). Out of these core genes, 25 could be annotated. They represented almost all functional categories, including "DNA, RNA and nucleotide metabolism", to "lysis", "head and packaging", "tail", "connector", "anti-host defense" and "other" (Table 43). By comparison, in chapter 4.3.5.4, only 63 core PCs were determined for the six sulfiviruses in the "Hayaniviridae" family. The discrepancy lies in the different protein clustering parameters used between the two chapters, with the parameters used in chapter 4.3.5.4 being more stringent and grouping proteins that are more similar. Likewise, some

functional annotations differed between the two chapters, reflecting the fact that the different databases queried for protein annotation sometimes return slightly different results and even with manual review, it can be difficult to find the most accurate consensus annotation. However, most of these differences did not affect the overall functional category of the respective protein (Table 43).

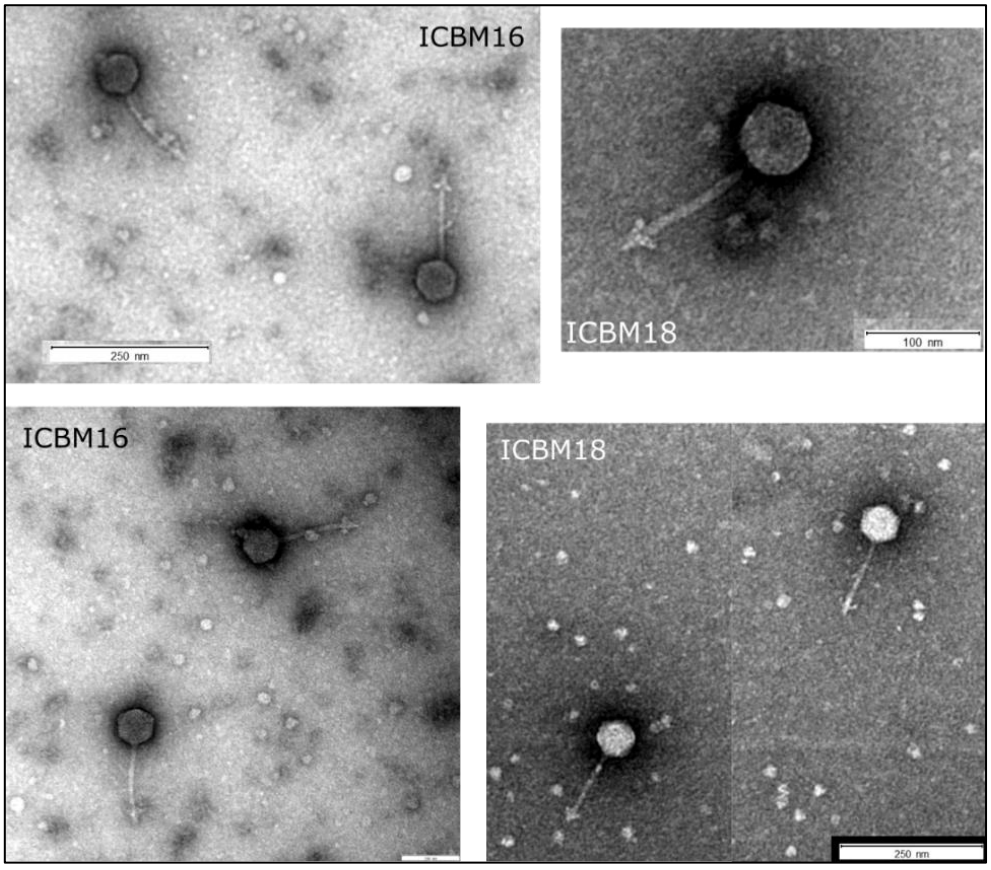


Fig. 55: TEM image of uranyl-acetate stained virions of *Sulfitobacter* phages ICBM16 and ICBM18.

Table 42: Phages of the "Sulfivirus" genus. *host strain was genome sequenced. **species level clusters derived from nucleotide-based intergenomic identities calculated with VIRIDIC, IDs differ from those assigned in chapter 4.3.4. ***different North Sea water samples, from the shore (NHS), a mesocosm experiment (P1, P2 and P4), and from the open sea (HE504-33, HE396-6 and HE440-S).

Isolation host	Phage name (full)	Phage name (short)	Species cluster ID**	Species assignment in chapter 4.3.4.	Isolation source***	Isolation procedure	Genome status	Genome size [kb]	GC content [%]
Sulfitobacter sp. M43	Sulfitobacter phage ICBM9	ICBM9	46		NHS	Direct plating	partial	52.743	44.80%
Sulfitobacter sp. M43	Sulfitobacter phage ICBM10	ICBM10	1		NHS	Direct plating	partial	51.641	44.80%
Sulfitobacter sp. M45*	Sulfitobacter phage ICBM12	ICBM12	9	21, "vif"	P2	Direct plating	complete	54.842	44.80%
Sulfitobacter sp. M45*	Sulfitobacter phage ICBM13	ICBM13	9	21, "vif"	P2	Direct plating	complete	53.694	44.80%
Sulfitobacter sp. M47*	Sulfitobacter phage ICBM15	ICBM15	24		P2	Direct plating	partial	51.401	44.70%
Sulfitobacter sp. M53*	Sulfitobacter phage ICBM16	ICBM16	15	28, "vier"	P1	Direct plating	complete	53.115	45.00%
Sulfitobacter sp. M53*	Sulfitobacter phage ICBM17	ICBM17	25		P2	Direct plating	partial	50.382	44.70%
Sulfitobacter sp. M53*	Sulfitobacter phage ICBM18	ICBM18	26	37	P2	Direct plating	complete	52.071	44.80%
Sulfitobacter sp. M53*	Sulfitobacter phage ICBM21	ICBM21	5	15, "zes"	NHS	Direct plating	complete	52.607	44.60%
Sulfitobacter sp. M54*	Sulfitobacter phage ICBM22	ICBM22	9	21, "vif"	P2	Direct plating	complete	54.842	44.80%
Sulfitobacter sp. M55*	Sulfitobacter phage ICBM23	ICBM23	21	33, "een"	P2	Direct plating	complete	53.772	44.90%
Sulfitobacter sp. M55*	Sulfitobacter phage ICBM24	ICBM24	21	33, "een"	NHS	Direct plating	complete	53.772	44.90%
Sulfitobacter sp. M63*	Sulfitobacter phage ICBM25	ICBM25	7	17, "twee"	NHS	Direct plating	complete	55.591	45.00%
Sulfitobacter sp. M66*	Sulfitobacter phage ICBM26	ICBM26	27		P2	Direct plating	partial	51.329	44.70%
Sulfitobacter sp. M66*	Sulfitobacter phage ICBM32	ICBM32	25		NHS	Direct plating	partial	49.909	44.60%
Sulfitobacter sp. M68	Sulfitobacter phage ICBM38	ICBM38	3	13	P4	Direct plating	complete	51.389	44.90%
Sulfitobacter sp. M69*	Sulfitobacter phage ICBM39	ICBM39	28	38	NHS	Direct plating	complete	53.256	44.60%
Sulfitobacter sp. M71	Sulfitobacter phage ICBM40	ICBM40	21	33, "een"	P2	Direct plating	complete	53.772	44.90%
Sulfitobacter sp. M72*	Sulfitobacter phage ICBM41	ICBM41	8	20	NHS	Direct plating	complete	53.787	44.80%
Sulfitobacter sp. M81*	Sulfitobacter phage ICBM42	ICBM42	29	39	NHS	Direct plating	complete	54.434	45.00%
Sulfitobacter sp. M83*	Sulfitobacter phage ICBM43	ICBM43	2	12	P2	Direct plating	complete	54.119	44.70%
Sulfitobacter sp. M85	Sulfitobacter phage ICBM45	ICBM45	30	40	P4	Direct plating	complete	53.766	44.60%

Sulfitobacter sp. M86	Sulfitobacter phage ICBM46	ICBM46	31		NHS	Direct plating	partial	45.405	44.10%
Sulfitobacter sp. M86	Sulfitobacter phage ICBM47	ICBM47	4	14, “drie”	P4	Direct plating	complete	52.388	44.80%
Sulfitobacter sp. M86	Sulfitobacter phage ICBM48	ICBM48	30	40	P4	Direct plating	complete	53.624	44.60%
Sulfitobacter sp. M90*	Sulfitobacter phage ICBM49	ICBM49	9	21, “vif”	P2	Direct plating	complete	54.336	44.80%
Sulfitobacter sp. M92	Sulfitobacter phage ICBM51	ICBM51	29	39	P2	Direct plating	complete	55.142	44.90%
Sulfitobacter sp. M92	Sulfitobacter phage ICBM52	ICBM52	7	17, “twee”	P4	Direct plating	complete	54.694	44.90%
Sulfitobacter sp. M105	Sulfitobacter phage ICBM53	ICBM53	32	41	P2	Direct plating	complete	53.465	45.00%
Sulfitobacter sp. M105	Sulfitobacter phage ICBM54	ICBM54	15	28, “vier”	NHS	Direct plating	complete	53.931	45.00%
Sulfitobacter sp. M157	Sulfitobacter phage ICBM56	ICBM56	33		NHS	Direct plating	partial	54.001	44.70%
Sulfitobacter sp. M165	Sulfitobacter phage ICBM57	ICBM57	9	21, “vif”	P2	Direct plating	complete	54.46	44.80%
Sulfitobacter sp. M165	Sulfitobacter phage ICBM58	ICBM58	9	21, “vif”	P2	Direct plating	complete	54.373	44.90%
Sulfitobacter sp. M165	Sulfitobacter phage ICBM59	ICBM59	9	21, “vif”	NHS	Direct plating	complete	54.967	44.80%
Sulfitobacter sp. M170	Sulfitobacter phage ICBM60	ICBM60	32	41	NHS	Direct plating	complete	53.548	45.00%
Sulfitobacter sp. M170	Sulfitobacter phage ICBM61	ICBM61	34	42	NHS	Direct plating	complete	52.642	44.70%
Sulfitobacter sp. M171*	Sulfitobacter phage ICBM62	ICBM62	20	32	P2	Direct plating	complete	50.122	44.90%
Sulfitobacter sp. M173*	Sulfitobacter phage ICBM63	ICBM63	35		NHS	Direct plating	partial	53.474	44.90%
Sulfitobacter sp. M176*	Sulfitobacter phage ICBM65	ICBM65	9	21, “vif”	NHS	Direct plating	complete	54.288	44.80%
Sulfitobacter sp. M180	Sulfitobacter phage ICBM67	ICBM67	36	43	NHS	Direct plating	complete	51.759	44.60%
Sulfitobacter sp. M180	Sulfitobacter phage ICBM68	ICBM68	18	30	NHS	Direct plating	complete	52.277	44.70%
Sulfitobacter sp. M183*	Sulfitobacter phage ICBM69	ICBM69	3	13	P2	Direct plating	complete	51.379	44.90%
Sulfitobacter sp. M186	Sulfitobacter phage ICBM70	ICBM70	9	21, “vif”	P2	Direct plating	complete	53.694	44.80%
Sulfitobacter sp. M187*	Sulfitobacter phage ICBM71	ICBM71	35		P2	Direct plating	partial	52.603	45.00%
Sulfitobacter sp. M191*	Sulfitobacter phage ICBM72	ICBM72	37		P2	Direct plating	partial	51.034	44.80%
Sulfitobacter sp. M192	Sulfitobacter phage ICBM74	ICBM74	38		NHS	Direct plating	partial	54.605	44.80%
Sulfitobacter sp. M196	Sulfitobacter phage ICBM76	ICBM76	20	32	P2	Direct plating	complete	50.122	44.90%
Sulfitobacter sp. M196	Sulfitobacter phage ICBM77	ICBM77	20	32	NHS	Direct plating	complete	50.195	44.90%

Sulfitobacter sp. M197	Sulfitobacter phage ICBM78	ICBM78	6	16	NHS	Direct plating	complete	53.429	44.90%
Sulfitobacter sp. M197	Sulfitobacter phage ICBM79	ICBM79	21	33, “een”	NHS	Direct plating	complete	53.781	44.90%
Sulfitobacter sp. M199*	Sulfitobacter phage ICBM80	ICBM80	39		NHS	Direct plating	partial	52.003	44.70%
Sulfitobacter sp. M200	Sulfitobacter phage ICBM82	ICBM82	28	38	NHS	Direct plating	complete	53.759	44.60%
Sulfitobacter sp. M201	Sulfitobacter phage ICBM83	ICBM83	40		NHS	Direct plating	partial	51.497	44.60%
Sulfitobacter sp. M201	Sulfitobacter phage ICBM84	ICBM84	41		NHS	Direct plating	partial	52.69	44.70%
Sulfitobacter sp. M207	Sulfitobacter phage ICBM86	ICBM86	42	44	NHS	Direct plating	complete	52.921	44.70%
Sulfitobacter sp. M207	Sulfitobacter phage ICBM87	ICBM87	43	45	NHS	Direct plating	complete	53.362	44.50%
Sulfitobacter sp. M207	Sulfitobacter phage ICBM88	ICBM88	44	46	NHS	Direct plating	complete	53.044	44.70%
Sulfitobacter sp. M244	Sulfitobacter phage ICBM89	ICBM89	45	47	P2	Direct plating	complete	53.957	44.90%
Sulfitobacter sp. M260	Sulfitobacter phage ICBM90	ICBM90	47	48	P2	Direct plating	complete	51.373	44.70%
Sulfitobacter sp. M260	Sulfitobacter phage ICBM91	ICBM91	3	13	P4	Direct plating	complete	51.389	44.90%
Sulfitobacter sp. M271*	Sulfitobacter phage ICBM94	ICBM94	48	49	P2	Direct plating	complete	54.72	44.70%
Sulfitobacter sp. M283*	Sulfitobacter phage ICBM95	ICBM95	9	21, “vif”	P2	Direct plating	complete	54.142	44.90%
Sulfitobacter sp. M300*	Sulfitobacter phage ICBM99	ICBM99	26	37	P2	Direct plating	complete	52.071	44.80%
Sulfitobacter sp. M303	Sulfitobacter phage ICBM100	ICBM100	2	12	P2	Direct plating	complete	54.118	44.70%
Sulfitobacter sp. M68	Sulfitobacter phage ICBM102	ICBM102	3	13	P2	Direct plating	complete	50.475	44.90%
Sulfitobacter sp. M86	Sulfitobacter phage ICBM103	ICBM103	4	14, “drie”	P2	Direct plating	complete	50.611	44.90%
Sulfitobacter sp. M276	Sulfitobacter phage ICBM105	ICBM105	3	13	P2	Direct plating	complete	51.379	44.90%
Sulfitobacter sp. M53*	Sulfitobacter phage ICBM107	ICBM107	5	15, “zes”	P2	Direct plating	complete	52.423	44.70%
Sulfitobacter sp. M55*	Sulfitobacter phage ICBM109	ICBM109	6	16	NHS	Direct plating	complete	51.998	45.00%
Sulfitobacter sp. M63*	Sulfitobacter phage ICBM110	ICBM110	7	17, “twee”	NHS	Direct plating	complete	55.587	45.00%
Sulfitobacter sp. M68	Sulfitobacter phage ICBM113	ICBM113	3	13	P4	Direct plating	complete	52.367	44.90%
Sulfitobacter sp. M72*	Sulfitobacter phage ICBM119	ICBM119	8	20	NHS	Direct plating	complete	53.787	44.80%
Sulfitobacter sp. M72*	Sulfitobacter phage ICBM120	ICBM120	10		P4	Direct plating	partial	53.159	44.60%
Sulfitobacter sp. M85	Sulfitobacter phage ICBM124	ICBM124	11	23	P4	Direct plating	complete	54.296	44.80%

Sulfitobacter sp. M86	Sulfitobacter phage ICBM126	ICBM126	12	24	NHS	Direct plating	complete	53.881	44.80%
Sulfitobacter sp. M86	Sulfitobacter phage ICBM127	ICBM127	13	25	NHS	Direct plating	complete	52.781	44.90%
Sulfitobacter sp. M91*	Sulfitobacter phage ICBM128	ICBM128	14	26	NHS	Direct plating	complete	54.612	44.70%
Sulfitobacter sp. M105	Sulfitobacter phage ICBM131	ICBM131	15	28, “vier”	NHS	Direct plating	complete	52.646	45.00%
Sulfitobacter sp. M165	Sulfitobacter phage ICBM133	ICBM133	9	21, “vif”	NHS	Direct plating	complete	54.967	44.80%
Sulfitobacter sp. M170	Sulfitobacter phage ICBM134	ICBM134	9	21, “vif”	NHS	Direct plating	complete	54.461	44.80%
Sulfitobacter sp. M172	Sulfitobacter phage ICBM135	ICBM135	16		NHS	Direct plating	partial	51.877	45.00%
Sulfitobacter sp. M176*	Sulfitobacter phage ICBM137	ICBM137	17	29	NHS	Direct plating	complete	53.554	44.80%
Sulfitobacter sp. M180	Sulfitobacter phage ICBM138	ICBM138	18	30	NHS	Direct plating	complete	52.782	44.80%
Sulfitobacter sp. M186	Sulfitobacter phage ICBM139	ICBM139	19	31	NHS	Direct plating	complete	53.722	44.70%
Sulfitobacter sp. M196	Sulfitobacter phage ICBM143	ICBM143	20	32	NHS	Direct plating	complete	52.012	44.70%
Sulfitobacter sp. M197	Sulfitobacter phage ICBM145	ICBM145	21	33, “een”	P2	Direct plating	complete	53.772	44.90%
Sulfitobacter sp. M242*	Sulfitobacter phage ICBM146	ICBM146	22	34	P2	Direct plating	complete	52.265	44.60%
Sulfitobacter sp. M260	Sulfitobacter phage ICBM147	ICBM147	23	35	P2	Direct plating	complete	51.69	44.80%
Sulfitobacter sp. M351*	Sulfitobacter phage ICBM154	ICBM154	22	34	P2	Direct plating	complete	52.265	44.60%

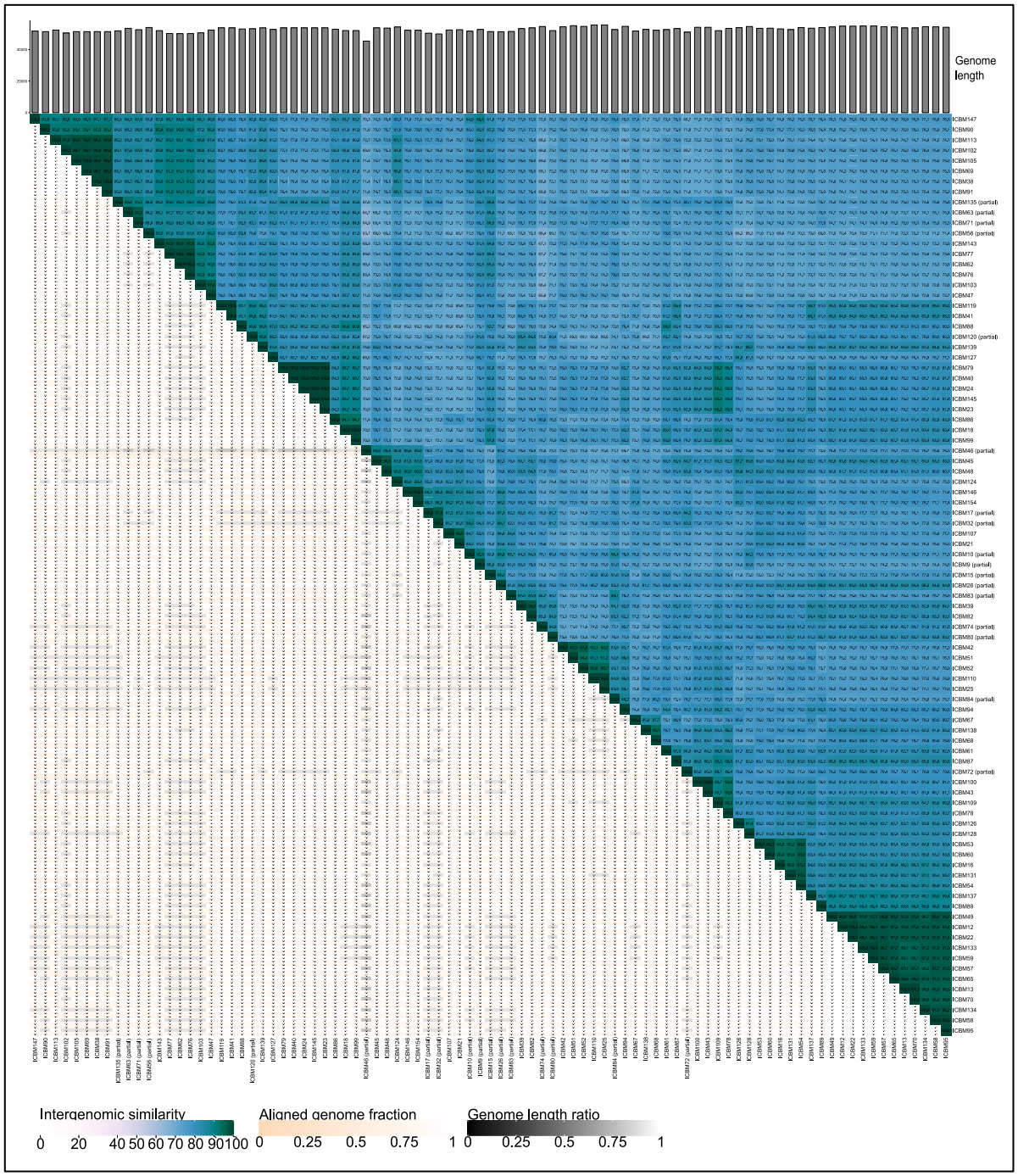


Fig. 56: Nucleotide-based intergenomic identities within the “Sulfivirus” genus, as calculated with VIRIDIC. Identies higher than 99.9 % are rounded up to 100 %.

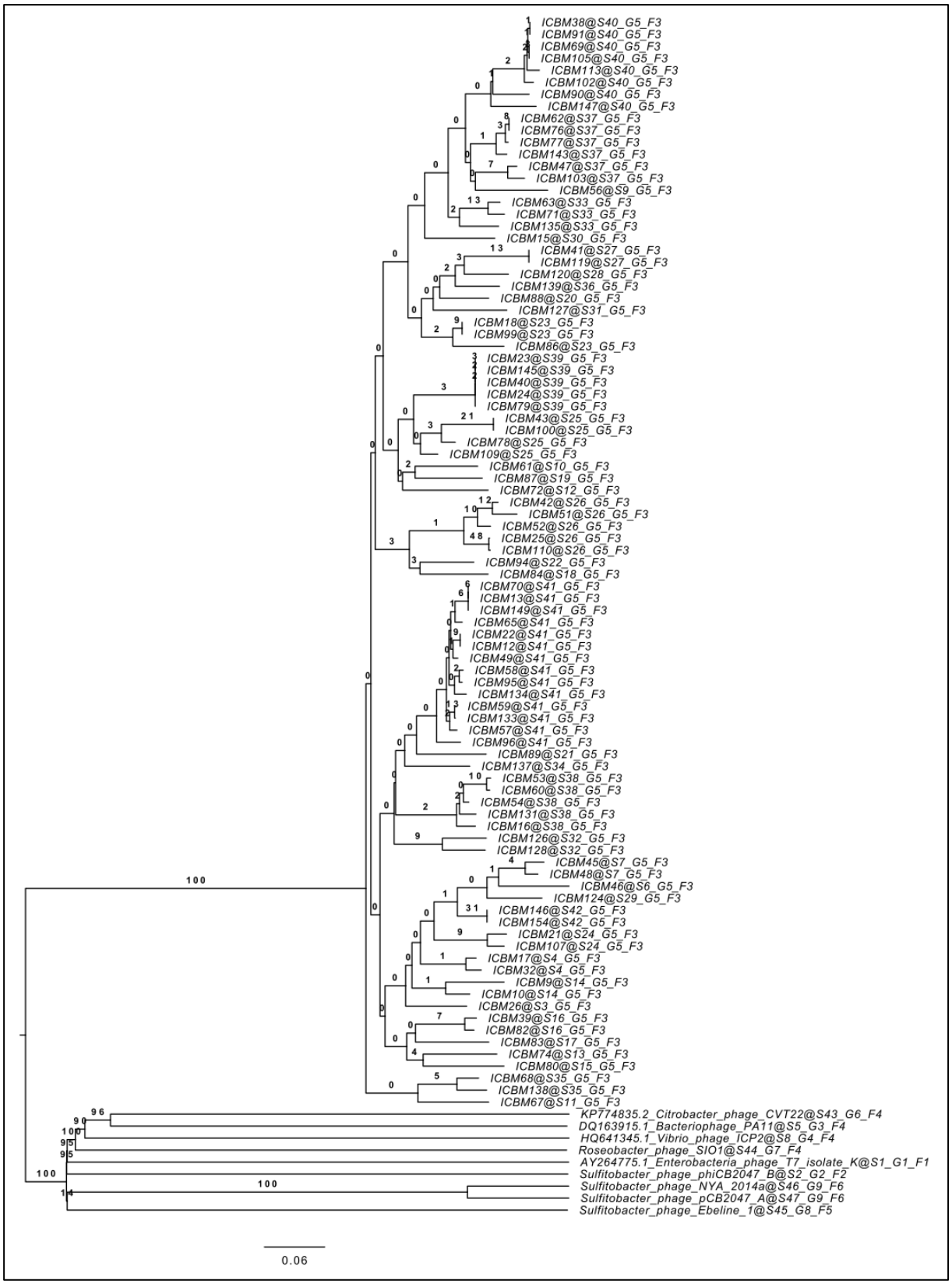


Fig. 57: Whole genome-based phylogeny of the 89 sulfiviruses and reference genomes, as obtained with the Genome-BLAST Distance Phylogeny method implemented in the VICTOR web service (nucleic acid data). Phages ICBM96 and ICBM149 turned out to be contaminated and were later removed from the collection. OPTSIL clustering into family (F), genus (G) and species (S) is included in the label after the phage ID.

Table 43: Core proteins of the "Sulfivirus" genus.

Category	PC ID	Protein function	PC ID in chapter 4.3.5.4.	Alternative functional annotation in chapter 4.3.5.4.
DNA, RNA and nucleotide metabolism	1	DNA polymerase	94	
	19	Exonuclease	90	
	18	Thymidylate synthase ThyX	1	
	2	HNH-endonuclease	305	
	10	Primase/helicase	200	
	28	Ribonucleotide reductase	300	
	9	RNA polymerase sigma factor	167	Hypothetical protein
	63	Endonuclease	179	HNH-endonuclease
Lysis	40	Lysozyme	3	Endolysin (lysozyme-peptidase)
Head and packaging	60	Capsid decoration protein	151	Capsid protein
	61	Capsid maturation protease	196	
	59	Major capsid protein	104	
	62	Portal protein	105	
	54	Terminase large subunit	82	
	20	Terminase small subunit	175	HNH endonuclease
Tail	53	Baseplate protein	368	Concanavalin A-like lectins/glucanases
	47	Major tail tube protein	99	Minor tail protein
	42	Putative tail fiber adhesion protein	2	Tail fiber protein
	44	Tail tape measure protein	310	
Connector	50	Putative head-tail joining protein / tail attachment protein	101	Head closure Hc1
	48	Putative tail terminator protein	100	Tail completion Tc1
Anti-host defense	25	DNA adenine methylase	180	
Other	24	Metallo-phosphoesterase	96	
	26	phosphoribosyl-ATP pyrophosphohydrolase MazG	181	Nucleotide pyrophosphohydrolase
Unknown	57, 3, 4, 5, 6, 7, 58, 8,	Hypothetical protein	158, 160, 161, 162, 163, 164, 165, 166, 168, 152,	

11, 43, 66, 67, 12, 13, 14, 15, 16, 17, 21, 22, 23, 27, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 45, 46, 49, 51, 52, 55, 56	259, 169, 170, 147, 171, 172, 173, 174, 176, 177, 178, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 97, 98, 195, 102, 103, 106, 197

Table 44: Non-core protein clusters of the "Sulfivirus" genus.

PC ID	Category	Protein function	Number of sulfivirus genomes containing the PC
86	anti-host defense	DNA_adenine_methylase	40
92		DNA_adenine_methylase	31
80		DNA_cytosine_methylase	50
119		putative_anti-restriction_nuclease	8
91	DNA, RNA and nucleotide metabolism	HNH_endonuclease	33
130		HNH_endonuclease	6
138		HNH_endonuclease	4
65		HNH_homing_endonuclease	86
110		Polynucleotide_kinase_/_phosphatase	11
90		putative_HNH_endonuclease	34
89		ribonucleotide_reductase_cobalamin_dependent	34
109		RNA_ligase	11
73		tRNA_pseudouridine_synthase_D	70
153	moron, auxiliary metabolic gene and host takeover	Phosphoadenosine_phosphosulfate_reductase	2
82	tail	putative tail fiber	49
116		putative_tail_protein	8
105		tail_fiber_assembly_protein	18
111	unknown	GLTT_repeat_(6_copies)	11

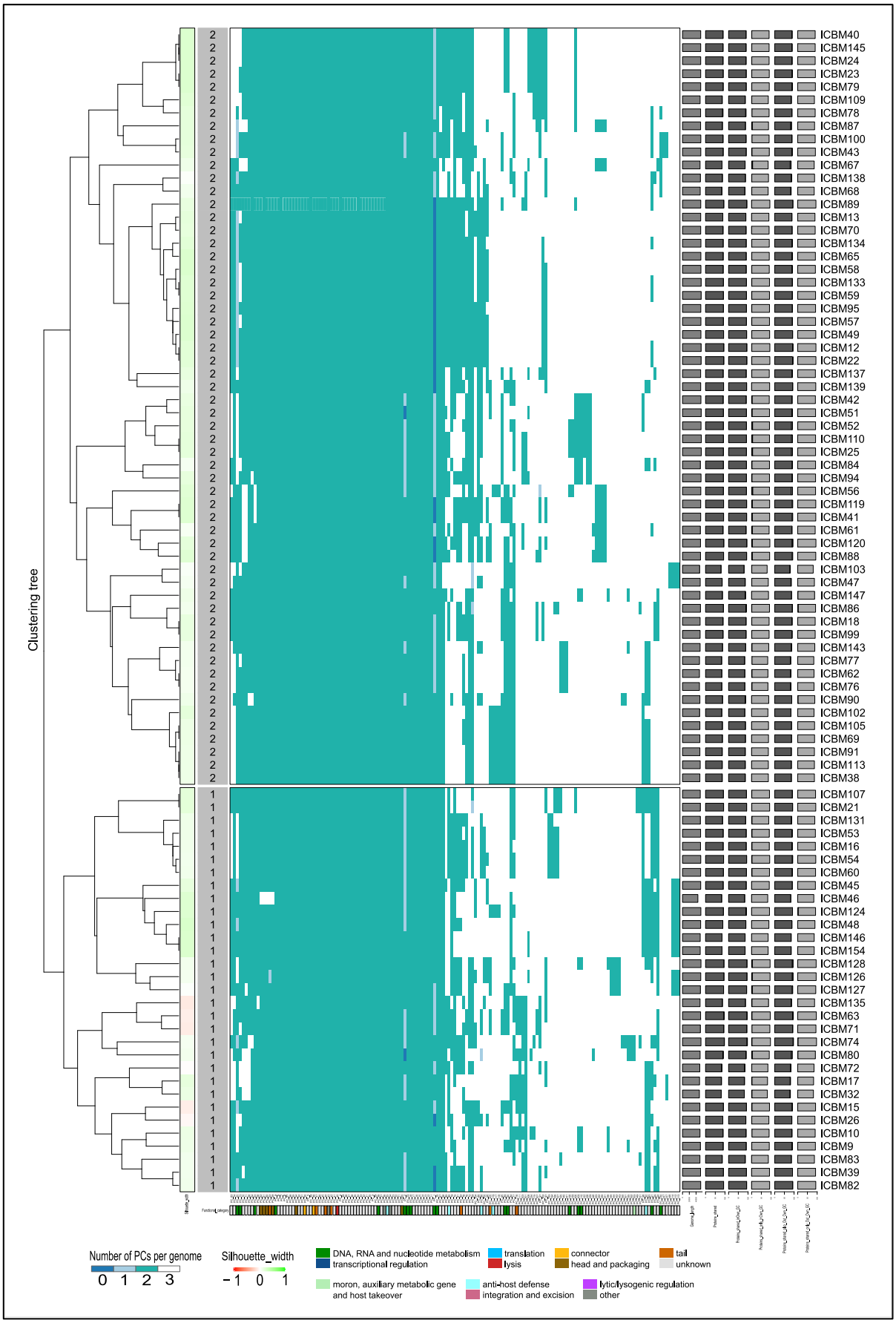


Fig. 58: Hierarchical clustering of the sulfivirus genomes when using relaxed parameters for protein clusterings. Clustering based on log evalues. Thresholds for matches being removed: evalue >0.00001, bitscore <50, coverage <0, percentage identity <0. Clustering distance of 0.9. Affiliation of protein clusters to functional categories is marked by different colors.

5.3.2.1. Composition of sulfiviral genomes

As shortly described already in chapter 4.3.5.4, the genomic architecture of the sulfiviruses comprised two arms, with genes organized in functional modules and encoded in forward direction on the left arm and in reverse direction on the right arm (Fig. 59). Apart from the two arms, 5 - 11 genes annotated as “hypothetical” were encoded in reverse direction at the very beginning of the genome. The genome of Sulfitobacter phage ICBM128 was the one with the most ORFs predicted among the complete genomes. Its architecture is shown in figure 59 and Table 46, as a representative of the “Sulfivirus” genus. For the genome maps and gene annotations of all complete sulfivirus genomes, see figure S16 in the appendix and SI files S5-4 and S5-5.

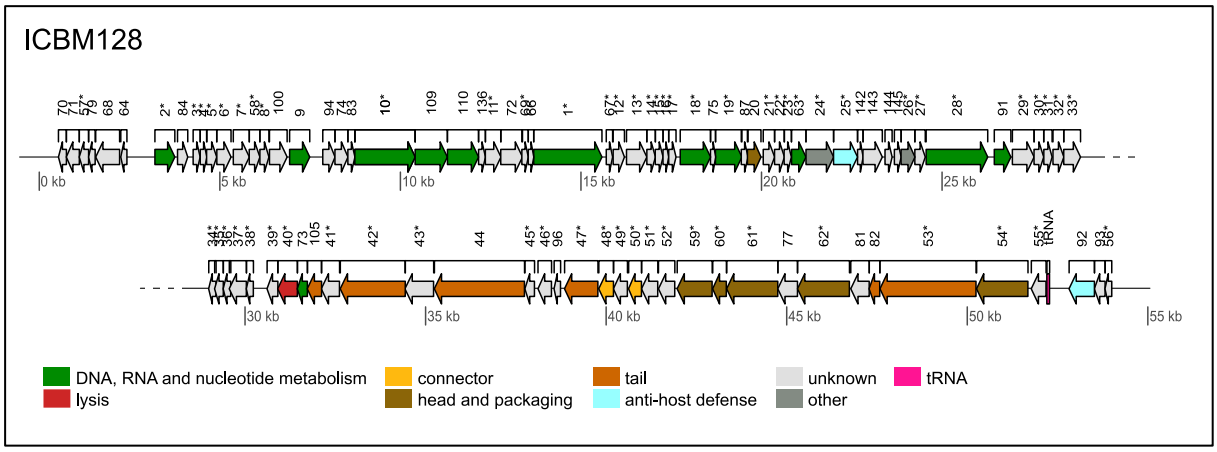


Fig. 59: Genome map of Sulfitobacter phage ICBM128. *core gene.

On the left arm of the genome, genes of the functional category “DNA, RNA and nucleotide metabolism” were encoded (Table 46, Fig. 59). In this module, the protein-encoding genes with functional annotations were: an HNH-endonuclease (PC_2), an RNA polymerase sigma factor (PC_9), a bifunctional primase/helicase gene (PC_10), an RNA ligase gene (PC_109), a polynucleotide kinase/phosphatase gene (PC_110), the DNA polymerase gene (PC_1), a thymidylate synthase ThyX gene (PC_18), an exonuclease gene (PC_19), a ribonucleotide reductase gene (PC_28) and finally another HNH-endonuclease gene (PC_91). The polynucleotide kinase/phosphatase gene and the RNA ligase gene were only present in (the same) eleven sulfivirus genomes. In addition, the three core genes coding for the terminase small subunit (PC_20) (category “head and packaging”), a metallo-phosphoesterase (PC_24) and a phosphoribosyl-ATP pyrophosphohydrolase MazG (PC_26, category “other”) were located on the left genomic arm. Furthermore, DNA methylases were encoded on this arm, with their number (1 - 3) and position varying between sulfiviruses (PC_25, PC_86 = DNA adenine methylase, PC_80 = DNA cytosine methylase). In some genomes, another DNA adenine

methylase gene was encoded in reverse direction at the very end of the genome (PC_92). Methylases can serve for protection against restriction-modification systems. Thus, they were assigned to the functional category “anti-host defense”. Seven sulfiviruses (species 3 and 11) additionally had a putative anti-restriction nuclease encoded at the very end of the left arm, in the middle of the genome (PC_119). In general, the sulfivirus genomes differed in the number and position of genes annotated as encoding “hypothetical proteins”. In contrast, the structure of the morphology module on the right arm was more conserved, also regarding the hypothetical proteins. The module started with the gene coding for the terminase large subunit (PC_54, functional category “head and packaging”), followed by the baseplate protein (PC_53) and a putative tail fiber (PC_82). The tail fiber gene was annotated only in about half of the genomes. Subsequently, the other genes of the “head and packaging” category were clustered together, coding for the portal protein (PC_62), the capsid maturation protease (PC_61), the capsid decoration protein (PC_60) and the major capsid protein (PC_59). They were followed by two genes coding for the “connectors”, the tail terminator protein (PC_48) and the head-tail joining protein (PC_50). The genes coding for the tail components were again grouped together, i.e. the major tail tube protein (PC_47), the tail tape measure protein (PC_44), a putative tail fiber adhesion protein (PC_42) and a tail fiber assembly protein (PC_105). In 15 of the complete genomes, another tail fiber assembly protein (PC_105) was annotated. At the end of the right arm, towards the middle of the genome, a lysozyme encoding gene was detected (PC_40).

As mentioned above, the majority of protein clusters (PCs) was present in all sulfivirus genomes, resembling the core proteome. Out of 94 PCs that were non-core proteins, 20 could be functionally annotated (Table 44). Most of them belonged to the categories “anti-host defense”, “DNA, RNA and nucleotide metabolism” and “tail”, indicating that major differences between the sulfivirus genomes were related to the adaptation to different host strains. While the different DNA methylases and the anti-restriction nuclease act against bacterial defense on DNA level, varying tail fiber proteins could serve for the attachment to differential hosts. Among the non-core PCs of the category “DNA, RNA and nucleotide metabolism” there were five PCs annotated as HNH homing endonucleases. Homing endonucleases are mobile elements that recognize DNA sequences lacking their gene at a specific site, cleave it and insert themselves into it by gene conversion (Belfort 2005; Stoddard 2011). HNH endonucleases were first known for the mediation of intron and intein mobility, in a process called “homing” (Dujon 1989). However, “intron-less homing” has been observed between phage genomes, in which freestanding endonucleases are transferred, with co-conversion of parts of the flanking DNA (Belle et al. 2002; Liu et al. 2003). Often, HNH endonucleases are present as insertions in other

genes, especially frequently in DNA polymerases and RNR genes (Novikova et al. 2016; Bellas et al. 2020). The sulfivirus genomes differed in number (1 - 6) and position of these HNH homing endonucleases (PC_2, PC_65, PC_90, PC_91, PC_130, and PC_138). Some of them had an HNH module inserted in the primase/helicase gene (PC_10). Furthermore, the ribonucleotide reductase (RNR) was either encoded by one large gene (PC_28) or split in two shorter genes (PC_28 and PC_89), interrupted by a putative HNH endonuclease. Kala et al. (2014) discovered that in *E. coli* phage HK97, an HNH protein is associated to the large terminase protein and plays an essential role in capsid morphogenesis. They found the co-location of HNH and terminase genes to be widespread among long-tailed phages and suggested that the terminase-associated HNH proteins represent a distinct subfamily. Also in many of the sulfivirus genomes, there were HNH endonuclease genes (PC_130, PC_138, and PC_65) encoded in close proximity to the terminase large subunit gene (PC_54).

The nature of the DNA polymerase gene (PC_1) also varied between sulfivirus genomes. Some had one large gene with both domains, while others had the DNA polymerase domain and the exonuclease domain divided into two genes. In some genomes, there was a third DNA polymerase gene, separated from the others by a methylase gene. Two sulfiviruses (ICBM32 and ICBM63, both with partial genomes) possessed a phosphoadenosine phosphosulfate (PAPS) reductase. As mentioned also in chaper 4.3.5.5 about the “Diferiteviridae” family, this enzyme has been found in phages before and could potentially promote host growth under sulfur limited conditions (Summer et al. 2006; Summer et al. 2007b).

Almost half of the sulfivirus genomes had a tRNA gene (Table 45). The tRNA encoded for asparagine, cysteine or arginine, and was located at the very end of the genome, directly upstream of the hypothetical protein PC_55 and the terminase large subunit (PC_54). Almost all genomes possessing a tRNA gene also encoded the gene coding for a tRNA pseudouridine synthase D (PC_73), located next to the lysis gene in the middle of the genome. As mentioned earlier in chapter 4.3.6, phage tRNAs can enhance translation efficiency and progeny production. For some phages, a correlation between the number of tRNAs encoded in the viral genome and the number of infected host strains has been reported (Holmfeldt et al. 2013). However, Holmfeldt et al. (2013) compared the host ranges of phages with 16 and 24 tRNAs encoded with those having one or none tRNAs. When comparing sulfiviruses with one tRNA and those without a tRNA, no pattern in host range size could be observed (Fig. 63 in chapter 5.3.4.). Xu et al. (2018) hypothesized that (cyano-) phages with high numbers of tRNA genes were prevailing in nutrient-rich environments, whereas in oligotrophic habitats phages had less

tRNA genes. At least for our sulfiviruses and roseophage isolates, which originate from algal bloom samples, this hypothesis does not apply.

Hierarchical clustering of the genomes based on individual proteins instead of protein clusters (using VirClust, identity and coverage thresholds = 100%) revealed that if smaller differences such as point mutations are considered, the sulfivirus genomes were highly diverse (Fig. 60). They displayed a high degree of microdiversity. In order to investigate how this microdiversity could evolve, the next analytical step was to look for traces of recombination between the phage genomes.

Table 45: tRNAs predicted in the genomes of the "Sulfivirus" genus.

Sequence name	tRNA start	tRNA end	tRNA type
ICBM9	5330	5406	Asn
ICBM10	46368	46292	Asn
ICBM12	52784	52710	Cys
ICBM18	50530	50454	Asn
ICBM21	50164	50088	Asn
ICBM22	52784	52710	Cys
ICBM38	49008	48932	Asn
ICBM41	52293	52217	Asn
ICBM45	51357	51281	Asn
ICBM47	50941	50865	Asn
ICBM48	51207	51131	Asn
ICBM49	52276	52202	Cys
ICBM53	51822	51748	Cys
ICBM56	8044	8120	Asn
ICBM61	51137	51061	Asn
ICBM62	47775	47699	Asn
ICBM63	48539	48463	Asn
ICBM69	48997	48921	Asn
ICBM71	47664	47588	Asn
ICBM76	47775	47699	Asn
ICBM77	47852	47776	Asn
ICBM82	27027	27103	Arg
ICBM86	51372	51296	Asn
ICBM88	51543	51467	Asn
ICBM89	52342	52268	Cys
ICBM90	49029	48953	Asn
ICBM91	49008	48932	Asn
ICBM99	50530	50454	Asn
ICBM102	49005	48929	Asn
ICBM103	49146	49070	Asn
ICBM105	48997	48921	Asn

ICBM113	49986	49910	Asn
ICBM119	52293	52217	Asn
ICBM120	48226	48150	Asn
ICBM124	51913	51837	Asn
ICBM126	52433	52357	Asn
ICBM127	51333	51257	Asn
ICBM128	52259	52183	Asn
ICBM137	51921	51847	Cys
ICBM143	49659	49583	Asn

Table 46: Gene annotations of Lentibacter phage ICBM128. Strand 1: forward orientation. Strand -1: reverse orientation. Hp = hypothetical protein. *Core proteins of sulfiviruses. (Note: PC-IDs are different from those in chapter 4.3.5.4.).

Gene ID	Gene start	Gene end	Gene length [bp]	Strand	PC ID	Protein function	Category
gene_1	533	754	222	-1	70	hp	unknown
gene_2	751	1110	360	-1	71	hp	unknown
gene_3	1107	1367	261	-1	57*	hp*	unknown
gene_4	1367	1549	183	-1	79	hp	unknown
gene_5	1565	2227	663	-1	68	hp	unknown
gene_6	2259	2429	171	-1	64	hp	unknown
gene_7	3204	3755	552	1	2*	HNH-endonuclease*	DNA, RNA and nucleotide metabolism
gene_8	3830	4114	285	1	84	hp	unknown
gene_9	4262	4456	195	1	3*	hp*	unknown
gene_10	4459	4644	186	1	4*	hp*	unknown
gene_11	4634	4915	282	1	5*	hp*	unknown
gene_12	4917	5312	396	1	6*	hp*	unknown
gene_13	5373	5819	447	1	7*	hp*	unknown
gene_14	5816	6112	297	1	58*	hp*	unknown
gene_15	6109	6360	252	1	8*	hp*	unknown
gene_16	6375	6866	492	1	100	hp	unknown
gene_17	6933	7490	558	1	9*	RNA polymerase sigma factor*	DNA, RNA and nucleotide metabolism
gene_18	7851	8192	342	1	94	hp	unknown
gene_19	8189	8548	360	1	74	hp	unknown
gene_20	8545	8742	198	1	83	hp	unknown
gene_21	8739	10409	1671	1	10*	Primase/helicase*	DNA, RNA and nucleotide metabolism
gene_22	10409	11302	894	1	109	RNA ligase	DNA, RNA and nucleotide metabolism
gene_23	11303	12166	864	1	110	Polynucleotide kinase / phosphatase	DNA, RNA and nucleotide metabolism
gene_24	12167	12355	189	1	136	hp	unknown
gene_25	12352	12786	435	1	11*	hp*	unknown
gene_26	12779	13381	603	1	72	hp	unknown
gene_27	13365	13544	180	1	69	hp	unknown
gene_28	13541	13696	156	1	66*	hp*	unknown
gene_29	13693	15582	1890	1	1*	DNA polymerase*	DNA, RNA and nucleotide metabolism
gene_30	15703	15876	174	1	67*	hp*	unknown

gene_31	15891	16205	315	1	12*	hp*	unknown
gene_32	16263	16826	564	1	13*	hp*	unknown
gene_33	16826	17065	240	1	14*	hp*	unknown
gene_34	17058	17276	219	1	15*	hp*	unknown
gene_35	17273	17431	159	1	16*	hp*	unknown
gene_36	17431	17634	204	1	17*	hp*	unknown
gene_37	17750	18589	840	1	18*	Thymidylate synthase ThyX*	DNA, RNA and nucleotide metabolism
gene_38	18586	18729	144	1	75	hp	unknown
gene_39	18726	19457	732	1	19*	Exonuclease*	DNA, RNA and nucleotide metabolism
gene_40	19439	19627	189	1	87	hp	unknown
gene_41	19611	19985	375	1	20*	Terminase, small subunit*	head and packaging
gene_42	20047	20364	318	1	21*	hp*	unknown
gene_43	20361	20636	276	1	22*	hp*	unknown
gene_44	20626	20826	201	1	23*	hp*	unknown
gene_45	20833	21231	399	1	63*	Endonuclease*	DNA, RNA and nucleotide metabolism
gene_46	21231	21998	768	1	24*	Metallo-phosphoesterase*	other
gene_47	21995	22660	666	1	25*	DNA adenine methylase*	anti-host defense
gene_48	22660	22815	156	1	142	hp	unknown
gene_49	22808	23344	537	1	143	Pentapeptide repeats	unknown
gene_50	23419	23616	198	1	144	hp	unknown
gene_51	23679	23867	189	1	145	hp	unknown
gene_52	23864	24250	387	1	26*	phosphoribosyl-ATP pyrophosphohydrolase MazG*	other
gene_53	24247	24534	288	1	27*	hp*	unknown
gene_54	24560	26263	1704	1	28*	Ribonucleotide reductase*	DNA, RNA and nucleotide metabolism
gene_55	26447	26905	459	1	91	HNH-endonuclease	DNA, RNA and nucleotide metabolism
gene_56	26946	27548	603	1	29*	hp*	unknown
gene_57	27545	27814	270	1	30*	hp*	unknown
gene_58	27817	28047	231	1	31*	hp*	unknown
gene_59	28071	28373	303	1	32*	hp*	unknown
gene_60	28370	28834	465	1	33*	hp*	unknown
gene_61	28973	29176	204	-1	34*	hp*	unknown
gene_62	29139	29378	240	-1	35*	hp*	unknown
gene_63	29375	29584	210	-1	36*	hp*	unknown
gene_64	29544	30017	474	-1	37*	hp*	unknown
gene_65	30017	30211	195	-1	38*	hp*	unknown
gene_66	30590	30892	303	-1	39*	hp*	unknown
gene_67	30885	31430	546	-1	40*	Lysozyme*	lysis
gene_68	31427	31708	282	-1	73	tRNA pseudouridine synthase D	DNA, RNA and nucleotide metabolism
gene_69	31705	32100	396	-1	105	Tail fiber assembly protein	tail
gene_70	32093	32596	504	-1	41*	hp*	unknown
gene_71	32618	34417	1800	-1	42*	Putative tail fiber adhesion protein*	tail
gene_72	34410	35201	792	-1	43*	hp*	unknown
gene_73	35223	37718	2496	-1	44*	Tail tape measure protein*	tail

gene_74	37730	37996	267	-1	45*	hp*	unknown
gene_75	38092	38466	375	-1	46*	hp*	unknown
gene_76	38538	38717	180	-1	96	hp	unknown
gene_77	38828	39751	924	-1	47*	Major tail tube protein*	tail
gene_78	39776	40183	408	-1	48*	Putative tail terminator protein*	connector
gene_79	40180	40569	390	-1	49*	hp*	unknown
gene_80	40600	40959	360	-1	50*	Putative head-tail joining protein / tail attachment protein*	connector
gene_81	40959	41408	450	-1	51*	hp*	unknown
gene_82	41428	41880	453	-1	52*	hp*	unknown
gene_83	41946	42914	969	-1	59*	Major capsid protein*	head and packaging
gene_84	42929	43309	381	-1	60*	Capsid decoration protein*	head and packaging
gene_85	43317	44732	1416	-1	61*	Capsid maturation protease*	head and packaging
gene_86	44732	45277	546	-1	77	hp	unknown
gene_87	45280	46716	1437	-1	62*	Portal protein*	head and packaging
gene_88	46753	47265	513	-1	81	hp	unknown
gene_89	47262	47558	297	-1	82	Putative tail fiber	tail
gene_90	47561	50221	2661	-1	53*	Baseplate protein*	tail
gene_91	50237	51658	1422	-1	54*	Terminase, large subunit*	head and packaging
gene_92	51750	52160	411	-1	55*	hp*	unknown
gene_93	52798	53502	705	-1	92	DNA adenine methylase	anti-host defense
gene_94	53499	53795	297	-1	93	hp	unknown
gene_95	53792	53980	189	-1	56*	hp*	unknown

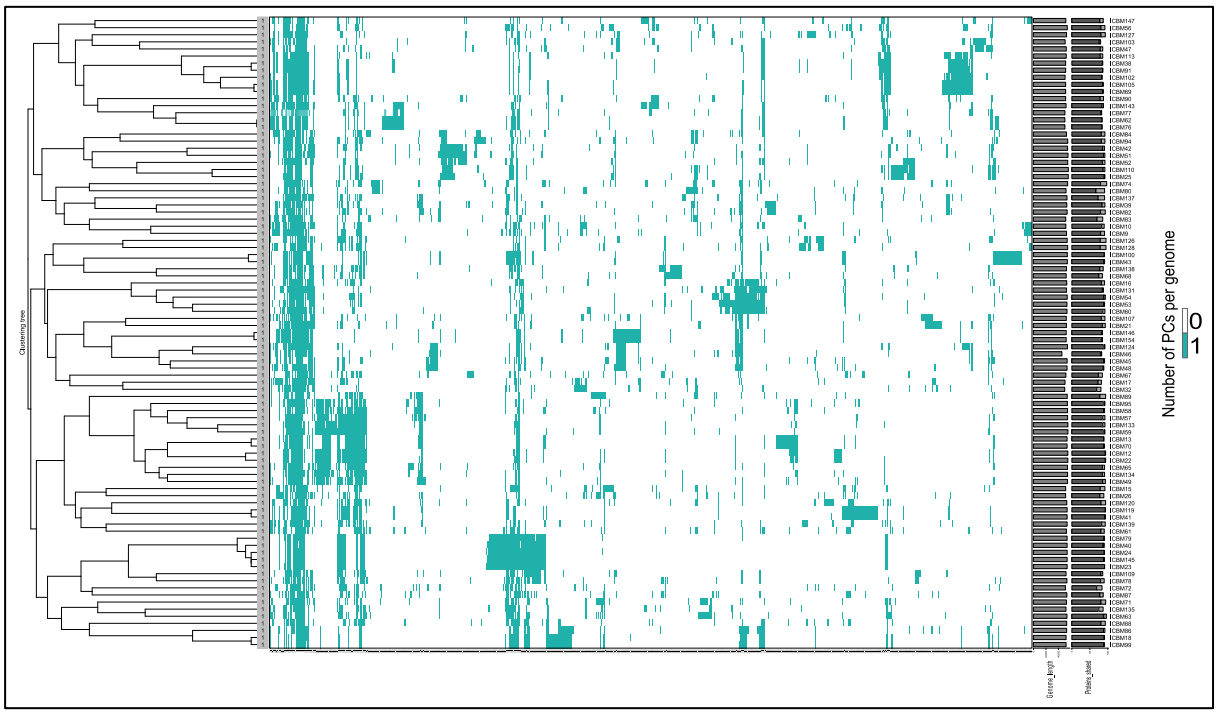


Fig. 60: Hierarchical clustering of the sulfivirus genomes when the intergenomic distances were calculated using protein clusters containing identical proteins (evalue >0.00001, bitscore <50, coverage <100, identity <100%).

5.3.3. Sulfivirus genomes are strongly influenced by recombination

We further investigated the genomic microdiversity of the sulfiviruses, aiming to test the hypothesis that sulfivirus microdiversity is influenced by horizontal gene transfer processes such as recombination. For this purpose, the core regions of the genomes were aligned and used to calculate a maximum likelihood phylogenetic tree. Recombination and mutation sites were determined using ClonalFrameML (Didelot and Wilson 2015) and the phylogeny was reconstructed accounting for recombination (Fig. 61). Overall, many recombination sites were detected in the sulfivirus genomes, having an average length of 55 bp. A higher frequency of recombination was spotted in the region between position 60000 and 70000. It will be a future task to determine which genes are encoded in this region. For this, the positions on the strand of concatenated core genes would need to be translated in actual positions on the genome. The fact that there were overlapping recombination sites, i.e. recombination determined at the same position in different genomes, indicated that there was recombination happening between the sulfivirus genomes. The relative effect of recombination to mutation was equal to r/m = 15.84, suggesting that recombination as a horizontal process contributed more to sulfiviral genome diversification than mutations, so vertically received nucleotide differences. In a follow-up project of this study, that is not included in this dissertation, Ismail Hayani showed in his master thesis that two sulfiviruses (ICBM16 and ICBM18) co-infect their original host and are present at the same time in the same cell. Future experiments should bring proof of intergenomic recombination between sulfivirus strains.

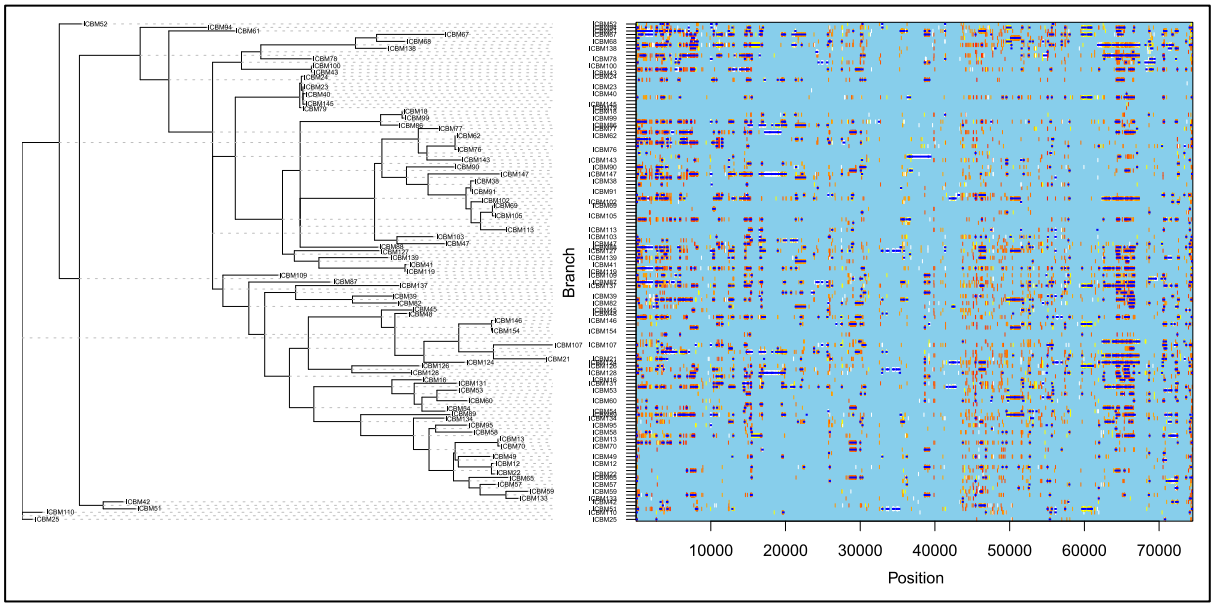


Fig. 61: Reconstructed phylogenetic tree of sulfiviral core regions and sites of recombination and polymorphism detected with ClonalFrameML. The tree reflects the true phylogeny, only branch lengths differ from the original maximum likelihood tree, as they account for recombination. For each branch of the tree and any position in the respective genome, recombination is

marked in blue. Polymorphic sites are marked either in white (no homoplasy) or in yellow to red (increasing levels of homoplasy).

5.3.4. Sulfiviruses display wide range of host specificity

The host range of the sulfiviruses was determined by testing all 89 phages against 60 *Sulfitobacter* host strains, which represented the original isolation hosts of the sulfiviruses. The host range tests were performed using the streak assay method and the results were confirmed by triplicates. The sulfiviruses displayed a wide range of host specificity, with numbers of infected host strains between 2 and 25. Many sulfiviruses infected 6 to 10 strains (Fig. 62 and 63). Clustering of the sulfiviruses based on the infection pattern revealed that phages of the same species had a rather similar host range (Fig. 63). The members of all species clusters created with VIRIDIC (Moraru et al. 2020) clustered together in the heatmap, when infectivity patterns were used as clustering criterium. The only exception was species 9, with one phage clustering slightly apart. However, also for this phage the majority of the infected hosts overlapped with those of the other species members. Still, infectivity patterns varied even within sulfivirus species. Only in four cases did two phages have the exact same host range, respectively. They had very high intergenomic sequence identities: ICBM69 and ICBM91 (99.7%), ICBM62 and ICBM77 (99.4%), ICBM146 and ICBM154 (99.998%), and ICBM43 and ICBM100 (99.998%).

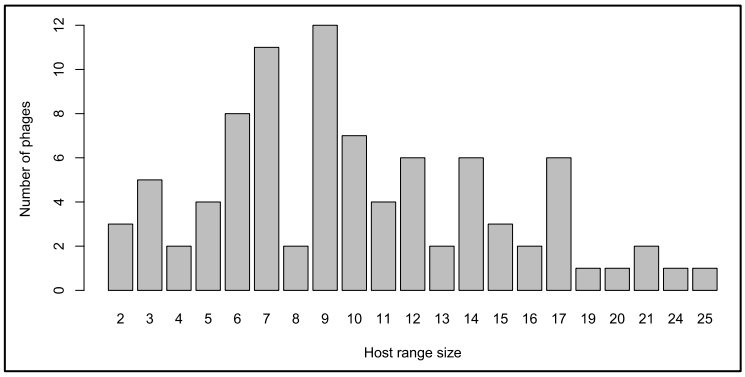


Fig. 62: Distribution of number of infected hosts among sulfiviruses.

Reordering the phage columns in the heatmap according to the core region-based phylogenetic tree obtained with ClonalFrameML resulted in an even better clustering of the infectivity patterns (Fig. 64). This new heatmap showed that the host range patterns depend on phylogenomic proximity: strains belonging to the same species have similar infectivity patterns, and closely related species share more hosts than further related species. This illustrates that the intra-species microdiversity of the sulfiviruses influences the host range (Fig. 64). In figure 65, several metadata, including the isolation hosts and sources, i.e. the seawater samples, of the

sulfiviruses are displayed on top. Phages with the same isolation host had similar host ranges, even if they did not belong to the same species (Fig. 65). This indicates that the genomic identity of the original isolation host also plays a role for the infectivity of the virus on other strains. However, the isolation host was also the strain the respective phage was cultivated with during preparation for the host range assay. Thus, phages with the same isolation and cultivation host strain could have been adapted to this strain at the moment of testing, creating a bias and leading to similar host range results. No clear correlation between isolation source and host range could be detected. For example, all three phages isolated from strain M105 had a similar host range, even though they belonged to different species and originated from different water sources (P2 and NHS). The isolation host seemed to be more relevant for the host range than the seawater sample. However, this could again be an indication, that the determined host range was influenced by the cultiviation conditions and may not reflect the true situation in the environment. After all, although infection was tested in triplicates and with the method producing single plaques, the results still need to be treated with caution, having in mind the bias of cultivation conditions, adaptation to the cultivation host and possibly undetected infections of low efficiency.

If we refer to the definition of de Jonge et al. (2019) (see introductory chapter 1.4.1.), the sulfiviruses can be described as broad-host-range phages, in the sense that together, as a species-level population they are able to infect multiple host strains of the same species. Of course, there are other opinions, defining “broad host range” by infection of different species or even genera. Since bacteria of different higher-level taxa were not included in the host range assay, it remains to be investigated whether the sulfiviruses are able to infect bacteria of different species or genera.

Evaluating the results from the perspective of the host strains revealed that their susceptibility to the sulfiviruses also differed greatly (Fig. 63). Some strains were only infected by three sulfiviruses, while *Sulfitobacter* sp. M283 was even infected by 63 sulfiviruses. Moreover, strains with identical ITS sequence were very differently infected by the sulfiviruses. To search for potential reasons, 30 of these strains were chosen for whole genome sequencing. They were selected in such a way that they had very different infection patterns and numbers of infecting viruses (Fig. 63 and 66).

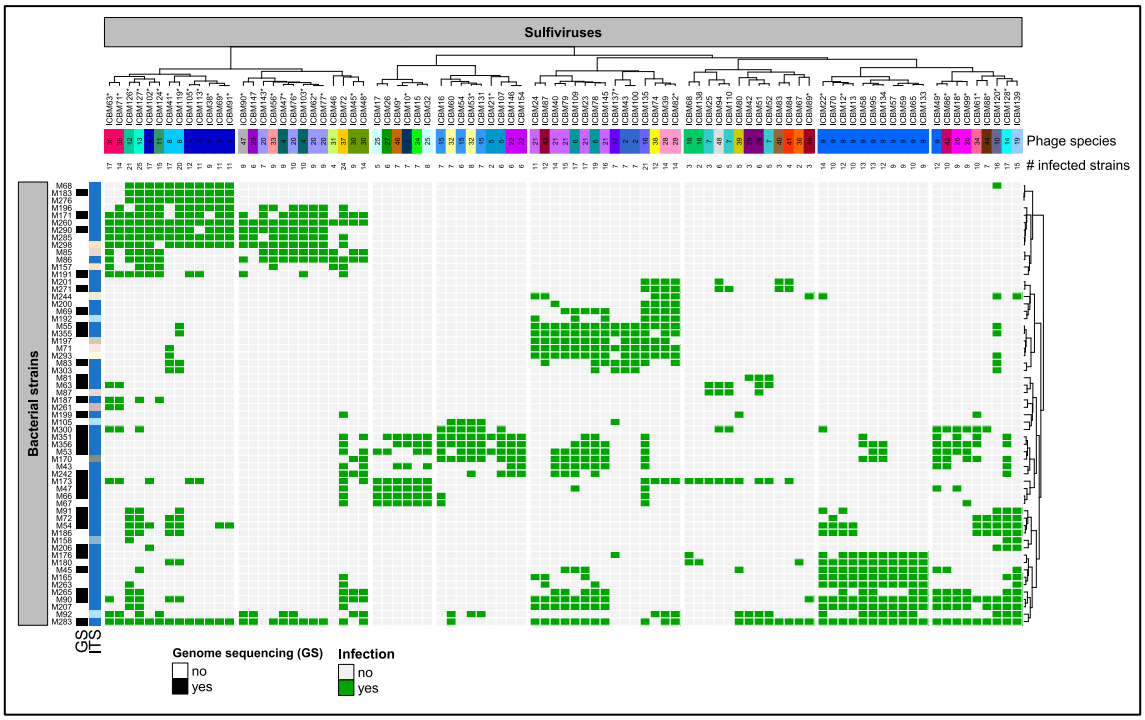


Fig. 63: Host ranges of 89 sulfiviruses tested against their original isolation hosts (60 strains). Positive infection is indicated by green squares. Dendrograms visualize clustering of phages and bacteria based on infection pattern. Assignment of phages to species (VIRIDIC) is displayed on top by colors and numbers. Sum of infected strains for each is phage is displayed on top. On the left, black and white squares indicated whether the bacterial strain was whole genome sequenced. Clustering of the host strains into groups based on ITS sequence similarity is represented by colors. *phage genome encoding a tRNA gene.

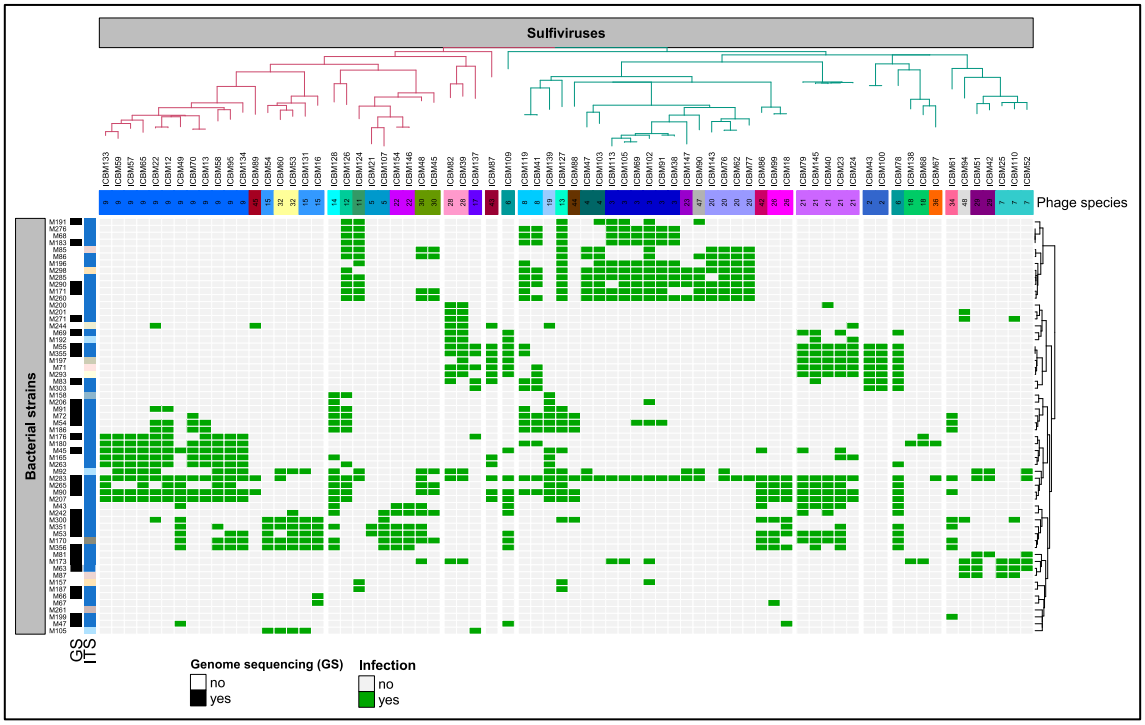


Fig. 64: Host ranges clustered by ClonalFrameML phylogenetic tree. 72 sulfiviruses (with complete genomes) tested against their original isolation hosts (60 strains). Positive infection is indicated by green squares. Dendrogram visualizes clustering of bacteria based on infection pattern. Assignment of phages to species (VIRIDIC) is displayed on top by colors and numbers. On the left, black and white squares indicated whether the bacterial strain was whole genome sequenced. Clustering of the host strains into groups based on ITS sequence similarity is represented by colors.

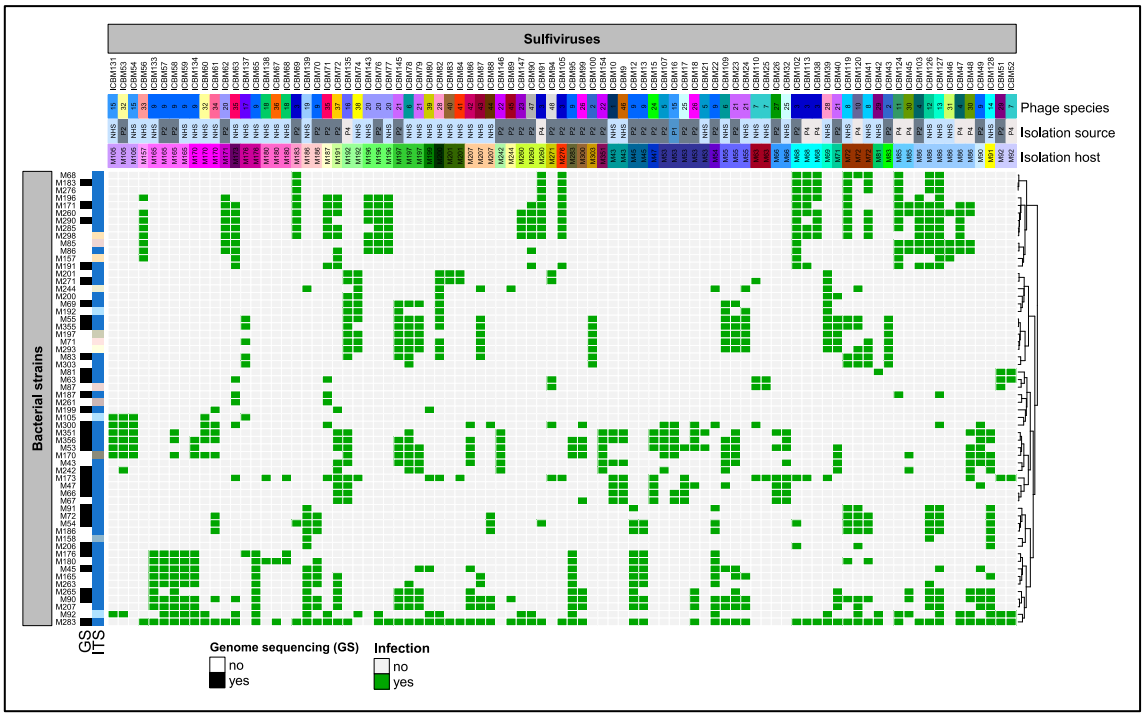


Fig. 65: Host ranges of 89 sulfiviruses tested against their original isolation hosts (60 strains), clustered based on isolation hosts. Positive infection is indicated by green squares. Assignment of phages to species (VIRIDIC), isolation source and isolation host strain are displayed on top. Dendrogram visualizes clustering of bacteria based on infection pattern. On the left, black and white squares indicated whether the bacterial strain was whole genome sequenced. Clustering of the host strains into groups based on ITS sequence similarity is represented by colors.

5.3.5. Genome analysis of 30 sulfivirus host strains

As mentioned before, thirty sulfivirus host strains were chosen for whole-genome sequencing. Genome sizes ranged from 3.85 to 4.41 Mb, with 3721 to 4293 predicted ORFs (Table 47, gen ome sequences will be available in the NCBI database). The chromosomes were 3.55 to 3.95 Mb in size, with 58.1 - 58.3 % G+C content. The strains possessed up to six extrachrom osomal elements (plasmids) of different sizes ranging from 5,342 bp to 338,691 bp. The plasm ids could be grouped into 16 clusters based on shared protein content (Table 48 and SI file S5-6). One plasmid with a size of 94 kb (cluster 3) was present in all strains. In order to determine if there was a connection between plasmid distribution and susceptibility to sulfivirus infection , strains were clustered based on the presence and absence of the plasmid groups (Fig. 66). No clear trend could be observed. Perhaps the plasmid clusters were too broad and a more det ailed distinction would reveal correlations. In addition, the heatmap in figure 66 shoul d be recalculated using the phage clustering from the reconstructed phylogenetic tree that con siders recombination, as it showed a better correlation of host range and viral species.

Prophage prediction on the plasmids led to the discovery of seven intact/active prophages (IPP) in the plasmids of cluster 1 (*Sulfitobacter* sp. M55, M83, M300 and M355), cluster 5 (*Sulfitobacter* sp. M72 and M91), and cluster 9 (*Sulfitobacter* sp. M191) (Table 49). The

prophages had sequence lengths ranging from 10.7 kb to 26.1 kb and encoded 13 to 32 proteins. In addition, several questionable prophage sequences were detected, also in other plasmids (SI files S5-7a-c). Three of the seven detected active prophages were 100% identical to each other (IPP1, IPP3, and IPP7), IPP2 and IPP4 shared 84.1% nucleotide-based intergenomic identity (Fig. 67). Another seven prophages were predicted in the chromosomes of some *Sulfitobacter* strains, with sequence lengths of 31.0 - 65.1 kb encoding 38 - 57 proteins. Two prophages were 100% identical to another one, respectively (IPP11 and IPP12, IPP10 and IPP13). Altogether, they were more similar to each other than the prophages in the plasmids, with intergenomic identity higher than 16.9% (Fig. 67). IPP9 and IPP14 clustered together with 74.7% nucleotide-based intergenomic identity, IPP8 was also similar with more than 50%. In addition, IPP5 encoded in the plasmid of strain M191 was somewhat similar to these prophages (≥38.1%). Interestingly, none of the predicted intact prophages showed any sequence similarity to the lytic sulfivirus isolates (Fig. 67).

In figure 66, the chromosomes and plasmids containing active prophages are marked with a phage symbol. The prophages were grouped based on their intergenomic identity and are displayed accordingly in different colors. For different host strains, divergent correlations were observed between the presence of prophages or plasmids and their susceptibility to sulfivirus infection. *Sulfitobacter* strains M271 and M69 carried highly similar prophages on their chromosomes (IPP9 and IPP14) and had plasmids of the same clusters. However, they were only partially infected by the same sulfiviruses. In contrast, strains M55 and M355 carried the exact same two prophages on their chromosome (IPP11 and IPP12) and on their cluster 1 plasmids (IPP1 and IPP7) and had highly similar infection pattern. Also strain 83 carrying the exact same prophage on the cluster 1 plasmid (IPP3) was infected by similar, but less sulfiviruses. In comparison to those three, strain M90 that had the same plasmid clusters but did not carry the prophage, was infected by many more sulfiviruses. This could indicate that the prophage on the cluster 1 plasmid might be somehow involved in the defense of strains M55, M355 and M83 against the respective sulfiviruses. However, strain M63 also had the same plasmid clusters and no prophage on the cluster 1 plasmid, but was susceptible to even less sulfiviruses. This could in turn be connected to the other prophage encoded in its chromosome (IPP13). However, the exact same prophage was present in the chromosome of strain M300 (IPP10), which was again infected by more sulfiviruses than strain M63, although it carried another prophage (IPP6) on one of its plasmids. Finally, it is noteworthy, that the strain infected by the most phages had no intact prophage. Of course, all these correlations are purely speculative. However, their contradictions make it quite clear that the mere presence or

absence of similar plasmids and prophages in the bacterial genomes is obviously not sufficient to explain the different infection patterns. On the one hand side, this analysis should be repeated in a more structured way and include the non-active prophages. This might reveal more profound patterns and correlations. On the other hand, the (intact) prophages might be too different from the lytic sulfivirus isolates to have an influence on their ability to infect. Even though not much is known about the exact mechanisms of superinfection immunity mediated by prophages and a large variety of mechanisms can be assumed, most studies so far have demonstrated this phenomenon between closely related phages (Bondy-Denomy et al. 2016; Mavrich and Hatfull 2019).

A more detailed analysis of the host genomes on gene and protein level is surely necessary. It might provide an explanation for the differential sulfivirus infections, especially if certain anti-phage defense systems can be found in strains infected or not infected by the same sulfiviruses. The (potential) presence of anti-defense systems in the viral genomes should be taken into consideration as well. Finally, the activity of specific defense and anti-defense genes could be investigated in infection experiments accompanied with transcriptomic analysis.

Table 47: Genome characteristics of 30 sulfivirus host strains.

Genome	Sequencing ID	Complete genome length [bp]	No. of ORFs	Chromosome size [bp]	Chromosome G+C content [%]	No. of plas-mids
Sulfitobacter sp. M45	CRMO_n_1	4,131,342	4,036	3,682,501	58.3	3
Sulfitobacter sp. M47	CRMO_n_2	4,046,630	3,884	3,643,977	58.1	3
Sulfitobacter sp. M53	CRMO_n_3	4,099,107	3,945	3,620,052	58.2	3
Sulfitobacter sp. M54	CRMO_n_4	4,185,969	4,043	3,635,973	58.2	4
Sulfitobacter sp. M55	CRMO_n_5	4,128,924	4,023	3,608,282	58.2	3
Sulfitobacter sp. M63	CRMO_n_6	4,179,797	4,084	3,748,437	58.2	3
Sulfitobacter sp. M66	CRMO_n_7	3,995,181	3,876	3,617,848	58.1	4
Sulfitobacter sp. M69	CRMO_n_8	4,049,375	3,949	3,767,683	58.2	2
Sulfitobacter sp. M72	CRMO_n_9	4,307,537	4,172	3,722,005	58.2	6
Sulfitobacter sp. M81	CRMO_n_10	4,410,327	4,293	3,949,596	58.3	4
Sulfitobacter sp. M83	CRMO_n_11	4,167,213	4,036	3,548,707	58.2	4
Sulfitobacter sp. M90	CRMO_n_12	4,008,238	3,894	3,634,983	58.2	3
Sulfitobacter sp. M91	CRMO_n_13	4,037,163	3,928	3,643,305	58.2	4
Sulfitobacter sp. M171	CRMO_n_14	4,154,620	4,002	3,553,525	58.2	6
Sulfitobacter sp. M173	CRMO_n_15	3,845,438	3,721	3,656,377	58.1	2
Sulfitobacter sp. M176	CRMO_n_16	4,082,011	3,959	3,598,080	58.1	4
Sulfitobacter sp. M183	CRMO_n_17	4,096,768	3,990	3,554,444	58.2	5
Sulfitobacter sp. M187	CRMO_n_18	4,149,573	3,988	3,654,650	58.1	4
Sulfitobacter sp. M191	CRMO_n_19	4,305,228	4,158	3,658,606	58.1	5
Sulfitobacter sp. M199	CRMO_n_20	4,252,012	4,115	3,626,458	58.1	4
Sulfitobacter sp. M206	CRMO_n_21	4,091,340	3,951	3,611,016	58.2	4
Sulfitobacter sp. M242	CRMO_n_22	4,059,199	3,895	3,590,381	58.2	3
Sulfitobacter sp. M265	CRMO_n_23	4,204,571	4,114	3,602,912	58.2	5
Sulfitobacter sp. M271	CRMO_n_24	4,026,887	3,901	3,757,359	58.2	2
Sulfitobacter sp. M283	CRMO_n_25	3,893,771	3,742	3,629,652	58.2	3
Sulfitobacter sp. M290	CRMO_n_26	4,152,470	3,997	3,673,615	58.2	4
Sulfitobacter sp. M300	CRMO_n_27	4,235,078	4,090	3,616,378	58.2	5
Sulfitobacter sp. M351	CRMO_n_28	4,053,755	3,888	3,587,296	58.1	3
Sulfitobacter sp. M355	CRMO_n_29	4,141,755	4,036	3,621,113	58.2	3
Sulfitobacter sp. M356	CRMO_n_30	4,078,751	3,930	3,623,626	58.2	3

Table 48: Plasmids of 30 sulfivirus host strains.

Genome	Sequencing ID	Plasmid size [bp] (contig ID CRMO_n_x_...)				
		Plasmid cluster 1	Plasmid cluster 2	Plasmid cluster 3	Plasmid cluster 4	Further clusters
Sulfitobacter sp. M45	CRMO_n_1	259,780 (2)	95,055 (3)	94,006 (4)		
Sulfitobacter sp. M47	CRMO_n_2			94,006 (4)	98,013 (3)	210,634 (2) (cluster 8)
Sulfitobacter sp. M53	CRMO_n_3			94,006 (4)	98,013 (3)	287,036 (2) (cluster 8)
Sulfitobacter sp. M54	CRMO_n_4	262,922 (2)	95,055 (4)	94,006 (5)	98,013 (3)	
Sulfitobacter sp. M55	CRMO_n_5	331,581* (2)	95,055 (3)	94,006 (4)		
Sulfitobacter sp. M63	CRMO_n_6	242,299 (2)	95,055 (3)	94,006 (4)		
Sulfitobacter sp. M66	CRMO_n_7			94,006 (3)		274,503 (2) (cluster 8), 7,624 (4) (cluster 11), 5,342 (5) (cluster 12)
Sulfitobacter sp. M69	CRMO_n_8	187,686 (2)		94,006 (3)		
Sulfitobacter sp. M72	CRMO_n_9	262,922 (2)	95,055 (4)	94,006 (5)	98,013 (3)	12,627* (7) (cluster 5), 22,909 (6) (cluster 16)
Sulfitobacter sp. M81	CRMO_n_10	173,820 (2)	94,892 (4)	94,006 (5)	98,013 (3)	
Sulfitobacter sp. M83	CRMO_n_11	331,432* (2)	95,055 (4)	94,006 (5)	98,013 (3)	
Sulfitobacter sp. M90	CRMO_n_12	184,400 (2)	94,849 (3)	94,006 (4)		
Sulfitobacter sp. M91	CRMO_n_13	192,172 (2)	95,055 (3)	94,006 (4)		12,625* (5) (cluster 5)
Sulfitobacter sp. M171	CRMO_n_14	279,354 (2)	81,358 (5)	94,006 (4)	98,013 (3)	24,674 (6) (cluster 6), 23,690 (7) (cluster 7)
Sulfitobacter sp. M173	CRMO_n_15		95,055 (2)	94,006 (3)		
Sulfitobacter sp. M176	CRMO_n_16	114,433 (3)		94,006 (5)	98,013 (4)	177,478 (2) (cluster 8)
Sulfitobacter sp. M183	CRMO_n_17	279,586 (2)	80,823 (4)	94,006 (3)		12,137 (6) (cluster 5), 75,772 (5) (cluster 9)
Sulfitobacter sp. M187	CRMO_n_18	207,849 (2)	95,055 (4)	94,006 (5)	98,013 (3)	
Sulfitobacter sp. M191	CRMO_n_19	207,849 (2)	95,055 (5)	94,006 (6)	98,013 (4)	151,699* (3) (cluster 9)
Sulfitobacter sp. M199	CRMO_n_20	338,691 (2)	94,844 (4)	94,006 (5)	98,013 (3)	
Sulfitobacter sp. M206	CRMO_n_21	193,456 (2)	94,849 (4)	94,006 (5)	98,013 (3)	
Sulfitobacter sp. M242	CRMO_n_22	276,799 (2)		94,006 (4)	98,013 (3)	

Sulfitobacter sp. M265	CRMO_n_23	267,941 (2)	95,044 (3)	94,006 (4)		73,928 (5) (cluster 10), 70,740 (6) (cluster 10)
Sulfitobacter sp. M271	CRMO_n_24	175,522 (2)		94,006 (3)		
Sulfitobacter sp. M283	CRMO_n_25		72,100 (4)	94,006 (3)	98,013 (2)	
Sulfitobacter sp. M290	CRMO_n_26	208,836 (2)	78,000 (5)	94,006 (4)	98,013 (3)	
Sulfitobacter sp. M300	CRMO_n_27	318,999* (2)	95,055 (4)	94,006 (5)	98,013 (3)	12,627 (6) (cluster 5)
Sulfitobacter sp. M351	CRMO_n_28			94,006 (4)	98,013 (3)	274,440 (2) (cluster 8)
Sulfitobacter sp. M355	CRMO_n_29	331,581* (2)	95,055 (3)	94,006 (4)		
Sulfitobacter sp. M356	CRMO_n_30			94,006 (4)	98,013 (3)	263,106 (2) (cluster 8)

Table 49: Intact prophages predicted in the plasmids and chromosomes of 30 sulfiviruses strains. For the complete results, see SI tables S5-7a-c.

Intact prophage ID	Prophage Candidate	Genome	Plasmid ID	Plasmid cluster	Start	End	Length	Prediction tool	Score	Gene number
IPP1	Canditate_217	*Sulfitobacter* sp. M55	CRMOn5_2	1	32945	47756	14812	Prophage Hunter	0.88	20
IPP2	Candidate_32	*Sulfitobacter* sp. M72	CRMOn9_7	5	1484	12563	11080	Prophage Hunter	0.89	16
IPP3	Candidate_9	*Sulfitobacter* sp. M83	CRMOn11_2	1	32922	47733	14812	Prophage Hunter	0.88	20
IPP4	Candidate_1	*Sulfitobacter* sp. M91	CRMOn13_5	5	1233	11940	10708	Prophage Hunter	0.98	15
IPP5	Candidate_85	*Sulfitobacter* sp. M191	CRMOn19_3	9	107522	133622	26101	Prophage Hunter	0.86	32
IPP6	Candidate_158	*Sulfitobacter* sp. M300	CRMOn27_2	1	130011	142349	12339	Prophage Hunter	0.86	13
IPP7	Candidate_173	*Sulfitobacter* sp. M355	CRMOn29_2	1	32945	47756	14812	Prophage Hunter	0.88	20
IPP8		*Sulfitobacter* sp. M81	CRMOn10_1		1527086	1558073	30987	PHASTER	150	44
IPP9		*Sulfitobacter* sp. M271	CRMOn24_1		1973673	2031566	57893	PHASTER	100	57
IPP10		*Sulfitobacter* sp. M300	CRMOn27_1		1888633	1922128	33495	PHASTER	100	38
IPP11		*Sulfitobacter* sp. M355	CRMOn29_1		418393	469993	51600	PHASTER	120	44
IPP12		*Sulfitobacter* sp. M55	CRMOn5_1		3429471	3481072	51601	PHASTER	120	44
IPP13		*Sulfitobacter* sp. M63	CRMOn6_1		519531	553026	33495	PHASTER	100	38
IPP14		*Sulfitobacter* sp. M69	CRMOn8_1		436949	502080	65131	PHASTER	100	57

Fig. 66: Susceptibility of *Sulfitobacter* strains to sulfivirus infection clustered based on the presence/absence of plasmid clusters. Only plasmid clusters 1 - 10 are displayed. Presence of a plasmid cluster is indicated in blue. Plasmids and chromosomes carrying an active prophage are marked with circles. Identical prophages are indicated by the same color. Positive infection is indicated by green squares. Dendrograms visualize clustering of phages (based on host range) and bacteria (based on plasmid presence). Assignment of phages to species (VIRIDIC) is displayed on top by colors and numbers.

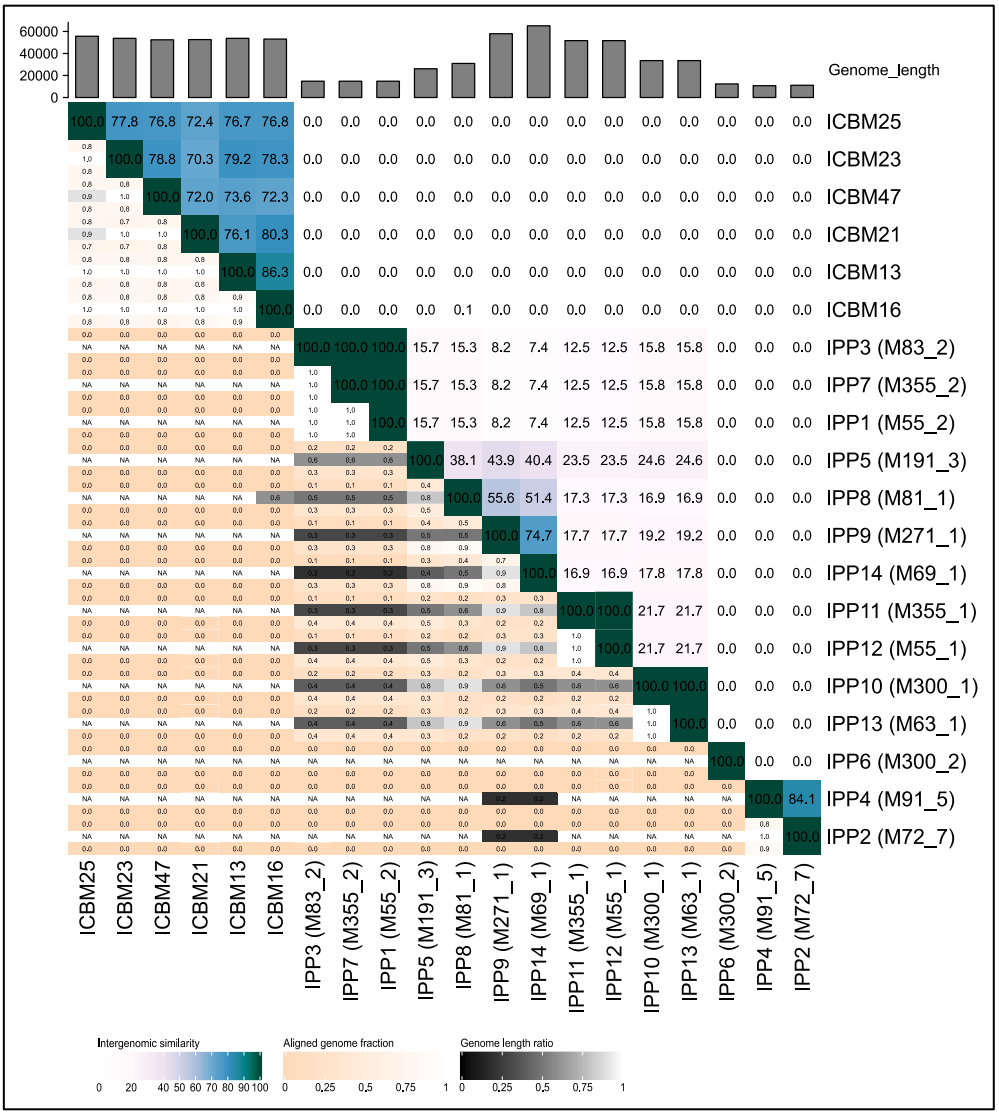


Fig. 67: Nucleotide-based intergenomic identity of the active prophages predicted in the genomes of 30 *Sulfitobacter* strains and representatives of sulfiviruses, calculated with VIRIDIC.

5.3.6. Conclusions

The large-scale roseophage isolation campaign provided us with the collection of 89 closely related dsDNA phages that have a siphoviral shape and occur in the same habitat, the North Sea. They belong to 48 species within one single genus, the “Sulfivirus” genus. By phylogenetic analysis of the 16S rRNA gene and the ITS region, we could show that their isolation hosts are 48 different *Sulfitobacter* strains of a microdiverse species-level population in the North Sea. Testing the host range of the sulfiviruses revealed a complex infection network between these phage and host collections comprising very different sizes of host range on the phage side and different degrees of susceptibility on the side of the bacterial strains. This prompted us to have a closer look at the genomic differences between the phages and hosts.

By analysing the flexible gene content (non-core proteins) in the genomes of both, phages and hosts, by detecting indications of recombination in the phage genomes and by investigating plasmid and prophage occurrences in the host genomes, we gained first impressions on how horizontal gene transfer nurtures diversity in this phage-host community. We illustrated a fascinating microdiversity, which is created on both sides, through the interplay and eternal competition between bacteriophages and their hosts. The next future research task will be to investigate in more detail the genomic dispositions that underlie this arms race by searching for defense systems in the bacterial genomes and counterdefense genes in the viral genomes, eventually detecting new, so far unknown ones.

6. Isolation and preliminary characterization of ICBM5, a ssDNA phage infecting *Sulfitobacter* sp. SH24-1b

6.1. Chapter summary

A new member of the *Microviridae* family, phage ICBM5, was isolated from a North Sea water sample infecting *Sulfitobacter dubius* SH24-1b. The ssDNA phage has a small, icosahedral capsid of 28.68 ± 1.95 nm in diameter and no tail. It shows a narrow host range infecting only two strains of the species *Sulfitobacter dubius*. We used proteins from ICBM5 as query and detected 65 ICBM5-related prophages and episomes in publicly available bacterial genomes. Most bacteria carrying ICBM5-like phages belong to the phyla *Proteobacteria* and *Bacteroidetes*, with many of them being members of the order *Rhodobacterales*. The isolation and preliminary characterization of this new phage is described in the following chapter. It was the starting point of a study in which we investigated in more detail the lifestyle of phage ICBM5 and the taxonomic classification and biogeographical distribution of ICBM5-like phages. The discovery of phage ICBM5 is an example that there is a diversity of ssDNA phages yet to be revealed.

Corresponding / subsequent publication:

Zucker, Falk; Bischoff, Vera; Olo Ndela, Eric; Heyerhoff, Benedikt; Poehlein, Anja; Freese, Heike M.; Roux, Simon; Simon, Meinhard; Enault, Francois; Moraru, Cristina (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). DOI: 10.1093/ve/veac070.

Falk Zucker and I are both first authors of this publication. In this dissertation, only the isolation and preliminary characterization of phage ICBM5 are included, because they mostly correspond to my contribution to this study. I performed the transmission electron microscopy (TEM) and the host range assays for phage ICBM5 and I calculated the 16S tree of the host strains. In addition, I contributed to manuscript writing.

6.2. Materials and Methods

6.2.1. Cultivation media

Marine broth (MB) was used for the liquid cultures and agar plates needed for plaque and spot assays. It had the following recipe: 5.0 g/l peptone, 1.0 g/l yeast extract, 0.1 g/l $C_6H_8FeO_7$, 12.6 g/l $MgCl_2 \times 6H_2O$, 3.24 g/l Na_2SO_4, 19.45 g/l NaCl, 2.38 g/l $CaCl_2 \times 2H_2O$, 0.55 g/l KCl, 0.16 g/l $NaHCO_3$, 0.01 g/l $Na_2HPO_4 \times 2H_2O$, 0.08 g/l KBr, 0.034 g/l $SrCl_2 \times 6H_2O$, 0.022 g/l H_3BO_3, 0.004 g/l $Na_2SiO_3 \times 3H_2O$, 0.0024 g/l NaF, and 0.0016 g/l NH_4NO_3. After autoclavation, the media was completed by addition of 1 ml/l of a multivitamin solution (Balch et al. 1979). Furthermore, artificial saltwater (ASW) base medium was used for plaque assays and purification after cesium chloride gradient centrifugation. It had the following recipe: 24.32 g/l NaCl, 10 g/l $MgCl_2 \times 6H_2O$, 1.5 g/l $CaCl_2 \times 6H_2O$, 0.66 g/l KCl, 4 g/l Na_2SO_4, 2.38 g/l (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (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), and 1 ml NaF (0.07 M).

6.2.2. Isolation of phage ICBM5

Phage ICBM5 was isolated from the coastal North Sea using a phage enrichment procedure, followed by plaque picking and purification. For this purpose, surface seawater was collected in June 2015 from the shoreline near Neuharlingersiel (53°42'09.8"N 7°41'58.9"E) during high tide, transported to the lab on ice, and then filtered through a 0.2-µm filter (Rotilabo-syringe filters, Carl Roth). A phage enrichment was set up by mixing nine parts of freshly filtered seawater with one part of 10x MB and adding an inoculum of exponentially growing *Sulfitobacter dubius* SH24-1b (Hahnke et al. 2013). After overnight incubation at 20 °C and 100 rpm, cells and debris were removed from the enrichment by centrifugation (15 min, 4000 x g, 20 °C) and 0.2-µm filtration of the supernatant. To test for the presence of phages, 20 µl of filtrate were spotted on a lawn of *S. dubius* SH24-1b. The clearing zone was then collected, passed through a 0.2-µm filter to remove cells, and used further in plaque assays, to obtain single plaques. For this purpose, serial dilutions (10^0, 10^{-1}, etc.) were prepared from the phage fractions by mixing with MB medium. Further, 100 µl of phage dilution were mixed with 280 µl of exponentially growing host culture (OD_{600} = 0.2 - 0.3) and incubated for 15 min on ice. The mixture was transferred to 3 ml MB-soft agar (0.6% low melting point Biozym Plaque GeneticPure agarose, Biozym, kept warm at 37 °C), mixed by brief vortexing, and poured onto the bottom MB agar layer (1.8% agar). After drying, the plates were incubated at 20 °C. Phage

plaques were picked and incubated overnight in 500 µl ASW base at 4 °C. After centrifugation (10 min, 10,000 x g, 4 °C), the supernatant was used for the next round of plaque assays. The procedure of plaque assay, picking of plaques, and re-plating was repeated three times to ensure the purity of the newly isolated phages. Finally, one plaque was picked and used to infect a liquid culture of *S. dubius* SH24-1b. After overnight incubation at 20 °C and 100 rpm, the phage lysate was obtained by removing cells and debris by centrifugation (15 min, 4000 x g, and 4 °C) and 0.2-µm filtration. The phage lysate was stored at 4 °C. For long-term storage, two types of glycerol stocks were prepared: (i) stock of free phage particles (one part phage fraction and one part MB media with 50% glycerol) and (ii) stock of infected host cells (one part infected cells – 375 µl phage fraction added to 375 µl host culture, 15 min on ice for absorption – and one part MB media with 50% glycerol).

6.2.3. Host range of phage ICBM5

To determine the host range of ICBM5, ninety-four different strains covering the phylogenetic diversity of *Rhodobacteraceae* (Table 50) were challenged with the purified ICBM5 phage by spot assay. For the spot assay, 280 µl of exponentially growing host culture (OD_{600} = 0.2 - 0.3) were mixed with 3 ml MB-soft agar and poured onto the bottom MB agar layer (1.8% agar). After drying the top layer, 15 µl of phage fraction, obtained from a liquid infection as described above, were spotted in triplicates onto the top layer. For each strain, three plates were prepared and incubated at 15 °C, 20 °C, or 28 °C. For those hosts showing clearing zones, infection by ICBM5 was further confirmed by plaque assays.

6.2.4. Purification of phage ICBM5 via CsCl gradient ultracentrifugation

To generate a high volume of lysate, we prepared sixty doublelayer agar plates with confluent ICBM5 lysis. After plaque formation, 5 ml of Sodium chloride Magnesium sulphate (SM) buffer (100 mM NaCl, 8 mM $MgSO_4$, and 50 mM Tris-HCl pH 7.4) were added to each plate, followed by incubation at 4 °C for 6 h. The phage-containing buffer was then collected and centrifuged for 15 min at 4,000 x g and 4 °C, to remove cells and cell debris. Then, phages were precipitated by adding polyethylene glycol (PEG) (Promega) (final concentration 10%) and NaCl (final concentration 0.6 mM) and incubating at 4 °C for 2 h. After centrifuging for 2 h at 7197 x g and 4 °C, the phage pellet was resuspended in 500 µl SM buffer (100 mM NaCl, 8 mM MgSO4, 50 mM Tris-HCl pH 7.4), followed by 30 min incubation at 4 °C. Further phage concentration and purification were done by cesium chloride (CsCl) gradient ultracentrifugation. A density gradient was set up by layering 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. The PEG concentrated phage fraction was added on top, followed by ultracentrifugation for 4 h at 20 °C and 25,000 rpm (Beckman, SW 41 Ti). Afterwards, the visible band corresponding to the phages was collected with a syringe and needle through the sidewall of the ultracentrifuge tube (~500 µl). Removal of CsCl was done by dialysis in Slide-A-Lyzer G2 Dialysis Cassettes 10K MWCO (Thermo Fisher Scientific) against ASW base, for a total of 21 h, with buffer exchange after 3 h and 18 h. The selected phage fraction was tested for lysis by spot assay.

6.2.5. Transmission electron microscopy of phage ICBM5

To prepare for TEM, 30 µl of CsCl-purified ICBM5 stock were pipetted on top of a carbon-coated grid (Formvar 162, 200 mesh) and phages were allowed to absorb for 3 min. This was followed by staining with 30 µl 2% uranyl acetate for 45 s and gentle removal of the liquid with filter paper. After air-drying for 15 min, the grids were visualized with the transmission electron microscope Zeiss EM902A. Images were documented with the Proscan High-Speed Slow Scan Charge Coupled Device (SSCCD) camera and analyzed using the software ImageSP viewer (Version 1.2.5.16). Negatively stained phages were used for capsid size measurements.

6.2.6. Testing the ssDNA nature of the ICBM5 phage genome

Phage genomic DNA was extracted from a CsCl-concentrated phage stock by mixing with the same amount of phenol:chloroform:isoamyl (Roth) solution and then gently inverting and centrifuging for 15 min, at 12,000 x g and 4 °C. The aqueous phase was then mixed with an equal amount of ice-cold absolute ethanol (Th.Geyer) and the DNA was precipitated at -80 °C for 30 min. The DNA was pelleted by centrifugation (20 min, 12,000 x g and 4 °C) and resuspended in nuclease-free water (Thermo Fisher Scientific). Afterwards, the DNA was purified with the NucAway spin column kit (Thermo Fisher Scientific) and quantified using the Nanodrop 2000 spectrophotometer.

To determine the genomic architecture of ICBM5, the phage DNA was exposed to four different enzymes: S1 nuclease (Thermo Fisher Scientific), TURBO DNase (Thermo Fisher Scientific), Exonuclease VII (New England Biolabs), and Hind III (New England Biolabs). Exonuclease VII and S1 strictly target ssDNA, while TURBO DNase digests both ssDNA and dsDNA. Hind III targets only dsDNA. For each enzyme, a 50-µl reaction was set up, by adding 1 µl of enzyme, 1 µg of extracted phage DNA, corresponding reaction buffers, and water. The four reactions were incubated for 30 min at 37 °C, followed by 10 min at 95 °C, for enzyme inactivation. For visualization of the digestion products, 2 µl of digested DNA were mixed with

5 µl loading buffer (BlueJuice Gel Loading Buffer, Thermo Fisher Scientific) 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). The gel was analyzed with the FAS Digi Gel Documentation System (NIPPON Genetics Europe) and evaluated using the BioDocAnalyze software (Biometra GmbH).

6.2.7. Sequencing of the ICBM5 phage genome via Illumina sequencing

The phage lysate from plates with confluent plaques was first concentrated using 15-ml Amicon ultracentrifugal filter columns (Merck Millipore), then 0.2 µm filtered to remove bacteria and cell debris, and finally purified on an OptiPrep density gradient (Sigma Aldrich). The gradient was set up by layering OptiPrep solutions in a concentration range from 10% to 50%, with an incremental step of 5%. After allowing the gradient to settle for 2 h at room temperature, 1 ml of phage solution was added, followed by ultracentrifugation for 12 h, at 40,000 x g and 20 °C (Beckman, SW 41 Ti). The gradient was divided into 1-ml fractions, which were then tested for the presence of phages by spot assays. The fraction with the highest concentration of ICBM5 was then washed and concentrated using 0.5-ml Amicon columns, during which the OptiPrep was replaced by SM buffer. Extracellular DNA was removed by incubating the phage concentrate with 0.043 units/µl of TURBO DNase (Thermo Fisher Scientific) for 30 min at 37 °C, followed by enzyme inactivation for 10 min at 75 °C with 15 mM ethylenediaminetetraacetic acid (EDTA). Further, the phage DNA was extracted using the ChargeSwitch gDNA Mini Bacteria Kit (Thermo Fisher Scientific), according to the instructions manual, but without using lysozyme in the first step. The ICBM5 ssDNA genome was converted to dsDNA by using the REPLI-g Mini kit (Qiagen), following the manufacturer's instructions. Throughout these procedures, the concentration and quality of the DNA were checked fluorometrically with Qubit 2.0 and the Qubit dsDNA HS Assay, spectrophotometrically with Nanodrop 2000 spectrophotometer, and visually by regular gel electrophoresis (0.7% agarose gel, 50 V, SYBR Gold staining).

An Illumina shotgun library was prepared using the Nextera XT DNA Sample Preparation Kit (Illumina). To assess the quality and size of the library, the samples were run on an Agilent Bioanalyzer 2100 using an Agilent High Sensitivity DNA Kit as recommended by the manufacturer (Agilent Technologies). Library DNA concentration was determined using the Qubit dsDNA HS Assay Kit as recommended by the manufacturer (Life Technologies GmbH). Sequencing was performed on a MiSeq system with the reagent kit v3 with 600 cycles (Illumina) as recommended by the manufacturer, resulting in 785.119 paired-end reads.

6.2.8. Assembly and annotation of the ICBM5 phage genome

The Illumina raw reads were cleaned with BBDuk in two steps. In the first step, the adaptors were removed, using the following parameters for BBDuk: "ktrim=r k=21 mink=8 tbo tpe ftm=5 rcomp=t ordered t=8". In the second step, any contaminating reads (from the host or from phiX174), as well as low-quality ends, were removed, using the following parameters for BBDuk: "k=31 rcomp=t hdist=1 qtrim=rl trimq=20, maq=20 minlen=30 ordered t=8". Afterward, the cleaned reads were assembled with Tadpole (parameters "k=50 t=8"). Both BBDuk and Tadpole are part of the BBTools package (https://jgi.doe.gov/data-and-tools/bbtools/). 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. For further analyses, the genome was linearized and one of the repeats was removed. Open reading frames (ORFs) were predicted using the MetaGeneAnnotator (Noguchi, Taniguchi, and Itoh 2008) implemented in VirClust (Moraru 2023). A first ORF annotation was done by using Domain Enhanced Lookup Time Accelerated Basic Local Alignment Search Tool (DELTA-BLAST) to search for homologous proteins in the nonredundant (NR) database (http://ncbi.nlm.nih.gov/). The ICBM5 phage genome is available in the NCBI GenBank database under the following accession number: OM782324. The sequences of the complete genome and the encoded proteins can also be found in the appendix.

6.2.9. Detection of ICBM5-like regions in bacterial genomes

Proteins from phage ICBM5 were used to query the NR database from NCBI, using DELTA-BLAST, with two iterations. Proteins detected as similar were downloaded in GenBank format, imported into Geneious v 9.1.5 (http://www.geneious.com, Kearse et al. (2012)), and identified as part of a viral or bacterial genome based on their organism name and 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.

6.2.10. Phylogenetic analysis of all host 16S rRNA genes and species assignment for *Sulfitobacter dubius* SH24-1b

A neighbor-joining tree of the 16S ribosomal RNA (rRNA) gene sequences from all phage hosts for which we could find the 16S rRNA gene (Table 51) in this study was constructed with the ARB software package (Ludwig et al. 2004). Tree calculation was performed using the reference dataset SSU Ref NR 111, with Jukes–Cantor correction, termini filter, and 1,000 bootstrap replicates. Members of the genus *Acidobacterium* served as an outgroup. For species

assignment of *S. dubius* SH24-1b, the average nucleotide identity (ANI) value between SH24-1b and the *S. dubius* type strain DSM 16472T was calculated with FastANI (Jain et al. 2018) and the digital DNA–DNA hybridization (dDDH) value was calculated with the genome-to-genome distance calculator (GGDC) (applying Formula 2) (Meier-Kolthoff et al. 2022).

6.3. Results and discussion

Phage ICBM5 was isolated from surface seawater, which was collected directly from the coast of the North Sea (53.702722 N, 7.699695 E) in June 2015. It infected *Sulfitobacter dubius* SH24-1b, which had been isolated from a seawater sample taken during a phytoplankton bloom on 12 May 2007 in the southern North Sea (54.7 N, 6.8 E) (Hahnke et al. 2013). Comparison of its 16S rRNA gene sequence revealed a 99.8% sequence identity with *Sulfitobacter dubius* type strain DSM 16472T. Showing a dDDH value of 70% and an ANI value of 96.9% with the type strain, strain SH24-1b could be assigned to the species *S. dubius*. The host range of phage ICBM5 was determined testing almost 100 bacterial strains of the *Rhodobacteraceae* family. Positive infection was only detected on the original host *S. dubius* SH24-1b and *S. dubius* DSM 16472T, revealing a narrow host range of phage ICBM5 (Table 50). Observation of uranyl-acetate-stained samples of phage ICBM5 with TEM revealed a morphology with an icosahedral capsid of 28.68 ±1.95 nm in diameter and no tail (100 phages measured and three measurements per phage) (Fig. 68a). The infection of *S. dubius* SH24-1b produced turbid plaques. Enzymatic digestion of the viral DNA showed that ICBM5 has an ssDNA genome (Fig. 68b). The sequenced genome was 5,581 bp in size and circularly closed, harboring six protein-coding genes (Fig. 68c). Out of these genes, four could be functionally annotated as a pilot protein, the major capsid protein (MCP), a lysis protein, and a replication initiation protein (Rep). Only the replication protein could be annotated using BLASTp and was highly similar to proteins from previously known *Microviridae*. The remaining three proteins were more distanly related to *Microviridae* proteins, as the annotation was only possible using DELTA-BLAST, a remote homology tool. Phage ICBM5 can be assigned as a new member of the *Microviridae* family, as it has the corresponding genome characteristics, virion morphology and core genes. Nevertheless, as the major capsid protein, which is highly consevered among known microviruses and thus used for phylogenetic analysis, was only found with a remote homology tool, phage ICBM5 seemed to be rather distantly related to known *Microviridae*.

To investigate how widespread ICBM5-like phages are, we used ICBM5 proteins to search for potential prophages in prokaryotic genomes from the NCBI NR database. Most of

the 72 detected ICBM5-like genomic regions were located on bacterial chromosomes. However, some were found on large plasmids and some were present as small separate contigs, which could be episomes and indicate a carrier-state lifestyle. The majority of the ICBM5-like genomic regions occurred in bacteria from *Alphaproteobacteria* (53.5%), *Bacteroidia* (29.5%) and *Gammaproteobacteria* (5.6%). Further host genomes belonged to the classes *Bacilli, Clostridia, Erysipelotrichia, Negativicutes, Cyanophycea*, and *Flavobacteriia* (Fig. 69). Among the *Alphaproteobacteria*, the orders *Hyphomicrobiales* and *Rhodobacterales* were represented, the latter comprising *Sulfitobacter dubius* SH24-1b and other members of the *Rhodobacteraceae* family. From the detected *Microviridae*-like regions, seven had been described earlier (Krupovic and Forterre 2011; Quaiser et al. 2015; Zheng et al. 2018). The strains with potential prophages originated from diverse habitats, ranging from marine and freshwater samples to animal-, plant- and human- associated samples (Table 51).

The presence of ICBM5-like prophages in so many different host taxa from such a variety of habitats indicated a diversity and importance of these microviruses that provided the impetus for a more comprehensive taxonomic analysis of ICBM5-related phages. Furthermore, the turbid plaques of phage ICBM5 on *S. dubius* SH24-1b suggested a lysogenic potential, which was supposed to be investigated more closely by means of one-step growth curves and direct-geneFISH (fluorescence-in-vito-hybridisation). The results of these investigations and the delineating of two new *Microviridae*-subfamilies can be read in the corresponding publication.

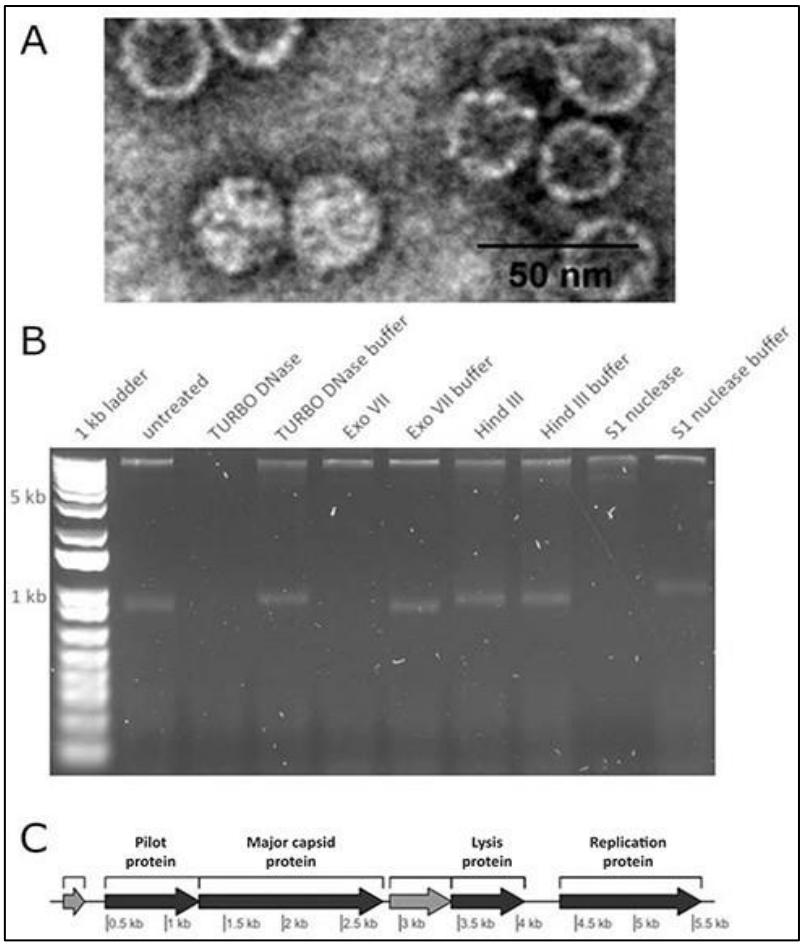


Fig. 68: (A) ICBM5 morphology determined by TEM of uranyl-acetate-stained virions. **(B)** Agarose gel shows enzymatic digestion of ICBM5 ssDNA phage. The DNA was digested by TURBO DNase, Exo VII, and S1 nuclease, but was not affected by treating it via restriction enzyme Hind III, which only targets dsDNA, or the exclusion of nucleases (usage of only buffer). The 1 kb 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 a circular, ssDNA molecule. **(C)** Genome map of ICBM5. In dark gray–identified proteins, with labels on top of each gene. In light gray–hypothetical proteins.

Table 50: List of *Rhodobacteraceae* strains used for the host range assay. Strains infected by phage ICBM5 are written in bold.

Name	Strain designation	Strain	Infected (- no, + yes)
Aliiroseovarius crassostreae	CV919-312, CVSP	DSM 16950T	-
Aliiroseovarius halocynthiae	MA1-10	DSM 27840T	-
Antarctobacter heliothermus	EL-219	DSM 11445T	-
Celeribacter baekdonensis	L-6	DSM 27375T	-
Celeribacter halophila	ZXM137	DSM 26270T	-
Celeribacter indicus	P73	DSM 27257T	-
Celeribacter marinus	IMCC12053	DSM 100036T	-
Celeribacter neptunius	H 14	DSM 26471T	-
Cognatishimia maritimus	GSW-M6	DSM 28223T	-
Cognatiyoonia koreensis	GA2-M3	DSM 17925T	-
Dinoroseobacter shibae	DFL 12	DSM 16493T	-
Hwanghaeicola aestuarii	Y26	DSM 22009T	-
Jannaschia donghaensis	DSW-17	DSM 102233T	-
Jannaschia helgolandensis	Hel10	DSM 14858T	-
Jannaschia pohangensis	H1-M8	DSM 19073T	-
Jannaschia rubra	4SM3	DSM 16279T	-
Leisingera aquimarina	R-26159	DSM 24565T	-
Leisingera caerulea	13	DSM 24564T	-
Leisingera daeponensis	TF-218	DSM 23529T	-
Leisingera methylohalidivorans	MB2	DSM 14336T	-
Limimaricola cinnabarinus	LL-001	DSM 29954T	-
Limimaricola hongkongensis	UST950701-009P	DSM 17492T	-
Limimaricola pyoseonensis	JJM85	DSM 21424T	-
Litoreibacter albidus	KMM 3851	DSM 26922T	-
Litoreibacter arenae	GA2-M15	DSM 19593T	-
Litoreibacter janthinus	KMM 3842	DSM 26921T	-
Loktanella fryxellensis	R-7670	DSM 16213T	-
Loktanella salsilacus	R-8904	DSM 16199T	-
Maribius pelagius	B5-6	DSM 26893T	-
Maribius salinus	CL-SP27	DSM 26892T	-
Marinovum algicola	FF3	DSM 10251T	-
Marinovum algicola	DG898	DSM 27768	-
Maritimibacter alkaliphilus	HTCC2654	DSM 100037T	-
Oceanicola granulosus	HTCC2516	DSM 15982T	-
Octadecabacter temperatus	SB1	DSM 26878T	-
Pacificibacter marinus	HDW-9	DSM 25228T	-
Palleronia marisminoris	B33	DSM 26347T	-
Phaeobacter gallaeciensis	BS 107	DSM 26640T	-
Phaeobacter inhibens		DSM 17395	-
Phaeobacter inhibens	T5	DSM 16374T	-
Phaeobacter inhibens	2.10	DSM 24588	-
Phaeobacter italicus	R11	DSM 26436T	-
Ponticoccus litoralis	CL-GR66	DSM 18986T	-
Pseudooceanicola batsensis	HTCC2597	DSM 15984T	-
Pseudooceanicola nanhaiensis	SS011B1-20	DSM 18065T	-
Pseudophaeobacter arcticus	20188	DSM 23566T	-
Pseudoruegeria lutimaris	HD-43	DSM 25294T	-
Roseibacterium elongatum	Och 323	DSM 19469T	-
Roseivivax isoporae	sw2	DSM 22223T	-
Roseobacter denitrificans	Och 114	DSM 7001T	-
Roseobacter litoralis	Och 149	DSM 6996T	-
Roseovarius indicus	B108	DSM 26383T	-
Roseovarius lutimaris	112	DSM 28463T	-
Roseovarius mucosus	DFL-24	DSM 17069T	-
Roseovarius nubinhibens	ISM	DSM 15170T	-
Ruegeria atlantica	1480	DSM 5823T	-
Ruegeria conchae	TW15	DSM 29317T	-
Ruegeria marina	ZH17	DSM 24837T	-
Ruegeria pomeroyi	DSS-3	DSM 15171T	-
Sagittula stellata	EE-37	DSM 11524T	-
Salinihabitans flavidus	ISL-46	DSM 27842T	-
Salipiger bermudensis	HTCC2601	DSM 26914T	-
Salipiger aestuarii	AD8	DSM 22011T	-
Salipiger marinus	CK-I3-6	DSM 26424T	-

Salipiger mucosus	A3	DSM 16094T	-
Salipiger pacificus	DX5-10	DSM 26894T	-
Sedimentitalea nanhaiensis	NH52F	DSM 24252T	-
Sediminimonas qiaohouensis	YIM B024	DSM 21189T	-
Shimia aestuarii	JC2049	DSM 15283T	-
Shimia haliotis	WM35	DSM 28453T	-
Shimia marina	CL-TA03	DSM 26895T	-
Sulfitobacter delicatus	KMM 3584	DSM 16477T	-
Sulfitobacter dubius	**KMM 3554**	**DSM 16472T**	+
Sulfitobacter indolifex	HEL-45	DSM 14862T	-
Sulfitobacter litoralis	Iso 3	DSM 17584T	-
Sulfitobacter marinus	SW-265	DSM 23422T	-
Sulfitobacter mediterraneus	CH-B427	DSM 12244T	-
Sulfitobacter noctilucae	NB-68	DSM 100978T	-
Sulfitobacter noctilucicola	NB-77	DSM 101015T	-
Sulfitobacter pseudonitzschiae	H3	DSM 26824T	-
Sulfitobacter sp.	EE-36	DSM 11700	-
Sulfitobacter dubius	**SH24-1b**		+
Thalassobius taeanensis	G4	DSM 22007T	-
Thalassococcus halodurans	UST050418-052	DSM 26915T	-
Thioclava dalianensis	DLFJ1-1	DSM 29618T	-
Thioclava pacifica	TL 2	DSM 10166T	-
Tranquillimonas alkanivorans	A34	DSM 19547T	-
Tranquillimonas rosea	BH87090	DSM 23042T	-
Tritonibacter multivorans	MD5	DSM 26470T	-
Tropicibacter naphthalenivorans	C02	DSM 19561T	-
Tropicimonas isoalkanivorans	B51	DSM 19548T	-
Wenxinia marina	HY34	DSM 24838T	-
Yoonia tamlensis	SSW-35	DSM 26879T	-
Yoonia vestfoldensis	R-9477	DSM 16212T	-

Table 51: Bacterial strains containing predicted ICBM5-like regions. *no 16S rRNA gene sequence available. ND = not determined.

Lysogenic strain / Host	Host accession	DNA type	Prophage size (bps)	Contig size (bps)	True prophage	Potential episome	Not microviridae	Sequencing contaminants	Host lineage	Host habitat / isolation site	Reference host habitat
Acinetobacter lwoffii strain SU1904	NZ_JACWEU010000076	contig	5.673	5.673		+			*Bacteria/ Proteobacteria/ Gammaproteobacteria/ Pseudomonadales/ Moraxellaceae/ Acinetobacter/ Acinetobacter calcoaceticus/baumannii* complex	clinical isolate	Hayashi et al. (2021)
Agrobacterium larrymoorei strain CFBP5477	NZ_SWKE01000015	ND	6.301	657.589	+				*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Agrobacterium*	tumours growing on pruned branches of Ficus benjamina	Bouzar and Jones (2001)
Agrobacterium tumefaciens strain 1D1108*	NZ_CP032923	plasmid	5.86	502.274	+				*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Agrobacterium*	rhizosphere	Conn (1942)
Alistipes onderdonkii WAL 8169 = DSM 19147	NZ_KB894552	ND	5.793	100.104	+				*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Rikenellaceae/ Alistipes*	human specimens	Song et al. (2006)
Alistipes sp. isolate P1-1	SCPE01000001	ND	5.699	3.006.009	+				*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Rikenellaceae/ Alistipes*	human specimens	Song et al. (2006)
Aphanizomenon flos aquae WA102 3645	LJOW01000239	contig	4.055			+			*Bacteria/ Cyanobacteria/ Nostocales/ Aphanizomenonaceae/ Aphanizomenon/*	freshwater	NCBI entry
Bacillus altitudinis strain DSM 26896	NZ_JXAI01000008	whole genome shotgun sequence	13.312	156.97			+		*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Bacillaceae/ Bacillus*	gut of marine fish	Esakkiraj et al. (2012)
Bacillus sp. X1 strain DE0237 NODE_39	NZ_VECT01000039	whole genome shotgun sequence	7.309	33.618			+		*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Bacillaceae/ Bacillus*	environmental sample	Zhang et al. (2020a)
Bacillus thuringiensis serovar thuringiensis	NZ_CM000748	circular chromosome	11.378	6.323.123			+		*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Bacillaceae/ Bacillus/ Bacillus cereus group*	moth larvae, animal tissue	Ibrahim et al. (2010)
Bacillus wiedmannii strain FSL J3-0113	NZ_LXFN01000004	whole genome shotgun sequence	9.571	109.662			+		*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Bacillaceae/ Bacillus/ Bacillus cereus group*	raw milk	NCBI entry
Bacteroides caccae strain 2789STDY5834946	NZ_CZBL01000008	ND	5.571	221.42	+				*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	NCBI entry
Bacteroides caccae strain AF46-5GN	QRNA01000095	ND	5.37	19.734	+				*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	NCBI entry
Bacteroides eggerthii DSM 20697*	NZ_DS995510	ND	6.195	19.345	+				*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	NCBI entry

Bacteroides finegoldii strain D54t1_190329_G10	NZ_JADNKZ010000005	whole genome shotgun sequence	6.471	317.103	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	NCBI entry
Bacteroides ovatus isolate *Bacteroides ovatus* MC1	NZ_CAAKNR010000141	ND	6.402	59.93	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	Eggerth and Gagnon (1933)
Bacteroides plebeius DSM 17135, New name: *Phocaeicola plebeius**	NZ_ABQC02000012	whole genome shotgun sequence	5.059	569.841	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	NCBI entry
Bacteroides sp. 2_2_4 supercont1.3*	NZ_EQ973357	ND	6.303	605.506	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	NCBI entry
Bacteroides sp. AF29-11	NZ_QTLY01000002	ND	6.481	622.059	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	NCBI entry
Bacteroides thetaiotaomicron strain 19_BTHE 1034_17932_216065*	NZ_JVQR01000241	ND	6.056	17.932	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human sample	NCBI entry
Bacteroides xylanisolvens strain AF38-2	NZ_QROO01000017	ND	8.224	141.458	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Bacteroidaceae/ Bacteroides*	human feces	Chassard et al. (2008)
Candidatus Rhodobacter lobularis isolate IGS	NZ_LFTY01000002	ND	7.74	3.995.303	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Rhodobacter*	marine, sponge-associated	Jourda et al. (2015)
Clostridioides difficile CD129	NZ_AVHP01000907	contig	5.697	5.697		+	*Bacteria/ Firmicutes/ Clostridia/ Eubacteriales/ Peptostreptococcaceae/ Clostridioides*	human feces	Lawson et al. (2016)
Coprobacter fastidiosus isolate CIM:MAG 335 contig_33578,	QALP01000027	whole genome shotgun sequence	6.609	20.19	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Barnesiellaceae/ Coprobacter*	human feces	Shkoporov et al. (2013)
Devosia chinhatensis strain IPL18	NZ_JZEY01000054	ND	7.499	2.139.066	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Hyphomicrobiaceae/ Devosia*	terrestrial, soil (oil contaminated)	Hassan et al. (2015)
Devosia sp. YR412	NZ_FOFL01000005	ND	11.405	221.727	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Hyphomicrobiaceae/ Devosia.*	terrestrial, root associated	JGI entry
Dysgonomonas macrotermitis strain DSM 27370	NZ_FQUC01000005	ND	6.139	265.55	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Dysgonamonadaceae/ Dysgonomonas*	hindgut of a fungus-growing termite	Yang et al. (2014)
Dysgonomonas sp. 521	NZ_QVMO01000010	ND	6.816	73.398	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Dysgonamonadaceae/ Dysgonomonas*	alimentary canal	NCBI entry
Elizabethkingia anophelis strain E6809 2b	NZ_MAHS01000003	ND	4.392	1.150.786	+		*Bacteria/ Bacteroidetes/ Flavobacteriia/ Flavobacteriales/ Flavobacteriaceae/ Elizabethkingia*	human blood	Nicholson et al. (2016)
Epibacterium ulvae strain U95	NZ_PHJF01000001	ND	7.87	804.761	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Epibacterium*	marine, isolated from macro algae	Breider et al. (2019)
Erysipelatoclostridium sp. An15	NZ_NFLA01000001	ND	6.491	201.071	+		*Bacteria/ Firmicutes/ Erysipelotrichia/ Erysipelotrichales/ Erysipelotrichaceae/ Erysipelatoclostridium*	animal tissue (Rooster)	NCBI entry

Escherichia coli strain OLC1558*	NZ_NWRV01000002	whole genome shotgun sequence	4.477	351.548	+			*Bacteria/ Proteobacteria/ Gammaproteobacteria/ Enterobacterales/ Enterobacteriaceae/ Escherichia*	unknown	NCBI entry
Escherichia sp. MOD1-EC6163	NZ_PTSA01000028	ND	5.993	50.742	+			*Bacteria/ Proteobacteria/ Gammaproteobacteria/ Enterobacterales/ Enterobacteriaceae/ Escherichia*	animal associated (duck feces)	NCBI entry
Gramella jeungdoensis strain KCTC 23123	NZ_SNQI01000008	contig	6.302	6.302		+		*Bacteria/ Bacteroidetes/ Flavobacteriia/ Flavobacteriales/ Flavobacteriaceae/ Gramella*	solar saltern	Joung et al. (2011)
Kosakonia cowanii strain Esp_Z chromosome	NZ_CP022690	circular chromosome	5.002	5.077.975	+			*Bacteria/ Proteobacteria/ Gammaproteobacteria/ Enterobacterales/ Enterobacteriaceae/ Kosakonia*	human specimens	Inoue et al. (2000)
Labrys sp. WJW	NZ_LYXZ01000012	ND	9.368	94.779	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Xanthobacteraceae/ Labrys*	soil	NCBI entry
Listeria monocytogenes strain 2014L-6088	NZ_JABXNI010000003	whole genome shotgun sequence	6.184	593.093			+	*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Listeriaceae/ Listeria*	animal associated (rabbit)	ATCC
Listeria monocytogenes strain C5	NZ_MDQI01000007	whole genome shotgun sequence	10.033	1.240.180	+		+	*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Listeriaceae/ Listeria*	freshwater	NCBI entry
Mammaliicoccus sciuri strain GDM7P051A C-1_NDMS04239_1	NZ_WIVK01000025	contig	5.855	5.855		+		*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Staphylococcaceae/ Mammaliicoccus*	animal associated (squirrel)	Kloos et al. (1976)
Maritimibacter sp. LZ-17 NODE_6	NZ_SWKQ01000016	whole genome shotgun sequence	8.679	114.54	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Maritimibacter*	plant associated (dinoflagellate)	NCBI entry
Mesorhizobium composti strain CC-YTH430	NZ_SSNY01000001	ND	5.831	699.301	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	from a compost sample in Taiwan	Lin et al. (2019)
Mesorhizobium sp. 3P27G6	NZ_SCNN01000030	ND	6.025	67.438	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	alpine spring water	NCBI entry
Mesorhizobium sp. isolate N.Cr.TU.016.05.1 NODE_86	SASU01000086	ND	6.165	23.594	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	plant root nodule (*Cicer arietinum*)	NCBI entry
Mesorhizobium sp. M1A.F.Ca.ET.072.01.1.1 NODE_40	NZ_RZSF01000040	ND	8.388	32.967	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	plant root nodule (*Cicer arietinum*)	NCBI entry
Mesorhizobium sp. M1A.F.Ca.IN.020.30.1.1 NODE_39	NZ_RZSO01000039	ND	7.714	17.913	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	plant root nodule (*Cicer arietinum*)	NCBI entry
Mesorhizobium sp. M1A.F.Ca.IN.022.06.1.1*	CP034455	circular chromosome	7.165	6.278.406	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	plant root nodule (*Cicer arietinum*)	NCBI entry
Mesorhizobium sp. M5C.F.Ca.IN.020.29.1.1 NODE_361	NZ_RZSW01000361	contig	6.491	6.491		+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	plant root nodule (*Cicer arietinum*)	NCBI entry
Mesorhizobium sp. M7A.F.Ca.ET.027.02.1.1 NODE_4	NZ_RZOL01000004	ND	6.94	56.052	+			*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Phyllobacteriaceae/ Mesorhizobium*	plant root nodule (*Cicer arietinum*)	NCBI entry

Neorhizobium sp. T17_20	NZ_PVBG01000001	ND	6.427	571.6	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Neorhizobium*	terrestrial, soil	NCBI entry
Nioella nitratireducens strain SSW136	NZ_MNBW01000014	ND	5.734	95.12	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Nioella*	surface sediment of the Jiulong River transered into estuary water	Liu et al. (2017)
Nitrospira sp. isolate RSF5	SWDO01000020	whole genome shotgun sequence	6.25	96.14	+		*Bacteria/ Nitrospirae/ Nitrospirales/ Nitrospiraceae/ Nitrospira/*	rapid sand filter	NCBI entry
Novosphingobium tardaugens NBRC 16725, New name: *Caenibius tardaugens*	NZ_BASZ01000013	ND	4.443	102.866	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Hyphomicrobiaceae/ Novosphingobium*	marine	Zheng et al. (2018)
Oceanicola sp. S124	NZ_AFPM01000156	ND	9115	38.215	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Oceanicola.*	marine	Kwon et al. (2012)
Oscillibacter sp. PC13	NZ_FOXE01000003	ND	5.417	177.285	+		*Bacteria/ Firmicutes/ Clostridia/ Clostridiales/ Oscillospiraceae/ Oscillibacter*	isolated from sheep rumen	Seshadri et al. (2018)
Paenibacillus odorifer strain DSM 15391*	NZ_CP009428	circular chromosome	8.259	6.812.473	+		*Bacteria/ Firmicutes/ Bacilli/ Bacillales/ Paenibacillaceae/ Paenibacillus*	soil from wheat roots	Berge et al. (2002)
Parabacteroides distasonis str. 3999B T(B) 6*	NZ_JNHQ01000100	ND	4.626	85.604	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Tannerellaceae/ Parabacteroides*	human feces	Sakamoto and Benno (2006)
Parabacteroides sp. AF39-10AC*	NZ_QTLP01000027	ND	4.583	76.059	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Tannerellaceae/ Parabacteroides*	human feces	NCBI entry
Parabacteroides sp. AM27-42	NZ_QTNL01000010	ND	4.983	114.974	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Tannerellaceae/ Parabacteroides*	human feces	NCBI entry
Paracoccus alkenifer strain DSM 11593	NZ_FNXG01000002	ND	8.198	871.604	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Paracoccus.*	biofilter for waste gas treatment	Lipski et al. (1998)
Phaeobacter inhibens strain DOK1-1	NZ_CP019307	circular chromosome	5.922	3.718.082	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Phaeobacter.*	marine, seawater	NCBI entry
Phascolarctobacterium faecium DSM 14760	NZ_QLTS01000001	ND	4.517	796.791	+		*Bacteria/ Firmicutes/ Negativicutes/ Acidaminococcales/ Acidaminococcaceae/ Phascolarctobacterium*	faeces of koala	Del Dot et al. (1993)
Prevotella bergensis DSM 17361*	NZ_GG704781	whole genome shotgun sequence	5.337	1.108.685	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Prevotellaceae/ Prevotella*	human skin	Downes et al. (2006)
Prevotella buccalis ATCC 35310	NZ_ADEG01000016	ND	4.975	88.363	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Prevotellaceae/ Prevotella*	human vaginal cavity	NCBI entry
Prevotella salivae F0493 contig0002	NZ_AWGW01000011	contig	6.888	6.888		+	*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Prevotellaceae/ Prevotella*	human oral cavity	Sakamoto et al. (2004)
Prevotella sp. CAG:1185*	HF992559	whole genome shotgun sequence	6.349	34.922	+		*Bacteria/ Bacteroidetes/ Bacteroidia/ Bacteroidales/ Prevotellaceae/ Prevotella*	human gut	NCBI entry

Ralstonia solanacearum strain UW179	NZ_CDLZ01000001	whole genome shotgun sequence	6.503	5.426.414		+	*Bacteria/ Proteobacteria/ Betaproteobacteria/ Burkholderiales/ Burkholderiaceae/ Ralstonia*	plant associated	DSMZ entry
Rhizobium pusense strain CCGM10	NZ_KV878023	ND	8.944	486.269	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	terrestrial, root nodules, (*Phaseolus vulgaris*)	Aguilar et al. (2016)
Rhizobium sp. BG4 plasmid pRPC1	CP044126	circular plasmid	6.609	1.609.137	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	root nodules (*Prosopis cineraria*)	NCBI entry
Rhizobium sp. MHM7A NODE_12*	NZ_VCHT01000012	whole genome shotgun sequence	6.733	221.567	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	pea root nodules (*Pisum sativum*)	NCBI entry
Rhizobium sp. Root564	NZ_LMGN01000011	ND	10.526	1.029.951	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	terrestrial, root nodules, (*Arabidopsis thaliana*)	Bai et al. (2015)
Rhizobium sp. Root651	NZ_LMHB01000005	circular plasmid	8.427	167.271	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	terrestrial, root nodules, (*Arabidopsis thaliana*)	Bai et al. (2015)
Rhizobium sp. UR51a	NZ_JYFU01000029	whole genome shotgun sequence	5.942	277.936	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	terrestrial, root nodules (*Oryza sativa*)	Souza et al. (2015)
Rhizobium sp. YS-1r	NZ_JPYQ01000001	ND	8.57	657.089	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	terrestrial, decaying wood from thermal pond	Jackson et al. (2017)
Rhizobium straminoryzae strain SM12	NZ_VJMG01000008	whole genome shotgun sequence	7.361	175.733	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Hyphomicrobiales/ Rhizobiaceae/ Rhizobium/Agrobacterium group/Rhizobium*	plant associated Rhizosphere (*Oryza sativa*)	NCBI entry
Rhodobacter capsulatus R121	NZ_AYQC01000019	ND	6.611	418.689	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Rhodobacter.*	pond water	Ding et al. (2014)
Rhodobacter capsulatus YW2	NZ_AYPZ01000002	ND	6.427	316.953	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Rhodobacter.*	terrestrial, forest	Weaver et al. (1975)
Rhodobacter sp. SW2	NZ_ACYY01000005	ND	9.745	188.067	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Rhodobacter.*	terrestrial, freshwater mud samples	Ehrenreich and Widdel (1994)
Rhodovulum sp. MB263	NZ_CP020384	circular chromosome	6.583	3.860.570	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Rhodovulum.*	marine, coastal mud flat	Hiraishi and Ueda (1995)
Rhodovulum sulfidophilum strain DSM 2351	NZ_AP014801	circular plasmid	6.209	111.306	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Rhodovulum.*	marine, mud from intertidal flats	Hansen and Veldkamp (1973)
Roseovarius sp. GCL-8	NZ_QITQ01000003	ND	6.624	282.961	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Roseovarius*	sand	NCBI entry

Ruegeria mobilis S1942, New name: *Tritonibacter mobilis*	NZ_JXYG01000006	ND	5.802	603.96	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Ruegeria.*	marine, seawater, sediment and copepod	(Sonnenschein et al. 2017)
Ruegeria sp. THAF57	NZ_CAIWNQ010000009	whole genome shotgun sequence	13.152	210.182	+		*Bacteria/ Proteobacteria/ Alphaproteobacteria/ Rhodobacterales/ Rhodobacteraceae/ Ruegeria.*	marine	NCBI entry
Stenotrophomonas maltophilia strain AS012690	NZ_VLGL01000002	whole genome shotgun sequence	11.045	130.645		+	*Bacteria/ Proteobacteria/ Gammaproteobacteria/ Xanthomonadales/ Xanthomonadaceae/ Stenotrophomonas/ Stenotrophomonas maltophilia group*	human specimens	NCBI entry
Stenotrophomonas rhizophila strain JC1	CP050062	circular chromosome	6.961	4.268.161		+	*Bacteria/ Proteobacteria/ Gammaproteobacteria/ Xanthomonadales/ Xanthomonadaceae/ Stenotrophomonas*	rhizosphere of potato and rape	Wolf et al. (2002)
Xanthomonas citri strain LE3-1	NZ_LN647176	whole genome shotgun sequence	6.748	5.045.425		+	*Bacteria/ Proteobacteria/ Gammaproteobacteria/ Xanthomonadales/ Xanthomonadaceae/ Xanthomonas*	plant associated (*C. aurantifolia*)	NCBI entry

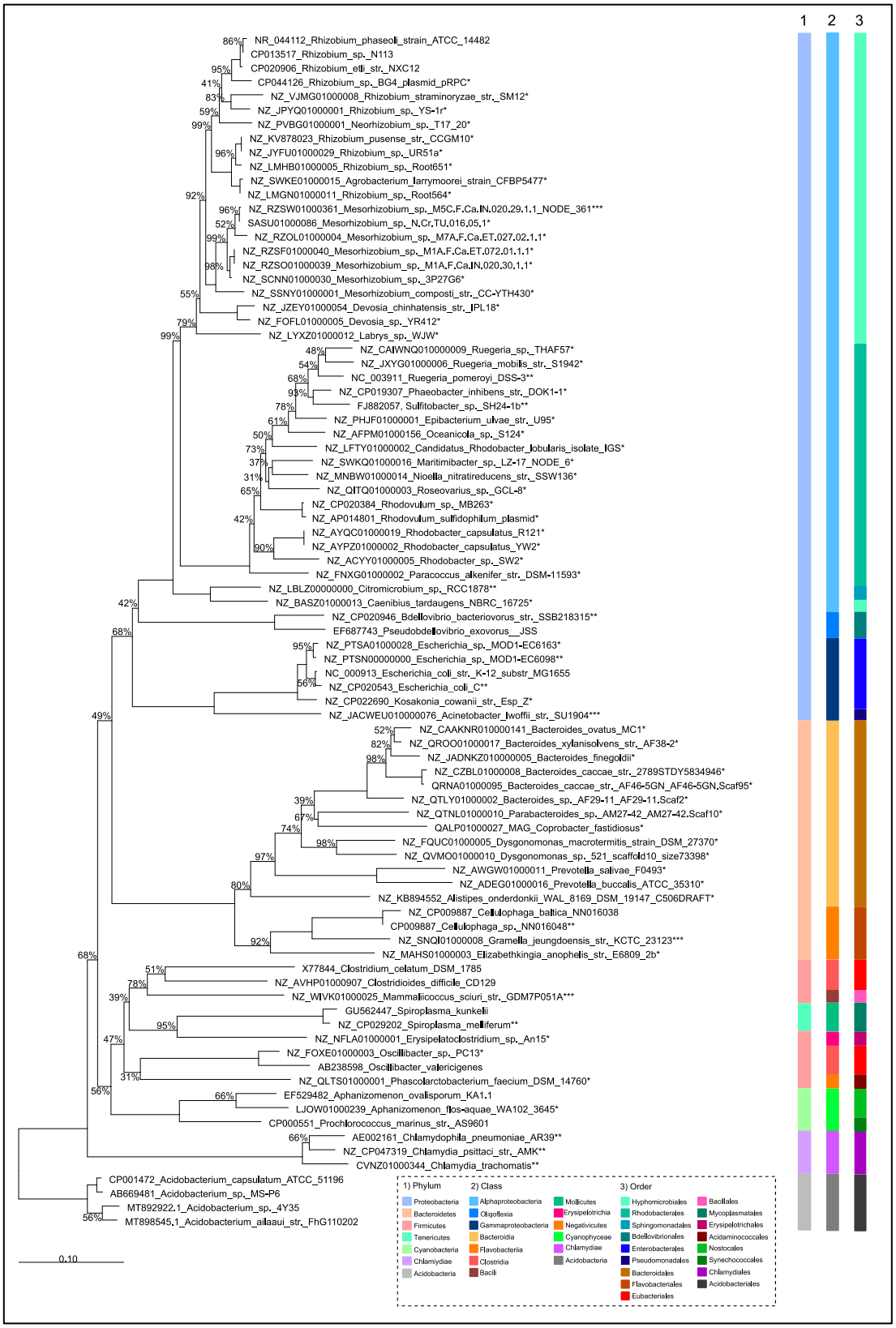


Fig. 69: 16S rRNA phylogenetic analysis of the (pro)-phage hosts. Neighbor-joining tree based on the 16S rRNA gene sequence similarity showing the position of *S. dubius* SH24-1b and other bacterial hosts for *Microviridae*-like (pro)-phages. Bootstrap values are derived from 1,000 replicates. GenBank accession numbers are given as prefixes, followed by species and strain names. The bar represents ten 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.

7. General discussion and outlook

More and more studies are uncovering the importance of phages and phage-host interactions for marine ecosystems. At the same time, a tremendous phage diversity is revealed by metagenomic studies (Mizuno et al. 2013; Gregory et al. 2019; Vila-Nistal et al. 2023). In these times of high-troughput sequencing, phage isolation studies are not so frequent, not in the least because of their much lower throughput and their dependence on host culturability. However, to fully understand the interactions between phages and hosts, and thus their true significance for biogeochemical cycles and bacterial evolution, isolation of specific phage-host systems their further characterization is necessary.

In this dissertation, our knowledge of phages infecting marine *Roseobacteraceae*, which are key players in the cycling of organic matter in marine ecosystems, was significantly extended. By combining phage isolation with database mining for environmental phage genomes, the cobaviruses were uncovered. This led to the delineation and later on, official acceptance by ICTV, of the first family of marine, tailed dsDNA viruses – the *Zobellviridae*, member of the *Caudoviricetes* class. Our biogeography survey showed that cobaviruses influence roseobacter populations at a global scale, from temperate to tropical marine waters, especially in coastal areas, and thus potentially influence the biogeochemical cycling in these environments (Chapter 2 & 3). Furthermore, the large-scale phage isolation campaign using direct plating yielded 128 genome-sequenced dsDNA phages. These belonged to twelve new genera and infected hosts of the genera *Sulfitobacter*, *Lentibacter* and *Octadecabacter*. Using hierarchical clustering of protein-clusters based intergenomic distances, we were able to classify these phages into four newly proposed viral families ("Hayaniviridae", "Schlingloffviridae", "Diferiteviridae", and "Woolleyviridae"), and four existing viral families (*Autographiviridae*, *Casjensviridae*, *Mesyanzhinoviridae*, and *Zobellviridae*), all in the *Caudoviricetes* class (Chapter 4). Apart from the genomic and taxonomic diversity revealed, this new collection provides many new phage-host systems for future detailed investigations. A start has been made with the preliminary analysis of the new sulfivirus group, a large genus of 89 phages infecting closely related *Sulfitobacter* strains, which pointed the spotlights on the arms race between phages and host bacteria and the genomic microdiversity arising on both sides (Chapter 5). Finally, while the focus of this dissertation was on dsDNA roseophages, the isolation and initial characterization of Sulfitobacter phage ICBM5 reminds that the

aforementioned enormous diversity of roseophages is complemented by a most likely great variety of ssDNA roseophages (Chapter 6).

7.1. The complex and eternally evolving field of phage classification and taxonomy

Any kind of system trying to classify biological entities is human made and inherently limited by our understanding of the diversity of the entities to be classified and the forces driving that diversity. Thus, quite often, increasing knowledge about the diversity will result in modifications of the classification system. In the case of phage taxonomic classification, new findings have led to substantial transformations in the recent years. Instead of the traditional Baltimore Classification, Koonin et al. (2020a) introduced the new "Megataxonomy of Viruses", which is based on evolutionary relationships. In this megataxonomy, viruses are grouped into six realms and within those, they are organized in 15 hierarchical ranks inspired by the Linnaean taxonomy.

These changes in virus classification over the past few years are also reflected in the different chapters of this dissertation. For example, the taxonomy of the cobaviruses changed between the initial publication of the study in February 2019 (Chapter 2) and the taxonomic proposal ratified by the ICTV in March 2021 (Chapter 3), due to the dissolution of the order Caudovirales and of the families Siphoviridae, Myoviridae, and Podoviridae, which previously contained all tailed dsDNA phages (Adriaenssens et al. 2021). In the new megataxonomy, the tailed dsDNA phages are classified in the *Duplodnaviria* realm and the *Caudoviricetes* class. Accordingly, the cobaviruses are currently classified in the *Zobellviridae* family within the *Caudoviricetes* order. While for the assignment of lower ranks based on intergenomic sequence identities, clear threshold recommendations (species ≥ 95%, genus ≥ 70%) have been compiled (Turner et al. 2021; Moraru 2023), the delineation of intermediate ranks suchs as family and order is not as standardized yet. However, it is suggested to use whole-proteome based clustering methods for this and to compare protein family profiles (Simmonds et al. 2023). The classification approach used later on in this dissertation (Chapter 4), using PC-based intergenomic similarities for family delineation, suggested that also the *Zobellviridae* family could be split into several families in the future, thus again changing the cobavirus taxonomy.

The ongoing changes in virus classification were a challenge for this dissertation. At the same time, they illustrate how exciting and at the cutting edge of virus research these projects

are. Finally, during our classification efforts, we actively took part in shaping phage classification criteria. Our studies revealed roseophage diversity at different levels (families, genera, and species) and proposed changes in some parts the current classification. On the one hand side, the proteomic tree-based classification of the new roseophage isolates (Chapter 4) layed the emphasis on family delineation and suggests that most likely existing phage families such as the aforementioned *Zobellviridae* and the *Autographiviridae* could be transformed into higher level taxa in the future. On the other hand side, the isolation of a large number of phages from a single genus, the sulfiviruses, placed the focus on virus classification at the genus, species and strain level (chapter 5). The revealed genomic microdiversity of the sulfiviruses, the cross-infection of their isolation host strains, and the co-infection of the same host cell by two sulfiviruses (not included in this dissertation) supports the re-thinking of the viral species definition.

Defining a species as members of populations that actually or potentially interbreed in nature, suggests that this group of phages, which co-occur in time and space and genetically exchange via recombination, could be regarded to as several strains of one species. This is in line with Bobay and Ochman (2018), who showed that the Biological Species Concept (BSC) could be applied for a viral species delineation based on the rate of gene exchange within populations. Instead of using rigid thresholds for phage species classification (e.g. 95%), we should consider using thresholds that define viral species as groups of individuals with higher recombination rates among them, even if it means using flexible thresholds for different virus groups. In the case of the sulfiviruses, which have a minimum of 69% intergenomic identity, and show increased recombination rates, this approach would challenge the 95% threshold used by the ICTV for species definition.

Also in other cases, our studies reflect the necessity and challenge for virus classification to include genomic information as well as aspects of lifestyle and ecology. For example, Lentibacter phage ICBM8 was included into the “Diferiteviridae” family, despite a PC-based intergenomic similarity below the 30% threshold. This decision was made based on congruent features with the family member Lentibacter phage ICBM165. Both phages have the same isolation host, the same predicted podoviral morphology and the same predicted DNA packaging strategy. Furthermore, they occur in the same habitat, as they both have been isolated from seawater sample HE504-33 (see chapter 4.3.5.5.). This example illustrates that classification thresholds especially at higher and intermediate levels sometimes need to be flexible in order to create ecologically meaningful taxa. Even though taxonomic classification is based on genetic information, especially for the fast evolving phages with their genomes

being highly influenced by horizontal gene transfer, their lifestyle and interactions might be more informative with regard to their recombinatory actitivies and allow flexible thresholding of the taxons.

7.2. The role of roseophages in the marine environment, more specifically during algal blooms

Viruses are the most abundant biological entities in the oceans. In surface waters, phages outnumber their hosts by an order of magnitude, with 20 - 40% of bacteria being lysed every day (Suttle 1994; Suttle 2007; Breitbart et al. 2018). They influence bacterioplankton dynamics and play an essential role in global carbon and nutrient cycling. Through viral lysis, bacterial community compositions are shifted as dominant members are lysed and thus decrease in numbers, as described by the "killing-the-winner" model (Thingstad 2000). At the same time, viral lysis leads to the release and recycling of carbon and intracellular nutrients instead of the unilateral biomass transfer to higher trophic levels. This phenomenon is termed "viral shunt" (Wilhelm and Suttle 1999). In contrast to the viral shunt, it was recently shown that phytoplankton can also graze on viruses, thereby redirecting biomass to higher trophic levels (DeLong et al. 2023). Furthermore, viral lysis and aggregation of the resulting cell debris can lead to carbon export to the deep ocean, known as the "viral shuttle" (Weinbauer 2004; Sullivan et al. 2017; Nissimov et al. 2018). Situations with increased carbon turnover, in which these interrelations can be well studied, are phytoplankton blooms.

Phytoplankton blooms are seasonal events mainly happening in coastal regions, in which microscopic algae accumulate in the photic zone, forming clouds sometimes being visible even from space (Kutser 2009; Dai et al. 2023). The breakdown of these blooms as a consequence of grazing, viral infection and limited nutrient availability releases large amounts of dissolved organic matter (DOM), which is consumed by heterotrophic bacteria (Biddanda and Benner 1997; Pinhassi et al. 2004; Thornton 2014; Buchan et al. 2014). The bacteria are subsequently infected by viruses, with lytic infections significantly contributing to the collapse of the bacterial bloom and the release of nutrients (Kuhlisch et al. 2021; Biggs et al. 2021). Algal blooms can be dangerous for marine life and also humans, with toxins produced by some algal species accumulating in so-called "harmful algal blooms" (HABs) and being transferred into the food web (Anderson 2009; Richlen et al. 2010). Globally, algal bloom frequency and extension have increased over the last two decades, even though on a regional scale trends are

more heterogeneous due to different climate developments and anthropogenic influences (Gobler 2020; Dai et al. 2023).

The waters around Helgoland Roads in the North Sea are characterized by phytoplankton blooms occurring annually during spring and summer (Gerdts et al. 2004; Wiltshire et al. 2010). Frequently, they have been a site for isolation of heterotrophic bacteria, including from the Roseobacter group, as well as of bacteriophages (Moebus and Nattkemper 1981; Wichels et al. 1998; Eilers et al. 2000b; Eilers et al. 2000a; Hahnke et al. 2013; Hahnke et al. 2015; Alejandre-Colomo et al. 2020; Heins et al. 2021; Heins and Harder 2022). Members of the *Roseobacteraceae* family are omnipresent in marine ecosystems inhabiting various ecological niches. They are present in pelagic waters, but also in sediment samples and attached to surfaces (Brinkhoff et al. 2008; Kanukollu et al. 2016; Simon et al. 2017). Often, they are associated to algae, which is why they are especially abundant during phytoplankton blooms (Eilers et al. 2001; Buchan et al. 2005; Lamy et al. 2009; Teeling et al. 2016).

The "Roseobacter group strains M#" used for our large-scale phage isolation campaign (chapter 4) were isolated from surface water samples taken at the Helgoland Roads time series station during the spring phytoplankton bloom in 2016 (Alejandre-Colomo et al. 2020). Alejandre-Colomo et al. (2020) isolated strains of the genus *Sulfitobacter* throughout the bloom, i.e. from early-bloom, bloom and post-bloom samples, indicating that these bacteria are generalists coping with all stages of the bloom and the related ecological conditions (Alejandre-Colomo et al. 2020). Also *Lentibacter* sp. SH36, host of the cobaviruses described in chapter 2 and several more phage isolates described in chapter 4, as well as *Sulfitobacter* sp. SH24-1b, host of the ssDNA phage ICBM5, were isolated from a phytoplankton spring bloom in the southern North Sea (Hahnke et al. 2013). Having isolated seven roseobacter strains with different physiological features from the same habitat, Hahnke et al. (2013) hypothesized that this metabolic versatility is what enables the Roseobacter group to occupy different ecological niches and to cope with changes in the supply of substrates during a phytoplankton bloom. Three of the host strains that were used for phage isolation in this dissertation were isolated from particle fractions: *Sulfitobacter* sp. SH24-1b (Hahnke et al. 2013), *Lentibacter* sp. MPI-62 (Helgoland, spring phytoplankton bloom 2017, Heins et al. (2021)) and *Octadecabacter* sp. MM282 (shore of Harlesiel, October 2017). This again highlights the inhabitation of different ecological niches by members of the Roseobacter group and shows that we have covered this habitat diversity in our phage isolation studies.

With phage abundance and activity increasing during phytoplankton blooms (Wiltshire et al. 2010; Bartlau et al. 2021), seawater samples taken at these times are ideal for phage isolation, as it has been shown before. Bartlau et al. (2021) isolated 44 new phages infecting *Flavobacteriia*, another important group of heterotrophic bacteria, from seawater samples taken at the Helgoland Roads time series station during the phytoplankton spring bloom in 2018. By use of direct plating for isolation and then read mapping against cellular metagenomes, they could prove active replication of these phages during the bloom. Furthermore, some of the phages could be re-isolated in the following years, supporting their hypothesis that these phages are permanently present in this environment and potentially modulate the flavobacterial population in the North Sea (Bartlau et al. 2021). Two of the environmental cobaviruses described in chapter 2 of this dissertation were detected in spring bloom metagenomes from the Helgoland Roads time series station (see chapter 2.3.6.). The fact that they were found in metagenomes of two successive years indicates that they are of long-term significance at this location as well. To search the metagenomes from Helgoland Roads also for the other new roseophages described in this dissertation would definitely be an interesting task for future research, especially as their isolation hosts originate from there.

Also in this current dissertation, the water samples used for phage isolation came from an algal bloom situation, a mesocosm experiment with an artificially induced algal bloom. In order to investigate the dynamics of phage-host interactions during a phytoplankton bloom in the North Sea, a large-scale mesocosm experiment was performed by an interdisciplinary team from the University of Oldenburg (Mori et al. 2021; Kerimoglu et al. 2022; Sutorius et al. 2022; Dlugosch et al. 2023). Multiple indoor 600-liter vessels, so-called planktotrons, were prepared with near-natural conditions and inoculated with North Sea water samples from which grazers had been removed by filtration. Over 38 days, the bacterial and viral composition as well as several other parameters were monitored during and after the artificially induced phytoplankton bloom (Mori et al. 2021). A biphasic phytoplankton bloom was observed, first dominated by diatoms (*Thalassiosira* spp., *Skeletonema marinoi*, and *Pseudonitzschia* sp.), and then by the haptophyte *Phaeocystis globosa*. The parallel mesocosms showed overall similar developments.

Heyerhoff et al. (in prep.) analyzed the virus-bacteria dynamics in the mesocosms by bacterial and viral metagenomic analyses and could show that in succession of the phytoplankton bloom, heterotrophic bacteria increased in abundance, dominated by members of the *Flavobacterales*, until being infected and lysed by viruses reflecting “killing-the-winner” dynamics. This way, phage infection contributed to the collapse of the bacterial bloom and at

the same time maintained bacterial diversity and stabilized the community composition, which was measureable as an increased Shannon diversity index (Heyerhoff et al. in prep.). Furthermore, an increase in dissolved organic carbon (DOC) concentration during the viral lysis phase could be observed, illustrating the impact of viral lysis on the organic carbon pool known as the viral shunt.

Members of the *Roseobacteraceae* family were not as dominant in the mesocosms (Heyerhoff et al. in prep.). Likewise, host prediction for metagenome-derived viral clusters based on CRISPR spacers revealed dominance of viruses infecting *Flavobacterales* while roseophages were not as abundant. Heyerhoff et al. (in prep.) found the numbers of virus-like particles to be very stable over the course of the experiment once they had increased in consequence of the bacterial bloom. The host affiliation of the viral contigs however showed succession patterns similar to those of the bacterial hosts. Thus, with the *Roseobacteraceae* increasing in relative abundance after the first diatom bloom and until the end of the monitoring, it could be that with a time delay also the roseophage increased in abundance, which was no longer recorded. Water samples for our phage isolation were taken a few days after the monitoring had ended. We could speculate that at that time point roseophage relative abundance was maybe higher.

As reviewed by (Silveira et al. 2021), viral infection strategies are dependent on host densities. At low host densities, lysogeny is thought to be dominating in order to persist over times of low nutrient availability (Coutinho et al. 2017; Silveira et al. 2021). However, also at high host densities, lysogeny is prevailing in order to prevent superinfection. The latter is described by the "piggyback-the-winner" hypothesis (Knowles et al. 2016; Silveira et al. 2021). On the contrary, at intermediate cell densities the lytic life cycle is expected to be predominant, with viruses lysing the most abundant and fastest-growing hosts according to the "killing-the-winner" hypothesis (Thingstad 2000). Our phage isolates were shown to infect their isolation hosts lytically, so they are potentially involved in the bacterial bloom collapse. For some of them, genetic analysis revealed a lysogenic potential, so they can also be part of the "piggyback-the-winner" dynamics.

In order to investigate the temporal occurrences of our new roseophages in the mesocosms, it would be a task for future research to perform read mapping of the planktotron metaviromes from different time points against our phage genomes. This would also give an insight into the question whether the isolated roseophage diversity was already present in the sea water inoculum or if it was created by the infection dynamics during the phytoplankton

bloom. Moreover, the 16S rRNA gene sequences of our isolation host strains should be searched in the amplicon metagenomes to find out at which time points of the bloom they occurred.

The identity or family affiliation of the roseophages isolated from the different planktotrons did not reveal any distinct patterns and did thus not allow any conclusions regarding the viral community compositions in the mesocosms (see chapter 4.3.5.). The differential yields of phage isolates between P1 (3) and P2 (54) could be explained by the fact that the number of virus-like particles (VLPs) in P1 was generally lower throughout the mesocosm experiment than in P2 (Heyerhoff et al. in prep.). Phages of the *Autographiviridae* family and the "Hayaniviridae" family originated from several planktotrons, which suggests that they are quite stable representatives during such a phytoplankton bloom and replicate actively also at lower total virus numbers. For the sulfiviruses of the "Hayaniviridae" family, this hypothesis is further supported by the fact that also the host ranges did not show any patterns regarding isolation sources (see chapter 5.3.4.).

7.3. Methodological considerations

In this dissertation, we isolated a remarkable diversity of roseophages. Using comparatively few host species, very different phages were obtained, even though they represent of course only a part of the diversity in the ocean, since we focused on lytic, dsDNA phages. For the same host strain, phages of different families, potentially even orders were isolated (e.g., phages ICBM4 and ICBM7 both infecting *Lentibacter* sp. SH36). We achieved this by using two different phage isolation techniques. The cobaviruses described in chapters 2 and 3 and some more *Lentibacter* phages described in chapter 4 were isolated using enrichment cultures. A drawback of this technique is that more virulent phages might outcompete others during cultivation, which results in a low number of phage isolates that not necessarily represent the most important phages in nature. To circumvent the bias of enrichment cultures, we used direct plating in the large-scale isolation campaign described in chapter 4. After concentration of viral particles in the pre-filtered seawater by cross-flow filtration, we plated an aliquot of this viral concentrate together with a small volume of exponentially growing host culture. Having a higher number of different phages in the concentrate and being able to detect individual phages as separate single plaques in the bacterial lawn, we hoped to catch a variety of phages that reflects more the true diversity in the seawater sample than an enrichment culture. This proved successful through the isolation of the microdiverse sulfivirus strains. Isolation of so many different sulfivirus strains might not have been possible in enrichment cultures, since the more

virulent strains would have prevailed. A disadvantage of the direct plating approach could be that there still is a physical enrichment of the phage particles by cross-flow filtration. This might not affect so much the diversity of the isolated phages, but it was previously shown to reduce the overall phage yield due to high shearing forces (Alonso et al. 2000; Castro-Mejía et al. 2015). At the same time, the direct plating approach allows a much higher throughput than enrichment cultures. This way, we achieved our goal and isolated phages with two levels of diversity. On the one hand side, phages of eight different families were isolated. On the other hand side, the large collection of highly similar, microdiverse sulfivirus strains was obtained.

Both isolation methods, direct plating and enrichment culture, were successful. For *Lentibacter* sp. SH36, eight phages belonging to four families were obtained, three of them from enrichment cultures and five from direct isolation. Furthermore, it is noteworthy that the two approaches gave different phages for the same host strain, indicating that different laboratory conditions and techniques can be favorable for different phages and thus a combination of both and maybe even other approaches can be useful. Nevertheless, there are limitations of both applied phage isolation techniques, which are not easily to be bypassed. In general, the incubation time of the enrichment, but also of the plaque assay could influence the phage yield. Slow growing phages might be lost if the incubation time is too short, but too long incubation could lead to contamination or takeover by few dominating viruses. This dilemma illustrates that laboratory conditions cannot easily mimic the natural situation, which is why the assessment of viral diversity by metagenomic analyses is very valuable and needs to go hand in hand with cultivation studies.

Another limitation of our phage isolation approaches is that we used pre-filtered seawater and thus only free phage particles were part of the viral concentrate while those attached to bacteria and algae were lost. This not only excludes phages of host bacteria that live in true association with algae. Also free phages are known to randomly attach to particles such as algal aggregates (Mari et al. 2007; Riemann and Grossart 2008; Bettarel et al. 2016). The fact that also bacteria and intracellular phage particles were filtered out prior to concentration, means that only those phages could be isolated, that were actively replicating and thus present in high numbers as free particles in the seawater sample. These considerations remind that the infection efficiency of phages on certain hosts most likely influence the chance of being isolated. Even if the direct plating approach is not as prone to the takeover of the most dominant and fastest replicating phage as an enrichment culture, still the infection efficiency must be high enough to produce visible plaques on the plate.

This requirement is also very important to be kept in mind when interpreting the results of host range assays. Host ranges tested with spot assays tend to be overestimated (Holmfeldt et al. 2007; Hyman and Abedon 2010). Thus, we applied a method adapted from the "Molten Streaking for Singles" technique described by Kauffman and Polz (2018) to determine the host range of the sulfiviruses. Successful infection of the respective host is indicated by single plaques, therefore false positive results due to spontaneous lysis of the bacteria can be excluded. The determined host range can be described as "plaquing host range" (Hyman and Abedon 2010). Nevertheless, this method does not consider varying infection efficiencies. Prerequisite for an infection to be recognized is the formation of a visible plaque within the given time frame, which in turn requires a certain level of infectivity. Thus, infections of low infectivity are not being detected and the "plaquing host range" might not reflect the complete scope of host strains. This limitation should not be underestimated when trying to elucidate the links between co-infection, gene exchange and the emergence of viral microdiversity. Kauffman et al. (2021) evaluated host range and recombination within a large dataset of *Vibrio* phages and detected that recombination is common even between phages without overlaps in the determined host range. They hypothesized that phages infect many strains without killing them, still enabling co-infection and recombination. This underlines again, how important it is to identify the determined host range as "plaquing" or "killing".

7.4. Outlook

The aim of this thesis was to isolate and characterize new phages infecting marine *Roseobacteraceae* and this way to shed more light on their diversity in the North Sea. This aim was fulfilled by the isolation and taxonomic classification of 113 new roseophages infecting strains from three roseobacter genera. With the creation of such a large roseophage collection, this dissertation offers many opportunities for further research. Through taxonomic classification and initial genome analyses, a fascinating diversity has been revealed that is just waiting to be explored in more detail.

The potential ecological relevance of the cobaviruses was already uncovered by examination of their biogeographical distribution using read mapping against marine metagenomes (chapter 2 and 3). A search in more than 5,000 metagenomes revealed that they occur worldwide in many marine habitats, which indicates their participation in shaping *Roseobacteraceae* communities all around the globe. In regards to the new roseophage isolates (chapter 4 and 5), their taxonomic classification into families together with other phages that

have been isolated from very diverse habitats, from soil to sea- and freshwater to anthropogenic environments (see chapter 4.3.5.), suggests that they could also be found worldwide in metagenomes. Furthermore, our new roseophage isolates were obtained from seawater samples from different years (2013, 2015, and 2018) and seasons (March and July 2018), illustrating that we captured part of the roseophage diversity in the North Sea at different times and spaces and indicating that these phages are of environmental relevance over a longer time period. This especially holds true for the two *Lentibacter* phages in the "Schlingloffviridae" family, which were isolated in different years.

Nevertheless, the environmental relevance of the new roseophages should be assessed more specifically in the future. This could be done by read mapping against metagenomes providing different types of information: (i) metagenomes from the planktotron mesocosm experiment have a background of fine scale metadata from a phytoplankton bloom (Mori et al. 2021; Dlugosch et al. 2023), (ii) metagenomes from time series stations such as Helgoland Roads reflect phytoplankton blooms in the natural environment and are available from several years and seasons (Teeling et al. 2012; Teeling et al. 2016; Chafee et al. 2018; Krüger et al. 2019), and (iii) metagenomes from different locations and habitats around the world can provide knowledge about their spatial distribution. If available, both viral and bacterial metagenomes should be investigated. Lytic phages can be captured in cellular metagenomes due to ongoing infections during the time point of sampling or because free virus particles were caught because of unspecific binding during filtration, as it was likely the case for the cobaviruses. However, several of the new roseophage isolates showed indications for lysogenic potential and thus could be found being present as prophages. In addition, metagenome searches could gain information about the habitat and lifestyle of the *Roseobacteraceae* host strains. As the cobaviruses were often found in metagenomes from protist size fractions, we hypothesized that their host bacteria could be protist-associated. This was supported by the fact that the cobaviruses possessed a class II cobalamin-dependent RNR gene, which was earlier hypothesized as an indication for infection of vitamin B12-producing bacteria living in association with phototrophic protists (Sakowski et al. 2014). Since cobalamin-dependent RNR genes were also present in some of the other new roseophages (*Mesyanzhinoviridae* and "Hayaniviridae", chapter 4.3.5) and since the host genera *Sulfitobacter* and *Octadecabacter* are known to associate with algae (Guannel et al. 2011; Dogs et al. 2017; Yang et al. 2021; Jin et al. 2023), the new roseophage genomes will likely be found in metagenomes from algal size fractions, as well.

Furthermore, the lifestyle of the new roseophages should be examined. On the one hand, their life cycle when infecting the respective isolation host strain should be described by means of one-step growth curves (temporal succession, burst size, etc.). On the other hand, their ability to infect other strains should be tested. Both can be done in culture but also *in situ* using the phageFISH technique (Allers et al. 2013; Barrero-Canosa and Moraru 2019). For the cobaviruses, the host range was already determined, however only using a collection of type strains from different *Roseobacteraceae* genera. It would be interesting to test their ability to infect more closely related strains. In contrast, the sulfiviruses were only tested for their infectivity against the highly similar, species-level related sulfivirus host strains. A future task could be to determine whether they can infect strains of different species and genera.

With the sulfivirus project being presented in this dissertation in the form of preliminary results (chapter 5), the main goal for the near future is to learn more about the relationship between this intriguing group of phages and their hosts, and how their interactions generate microdiversity on both sides. A start has been made by Ismail Hayani, who could show in his master thesis that sulfiviruses ICBM16 and ICBM18 co-infect their isolation host *Sulfitobacter* sp. M53. A future task will be to prove that this co-infection leads to recombination events between the viral genomes and thereby contributes to horizontal gene transfer and microdiversity. Furthermore, the infection cycle of the sulfiviruses should not only be analyzed in single infection cultures, but also in co-culture with multiple phages and maybe even multiple host strains. Following the infection dynamics by phageFISH and transcriptomics could give valuable information about potential competition between sulfiviruses, co-infection or superinfection exclusion, potential prophage induction and regulation of defense systems on both sides, phages and hosts.

Appendix

a) Supplementary tables

Table S 1: List of *Rhodobacteraceae* strains used for the host range assay.

Name	Strain designation	Strain	Infected (- no, + yes)
Antarctobacter heliothermus	EL-219	DSM 11445	-
Celeribacter baekdonensis	L-6	DSM 27375	-
Celeribacter indicus	P73	DSM 27257	-
Celeribacter marinus	IMCC12053	DSM 100036	-
Celeribacter neptunius	H 14	DSM 26471	-
Citreicella aestuarii	AD8	DSM 22011	-
Citreicella marina	CK-I3-6	DSM 26424	-
Dinoroseobacter shibae	5 Plasmids	DSM 16493	-
Huaishuia halophila	ZXM137	DSM 26270	-
Hwanghaeicola aestuarii	Y26	DSM 22009	-
Jannaschia donghaensis	DSW-17	DSM 102233	-
Jannaschia helgolandensis	Hel10	DSM 14858	-
Jannaschia pohangensis	H1-M8	DSM 19073	-
Jannaschia rubra	4SM3	DSM 16279	-
Leisingera aquimarina	R-26159	DSM 24565	-
Leisingera caerulea	13	DSM 24564	-
Leisingera daeponensis	TF-218	DSM 23529	-
Leisingera methylohalidivorans	MB2	DSM 14336	-
Lentibacter sp.	SH36		+
Litoreibacter albidus	KMM 3851	DSM 26922	-
Litoreibacter arenae	GA2-M15	DSM 19593	-
Litoreibacter janthinus	KMM 3842	DSM 26921	-
Litorimicrobium taeanense	G4	DSM 22007	-
Loktanella cinnabarina	LL-001	DSM 29954	-
Loktanella fryxellensis	R-7670	DSM 16213	-
Loktanella hongkongensis	UST950701-009P	DSM 17492	-
Loktanella koreensis	GA2-M3	DSM 17925	-
Loktanella pyoseonensis	JJM85	DSM 21424	-
Loktanella salsilacus	R-8904	DSM 16199	-
Loktanella tamlensis	SSW-35	DSM 26879	-
Loktanella vestfoldensis	R-9477	DSM 16212	-
Maribius pelagius	B5-6	DSM 26893	-
Maribius salinus	CL-SP27	DSM 26892	-
Marinovum algicola	FF3	DSM 10251	-
Marinovum algicola	DG898	DSM 27768	-
Maritimibacter alkaliphilus	HTCC2654	DSM 100037	-
Nautella italica	R11	DSM 26436	-
Oceanibulbus indolifex	HEL-45	DSM 14862	-
Oceanicola batsensis	HTCC2597	DSM 15984	-
Oceanicola granulosus	HTCC2516	DSM 15982	-
Oceanicola nanhaiensis	SS011B1-20	DSM 18065	-
Octadecabacter temperatus	SB1	DSM 26878	-
Palleronia marisminoris	B33	DSM 26347	-
Pelagibaca bermudensis	HTCC2601	DSM 26914	-
Phaeobacter gallaeciensis	BS 107	DSM 26640	-
Phaeobacter inhibens		DSM 17395	-
Phaeobacter inhibens	T5	DSM 16374	-
Phaeobacter inhibens	2.10	DSM 24588	-
Ponticoccus litoralis	CL-GR66	DSM 18986	-
Pseudophaeobacter arcticus	20188	DSM 23566	-
Pseudoruegeria lutimaris	HD-43	DSM 25294	-
Roseibacterium elongatum	Och 323	DSM 16469	-
Roseivivax isoporae	sw2	DSM 22223	-

Roseivivax roseus	BH87090	DSM 23042	-
Roseobacter denitrificans	Och 114	DSM 7001	-
Roseobacter litoralis	Och 149	DSM 6996	-
Roseovarius crassostreae	CV919-312, CVSP	DSM 16950	-
Roseovarius halocynthiae	MA1-10	DSM 27840	-
Roseovarius indicus	B108	DSM 26383	-
Roseovarius lutimaris	112	DSM 28463	-
Roseovarius marinus	HDW-9	DSM 25228	-
Roseovarius mucosus	DFL-24	DSM 17069	-
Roseovarius nubinhibens	ISM	DSM 15170	-
Ruegeria atlantica	1480	DSM 5823	-
Ruegeria conchae	TW15	DSM 29317	-
Ruegeria marina	ZH17	DSM 24837	-
Ruegeria pomeroyi	DSS-3	DSM 15171	-
Sagittula stellata	EE-37	DSM 11524	-
Salinihabitans flavidus	ISL-46	DSM 27842	-
Salipiger mucosus	A3	DSM 16094	-
Sedimentitalea nanhaiensis	NH52F	DSM 24252	-
Sediminimonas qiaohouensis	YIM B024	DSM 21189	-
Shimia haliotis	WM35	DSM 28453	-
Shimia marina	CL-TA03	DSM 26895	-
Sulfitobacter delicatus	KMM 3584	DSM 16477	-
Sulfitobacter dubius	KMM 3554	DSM 16472	-
Sulfitobacter litoralis	Iso 3	DSM 17584	-
Sulfitobacter marinus	SW-265	DSM 23422	-
Sulfitobacter mediterraneus	CH-B427	DSM 12244	-
Sulfitobacter noctilucae	NB-68	DSM 100978	-
Sulfitobacter noctilucicola	NB-77	DSM 101015	-
Sulfitobacter pseudonitzschiae	H3	DSM 26824	-
Sulfitobacter sp.	EE-36	DSM 11700	-
Thalassobius aestuarii	JC2049	DSM 15283	-
Thalassobius maritimus	GSW-M6	DSM 28223	-
Thalassococcus halodurans	UST050418-052	DSM 26915	-
Thioclava dalianensis	DLFJ1-1	DSM 29618	-
Thioclava pacifica	TL 2	DSM 10166	-
Tranquillimonas alkanivorans	A34	DSM 19547	-
Tropicibacter multivorans	MD5	DSM 26470	-
Tropicibacter naphthalenivorans	C02	DSM 19561	-
Tropicimonas isoalkanivorans	B51	DSM 19548	-
Wenxinia marina	HY34	DSM 24838	-
Yangia pacifica	DX5-10	DSM 26894	-

Table S 2: Sequencing and assembling phage genomes from isolates and enrichments.

	Phage isolates - sequenced by		**Enrichments - sequenced by**	
	Illumina	**PacBio**	**Illumina (S1 and S2)**	**PacBio (S2 only)**
ICBM1	yes	yes	not assembled	not assembled
ICBM2	yes	no	assembled in S2	assembled in S2
ICBM3	n.a.	n.a.	assembled in S1, S2	assembled in S2

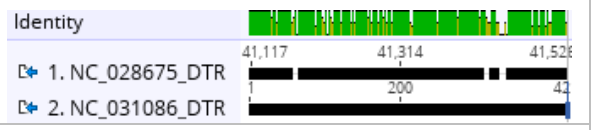
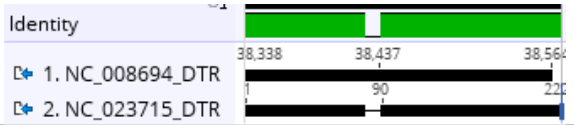
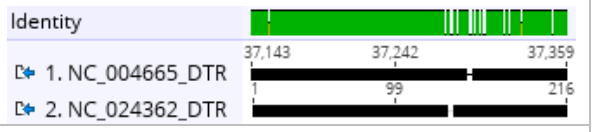
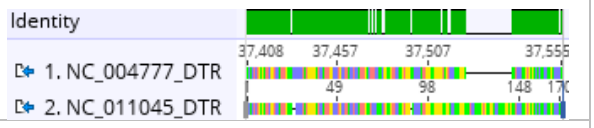
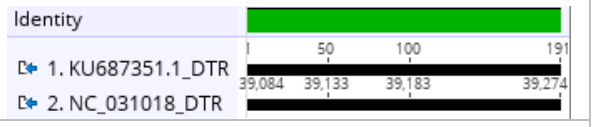
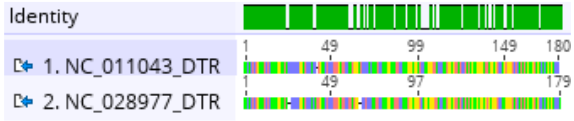
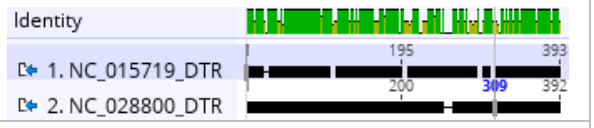
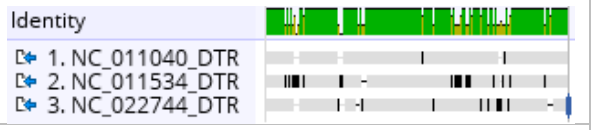
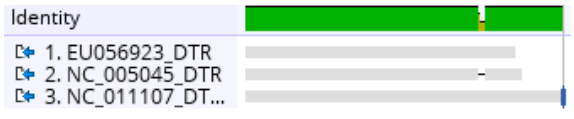
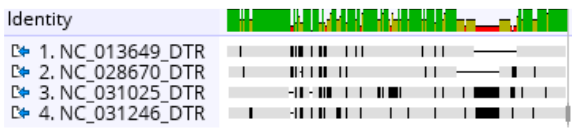
Table S 3: Cobaviral genomes - % GC and length (bases).

Name	G+C content	Sequence Length (bases)
vB_LenP_ICBM1	47.00%	40163
vB_LenP_ICBM2	47.80%	40907
vB_LenP_ICBM3	47.30%	40498
SIO1	46.20%	40072
P12053L	46.10%	39061
EnvX	40.20%	40752
EnvY	40.10%	36003
EnvZ	44.80%	35824
Env8	39.80%	38447
Env9	40.30%	41607
Env14	40.30%	35066

Table S 4: Bi-directional rho-independent transcriptional terminators in the genomes of the Cobavirus group (blue - stems, red - loops).

Name	Terminator name	Position	Sequence	Strand	Delta G
ICBM1	t1	15927-15979	CAAATAAGTAAAGCCCCCAAGGAGAAATCCAAGGGGGCTaTTTCTTTGTGTAT	-	-13.84
ICBM1	t2	15930-15981	CACAAAGAAATAGCCCCCTTGGATTTCTCCTTGGGGGCTTTACTTATTTGGA	+	-13.64
ICBM3	t1	15916-15968	CAAATAAGTAAAGCCCCCAAGGAGAAATCCAAGGGGGCTaTTTCTTTATGTAT	-	-13.84
ICBM3	t2	15919-15970	CATAAAGAAATAGCCCCCTTGGATTTCTCCTTGGGGGCTTTACTTATTTGGA	+	-13.64
ICBM2	t1	14733-14782	AACACAAAGAAGCCCCCAAGGAGAAATCCAAGGGGGCTTTTGCTTGTCTA	-	-12.64
ICBM2	t2	14735-14784	GACAAGCAAAAGCCCCCTTGGATTTCTCCTTGGGGGCTTCTTTGTGTTTA	+	-13.14
P12053L	t1	17261-17316	TAAACACAAAGAAGCCCCCAAGGATTTTACTCCAAGGGGGCTTTTGCTTGTTCATC	-	-14.09
P12053L	t2	17265-17316	AACAAGCAAAAGCCCCCTTGGAGTAAAATCCTTGGGGGCTTCTTTGTGTTTA	+	-12.99
SIO1	t1	15149-15204	TAAACACAAAGAAGCCCCCAAGGATTAATCTCCAAGGGGGCTTTTGTTTGTCTATA	-	-14.09
SIO1	t2	15153-15204	GACAAACAAAAGCCCCCTTGGAGATTAATCCTTGGGGGCTTCTTTGTGTTTA	+	-12.99
EnvX	t1	12131-12185	GAAATAAAGAAGAAGCCCCAAGGAGAAATCCTGAGGGGCTTTTTTATTACTCTTG	-	12.06
EnvX	t2	12134-12187	GAGTAATAAAAAAGCCCCTCAGGATTTCTCCTTGGGGCTTcTTCTTTATTTCTT	+	-11.26
EnvY	t1	12048-12102	GAAATAAAGAAGAAGCCCCAAGGAGAAATCCTGAGGGGCTTTTTTATTACTCTTG	-	-12.06
EnvY	t2	12051-12104	GAGTAATAAAAAAGCCCCTCAGGATTTCTCCTTGGGGCTTcTTCTTTATTTCTT	+	-11.26
Env8	t1	11722-11764	AGAAAAAGTAAGGGAGCCTAAGTAGCTCCCcTTTTTTATACCT	-	-10.60
Env8	t2	11724-11765	GTATAAAAAAGGGGAGCTACTTAGGCTCCCTTACTTTTTCTT	+	-12.70
Env9	t1	18115-18164	AAATAAAATAAACCCCCTTGGATTTCTCCTTGGGGGTTTTTTCTTACTTG	-	-10.44
Env9	t2	18115-18169	CAAGTAAGAAAAAACCCCCAAGGAGAAATCCAAGGGGGTTTaTTTTATTTCTTTT	+	-12.34
Env14	t1	12255-12297	TAAAAGAAGAAGGGAGCCTAAGTAGCTCCCcTTTTTTTATGC	-	-10.60
Env14	t2	12257-12298	ATAAAAAAAAGGGGAGCTACTTAGGCTCCCTTCTTCTTTTAA	+	-12.70

Table S 5: DTRs from related phages.

Group	phages	DTRs - length	DTRs - nucleotide identity (excluding gaps)	DTRs - alignment
1	Acinetobacter phage phiAB1	410	87%	
	Acinetobacter phage phiAB6	421		
2	Yersinia phage Berlin	227	95%	
	Yersinia phage Yep-phi	222		
3	Pseudomonad phage gh-1	217	92%	
	Pseudomonas phage phiPSA2	216		
4	Yersinia phage phiA1122	148	80%	
	Enterobacteria phage 13a	170		
5	Citrobacter phage SH5	191	100%	
	Citrobacter phage SH4	191		
6	Klebsiella phage K11	180	83%	
	Klebsiella phage vB_KpnP_KpV289	179		
7	Enterobacteria phage K30	393	80%	
	Klebsiella phage K5	392		
8	Enterobacteria phage BA14	194	83% - 93%	
	Kluyvera phage Kvp1	194		
	Erwinia phage FE44	193		
9	Pseudomonas phage PT5	413	98% - 100%	
	Pseudomonas phage phiKMV	414		
	Pseudomonas phage PT2	488		
10	Klebsiella phage KP34	216	71% - 87%	
	Klebsiella phage vB_KpnP_KpV41	214		
	Klebsiella phage vB_KpnP_KpV475	243		
	Klebsiella phage vB_KpnP_KpV71	246		
11	Enterobacteria phage EcoDS1	178	84% - 93%	

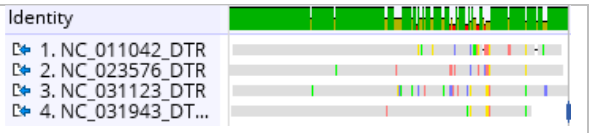
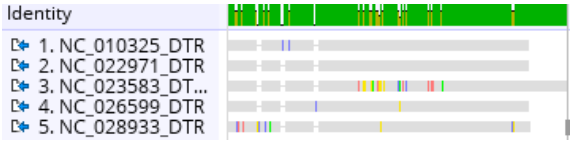
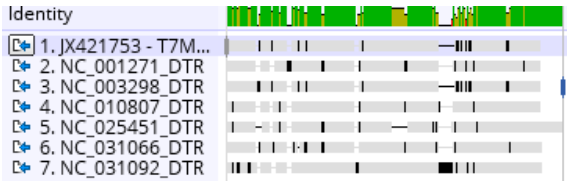
	Citrobacter phage CR44b	183		
	Citrobacter phage SH3	184		
	Escherichia phage vB_EcoP_GA2A	165		
12	Pseudomonas phage LUZ24	165	88% - 99%	
	Pseudomonas phage phiIBB-PAA2	183		
	Pseudomonas phage TL	207		
	Pseudomonas phage vB_PaeP_C2-10_Ab22	184		
	Pseudomonas phage PhiCHU	185		
13	Enterobacteria phage T7M	230	83% - 100%	
	Yersinia phage phiYeO3-12	232		
	Enterobacteria phage T3	231		
	Salmonella phage phiSG-JL2	230		
	Yersinia phage vB_YenP_AP5	235		
	Citrobacter phage SH1	231		
	Citrobacter phage SH2	243		

Table S 6: Prophage predictions using Phaster for the contigs containing the spanin genes.

Organism	Contig accession	Predicted prophage position	Prophage score	Spanin position	Spanin accession	Spanin associated to prophage
Thalassobius gelatinovorus strain DSM 5887	NZ_FOFW01000006	189621-213252	intact	208342..208716	WP_058264408.1	yes
Salinihabitans flavidus strain DSM 27842	FODS01000069	none, too short contig		1323..1676	SEP27667	unknown
Ruegeria mobilis strain DSM 23403	NZ_FNNK01000005	157938-194421	intact	178515..178898	WP_065332003	yes
Ruegeria mobilis strain M41-2.2	NZ_LNWW01000004	700765-722736	incomplete	721841..722224	WP_065329472	yes
Silicibacter sp. TM1040	NC_008044.1	1376648-1397748	intact	1396955..1397338	WP_011538643	yes
Rhodovulum sp. MB263	NZ_CP020384	862737-879872	incomplete	865553..865897	WP_080615430	yes
Rhodovulum sulfidophilum strain AB26	MSYQ01000001	many, not associated with the spanin		1904888..1905244	OLS44559.1	no
Phaeobacter sp. P97	NZ_CP016364	1925407-1946347	incomplete	1945236..1945619	WP_072504847	yes
Phaeobacter inhibens strain S4Sm	NZ_LOHU01000025	4300-27092	questionable	5028..5411	WP_061047696	yes
Phaeobacter sp. S26	NZ_JSWK01000008	125831-162508	questionable	140888..141271	WP_040172280	yes
Phaeobacter inhibens DSM 16374	NZ_KI421498	1832819-1867875	intact	1854039..1854422	WP_027247877	yes
Phaeobacter inhibens DSM 17395	NC_018290.1	1905561-1926725	intact	1925596..1925973	WP_014880219	yes
Leisingera sp. ANG-M7	NZ_JWLI01000008.1	60402-97630	intact	77885..78262	WP_052272404	yes

Table S 7: Environmental distribution of *Celeribacter marinus* IMCC 12053, based on the 16S rRNA blast hits in the NR Blast databases from NCBI, with minimum 99.0% identity.

Blast hit Accession	Isolation source	Country	coordinates	% nucleotide Identity	Alignment length	evalue
AJ391195.1	Adriatic Sea			100	1338	0
AJ391196.1	Adriatic Sea			100	1360	0
CP012023.1	coastal surface seawater of the Yellow Sea	South Korea		100	1471	0
AM990783.1	sea water North Western Mediterranean Sea	France	42.29 N 3.08 E	99.858	1407	0
AY697903.1	seawater	Antarctica		100	1410	0
NR_133717.1	seawater, Yellow Sea	South Korea		99.853	1359	0
HM140667.1	toxigenic diatom Pseudo-nitzschia, Puget Sound	WA (Main Basin)		99.29	1409	0
GU061042.1	Yellow Sea intertidal beach	Korea		99.283	1394	0
GU061048.1	Yellow Sea intertidal beach	Korea		99.211	1394	0
HM140672.1	Puget Sound	WA (Main Basin)		99.219	1409	0
HM140674.1	Puget Sound	WA (Main Basin)		99.148	1409	0

Table S 8: Environmental distribution of *Lentibacter* sp. SH36, based on the 16S rRNA blast hits in the NR Blast databases from NCBI, with minimum 99.0% identity.

Blast hit Accession		Isolation source	Country	coordinates	% nucleotide Identity	Alignment length	evalue
AJ391182.1		Adriatic Sea			100	1429	0
JQ269272.1		estuary zone of Jiulong River	China		99.508	1422	0
AF305498.1		German bight			99.502	1405	0
AY701455.1		Gymnodinium catenatum from Huon Estuary,	Tasmania (Australia)		99.71	1381	0
DQ234098.2		mangrove, Danshui river estuary	Northern Taiwan		99.862	1452	0
DQ234152.2		mangrove, Danshui river estuary	Northern Taiwan		99.931	1452	0
DQ234180.2		mangrove, Danshui river estuary	Northern Taiwan		99.931	1452	0
DQ234196.2		mangrove, Danshui river estuary	Northern Taiwan		99.931	1452	0
DQ234202.2		mangrove, Danshui river estuary	Northern Taiwan		99.862	1452	0
DQ234210.2		mangrove, Danshui river estuary	Northern Taiwan		99.931	1452	0
DQ234244.2		mangrove, Danshui river estuary	Northern Taiwan		99.862	1452	0
AY145564.1		marine section of the Weser estuary	Germany		99.778	1350	0
EU799040.1		Newport Harbour, RI	USA	41.486 N 71.351 W	99.86	1428	0
EU799044.1		Newport Harbour, RI	USA	41.486 N 71.351 W	99.93	1428	0
EU799171.1		Newport Harbour, RI	USA	41.486 N 71.351 W	99.72	1428	0
EU799468.1		Newport Harbour, RI	USA	41.486 N 71.351 W	100	1403	0
EU799546.1		Newport Harbour, RI	USA	41.486 N 71.351 W	99.79	1428	0
EU799658.1		Newport Harbour, RI	USA	41.486 N 71.351 W	99.86	1428	0
EU800046.1		Newport Harbour, RI	USA	41.486 N 71.351 W	99.719	1426	0
FJ882054.1		North Sea, 2 m depth		54.420 N 6.480 E	99.717	1415	0
FJ154967.1		ocean water from Bohai Bay	China		99.93	1422	0
HM057661.1		ocean water from the Yellow Sea			99.93	1422	0
JQ712107.1		oil-contaminated seawater			100	1422	0
KJ094194.1		polluted marine sediments			99.848	1319	0
EF659446.1		Poole Harbour seawater			99.854	1371	0
HM591431.1		pretreatment systems for seawater reverse osmosis process	South Korea		99.789	1423	0
HM591461.1		pretreatment systems for seawater reverse osmosis process	South Korea		99.93	1423	0
JF514256.1		Seawater from coast of Xiaomaidao Island, Qingdao	China		99.79	1426	0
AM945553.1		sea water, Adriatic sea	Italy	44.69 N 12.52 E	99.774	1329	0
AM945571.1		sea water, Adriatic sea	Italy	44.69 N 12.52 E	100	1269	0
AM945577.1		sea water, Adriatic sea	Italy	44.69 N 12.52 E	99.924	1315	0
AM945578.1		sea water, Adriatic sea	Italy	44.69 N 12.52 E	99.925	1325	0
AM945580.1		sea water, Adriatic sea	Italy	44.69 N 12.52 E	99.774	1330	0
AM945584.1		sea water, Adriatic sea	Italy	44.69 N 12.52 E	99.696	1317	0
AM945585.1		sea water, Adriatic sea	Italy	44.69 N 12.52 E	99.699	1328	0
EU930869.3		sea water sample, Gwangyang Bay	South Korea		99.928	1385	0
AB496659.1		Seawater, Shizuoka, Shimoda, Oura Bay	Japan		100	1361	0
FJ425223.1		seawater			99.783	1384	0
FJ436731.1		seawater, coast of Qingdao	China	36.0276 N 120.1846 E	99.925	1328	0
JQ195110.1		seawater; next to dolphin A, San Diego Bay	USA: San Diego, CA		99.844	1281	0
JQ195194.1		seawater; next to dolphin A, San Diego Bay	USA: San Diego, CA		99.844	1283	0
JQ195765.1		seawater; next to dolphin C, San Diego Bay	USA: San Diego, CA		99.922	1281	0

JQ195767.1		seawater; next to dolphin C, San Diego Bay	USA: San Diego, CA		100	1281	0
JQ196782.1		seawater; next to dolphin E, San Diego Bay	USA: San Diego, CA		100	1281	0
JQ712031.1		surface seawater offshore Qingdao	China		99.367	1422	0
JQ712040.1		surface seawater offshore Qingdao			99.93	1422	0
JQ712053.1		surface seawater offshore Qingdao			99.93	1422	0
JQ712057.1		surface seawater offshore Qingdao			100	1422	0
NR_108333.1		coastal region of Qingdao	China	36.027N 120.184E	100	1370	0

Table S 9: Environmental distribution of *Roseobacter* sp. GAI101, based on the 16S rRNA blast hits in the NR Blast databases from NCBI, with minimum 99.0% identity.

Blast hit Accession	Isolation source	Country	coordinates	% nucleotide Identity	Alignment length	evalue
JX529099.1	500m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.9533 S 15.1437 W	99.561	1367	0
JX529112.1	500m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.9533 S 15.1437 W	99.634	1367	0
JX529220.1	500m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.9533 S 15.1437 W	99.561	1367	0
JX531118.1	500m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.5015 S 15.4450 W	99.561	1367	0
JX531139.1	500m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.5015 S 15.4450 W	99.635	1368	0
JX531386.1	500m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.5015 S 15.4450 W	99.415	1368	0
AJ278782.1	Antarctic seawater isolate			99.571	1399	0
EU016167.1	Arctic deep-sea sediment			99.719	1424	0
FJ889522.1	Arctic Ocean			99.58	1430	0
FJ889535.1	Arctic Ocean			99.579	1424	0
FJ889536.1	Arctic Ocean			99.579	1424	0
FJ889527.1	Arctic Ocean			99.51	1430	0
FJ889528.1	Arctic Ocean			99.511	1431	0
FJ889530.1	Arctic Ocean			99.368	1424	0
FJ889531.1	Arctic Ocean			99.28	1388	0
FJ889542.1	Arctic Ocean			99.298	1424	0
EU365544.1	Arctic seawater			99.58	1430	0
EU365589.1	Arctic seawater			99.65	1430	0

JX525417.1	chlrophyll maxima (~30m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		48.0329 S 15.7856 W	99.561	1367	0
JX525429.1	chlrophyll maxima (~30m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		48.0329 S 15.7856 W	99.707	1367	0
JX525500.1	chlrophyll maxima (~30m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		48.0329 S 15.7856 W	99.489	1370	0
JX525579.1	chlrophyll maxima (~30m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		48.0329 S 15.7856 W	99.634	1367	0
JX527826.1	chlrophyll maxima (~30m) water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.9533 S 15.1437 W	99.488	1368	0
FN377731.1	clay-like sticky sediment with cylindrical plants/weeds on surface less than KB3	Svalbard, Kongsfjord region	78.59 N 11.56 E	99.52	1459	0
KJ939484.1	continental shelf sediments of Bay of Bengal			99.56	1364	0
NR_043547.1	East Sea	Korea		99.217	1404	0
FR684971.1	marine biome, fjord, coastal water	Norway, Raunefjord	60.27 N 5.22 E	99.508	1423	0
KJ475186.1	marine sediment, Ross Sea, Antarctica		76°06'17.10"S/ 169°12'45.36"E - 76°41'03.60"S/ 169°11'30.66"E	99.634	1365	0
AB733557.1	Nansei-Shotō Trench off Miyako Island, Okinawa Prefecture, southern Japan			99.719	1421	0
AY573043.1	Ny-Alesund, Svalbard, Norway	arctic	79° N, 12° E	99.579	1426	0
AY697915.1	seawater	Antarctica		99.37	1428	0
GQ358930.1	seawater from Kongsfjorden, Spitsbergen	Norway		99.579	1424	0
AF007257.2	seawater, Skidaway River			99.929	1414	0
AY794211.1	Soil from Ushuaia	Argentina	54°80' S and 68°13' W	99.37	1429	0
JX310158.1	Southern Ocean		47–48◦ South and 15–17◦ West	99.04	1458	0
GU584137.1	water sample	Antarctica	47–48°S and 15–17°W	99.588	1458	0
EF471650.1	whole surface water from Chesapeake Bay		39.13 N 76.33 W	99.719	1423	0
AY902203.1				99.489	1369	0
JX529304.1	500m depth water samples filtered on 0.22 micron filter paper; derived from the Southern ocean iron fertilization experiment (LOHAFEX)		47.9533 S 15.1437 W	99.122	1367	0

Table S 10: Environmental distribution of *Roseobacter* sp. SIO, based on the 16S rRNA blast hits in the NR Blast databases from NCBI, with minimum 99.6% identity.

Blast hit Accession	Isolation source	Country	coordinates	% nucleotide Identity	Alignment length	evalue
JX864948.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865037.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865038.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865039.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865040.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865041.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865044.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.688	320	1.16E-163
JX865047.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865055.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865056.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865060.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
JX865061.1	5m depth, the Gulf of Trieste, northernAdriatic Sea, NE Mediterranean			99.692	325	1.92E-166
AB703474.1	coastal hot spring, Kagoshima, Ibusuki hot spring	Japan		99.692	325	1.92E-166
AB703500.1	coastal hot spring, Kagoshima, Ibusuki hot spring	Japan		99.692	325	1.92E-166
AF007250.1	coastal seawater, Sapelo Island			99.686	318	1.50E-162
LT549272.1	corrosion biofilm Boothbay Harbor	USA:Maine	43.8443 N 69.6409 W	99.692	325	1.92E-166
LT549313.1	corrosion biofilm Boothbay Harbor	USA:Maine	43.8443 N 69.6409 W	99.692	325	1.92E-166
LT549329.1	corrosion biofilm Boothbay Harbor	USA:Maine	43.8443 N 69.6409 W	99.692	325	1.92E-166
LT549343.1	corrosion biofilm Boothbay Harbor	USA:Maine	43.8443 N 69.6409 W	99.692	325	1.92E-166
DQ234090.2	Danshui river estuary, mangrove	Northern Taiwan		99.692	325	1.92E-166
DQ234137.2	Danshui river estuary, mangrove	Northern Taiwan		99.692	325	1.92E-166
DQ234157.2	Danshui river estuary, mangrove	Northern Taiwan		99.692	325	1.92E-166
DQ234162.2	Danshui river estuary, mangrove	Northern Taiwan		99.692	325	1.92E-166
DQ234224.2	Danshui river estuary, mangrove	Northern Taiwan		99.692	325	1.92E-166
KC006265.1	estuary in middle of Jiulong River	China	24.26 N 117.127 E	99.69	323	2.48E-165
KC006295.1	estuary in middle of Jiulong River	China	24.26 N 117.129 E	99.692	325	1.92E-166
KT275136.1	Gut of Olive Flounder Paralichthys olivaceus			99.692	325	1.92E-166
EF215736.1	inert artificial surfaces submerged in marine water on Qingdao coast	China		99.692	325	1.92E-166
AY712068.1	lon=81.2699W, lat=31.3929N; surface water collected on Jul 18, 2001, Sapelo Island Microbial Observatory Dean Creek Marsh sampling site	USA: Georgia		99.692	325	1.92E-166
AY937021.1	marine sediment Nagasaki	Japan:		99.692	325	1.92E-166
AY941090.1	marine sediment Nagasaki	Japan:		99.692	325	1.92E-166
KC462965.1	nature reserve Kullaberg	Sweden		99.692	325	1.92E-166
KC462968.1	nature reserve Kullaberg	Sweden		99.692	325	1.92E-166

EU799217.1	Newport Harbour, RI	USA	41.486 N 71.351 W	99.692	325	1.92E-166
EU799316.1	Newport Harbour, RI	USA	41.486 N 71.351 W	99.692	325	1.92E-166
EU799841.1	Newport Harbour, RI	USA	41.486 N 71.351 W	99.692	325	1.92E-166
JQ432672.1	Northern Adriatic Sea, NE Mediterranean			99.692	325	1.92E-166
KF185598.1	Northern Adriatic Sea: Gulf of Trieste,coastal seawater sample 10m depth			99.692	325	1.92E-166
AY576710.1	northwestern Mediterranean Sea			99.692	325	1.92E-166
GQ347802.1	Saanich Inlet, 10 m depth	48.5883 N 123.5037 W		99.692	325	1.92E-166
AF190747.1	Scripps Institution of Oceanography Pier			100	325	4.13E-168
U64005.1	Scripps Pier			99.692	325	1.92E-166
AM945581.1	sea water	Italy:Adriatic sea	44.69 N 12.52 E	99.692	325	1.92E-166
HQ203887.1	seawater			99.692	325	1.92E-166
HM921237.1	seawater, Kiel Fjord	Germany	54°27′4N, 10°12′E	99.692	325	1.92E-166
FJ161222.1	Shandong coast	China		99.69	323	2.48E-165
FJ161239.1	Shandong coast, China	China		99.692	325	1.92E-166
EF491274.1	steel surfaces immerged in marine waterof Qingdao Coast			99.692	325	1.92E-166
AB362414.1	surface seawater, Niigata, Urahama coast	Japan		99.692	325	1.92E-166
HE981601.1	surface seawater, Mallorca Island, Palma Harbour	Spain		99.688	321	3.21E-164
KP198349.1	surface seawater, Cape Cod Bay	USA	41.744 N 70.219 W	99.692	325	1.92E-166
GU061006.1	Yellow Sea intertidal beach	Korea		99.692	325	1.92E-166
GU061090.1	Yellow Sea intertidal beach	Korea		99.692	325	1.92E-166
GU061106.1	Yellow Sea intertidal beach	Korea		99.692	325	1.92E-166
AF365990.1	off La Jolla, CA	USA		99.692	325	1.92E-166
AY536559.1				99.692	325	1.92E-166
JN653206.1	Bayboro Harbor, Florida	USA	27.76 N 82.63 W	99.692	325	1.92E-166

Table S 11: Prophage prediction using Phaster for the contigs containing the glutaredoxin gene.

Organism	**Contig accession**	**Predicted prophage position**	**Prophage score**	**Glutaredoxin position**	**Glutaredoxin accession**	**Glutaredoxin associated to prophage**
Rhizobium sp. Leaf383	NZ_LMQD01000008	1407-72420	intact	39353..39715	WP_082506598	yes
Sinorhizobium sp. LM21 (plasmid)	KM659098	11856-47482	questionable	26157..26360	AJW30150	yes
Hepatospora eriocheir strain GB1	LVKB01000119	none		3861..4241	ORD96177	
Hepatospora eriocheir strain GB2	LVKB01000120	none		3003..3317	ORD96175	

Table S 12: Prophage prediction using Phaster for the contigs containing the RNR genes.

Organism	Contig accession	Predicted prophage position	Prophage score	RNR position	RNR accession	RNR associated to prophage
Aminobacter sp. J41	NZ_JAGL01000009.1	44041-86811	intact	72083-74071	WP_024847845	yes
Inquilinus limosus MP06	NZ_JANX01000001.1	32458-86185	questionable	47409..49382	WP_034830439	yes
Azorhizobium caulinodans ORS 571	NC_009937.1	4097072-4138015	questionable	4122652..4124646	WP_012172112	yes
Microvirga flocculans ATCC BAA-817	NZ_JAEA01000002.1	156454-179652	incomplete	163007..164998	WP_027314944	yes
Streptomyces sp. RTd22	NZ_CP015726.1	8923271-8946147	incomplete	8923015..8924970	WP_063728949.1	yes
Streptomyces sp. CB03238	NZ_NBCN01000010	231547-250436	incomplete	227661..229613	WP_084899922	in vicinity
Streptomyces sp. NRRL F-4489	NZ_LLZI01000124	13535-24084	incomplete	19443..21374	WP_066977110	yes
Streptomyces rimosus subsp. rimosus strain NRRL WC-3904	NZ_JOCQ01000004	101674-119267	incomplete	111002..112933	WP_050514495	yes
Solirubrobacter sp. URHD0082	NZ_AUEK01000007	344542-357687	incomplete	307203..309158	WP_051323998	in vicinity
Labrenzia alexandrii DFL-11	EQ973124	3254-38475	questionable	15955..18009	EEE42840	yes
Methylobacterium sp. Leaf361	NZ_LMPY01000023.1	581118-616465	questionable	594636..596663	WP_082557943	yes
Methylobacterium aquaticum plasmid pMaq22A_3p	NZ_AP014707	65911-85601	incomplete	83571..85601	WP_060851403	yes
Marinobacter similis strain A3d10	CP007151.1	3645543-3690844	intact	3653541..3655502	AHI29708.1	yes
Protochlamydia amoebophila strain EI2 DB44_AM	NZ_JSAN01000012.1	none		231..2243	WP_039355956	no
Gonium pectorale isolate NIES-2863	LSYV01000168	none		31484..37421	KXZ42284.1	no
Volvox carteri f. nagariensis	NZ_JOCQ01000004.1	none		<195692..>202991	XP_002955713	no
Criblamydia sequanensis	NZ_CCEJ010000007	none		134416..136428	WP_041017866	no
Parachlamydia acanthamoebae strain Bn9	NZ_BAWW01000003	none		161269..163278	WP_006342605	no
Parachlamydia acanthamoebae strain OEW1 DB43_DW	NZ_JSAM01000017	none		15051..17060	WP_006342605	no
Parachlamydia acanthamoebae UV7	NC_015702	none		895276..897285	WP_006342605	no
Parachlamydia acanthamoebae str. Hall's coccus	NZ_ACZE01000088	none		17369..19378	WP_006342605	no
Estrella lausannensis	CWGJ01000019	none		124560..126572	CRX38802	no
Candidatus Rubidus massiliensis	CCSC01000001	1811588-1828589	incomplete	1412888..1414906	CDZ80757	no
Waddlia chondrophila	NC_014225	none		368424..370439	WP_013181434	no
Waddlia chondrophila 2032/99	FR872643	none		26560..28614	CCB90950	no
Neochlamydia sp. TUME1	NZ_JRXI01000032	none		32033..34045	WP_039383590	no
Neochlamydia sp. S13	NZ_BASK01001259	none		36271..38283	WP_042242257	no
Neochlamydia sp. EPS4	NZ_JSDQ01000098	none		4026..6038	WP_044882767	no
Parachlamydia sp. C2	NZ_FCNU01000032.1	none		238258..240270	WP_068471381	no
Candidatus Protochlamydia amoebophila UWE25	NC_005861	none		178729..180741	WP_011174661	no
Candidatus Protochlamydia sp. W-9	NZ_BCPZ01000156.1	none		50658..52670	WP_075883166	no
Candidatus Protochlamydia sp. R18	NZ_BASL01000738	none		1527..3539	WP_042281843	no

Table S 13: PhageTerm results of genome end determination for the sulfivirus genomes.

Phage ID	Göttingen ID	Genome status	Ends	Left	P-value left	Right	P-value right	Permuted	Orientation	Class	Type
ICBM9	CRMO_v_4	partial	Redundant	2976	4.09e-103	3304	1.19e-100	No	NA	DTR (short)	T7
ICBM10	CRMO_v_5	partial	Redundant	48386	3.60e-77	48714	7.21e-86	No	NA	DTR (short)	T7
ICBM12	CRMO_v_7	complete	Redundant	9141	4.42e-113	9468	1.60e-80	No	NA	DTR (short)	T7
ICBM13	CRMO_v_8	complete	Redundant	11460	1.45e-86	11788	6.81e-145	No	NA	DTR (short)	T7
ICBM15	CRMO_v_10	partial	Redundant	47666	4.23e-66	47993	5.91e-88	No	NA	DTR (short)	T7
ICBM16	CRMO_v_11	complete	Redundant	14465	1.60e-103	14792	1.97e-123	No	NA	DTR (short)	T7
ICBM17	CRMO_v_12	partial	Redundant	1773	7.19e-120	2100	6.30e-98	No	NA	DTR (short)	T7
ICBM18	CRMO_v_13	complete	Redundant	23206	4.11e-123	23534	3.71e-79	No	NA	DTR (short)	T7
ICBM21	CRMO_v_16	complete	Redundant	32674	1.70e-79	33001	1.55e-115	No	NA	DTR (short)	T7
ICBM22	CRMO_v_17	complete	Redundant	17299	2.33e-114	17626	2.03e-100	No	NA	DTR (short)	T7
ICBM23	CRMO_v_18	complete	Redundant	35965	2.39e-22	36294	6.67e-68	No	NA	DTR (short)	T7
ICBM24	CRMO_v_19	complete	Redundant	30504	4.31e-73	30833	3.43e-95	No	NA	DTR (short)	T7
ICBM25	CRMO_v_20	complete	Redundant	18266	2.44e-103	18593	1.79e-64	No	NA	DTR (short)	T7
ICBM26	CRMO_v_21	partial	Redundant	Multiple	4.44e-41	Multiple	1.17e-51	Yes	NA	-	-
ICBM32	CRMO_v_27	partial	Redundant	47760	3.78e-83	48087	6.65e-144	No	NA	DTR (short)	T7
ICBM38	CRMO_v_33	complete	Redundant	8079	1.18e-91	8407	8.93e-101	No	NA	DTR (short)	T7
ICBM39	CRMO_v_34	complete	Redundant	49276	1.07e-83	49582	2.33e-91	No	NA	DTR (short)	T7
ICBM40	CRMO_v_35	complete	Redundant	33751	4.32e-64	34080	3.60e-85	No	NA	DTR (short)	T7
ICBM41	CRMO_v_36	complete	Redundant	2721	1.90e-59	3081	2.22e-115	No	NA	DTR (short)	T7
ICBM42	CRMO_v_37	complete	Redundant	38583	6.62e-104	38911	6.42e-75	No	NA	DTR (short)	T7
ICBM43	CRMO_v_38	complete	Redundant	29385	1.30e-51	29690	3.67e-80	No	NA	DTR (short)	T7
ICBM45	CRMO_v_40	complete	Redundant	17421	7.91e-131	17749	1.49e-137	No	NA	DTR (short)	T7
ICBM46	CRMO_v_41	partial	Redundant	4290	1.29e-84	4618	3.28e-136	No	NA	DTR (short)	T7
ICBM47	CRMO_v_42	complete	Redundant	36702	1.21e-82	37007	5.13e-68	No	NA	DTR (short)	T7
ICBM48	CRMO_v_43	complete	Redundant	29170	1.84e-91	29498	2.16e-143	No	NA	DTR (short)	T7
ICBM49	CRMO_v_44	complete	Redundant	2406	3.31e-92	2733	8.23e-134	No	NA	DTR (short)	T7
ICBM51	CRMO_v_46	complete	Redundant	39028	1.52e-95	39355	3.63e-126	No	NA	DTR (short)	T7
ICBM52	CRMO_v_47	complete	Redundant	11342	3.73e-190	11670	1.02e-143	No	NA	DTR (short)	T7
ICBM53	CRMO_v_48	complete	Redundant	7024	5.27e-83	7351	1.65e-120	No	NA	DTR (short)	T7
ICBM54	CRMO_v_49	complete	Redundant	2055	7.37e-141	2382	5.16e-82	No	NA	DTR (short)	T7

ICBM56	CRMO_v_51	partial	Redundant	6521	5.23e-104	6881	8.19e-125	No	NA	DTR (short)	T7
ICBM57	CRMO_v_52	complete	Redundant	1414	1.13e-17	1741	6.12e-35	No	NA	DTR (short)	T7
ICBM58	CRMO_v_53	complete	Redundant	800	6.09e-138	1127	2.69e-138	No	NA	DTR (short)	T7
ICBM59	CRMO_v_54	complete	Redundant	4861	1.16e-99	5188	4.79e-129	No	NA	DTR (short)	T7
ICBM60	CRMO_v_55	complete	Redundant	27841	8.90e-103	28168	1.24e-75	No	NA	DTR (short)	T7
ICBM61	CRMO_v_56	complete	Redundant	24061	6.98e-83	24421	2.71e-144	No	NA	DTR (short)	T7
ICBM62	CRMO_v_57	complete	Redundant	11483	6.55e-82	11811	5.99e-71	No	NA	DTR (short)	T7
ICBM63	CRMO_v_58	partial	Redundant	49683	4.10e-79	50010	4.83e-110	No	NA	DTR (short)	T7
ICBM65	CRMO_v_60	complete	Redundant	34877	8.60e-76	35204	3.27e-88	No	NA	DTR (short)	T7
ICBM67	CRMO_v_62	complete	Redundant	3584	3.13e-109	3944	2.14e-133	No	NA	DTR (short)	T7
ICBM68	CRMO_v_63	complete	Redundant	29637	7.43e-104	29997	2.44e-78	No	NA	DTR (short)	T7
ICBM69	CRMO_v_64	complete	Redundant	29661	1.18e-18	29989	1.38e-33	No	NA	DTR (short)	T7
ICBM70	CRMO_v_65	complete	Redundant	31136	1.42e-80	31464	1.91e-77	No	NA	DTR (short)	T7
ICBM71	CRMO_v_66	partial	Redundant	48799	3.06e-133	49126	2.26e-112	No	NA	DTR (short)	T7
ICBM72	CRMO_v_67	partial	Redundant	48898	5.88e-106	49258	1.14e-136	No	NA	DTR (short)	T7
ICBM74	CRMO_v_69	partial	Redundant	3162	1.51e-111	3468	2.70e-71	No	NA	DTR (short)	T7
ICBM76	CRMO_v_71	complete	Redundant	45273	1.83e-106	45601	2.17e-88	No	NA	DTR (short)	T7
ICBM77	CRMO_v_72	complete	Redundant	13544	3.99e-143	13872	1.54e-114	No	NA	DTR (short)	T7
ICBM78	CRMO_v_73	complete	Redundant	15111	3.10e-63	15439	7.37e-28	No	NA	DTR (short)	T7
ICBM79	CRMO_v_74	complete	Redundant	18431	1.87e-112	18760	4.84e-106	No	NA	DTR (short)	T7
ICBM80	CRMO_v_75	partial	Redundant	49619	1.98e-72	49946	2.37e-115	No	NA	DTR (short)	T7
ICBM82	CRMO_v_77	complete	Redundant	49636	6.18e-119	49942	5.55e-111	No	NA	DTR (short)	T7
ICBM83	CRMO_v_78	partial	Redundant	3791	4.92e-116	4096	2.22e-111	No	NA	DTR (short)	T7
ICBM84	CRMO_v_79	partial	Redundant	47992	5.67e-88	48297	7.74e-78	No	NA	DTR (short)	T7
ICBM86	CRMO_v_81	complete	Redundant	20076	1.69e-113	20404	3.42e-106	No	NA	DTR (short)	T7
ICBM87	CRMO_v_82	complete	Redundant	52519	5.78e-99	52879	7.06e-72	No	NA	DTR (short)	T7
ICBM88	CRMO_v_83	complete	Redundant	49792	1.57e-90	50152	2.03e-102	No	NA	DTR (short)	T7
ICBM89	CRMO_v_84	complete	Redundant	50739	5.85e-47	51044	5.63e-103	No	NA	DTR (short)	T7
ICBM90	CRMO_v_85	complete	Redundant	9775	1.64e-100	10102	9.99e-91	No	NA	DTR (short)	T7
ICBM91	CRMO_v_86	complete	Redundant	33384	1.44e-62	33712	3.51e-105	No	NA	DTR (short)	T7
ICBM94	CRMO_v_89	complete	Redundant	16862	9.09e-147	17189	9.39e-80	No	NA	DTR (short)	T7
ICBM95	CRMO_v_90	complete	Redundant	16608	5.94e-105	16935	2.18e-78	No	NA	DTR (short)	T7
ICBM99	CRMO_v_94	complete	Redundant	19181	3.22e-90	19509	7.50e-79	No	NA	DTR (short)	T7

ICBM100	CRMO_v_95	complete	Redundant	49168	7.15e-75	49473	6.65e-81	No	NA	DTR (short)	T7
ICBM102	CRMO_v_97	complete	Redundant	1055	7.85e-94	1383	7.90e-38	No	NA	DTR (short)	T7
ICBM103	CRMO_v_98	complete	Redundant	24255	3.47e-73	24583	7.10e-57	No	NA	DTR (short)	T7
ICBM105	CRMO_v_100	complete	Redundant	23140	3.10e-112	23468	4.57e-68	No	NA	DTR (short)	T7
ICBM107	CRMO_v_102	complete	Redundant	14935	6.16e-268	15263	8.45e-71	No	NA	DTR (short)	T7
ICBM109	CRMO_v_104	complete	Redundant	38792	2.72e-109	39120	1.90e-70	No	NA	DTR (short)	T7
ICBM110	CRMO_v_105	complete	Redundant	6757	2.53e-136	7084	2.45e-158	No	NA	DTR (short)	T7
ICBM113	CRMO_v_108	complete	Redundant	40066	6.56e-130	40394	2.62e-92	No	NA	DTR (short)	T7
ICBM119	CRMO_v_114	complete	Redundant	29759	2.42e-169	30119	2.11e-79	No	NA	DTR (short)	T7
ICBM120	CRMO_v_115	partial	Redundant	3432	1.44e-88	3792	4.49e-18	No	NA	DTR (short)	T7
ICBM124	CRMO_v_119	complete	Redundant	38659	1.04e-47	38987	1.93e-85	No	NA	DTR (short)	T7
ICBM126	CRMO_v_121	complete	Redundant	21875	7.51e-107	22180	1.80e-59	No	NA	DTR (short)	T7
ICBM127	CRMO_v_122	complete	Redundant	10022	2.31e-158	10327	2.10e-92	No	NA	DTR (short)	T7
ICBM128	CRMO_v_123	complete	Redundant	32538	2.83e-34	32866	1.79e-57	No	NA	DTR (short)	T7
ICBM131	CRMO_v_126	complete	Redundant	9141	3.44e-76	9468	1.42e-124	No	NA	DTR (short)	T7
ICBM133	CRMO_v_128	complete	Redundant	46867	5.21e-115	47194	8.44e-115	No	NA	DTR (short)	T7
ICBM134	CRMO_v_129	complete	Redundant	33761	4.62e-125	34088	7.68e-286	No	NA	DTR (short)	T7
ICBM135	CRMO_v_130	partial	Redundant	2997	2.97e-65	3325	5.13e-122	No	NA	DTR (short)	T7
ICBM137	CRMO_v_132	complete	Redundant	50093	8.08e-39	50421	3.16e-55	No	NA	DTR (short)	T7
ICBM138	CRMO_v_133	complete	Redundant	23566	5.24e-103	23871	3.90e-64	No	NA	DTR (short)	T7
ICBM139	CRMO_v_134	complete	Redundant	18415	3.04e-45	18743	5.53e-52	No	NA	DTR (short)	T7
ICBM143	CRMO_v_138	complete	Redundant	11387	1.46e-80	11715	3.05e-55	No	NA	DTR (short)	T7
ICBM145	CRMO_v_140	complete	Redundant	2095	4.14e-61	2424	2.11e-66	No	NA	DTR (short)	T7
ICBM146	CRMO_v_141	complete	Redundant	32999	1.30e-60	33327	3.26e-149	No	NA	DTR (short)	T7
ICBM147	CRMO_v_142	complete	Redundant	50614	2.19e-71	50942	9.35e-64	No	NA	DTR (short)	T7
ICBM154	CRMO_v_149	complete	Redundant	42422	1.83e-86	42750	8.30e-55	No	NA	DTR (short)	T7

b) Supplementary figures

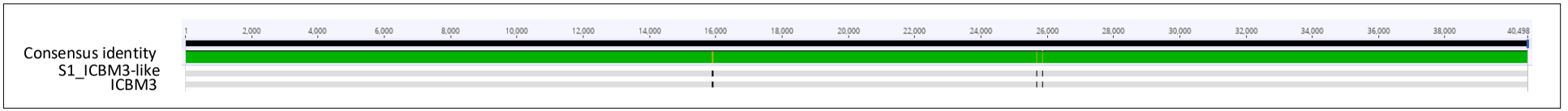
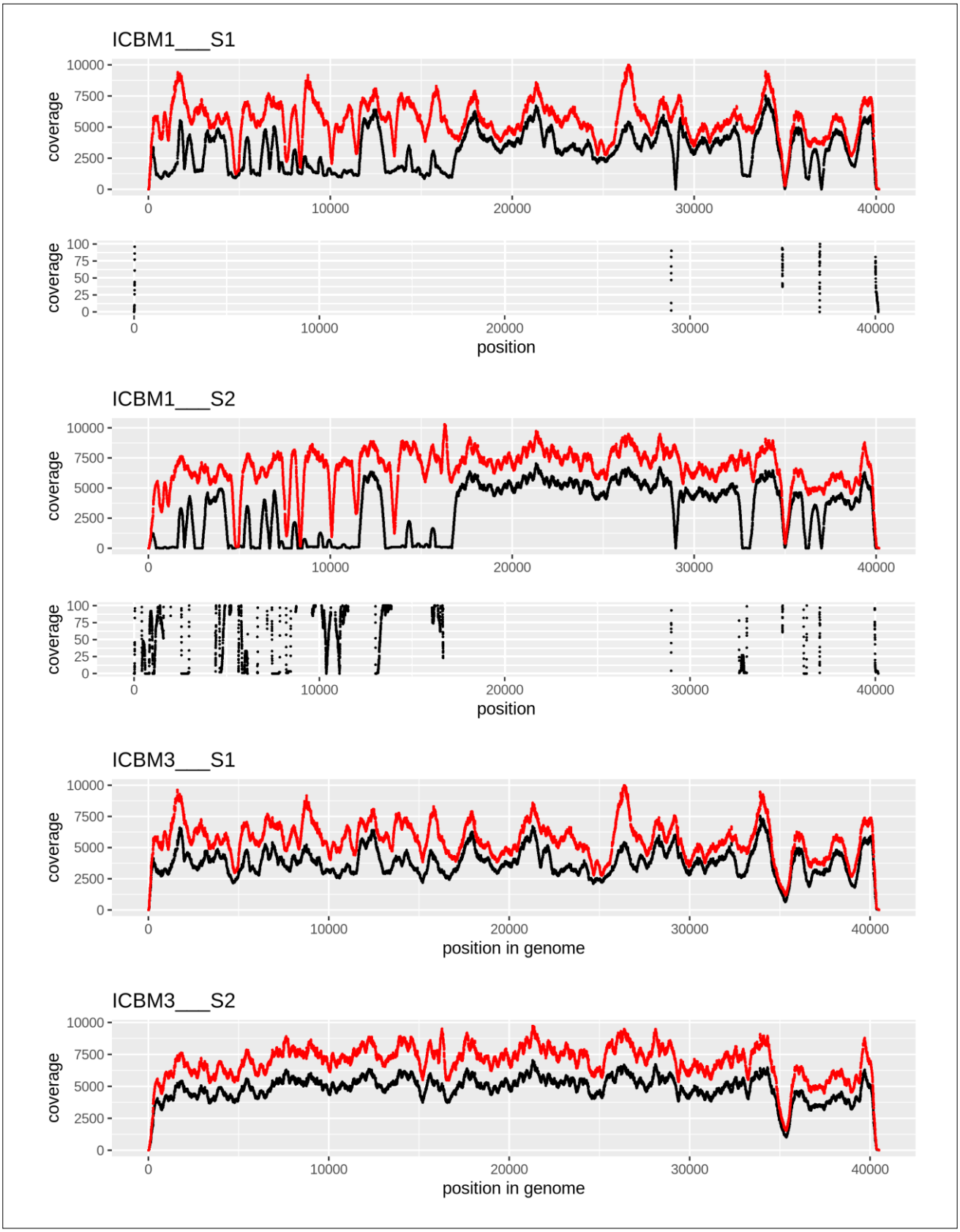


Fig. S 1: Pairwise genome comparison between ICBM3 and the ICBM3-like genome assembled from the S1 enrichment. In the alignment field, the differences are marked in black. In the consensus identity field, the presence of mismatches is signaled in orange.


ICBM1___S1
coverage
position
ICBM1___S2
coverage
position
ICBM3___S1
coverage
position in genome
ICBM3___S2
coverage
position in genome

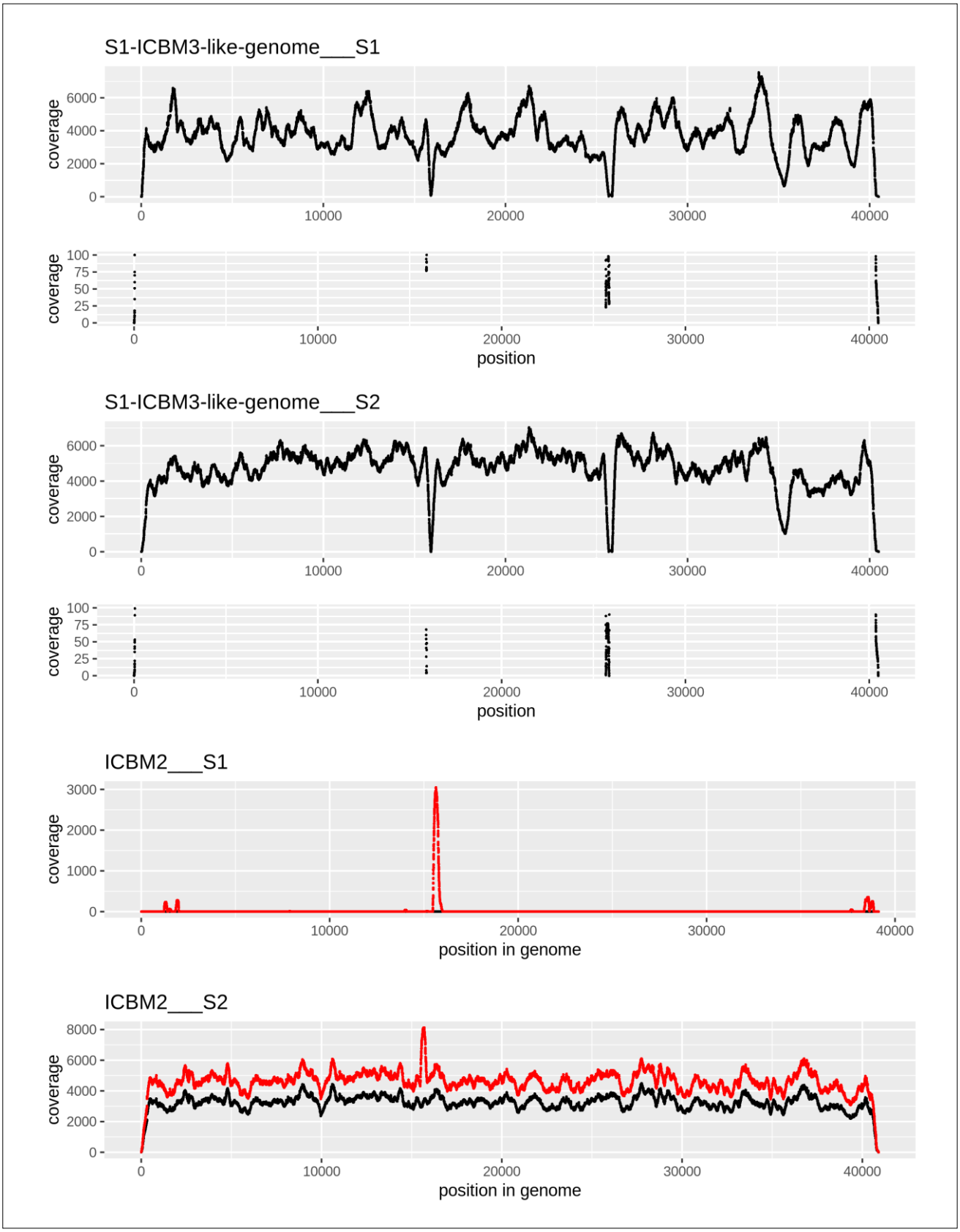


Fig. S 2: Plots showing the read coverage along the phage genomes for the S1 and S2 phage enrichments. In black– coverage of reads with 100% identity. In red – coverage of reads with >95% identity. The small differences between ICBM3 and the ICBM3-like genome from S1 enrichment could be due to sequencing errors, as the coverage in the variable regions dropped sharply for the genome retrieved from S1, but not for ICBM3 (see Fig. S1).

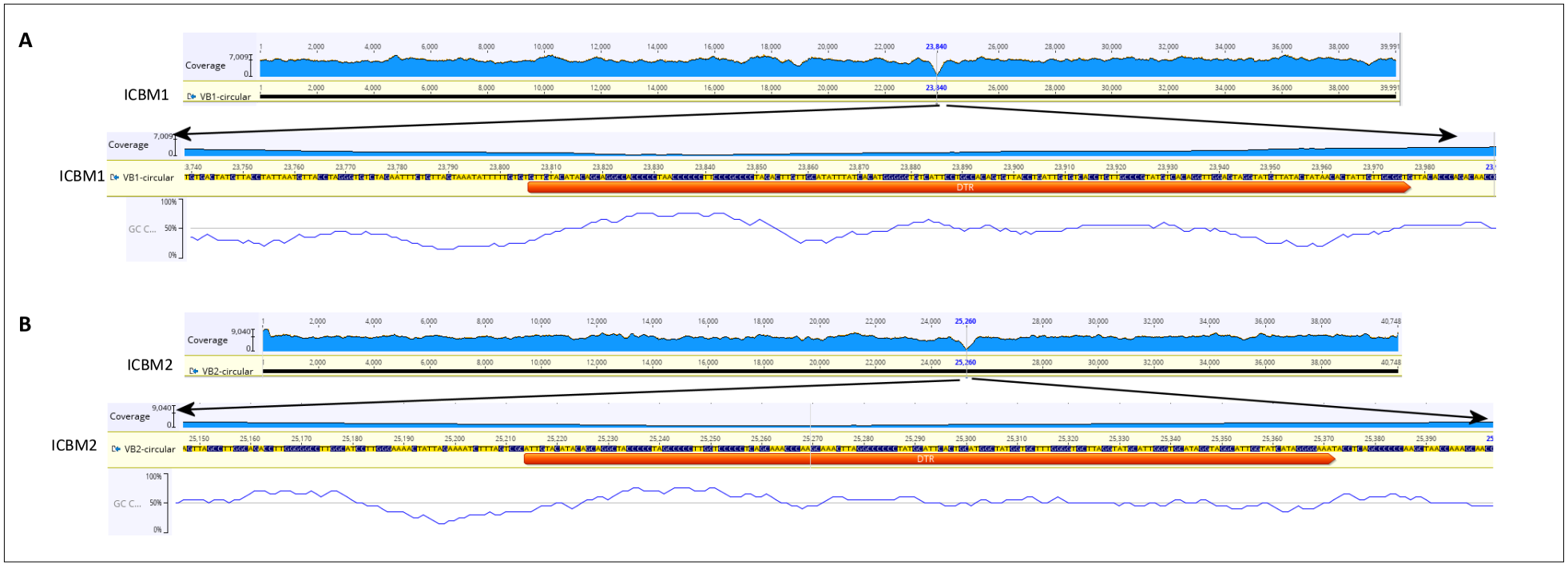


Fig. S 3: Illumina short read sequencing, distribution of reads along the assembled contigs for ICBM1 (A) and ICBM2 (B) phages. The position of the drop in coverage is indicated in blue numbers. The magnifications show the region corresponding to the drop in coverage, with the corresponding G+C content graph (blue line) and the DTR positions (as determined from PacBio read coverage) marked in red.

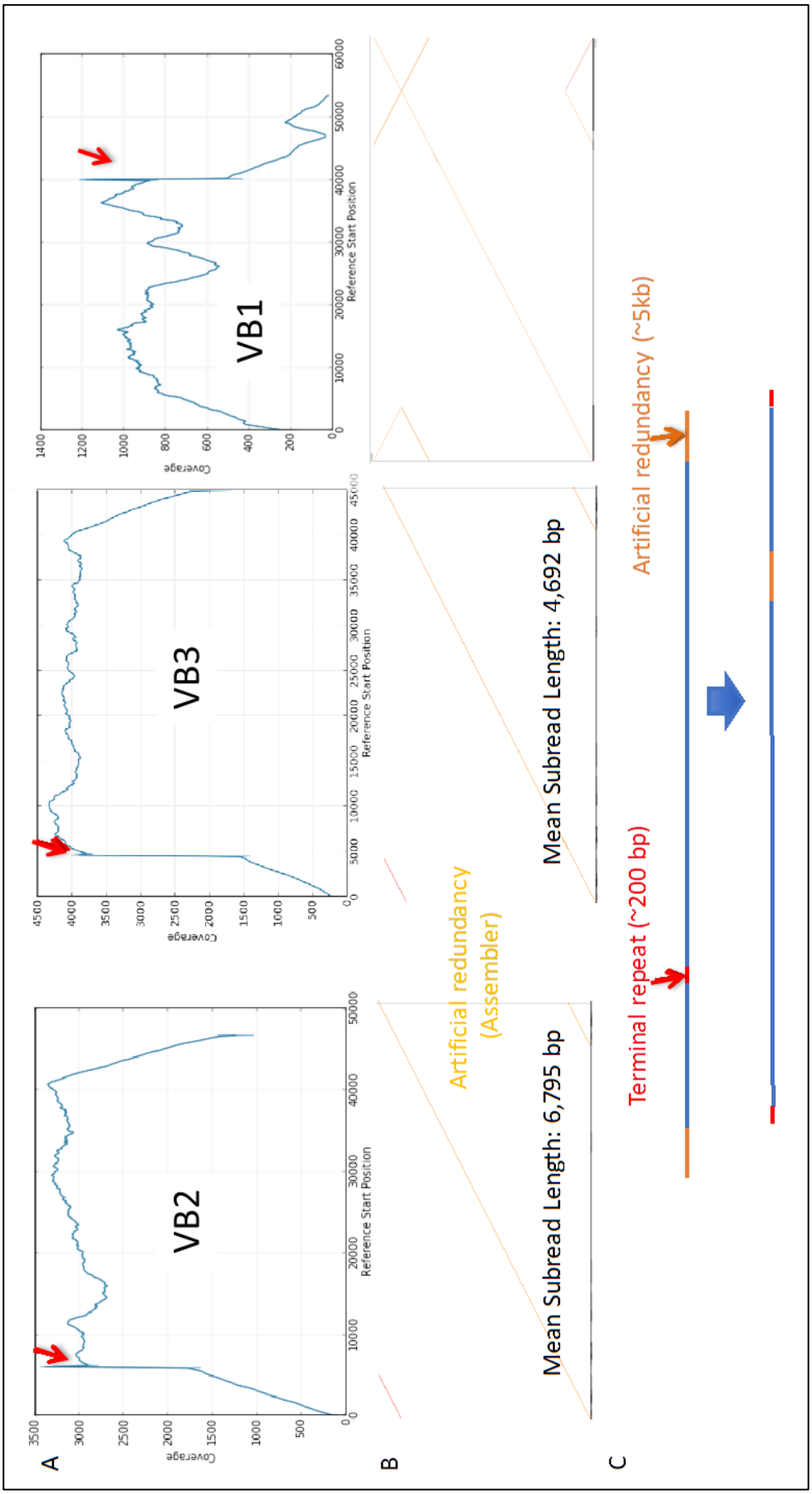


Fig. S 4: A. Coverage plots (window = 100 bp) of long read genome assemblies of both phages ICBM1 and ICBM2 contained within enrichment. Distinct coverage spikes of about 100 bp can be observed at position 5kb (terminal redundancies, red arrows). **B.** Artificial redundancies at the end of the contigs (shown here as dotplots) are produced by the genome assembler usually at the size of the mean subread length suggesting a circular genome structure. **C.** Delineation of the final phage genome structure applying the "linearization" process. The artificial redundancy (yellow) is removed and the terminal repeat (red) placed twice at the end of the phage genomes.

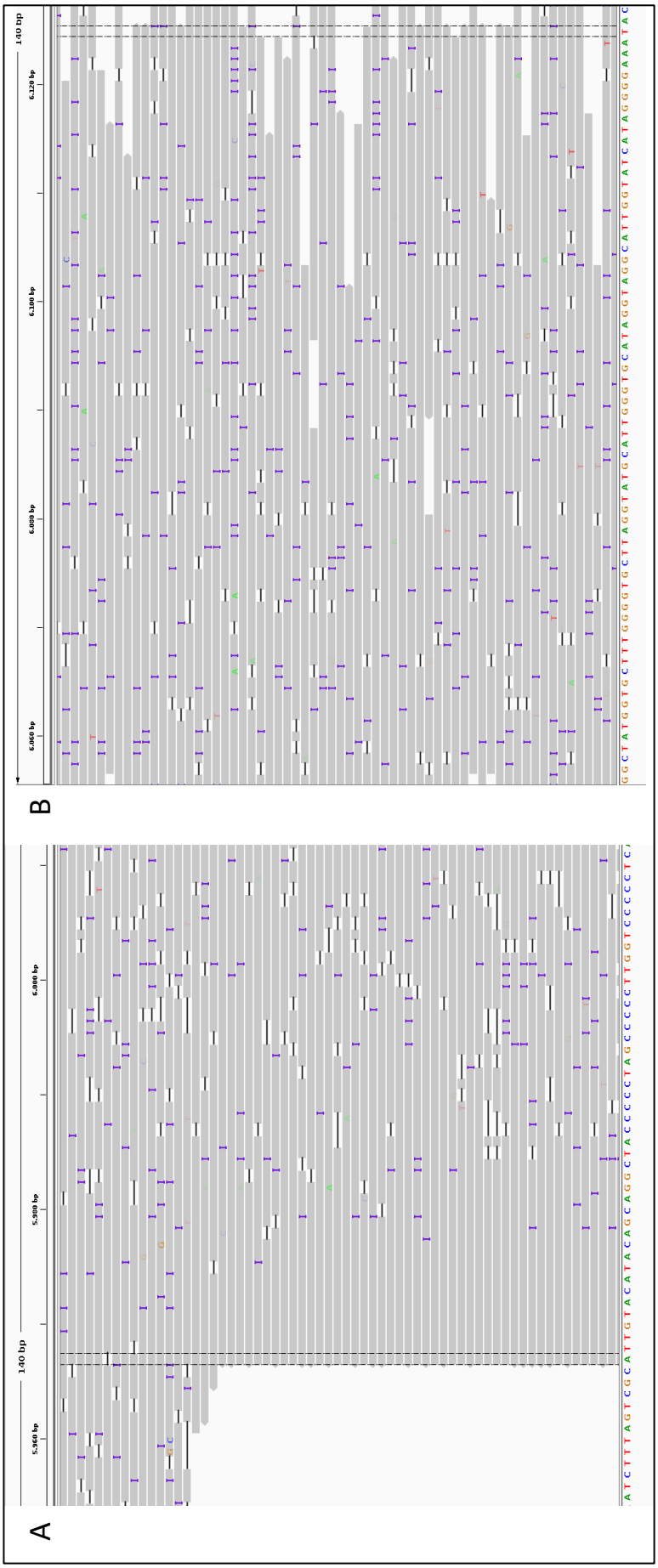


Fig. S 5: Distinct start **(A)** and end positions **(B)** of a linear phage genome as retrieved from a visualization of a long read genome mapping by IGV.

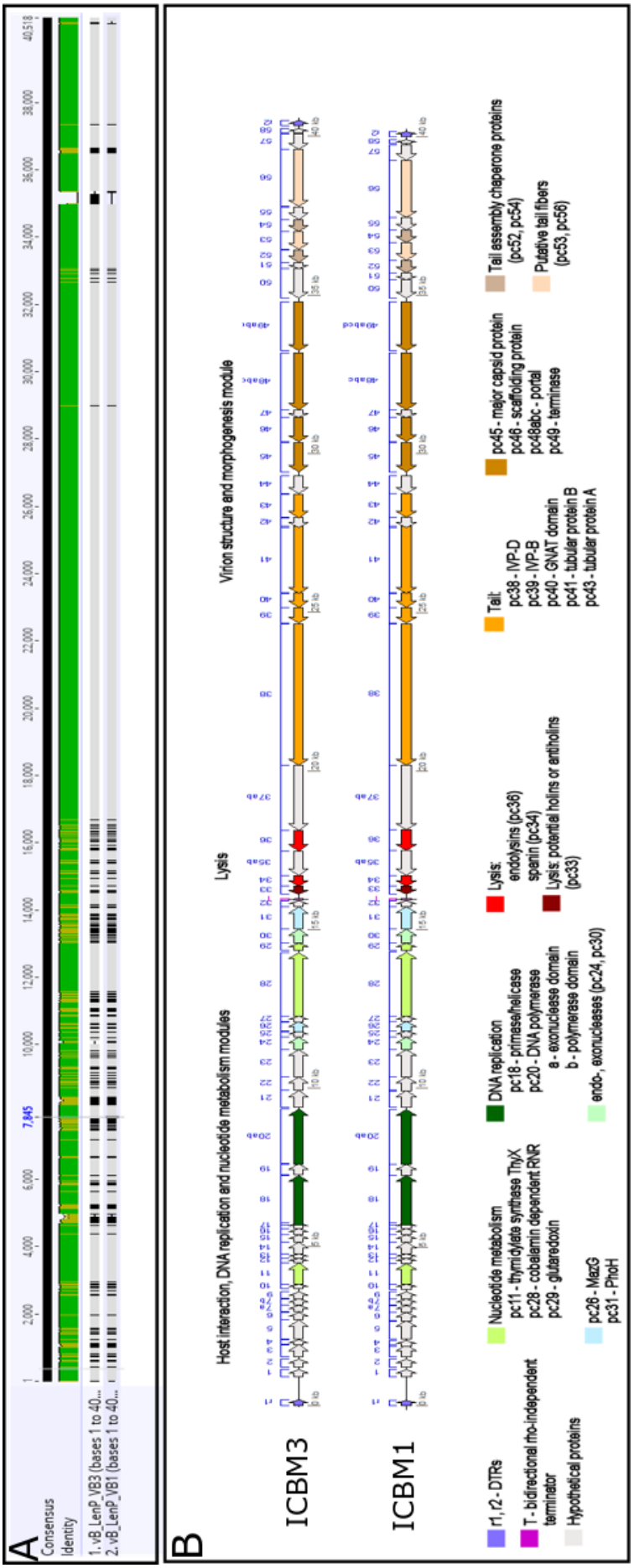


Fig. S 6: Genome comparison of ICBM1 and ICBM3 phages: **(A)** showing the location of mismatches and deletions and **(B)** showing the predicted genes.

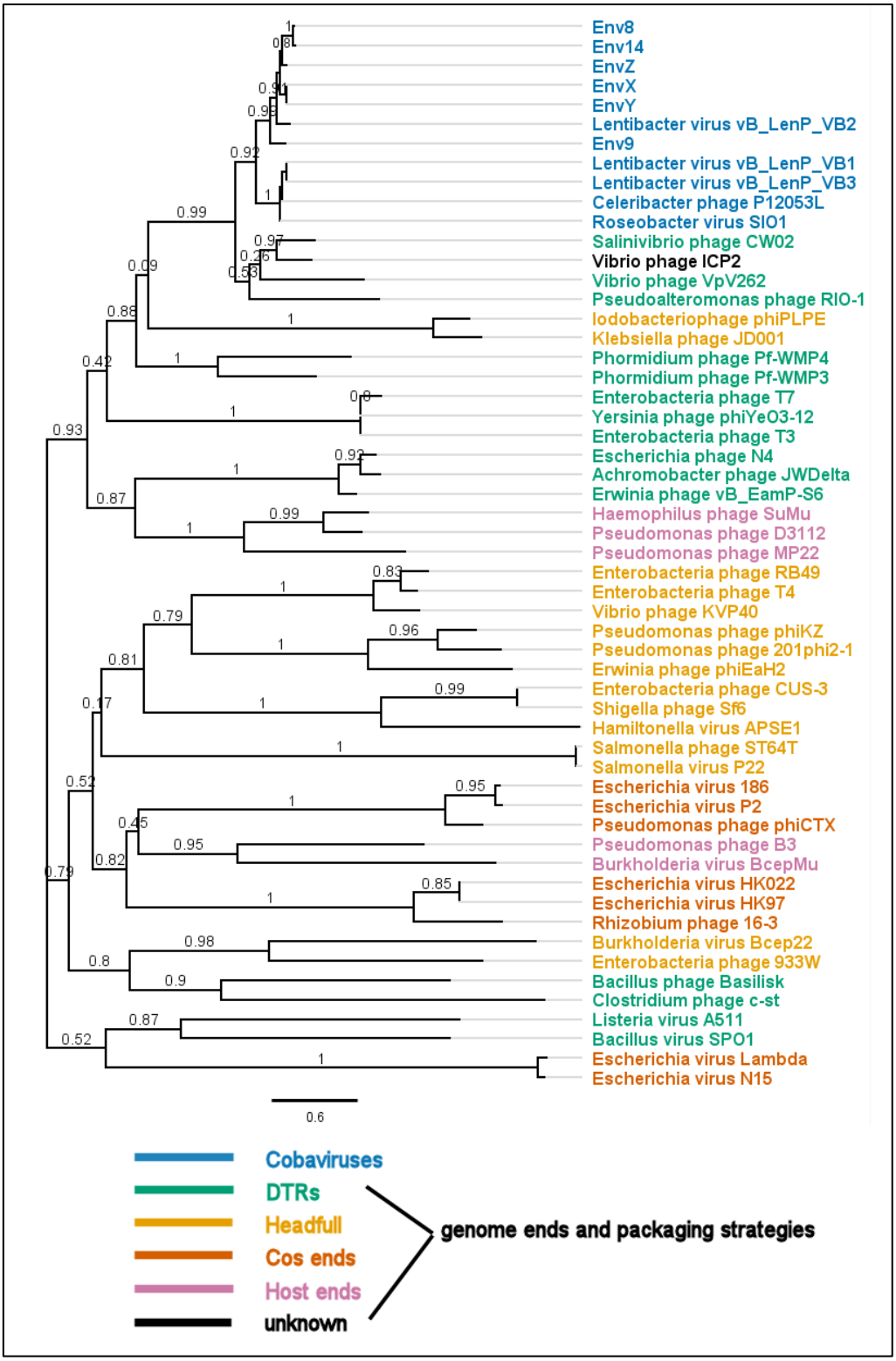


Fig. S 7: Phylogenetic analysis of the Terminase large subunit from cobaviruses and other phages with known genome ends and packaging strategies. The evolutionary history was inferred using the approximately-maximum-likelihood method implemented in FastTree 2.1.5. The node labels represent Fast Tree support values. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. The tree is unrooted.

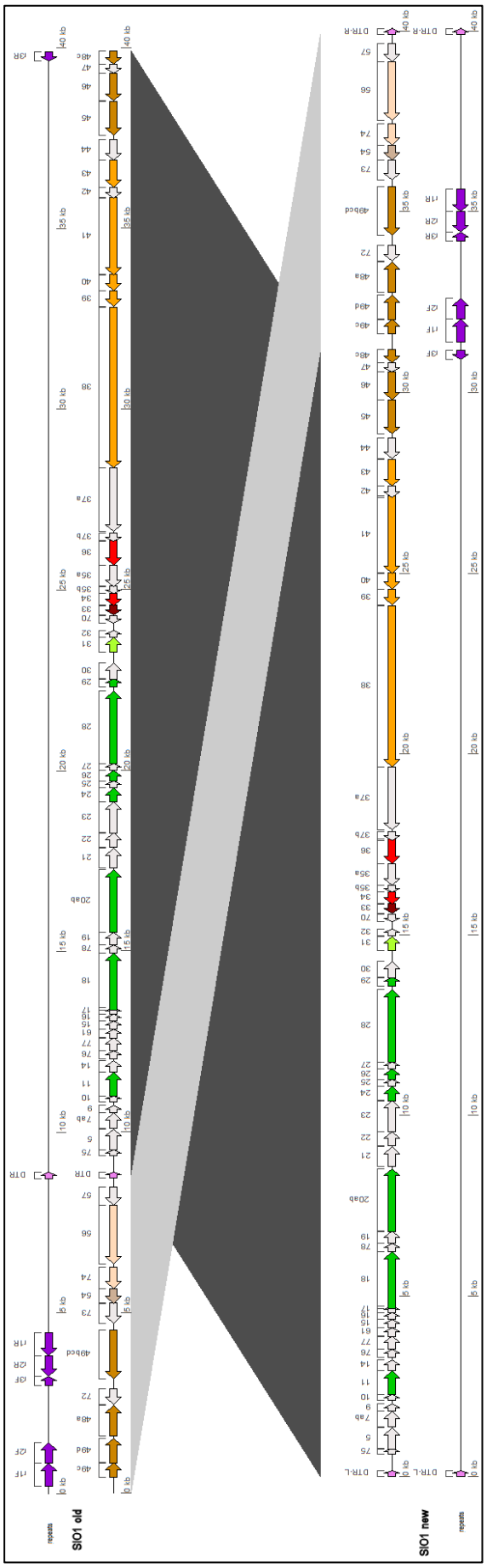


Fig. S 8: Genomic maps of the SIO1 phage, before and after genome end correction. Purple arrows: position of the inverted repeats. Pink arrows: position of DTRs.

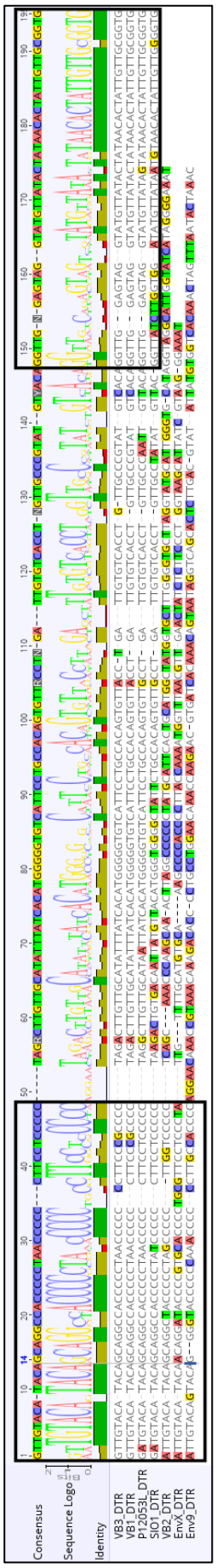


Fig. S 9: Multiple alignments of genomic ends from all cobaviruses with complete genome.

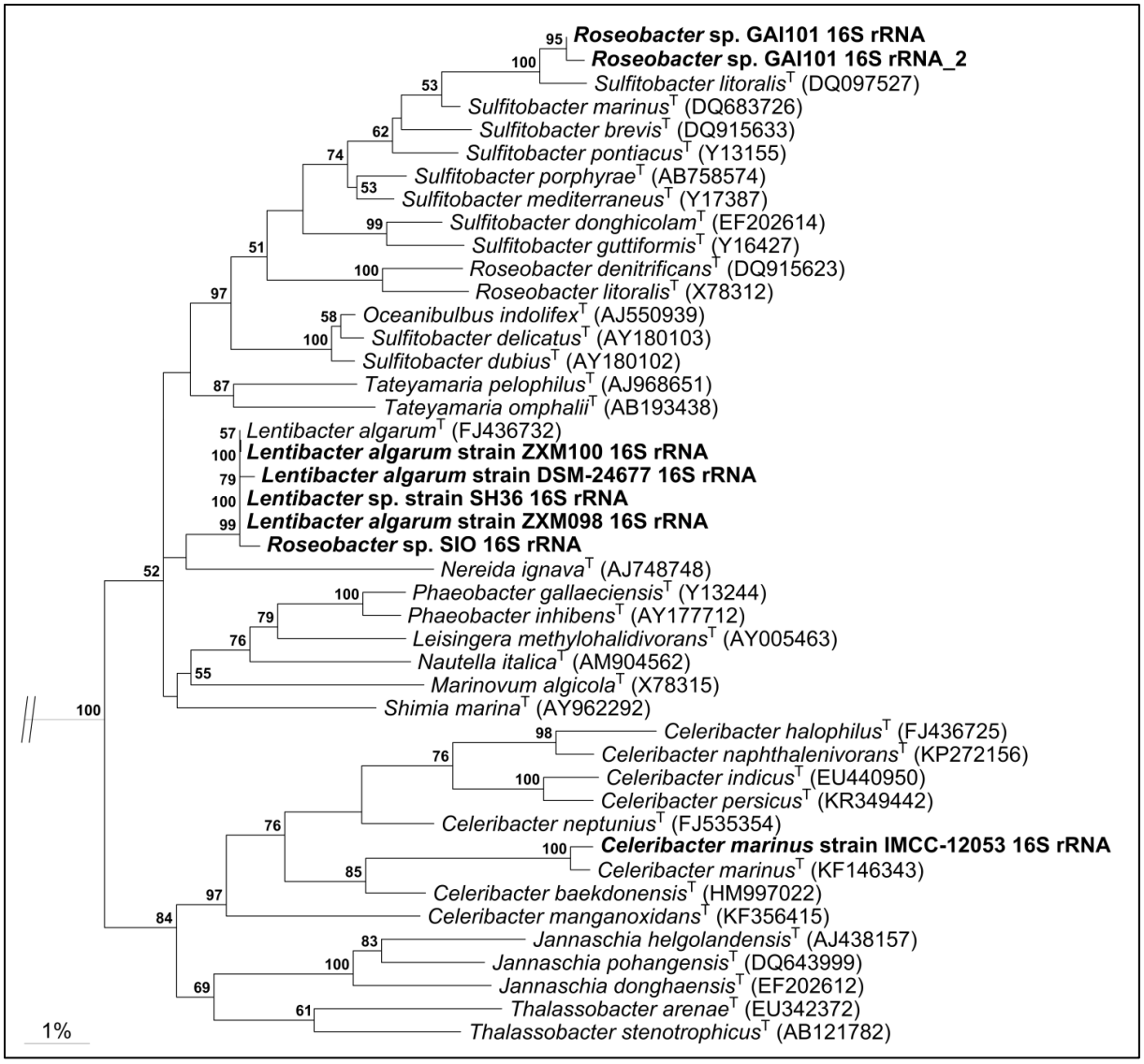


Fig. S 10: Neighbour joining tree based on 16S rRNA gene similarity showing the phylogenetic affiliation of bacterial hosts analysed in this study (bold) within the *Rhodobacteraceae*. Sequences of type material (>1300bp) were used to construct the backbone tree. Only bootstrap values ≥50% (derived from 1500 replicates) are shown. Selected sequences related to *Gammaproteobacteria* were used as outgroup to define the root of the tree (not shown). GenBank accession numbers are given in parentheses. Scale bar indicates percentage of sequence divergence.

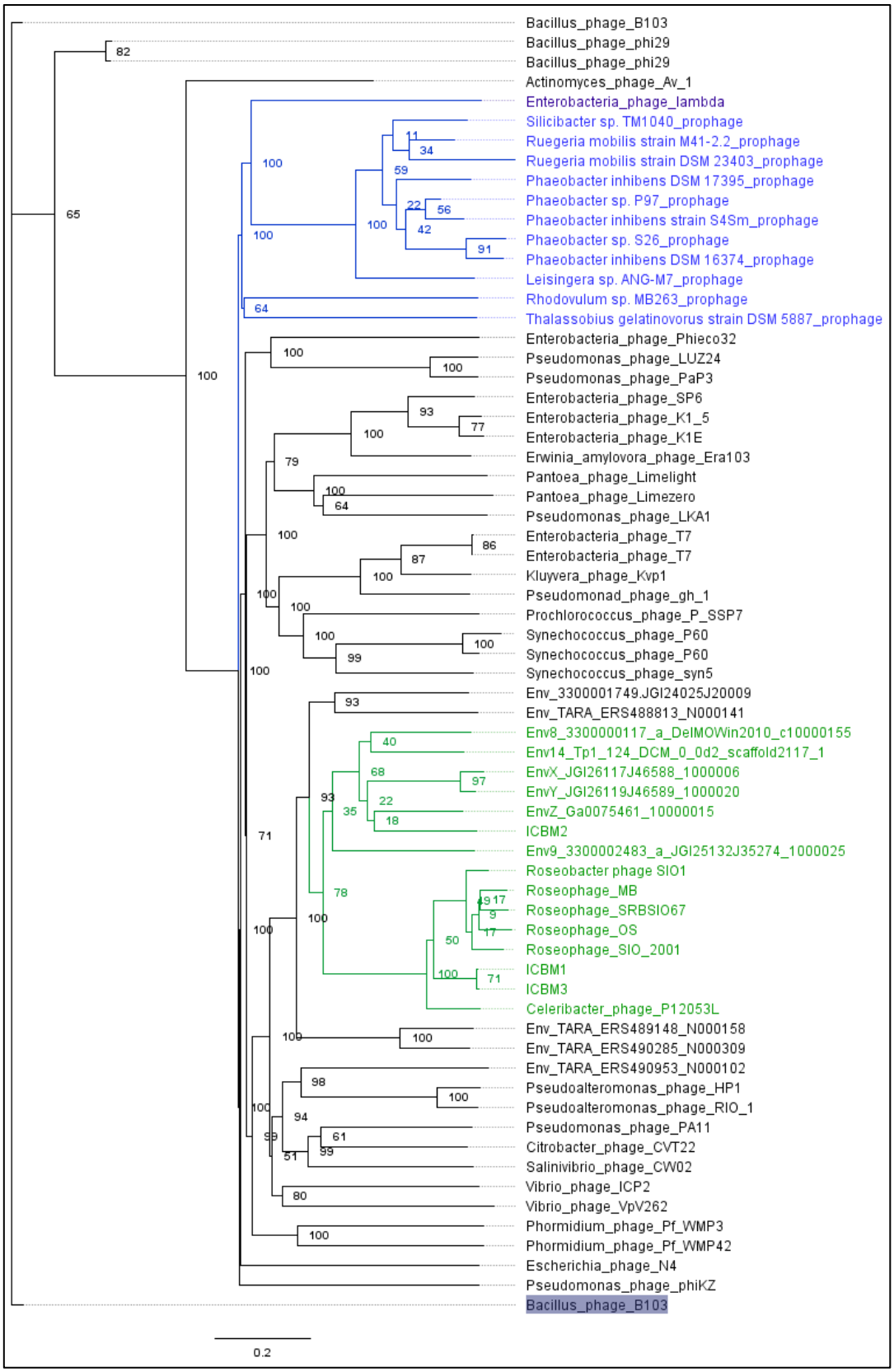


Fig. S 11: Phylogenetic positioning of the cobaviruses (green font) and spanin containing prophages from *Rhodobacteraceae* (blue font). The whole-genome-based phylogeny was inferred using the Genome-BLAST Distance Phylogeny method implemented in the VICTOR web service, using the amino acid data.

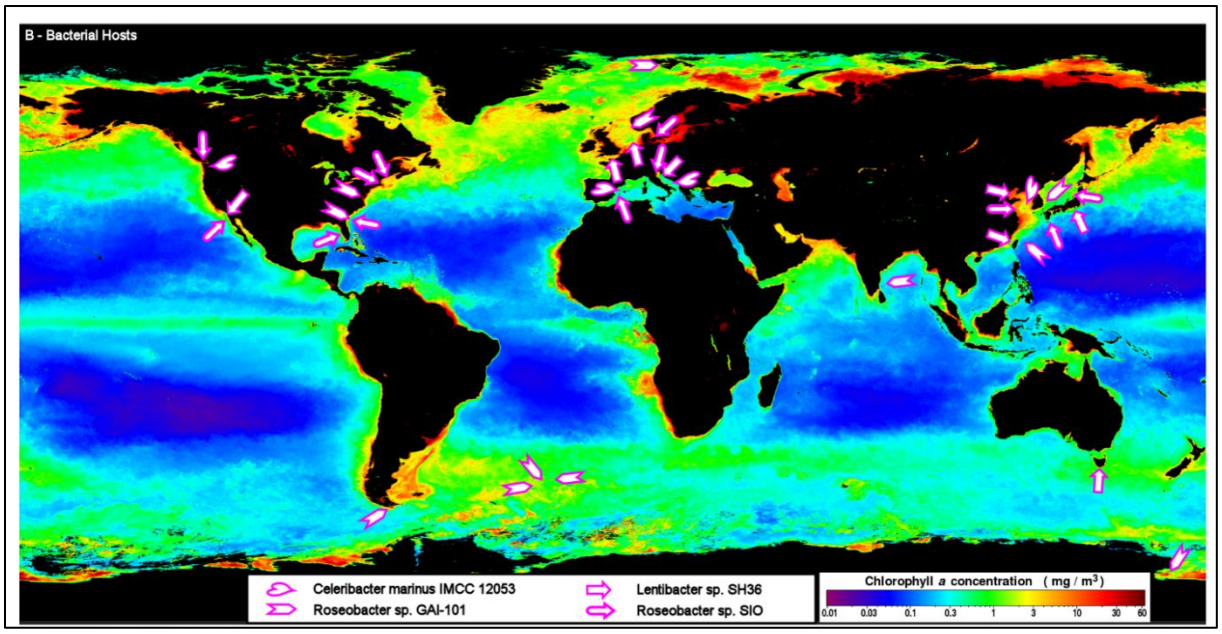


Fig. S 12: Locations in which cobavirus hosts were found based on a 16S rRNA survey in the NR Blast database and in the Tara Ocean samples.

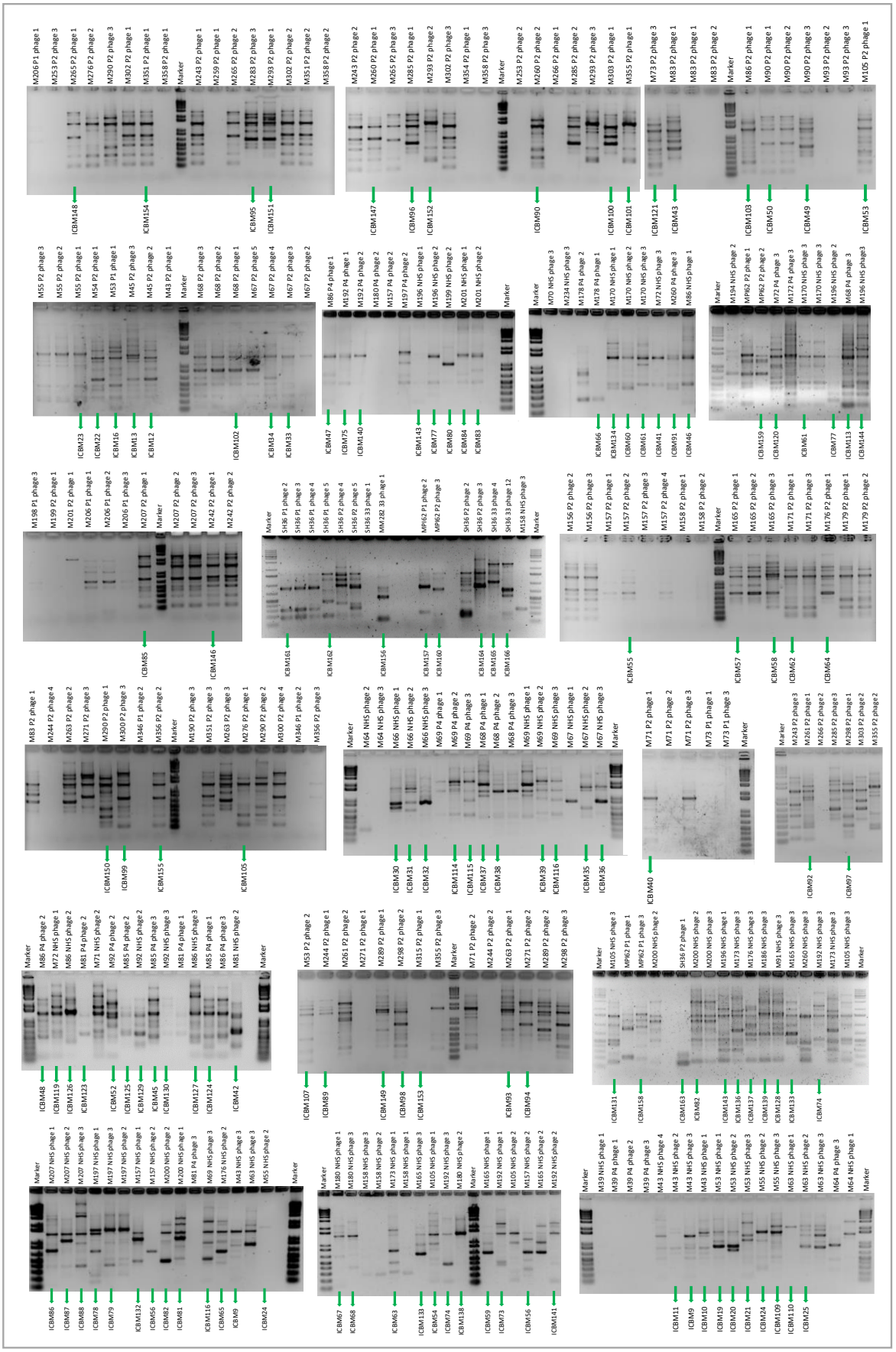


Fig. S 13: Gel electrophoresis of RAPD-PCR products for all phage isolates. For each unique pattern, one isolate was chosen to be genome sequenced (green arrows). Marker 1kb Plus DNA ladder (Invitrogen™).

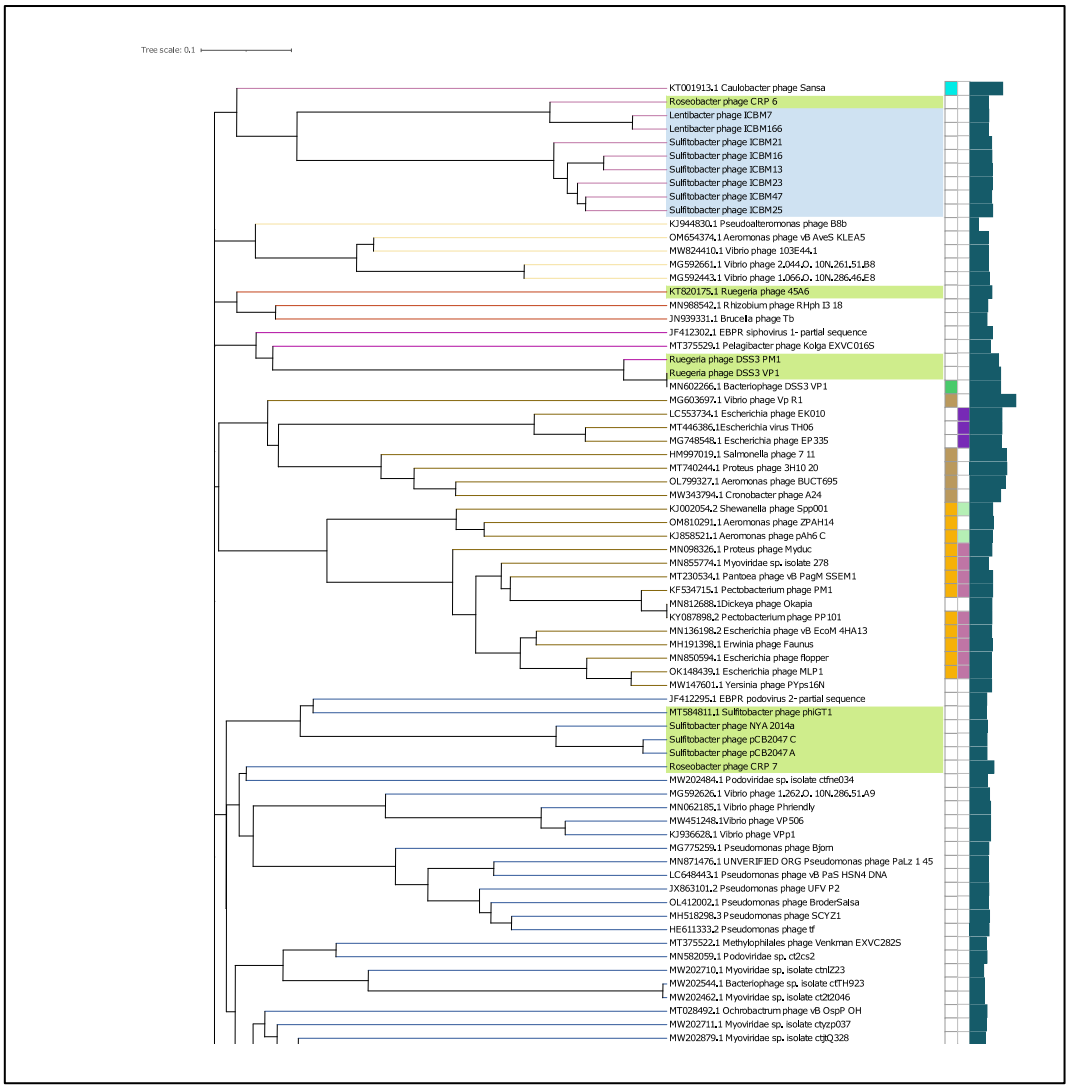


Fig. S 14a-i: Whole-genome based proteomic tree of 965 dsDNA phages. The tree is split into several figures, with overlaps. Names of cultivated roseophages are marked in green (previous) and blue (this study). Color-strips indicate affiliation to ICTV-recognized families and subfamilies, respectively. Genome lengths are displayed as bar chart. Colored branches indicate viral genome clusters (VGCs).

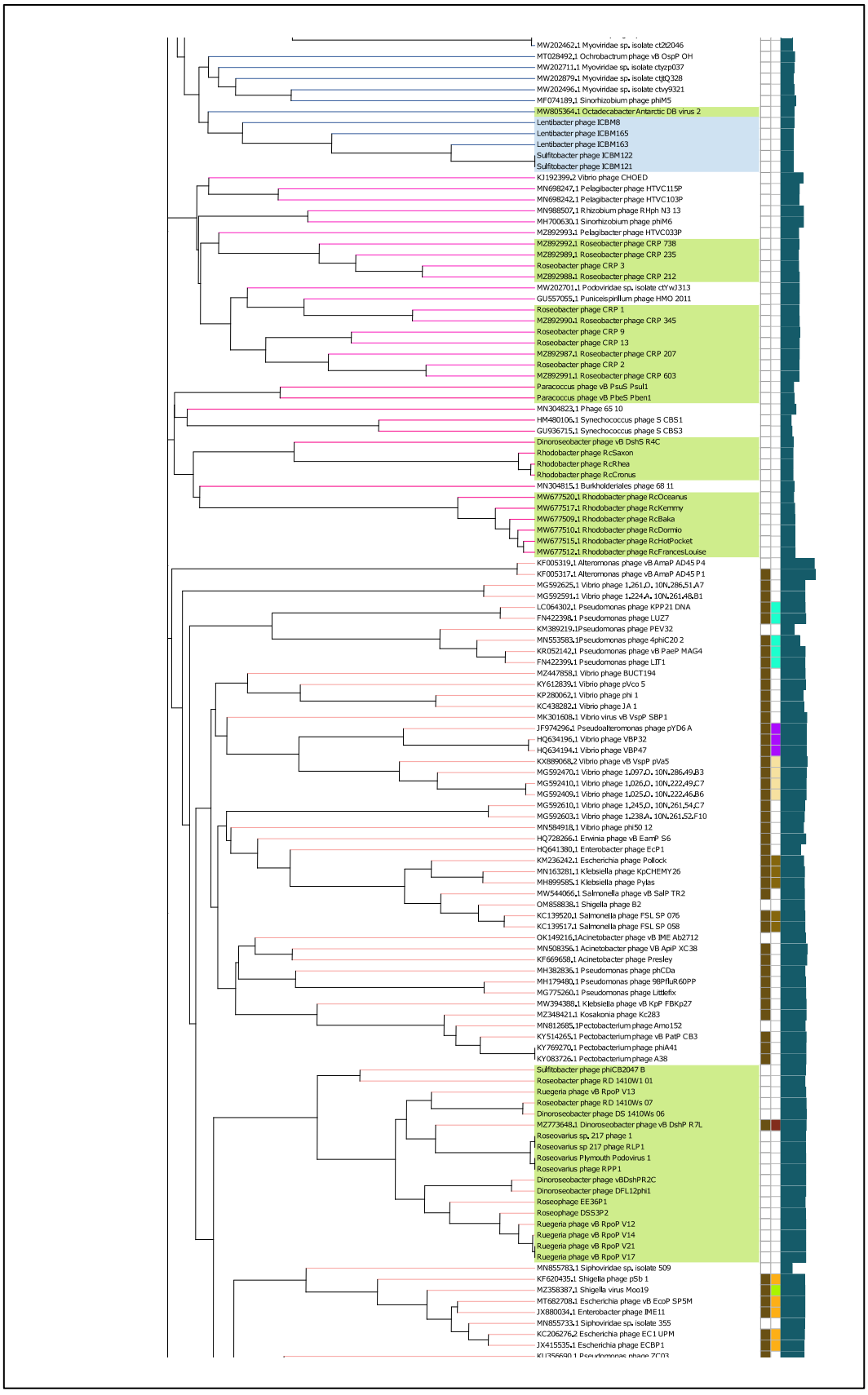

MW202462.1 Myoviridae sp. isolate ct2t2046
MT028492.1 Ochrobactrum phage vB OspP OH
MW202711.1 Myoviridae sp. isolate ctyzp037
MW202879.1 Myoviridae sp. isolate ctjtQ328
MW202496.1 Myoviridae sp. isolate ctvy9321
MF074189.1 Sinorhizobium phage phiM5
MW805364.1 Octadecabacter Antarctic DB virus 2
Lentibacter phage ICBM8
Lentibacter phage ICBM165
Lentibacter phage ICBM163
Sulfitobacter phage ICBM122
Sulfitobacter phage ICBM121
KJ192399.2 Vibrio phage CHOED
MN698247.1 Pelagibacter phage HTVC115P
MN698242.1 Pelagibacter phage HTVC103P
MN988507.1 Rhizobium phage RHph N3 13
MH700630.1 Sinorhizobium phage phiM6
MZ892993.1 Pelagibacter phage HTVC033P
MZ892992.1 Roseobacter phage CRP 738
MZ892989.1 Roseobacter phage CRP 235
Roseobacter phage CRP 3
MZ892988.1 Roseobacter phage CRP 212
MW202701.1 Podoviridae sp. isolate ctYwJ313
GU557055.1 Puniceispirillum phage HMO 2011
Roseobacter phage CRP 1
MZ892990.1 Roseobacter phage CRP 345
Roseobacter phage CRP 9
Roseobacter phage CRP 13
MZ892987.1 Roseobacter phage CRP 207
Roseobacter phage CRP 2
MZ892991.1 Roseobacter phage CRP 603
Paracoccus phage vB PsuS Psul1
Paracoccus phage vB PbeS Pben1
MN304823.1 Phage 65 10
HM480106.1 Synechococcus phage S CBS1
GU936715.1 Synechococcus phage S CBS3
Dinoroseobacter phage vB DshS R4C
Rhodobacter phage RcSaxon
Rhodobacter phage RcRhea
Rhodobacter phage RcCronus
MN304815.1 Burkholderiales phage 68 11
MW677520.1 Rhodobacter phage RcOceanus
MW677517.1 Rhodobacter phage RcKemmy
MW677509.1 Rhodobacter phage RcBaka
MW677510.1 Rhodobacter phage RcDormio
MW677515.1 Rhodobacter phage RcHotPocket
MW677512.1 Rhodobacter phage RcFrancesLouise
KF005319.1 Alteromonas phage vB AmeP AD45 P4
KF005317.1 Alteromonas phage vB AmeP AD45 P1
MG592625.1 Vibrio phage 1.261.O. 10N.286.51.A7
MG592591.1 Vibrio phage 1.224.A. 10N.261.48.B1
LC064302.1 Pseudomonas phage KPP21 DNA
FN422398.1 Pseudomonas phage LUZ7
KM389219.1Pseudomonas phage PEV32
MN553583.1Pseudomonas phage 4phiC20 2
KR052142.1 Pseudomonas phage vB PaeP MAG4
FN422399.1 Pseudomonas phage LIT1
MZ447858.1 Vibrio phage BUCT194
KY612839.1 Vibrio phage pVco 5
KP280062.1 Vibrio phage phi 1
KC438282.1 Vibrio phage JA 1
MK301608.1 Vibrio virus vB VspP SBP1
JF974296.1 Pseudoalteromonas phage pYD6 A
HQ634196.1 Vibrio phage VBP32
HQ634194.1 Vibrio phage VBP47
KX889068.2 Vibrio phage vB VspP pVa5
MG592470.1 Vibrio phage 1.097.O. 10N.286.49.B3
MG592410.1 Vibrio phage 1.026.O. 10N.222.49.C7
MG592409.1 Vibrio phage 1.025.O. 10N.222.46.B6
MG592610.1 Vibrio phage 1.245.O. 10N.261.54.C7
MG592603.1 Vibrio phage 1.238.A. 10N.261.52.F10
MN584918.1 Vibrio phage phi50 12
HQ728266.1 Erwinia phage vB EamP S6
HQ641380.1 Enterobacter phage EcP1
KM236242.1 Escherichia phage Pollock
MN163281.1 Klebsiella phage KpCHEMY26
MH899585.1 Klebsiella phage Pylas
MW544066.1 Salmonella phage vB SalP TR2
OM858838.1 Shigella phage B2
KC139520.1 Salmonella phage FSL SP 076
KC139517.1 Salmonella phage FSL SP 058
OK149216.1Acinetobacter phage vB IME Ab2712
MN508356.1 Acinetobacter phage VB ApiP XC38
KF669658.1 Acinetobacter phage Presley
MH382836.1 Pseudomonas phage phCDa
MH179480.1 Pseudomonas phage 98PfluR60PP
MG775260.1 Pseudomonas phage Littlefix
MW394388.1 Klebsiella phage vB KpP FBKp27
MZ348421.1 Kosakonia phage Kc283
MN812685.1Pectobacterium phage Amo152
KY514265.1 Pectobacterium phage vB PatP CB3
KY769270.1 Pectobacterium phage phiA41
KY083726.1 Pectobacterium phage A38
Sulfitobacter phage phiCB2047 B
Roseobacter phage RD 1410W1 01
Ruegeria phage vB RpoP V13
Roseobacter phage RD 1410Ws 07
Dinoroseobacter phage DS 1410Ws 06
MZ773648.1 Dinoroseobacter phage vB DshP R7L
Roseovarius sp. 217 phage 1
Roseovarius sp 217 phage RLP1
Roseovarius Plymouth Podovirus 1
Roseovarius phage RPP1
Dinoroseobacter phage vBDshPR2C
Dinoroseobacter phage DFL12phi1
Roseophage EE36P1
Roseophage DSS3P2
Ruegeria phage vB RpoP V12
Ruegeria phage vB RpoP V14
Ruegeria phage vB RpoP V21
Ruegeria phage vB RpoP V17
MN855783.1 Siphoviridae sp. isolate 509
KF620435.1 Shigella phage pSb 1
MZ358387.1 Shigella virus Moo19
MT682708.1 Escherichia phage vB EcoP SP5M
JX880034.1 Enterobacter phage IME11
MN855733.1 Siphoviridae sp. isolate 355
KC206276.2 Escherichia phage EC1 UPM
JX415535.1 Escherichia phage ECBP1

Fig. S 14b

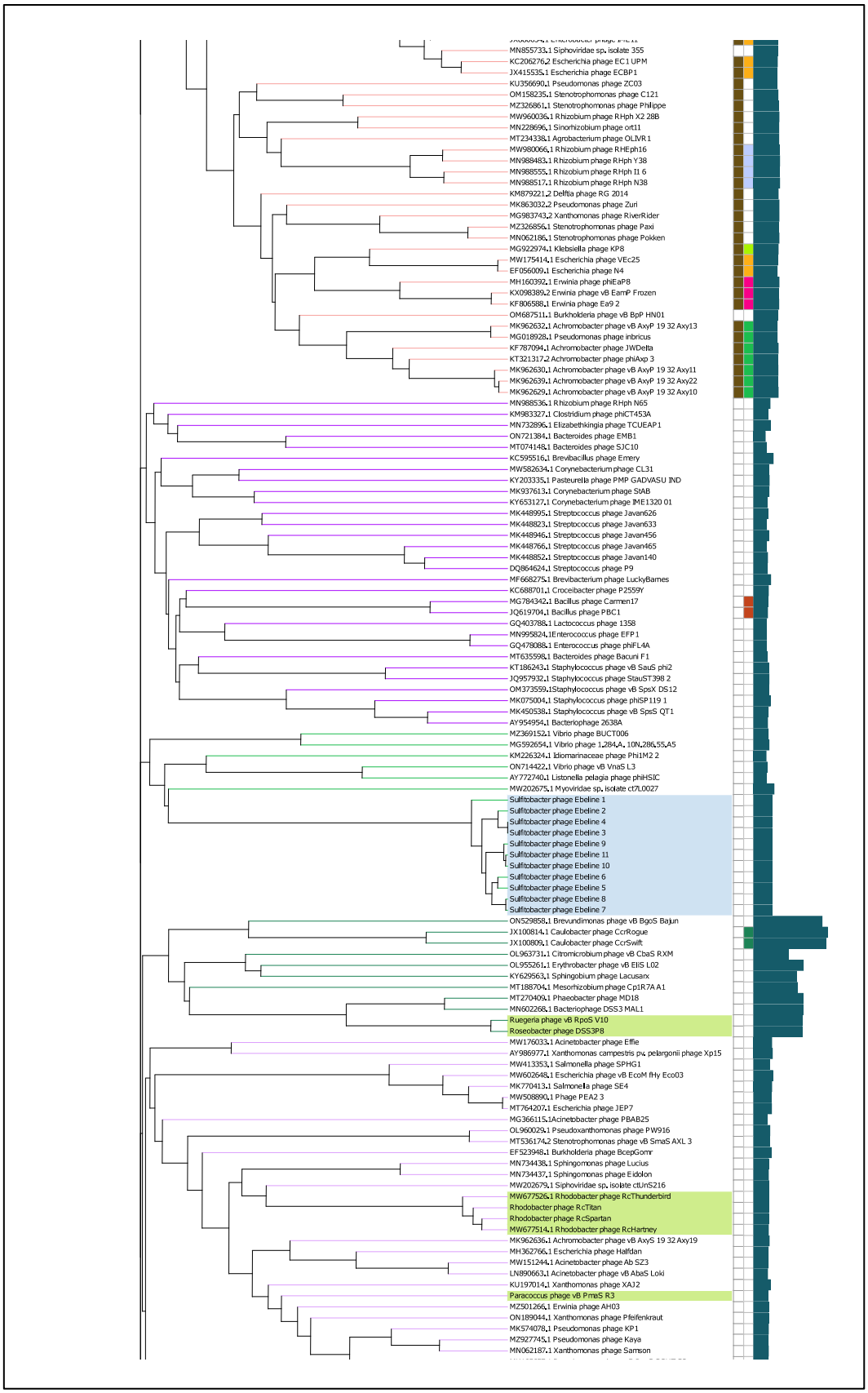

MN855733.1 Siphoviridae sp. isolate 355
KC206276.2 Escherichia phage EC1 UPM
JX415535.1 Escherichia phage ECBP1
KU356690.1 Pseudomonas phage ZC03
OM158235.1 Stenotrophomonas phage C121
MZ326861.1 Stenotrophomonas phage Philippe
MW960036.1 Rhizobium phage RHph X2 28B
MN228696.1 Sinorhizobium phage ort11
MT234338.1 Agrobacterium phage OLIVR1
MW980066.1 Rhizobium phage RHEph16
MN988483.1 Rhizobium phage RHph Y38
MN988555.1 Rhizobium phage RHph I1 6
MN988517.1 Rhizobium phage RHph N38
KM879221.2 Delftia phage RG 2014
MK863032.2 Pseudomonas phage Zuri
MG983743.2 Xanthomonas phage RiverRider
MZ326856.1 Stenotrophomonas phage Paxi
MN062186.1 Stenotrophomonas phage Pokken
MG922974.1 Klebsiella phage KP8
MW175414.1 Escherichia phage VEc25
EF056009.1 Escherichia phage N4
MH160392.1 Erwinia phage phiEaP8
KX098389.2 Erwinia phage vB EamP Frozen
KF806588.1 Erwinia phage Ea9 2
OM687511.1 Burkholderia phage vB BpP HN01
MK962632.1 Achromobacter phage vB AxyP 19 32 Axy13
MG018928.1 Pseudomonas phage inbricus
KF787094.1 Achromobacter phage JWDelta
KT321317.2 Achromobacter phage phiAxp 3
MK962630.1 Achromobacter phage vB AxyP 19 32 Axy11
MK962639.1 Achromobacter phage vB AxyP 19 32 Axy22
MK962629.1 Achromobacter phage vB AxyP 19 32 Axy10
MN988536.1 Rhizobium phage RHph N65
KM983327.1 Clostridium phage phiCT453A
MN732896.1 Elizabethkingia phage TCUEAP1
ON721384.1 Bacteroides phage EMB1
MT074148.1 Bacteroides phage SJC10
KC595516.1 Brevibacillus phage Emery
MW582634.1 Corynebacterium phage CL31
KY203335.1 Pasteurella phage PMP GADVASU IND
MK937613.1 Corynebacterium phage StAB
KY653127.1 Corynebacterium phage IME1320 01
MK448995.1 Streptococcus phage Javan626
MK448823.1 Streptococcus phage Javan633
MK448946.1 Streptococcus phage Javan456
MK448766.1 Streptococcus phage Javan465
MK448852.1 Streptococcus phage Javan140
DQ864624.1 Streptococcus phage P9
MF668275.1 Brevibacterium phage LuckyBarnes
KC688701.1 Crocelibacter phage P2559Y
MG784342.1 Bacillus phage Carmen17
JQ619704.1 Bacillus phage PBC1
GQ403788.1 Lactococcus phage 1358
MN995824.1Enterococcus phage EFP1
GQ478088.1 Enterococcus phage phiFL4A
MT635598.1 Bacteroides phage Bacuni F1
KT186243.1 Staphylococcus phage vB SauS phi2
JQ957932.1 Staphylococcus phage StauST398 2
OM373559.1Staphylococcus phage vB SpsX DS12
MK075004.1 Staphylococcus phage phiSP.119 1
MK450538.1 Staphylococcus phage vB SpsS QT1
AY954954.1 Bacteriophage 2638A
MZ369152.1 Vibrio phage BUCT006
MG592654.1 Vibrio phage 1.284.A. 10N.286.55.A5
KM226324.1 Idiomarinaceae phage Phi1M2 2
ON714422.1 Vibrio phage vB VnaS L3
AY772740.1 Listonella pelagia phage phiHSIC
MW202675.1 Myoviridae sp. isolate ct7L0027
Sulfitobacter phage Ebeline 1
Sulfitobacter phage Ebeline 2
Sulfitobacter phage Ebeline 4
Sulfitobacter phage Ebeline 3
Sulfitobacter phage Ebeline 9
Sulfitobacter phage Ebeline 11
Sulfitobacter phage Ebeline 10
Sulfitobacter phage Ebeline 6
Sulfitobacter phage Ebeline 5
Sulfitobacter phage Ebeline 8
Sulfitobacter phage Ebeline 7
ON529858.1 Brevundimonas phage vB BgoS Bajun
JX100814.1 Caulobacter phage CcrRogue
JX100809.1 Caulobacter phage CcrSwift
OL963731.1 Citromicrobium phage vB CbaS RXM
OL955261.1 Erythrobacter phage vB EliS L02
KY629563.1 Sphingobium phage Lacusarx
MT188704.1 Mesorhizobium phage Cp1R7A A1
MT270409.1 Phaeobacter phage MD18
MN602268.1 Bacteriophage DSS3 MAL1
Ruegeria phage vB RpoS V10
Roseobacter phage DSS3P8
MW176033.1 Acinetobacter phage Effie
AY986977.1 Xanthomonas campestris pv. pelargonii phage Xp15
MW413353.1 Salmonella phage SPHG1
MW602648.1 Escherichia phage vB EcoM fHy Eco03
MK770413.1 Salmonella phage SE4
MW508890.1 Phage PEA2 3
MT764207.1 Escherichia phage JEP7
MG366115.1Acinetobacter phage PBAB25
OL960029.1 Pseudoxanthomonas phage PW916
MT536174.2 Stenotrophomonas phage vB SmaS AXL 3
EF523948.1 Burkholderia phage BcepGomr
MN734438.1 Sphingomonas phage Lucius
MN734437.1 Sphingomonas phage Eidolon
MW202679.1 Siphoviridae sp. isolate ctUnS216
MW677526.1 Rhodobacter phage RcThunderbird
Rhodobacter phage RcTitan
Rhodobacter phage RcSpartan
MW677514.1 Rhodobacter phage RcHartney
MK962636.1 Achromobacter phage vB AxyS 19 32 Axy19
MH362766.1 Escherichia phage Halfdan
MW151244.1 Acinetobacter phage Ab SZ3
LN890663.1 Acinetobacter phage vB AbaS Loki
KU197014.1 Xanthomonas phage XAJ2
Paracoccus phage vB PmeS R3
MZ501266.1 Erwinia phage AH03
ON189044.1 Xanthomonas phage Pfeifenkraut
MK574078.1 Pseudomonas phage KP1
MZ927745.1 Pseudomonas phage Kaya
MN062187.1 Xanthomonas phage Samson

Fig. S 14c

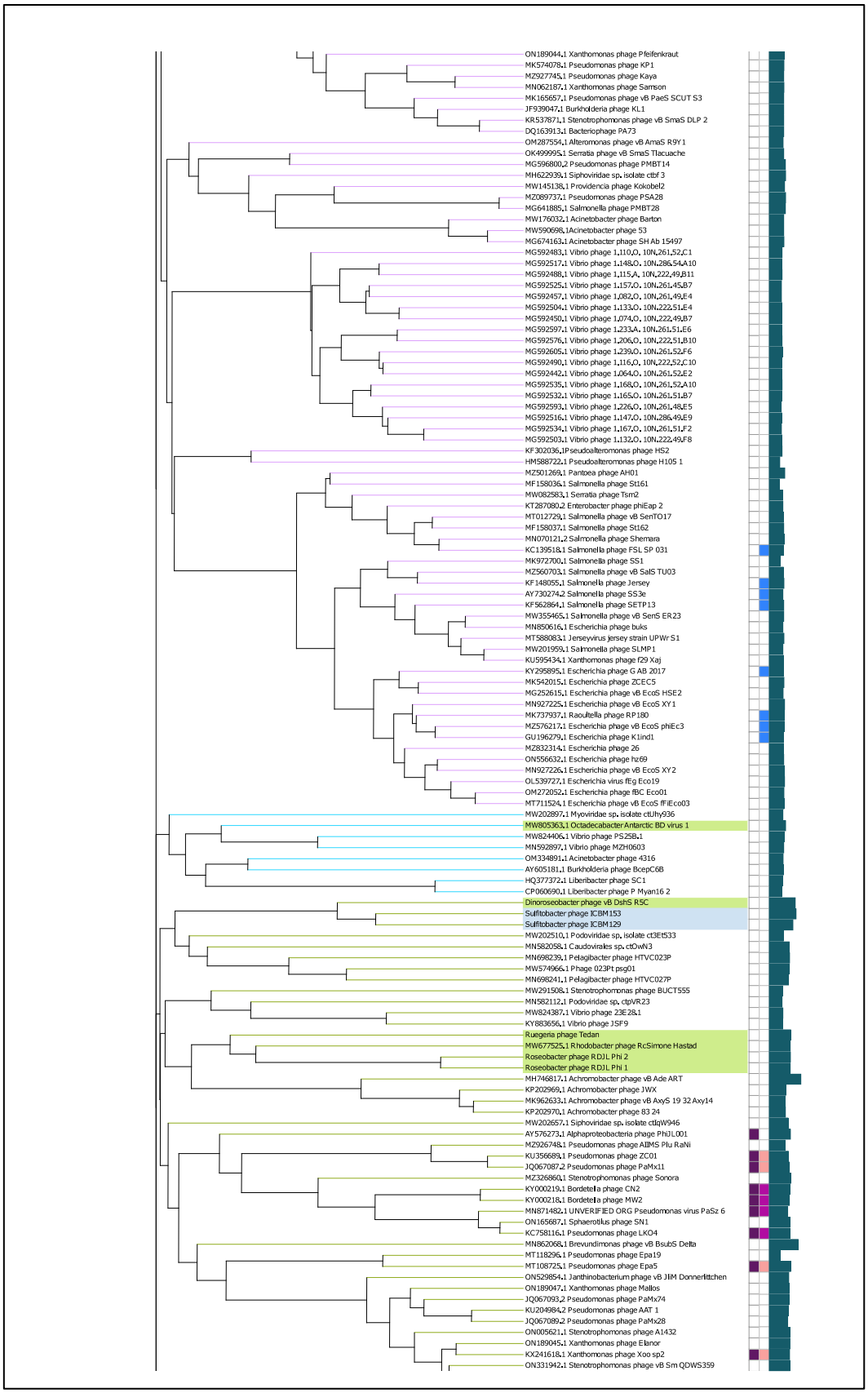

ON189044.1 Xanthomonas phage Pfeifenkraut
MK574078.1 Pseudomonas phage KP1
MZ927745.1 Pseudomonas phage Kaya
MN062187.1 Xanthomonas phage Samson
MK165657.1 Pseudomonas phage vB PaeS SCUT S3
JF939047.1 Burkholderia phage KL1
KR537871.1 Stenotrophomonas phage vB SmaS DLP 2
DQ163913.1 Bacteriophage PA73
OM287554.1 Alteromonas phage vB AmaS R9Y1
OK499995.1 Serratia phage vB SmaS Tlacuache
MG596800.2 Pseudomonas phage PMBT14
MH622939.1 Siphoviridae sp. isolate ctbf 3
MW145138.1 Providencia phage Kokobel2
MZ089737.1 Pseudomonas phage PSA28
MG641885.1 Salmonella phage PMBT28
MW176032.1 Acinetobacter phage Barton
MW590698.1Acinetobacter phage 53
MG674163.1 Acinetobacter phage SH Ab 15497
MG592483.1 Vibrio phage 1.110.O. 10N.261.52.C1
MG592517.1 Vibrio phage 1.148.O. 10N.286.54.A10
MG592488.1 Vibrio phage 1.115.A. 10N.222.49.B11
MG592525.1 Vibrio phage 1.157.O. 10N.261.45.B7
MG592457.1 Vibrio phage 1.082.O. 10N.261.49.E4
MG592504.1 Vibrio phage 1.133.O. 10N.222.51.E4
MG592450.1 Vibrio phage 1.074.O. 10N.222.49.B7
MG592597.1 Vibrio phage 1.233.A. 10N.261.51.E6
MG592576.1 Vibrio phage 1.206.O. 10N.222.51.B10
MG592605.1 Vibrio phage 1.239.O. 10N.261.52.F6
MG592490.1 Vibrio phage 1.116.O. 10N.222.52.C10
MG592442.1 Vibrio phage 1.064.O. 10N.261.52.E2
MG592535.1 Vibrio phage 1.168.O. 10N.261.52.A10
MG592532.1 Vibrio phage 1.165.O. 10N.261.51.B7
MG592593.1 Vibrio phage 1.226.O. 10N.261.48.E5
MG592516.1 Vibrio phage 1.147.O. 10N.286.49.E9
MG592534.1 Vibrio phage 1.167.O. 10N.261.51.F2
MG592503.1 Vibrio phage 1.132.O. 10N.222.49.F8
KF302036.1Pseudoalteromonas phage HS2
HM588722.1 Pseudoalteromonas phage H105 1
MZ501269.1 Pantoea phage AH01
MF158036.1 Salmonella phage St161
MW082583.1 Serratia phage Tsm2
KT287080.2 Enterobacter phage phiEap 2
MT012729.1 Salmonella phage vB SenTO17
MF158037.1 Salmonella phage St162
MN070121.2 Salmonella phage Shemara
KC13918.1 Salmonella phage FSL SP 031
MK972700.1 Salmonella phage SS1
MZ560703.1 Salmonella phage vB SalS TU03
KF148055.1 Salmonella phage Jersey
AY730274.2 Salmonella phage SS3e
KF562864.1 Salmonella phage SETP13
MW355465.1 Salmonella phage vB SenS ER23
MN850616.1 Escherichia phage buks
MT588083.1 Jerseyvirus jersey strain UPWr S1
MW201959.1 Salmonella phage SLMP1
KU595434.1 Xanthomonas phage f29 Xaj
KY295895.1 Escherichia phage G AB 2017
MK542015.1 Escherichia phage ZCEC5
MG252615.1 Escherichia phage vB EcoS HSE2
MN927225.1 Escherichia phage vB EcoS XY1
MK737937.1 Raoultella phage RP180
MZ576217.1 Escherichia phage vB EcoS phiEc3
GU196279.1 Escherichia phage K1ind1
MZ832314.1 Escherichia phage 26
ON556632.1 Escherichia phage hz69
MN927226.1 Escherichia phage vB EcoS XY2
OL539727.1 Escherichia virus fEg Eco19
OM272052.1 Escherichia phage fBC Eco01
MT711524.1 Escherichia phage vB EcoS fFiEco03
MW202897.1 Myoviridae sp. isolate ctUhy936
MW805363.1 Octadecabacter Antarctic BD virus 1
MW824406.1 Vibrio phage PS2SB.1
MN592897.1 Vibrio phage MZH0603
OM334891.1 Acinetobacter phage 4316
AY605181.1 Burkholderia phage BcepC6B
HQ377372.1 Liberibacter phage SC1
CP060690.1 Liberibacter phage P Myan16 2
Dinoroseobacter phage vB DshS R5C
Sulfitobacter phage ICBM153
Sulfitobacter phage ICBM129
MW202510.1 Podoviridae sp. isolate ct3Et533
MN582058.1 Caudovirales sp. ctOwN3
MN698239.1 Pelagibacter phage HTVC023P
MW574966.1 Phage 023Pt psg01
MN698241.1 Pelagibacter phage HTVC027P
MW291508.1 Stenotrophomonas phage BUCT555
MN582112.1 Podoviridae sp. ctpVR23
MW824387.1 Vibrio phage Z3E28.1
KY883656.1 Vibrio phage JSF9
Ruegeria phage Tedan
MW677525.1 Rhodobacter phage RcSimone Hastad
Roseobacter phage RDJL Phi 2
Roseobacter phage RDJL Phi 1
MH746817.1 Achromobacter phage vB Ade ART
KP202969.1 Achromobacter phage JWX
MK962633.1 Achromobacter phage vB AxyS 19 32 Axy14
KP202970.1 Achromobacter phage 83 24
MW202657.1 Siphoviridae sp. isolate ctIqW946
AY576273.1 Alphaproteobacteria phage PhiJL001
MZ926748.1 Pseudomonas phage AIIMS Plu RaNi
KU356689.1 Pseudomonas phage ZC01
JQ067087.2 Pseudomonas phage PaMx11
MZ326860.1 Stenotrophomonas phage Sonora
KY000219.1 Bordetella phage CN2
KY000218.1 Bordetella phage MW2
MN871482.1 UNVERIFIED ORG Pseudomonas virus PaSz 6
ON165687.1 Sphaerotilus phage SN1
KC758116.1 Pseudomonas phage LKO4
MN862068.1 Brevundimonas phage vB BsubS Delta
MT118296.1 Pseudomonas phage Epa19
MT108725.1 Pseudomonas phage Epa5
ON529854.1 Janthinobacterium phage vB JliM Donnerlittchen
ON189047.1 Xanthomonas phage Mallos
JQ067093.2 Pseudomonas phage PaMx74
KU204984.2 Pseudomonas phage AAT 1
JQ067089.2 Pseudomonas phage PaMx28
ON005621.1 Stenotrophomonas phage A1432
ON189045.1 Xanthomonas phage Elanor
KX241618.1 Xanthomonas phage Xoo sp2
ON331942.1 Stenotrophomonas phage vB Sm QDWS359

Fig. S 14d

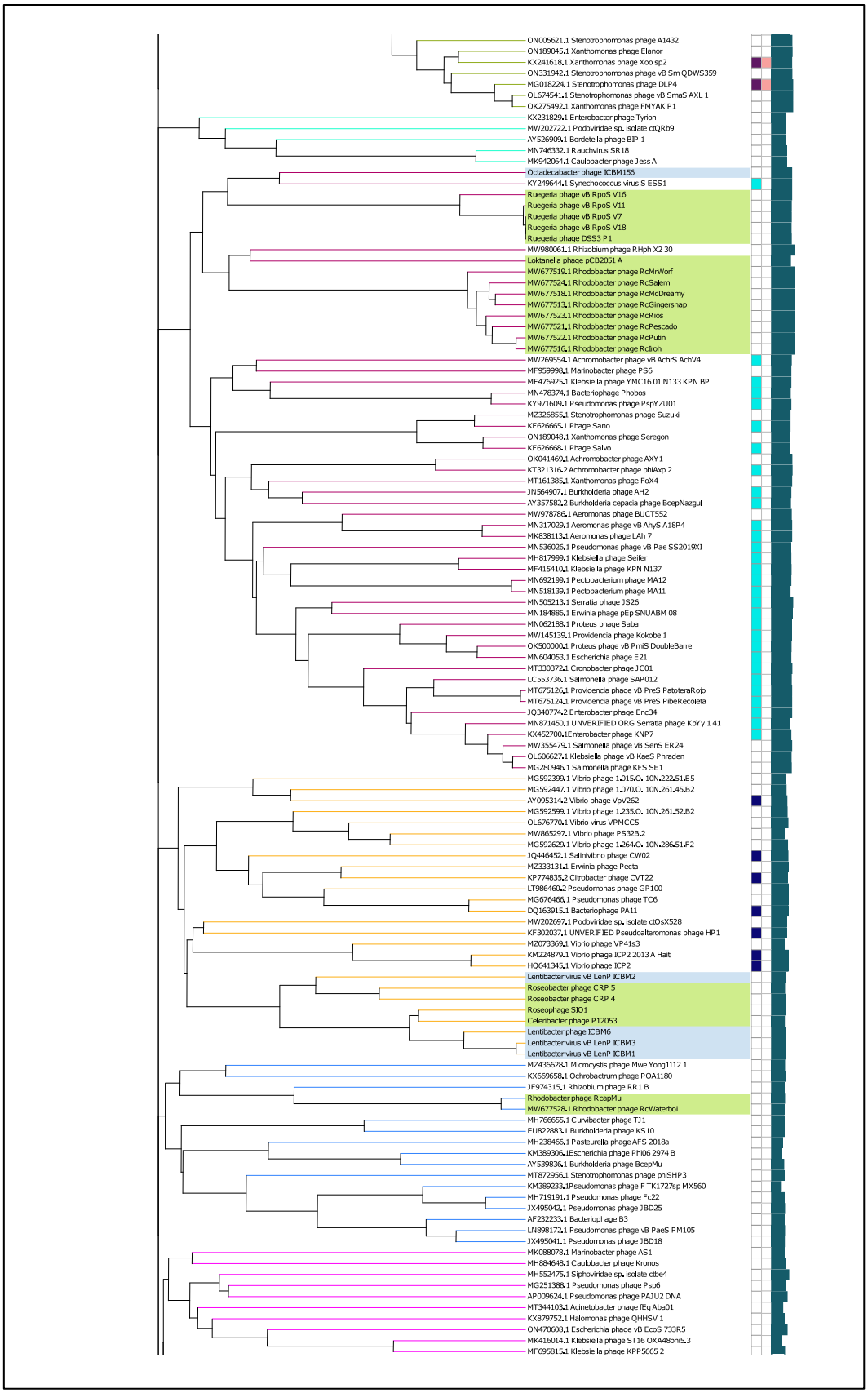

ON005621.1 Stenotrophomonas phage A1432
ON189045.1 Xanthomonas phage Elanor
KX241618.1 Xanthomonas phage Xoo sp2
ON331942.1 Stenotrophomonas phage vB Sm QDWS359
MG018224.1 Stenotrophomonas phage DLP4
OL674541.1 Stenotrophomonas phage vB SmaS AXL 1
OK275492.1 Xanthomonas phage FMYAK P1
KX231829.1 Enterobacter phage Tyrion
MW202722.1 Podoviridae sp. isolate ctQRb9
AY526909.1 Bordetella phage BIP 1
MN746332.1 Rauchvirus SR18
MK942064.1 Caulobacter phage Jess A
Octadecabacter phage ICBM156
KY249644.1 Synechococcus virus S ESS1
Ruegeria phage vB RpoS V16
Ruegeria phage vB RpoS V11
Ruegeria phage vB RpoS V7
Ruegeria phage vB RpoS V18
Ruegeria phage DSS3 P1
MW980061.1 Rhizobium phage RHph X2 30
Loktanella phage pCB2051 A
MW677519.1 Rhodobacter phage RcMrWorf
MW677524.1 Rhodobacter phage RcSalem
MW677518.1 Rhodobacter phage RcMcDreamy
MW677513.1 Rhodobacter phage RcGingersnap
MW677523.1 Rhodobacter phage RcRios
MW677521.1 Rhodobacter phage RcPescado
MW677522.1 Rhodobacter phage RcPutin
MW677516.1 Rhodobacter phage Rcltoh
MW269554.1 Achromobacter phage vB AchrS AchV4
MF959998.1 Marinobacter phage PS6
MF476925.1 Klebsiella phage YMC16 01 N133 KPN BP
MN478374.1 Bacteriophage Phobos
KY971609.1 Pseudomonas phage PspYZU01
MZ326855.1 Stenotrophomonas phage Suzuki
KF626665.1 Phage Sano
ON189048.1 Xanthomonas phage Seregon
KF626668.1 Phage Salvo
OK041469.1 Achromobacter phage AXY1
KT321316.2 Achromobacter phage phiAxp 2
MT161385.1 Xanthomonas phage FoX4
JN564907.1 Burkholderia phage AH2
AY357582.2 Burkholderia cepacia phage BcepNazgul
MW978786.1 Aeromonas phage BUCT552
MN317029.1 Aeromonas phage vB AhyS A18P4
MK838113.1 Aeromonas phage LAh 7
MN536026.1 Pseudomonas phage vB Pae SS2019XI
MH817999.1 Klebsiella phage Seifer
MF415410.1 Klebsiella phage KPN N137
MN692199.1 Pectobacterium phage MA12
MN518139.1 Pectobacterium phage MA11
MN505213.1 Serratia phage JS26
MN184886.1 Erwinia phage pEp SNUABM 08
MN062188.1 Proteus phage Saba
MW145139.1 Providencia phage Kokobel1
OK500000.1 Proteus phage vB PmiS DoubleBarrel
MN604053.1 Escherichia phage E21
MT330372.1 Cronobacter phage JC01
LC553736.1 Salmonella phage SAP012
MT675126.1 Providencia phage vB PreS PatoteraRojo
MT675124.1 Providencia phage vB PreS PibeRecoleta
JQ340774.2 Enterobacter phage Enc34
MN871450.1 UNVERIFIED ORG Serratia phage KpYy 1 41
KX452700.1Enterobacter phage KNP7
MW355479.1 Salmonella phage vB SenS ER24
OL606627.1 Klebsiella phage vB KaeS Phraden
MG280946.1 Salmonella phage KFS SE1
MG592399.1 Vibrio phage 1.015.O. 10N.222.51.E5
MG592447.1 Vibrio phage 1.070.O. 10N.261.45.B2
AY095314.2 Vibrio phage VpV262
MG592599.1 Vibrio phage 1.235.O. 10N.261.52.B2
OL676770.1 Vibrio virus VPMCC5
MW865297.1 Vibrio phage PS32B.2
MG592629.1 Vibrio phage 1.264.O. 10N.286.51.F2
JQ446452.1 Salinivibrio phage CW02
MZ333131.1 Erwinia phage Pecta
KP774835.2 Citrobacter phage CVT22
LT986460.2 Pseudomonas phage GP100
MG676466.1 Pseudomonas phage TC6
DQ163915.1 Bacteriophage PA11
MW202697.1 Podoviridae sp. isolate ctOsX528
KF302037.1 UNVERIFIED Pseudoalteromonas phage HP1
MZ073369.1 Vibrio phage VP41s3
KM224879.1 Vibrio phage ICP2 2013 A Haiti
HQ641345.1 Vibrio phage ICP2
Lentibacter virus vB LenP ICBM2
Roseobacter phage CRP 5
Roseobacter phage CRP 4
Roseophage SIO1
Celeribacter phage P12053L
Lentibacter phage ICBM6
Lentibacter virus vB LenP ICBM3
Lentibacter virus vB LenP ICBM1
MZ436628.1 Microcystis phage Mwe Yong1112 1
KX669658.1 Ochrobactrum phage POA1180
JF974315.1 Rhizobium phage RR1 B
Rhodobacter phage RcapMu
MW677528.1 Rhodobacter phage RcWaterboi
MH766655.1 Curvibacter phage TJ1
EU822883.1 Burkholderia phage KS10
MH238466.1 Pasteurella phage AFS 2018a
KM389306.1Escherichia phage Phi06 2974 B
AY539836.1 Burkholderia phage BcepMu
MT872956.1 Stenotrophomonas phage phiSHP3
KM389233.1Pseudomonas phage F TK1727sp MX560
MH719191.1 Pseudomonas phage Fc22
JX495042.1 Pseudomonas phage JBD25
AF232233.1 Bacteriophage B3
LN898172.1 Pseudomonas phage vB PaeS PM105
JX495041.1 Pseudomonas phage JBD18
MK088078.1 Marinobacter phage AS1
MH884648.1 Caulobacter phage Kronos
MH552475.1 Siphoviridae sp. isolate ctbe4
MG251388.1 Pseudomonas phage Psp6
AP009624.1 Pseudomonas phage PAJU2 DNA
MT344103.1 Acinetobacter phage fEg Aba01
KX879752.1 Halomonas phage QHHSV 1
ON470608.1 Escherichia phage vB EcoS 733R5
MK416014.1 Klebsiella phage ST16 OXA48phi5.3
MF695815.1 Klebsiella phage KPPS665 2

Fig. S 14e

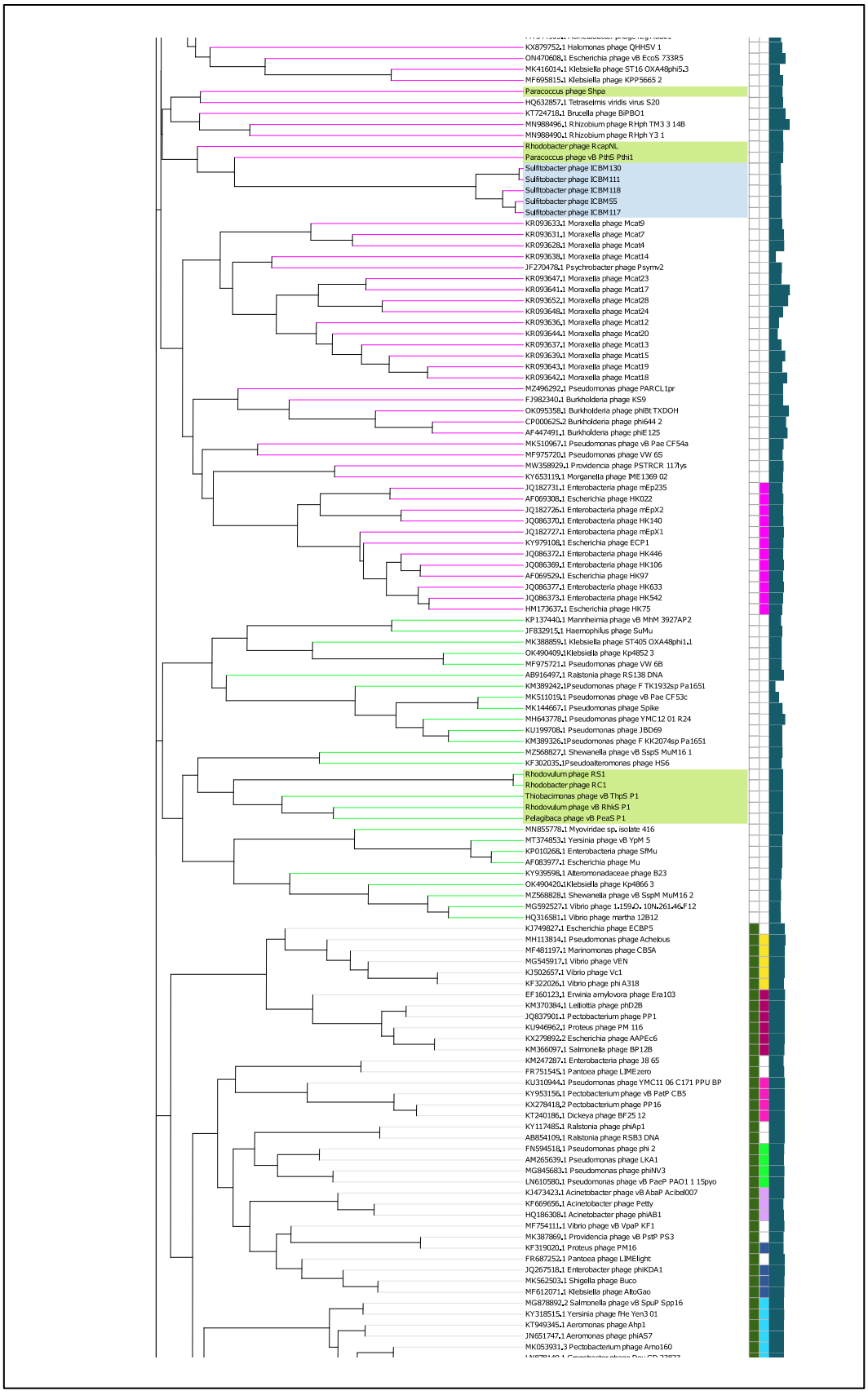

KX879752.1 Halomonas phage QHHSV 1
ON470608.1 Escherichia phage vB EcoS 733R5
MK416014.1 Klebsiella phage ST16 OXA48phi5.3
MF695815.1 Klebsiella phage KPP5665 2
Paracoccus phage Shpa
HQ632857.1 Tetraselmis viridis virus S20
KT724718.1 Brucella phage BiPBO1
MN988496.1 Rhizobium phage RHph TM3 3 14B
MN988490.1 Rhizobium phage RHph Y3 1
Rhodobacter phage RcapNL
Paracoccus phage vB PthS Pthi1
Sulfitobacter phage ICBM130
Sulfitobacter phage ICBM111
Sulfitobacter phage ICBM118
Sulfitobacter phage ICBM55
Sulfitobacter phage ICBM117
KR093633.1 Moraxella phage Mcat9
KR093631.1 Moraxella phage Mcat7
KR093628.1 Moraxella phage Mcat4
KR093638.1 Moraxella phage Mcat14
JF270478.1 Psychrobacter phage Psymv2
KR093647.1 Moraxella phage Mcat23
KR093641.1 Moraxella phage Mcat17
KR093652.1 Moraxella phage Mcat28
KR093648.1 Moraxella phage Mcat24
KR093636.1 Moraxella phage Mcat12
KR093644.1 Moraxella phage Mcat20
KR093637.1 Moraxella phage Mcat13
KR093639.1 Moraxella phage Mcat15
KR093643.1 Moraxella phage Mcat19
KR093642.1 Moraxella phage Mcat18
MZ496292.1 Pseudomonas phage PARCL1pr
FJ982340.1 Burkholderia phage KS9
OK095358.1 Burkholderia phage phiBt TXDOH
CP000625.2 Burkholderia phage phi644 2
AF447491.1 Burkholderia phage phiE125
MK510967.1 Pseudomonas phage vB Pae CF54a
MF975720.1 Pseudomonas phage VW 6S
MW358929.1 Providencia phage PSTRCR 117lys
KY653119.1 Morganella phage IME1369 02
JQ182731.1 Enterobacteria phage mEp235
AF069308.1 Escherichia phage HK022
JQ182726.1 Enterobacteria phage mEpX2
JQ086370.1 Enterobacteria phage HK140
JQ182727.1 Enterobacteria phage mEpX1
KY979108.1 Escherichia phage ECP1
JQ086372.1 Enterobacteria phage HK446
JQ086369.1 Enterobacteria phage HK106
AF069529.1 Escherichia phage HK97
JQ086377.1 Enterobacteria phage HK633
JQ086373.1 Enterobacteria phage HK542
HM173637.1 Escherichia phage HK75
KP137440.1 Mannheimia phage vB MhM 3927AP2
JF832915.1 Haemophilus phage SuMu
MK388859.1 Klebsiella phage ST405 OXA48phi1.1
OK490409.1Klebsiella phage Kp4852 3
MF975721.1 Pseudomonas phage VW 6B
AB916497.1 Ralstonia phage RS138 DNA
KM389242.1Pseudomonas phage F TK1932sp Pa1651
MK511019.1 Pseudomonas phage vB Pae CF53c
MK144667.1 Pseudomonas phage Spike
MH643778.1 Pseudomonas phage YMC12 01 R24
KU199708.1 Pseudomonas phage JBD69
KM389326.1Pseudomonas phage F KK2074sp Pa1651
MZ568827.1 Shewanella phage vB SspS MuM16 1
KF302035.1Pseudoalteromonas phage HS6
Rhodovulum phage RS1
Rhodobacter phage RC1
Thiobacimonas phage vB ThpS P1
Rhodovulum phage vB RhkS P1
Pelagibaca phage vB PgaS P1
MN855778.1 Myoviridae sp. isolate 416
MT374853.1 Yersinia phage vB YpM 5
KP010268.1 Enterobacteria phage SfMu
AF083977.1 Escherichia phage Mu
KY939598.1 Alteromonadaceae phage B23
OK490420.1Klebsiella phage Kp4866 3
MZ568828.1 Shewanella phage vB SspM MuM16 2
MG592527.1 Vibrio phage 1.159O. 10N.261.46.F12
HQ316581.1 Vibrio phage martha 12B12
KJ749827.1 Escherichia phage ECBP5
MH113814.1 Pseudomonas phage Achelous
MF481197.1 Marinomonas phage CB5A
MG545917.1 Vibrio phage VEN
KJ502657.1 Vibrio phage Vc1
KF322026.1 Vibrio phage phi A318
EF160123.1 Erwinia amylovora phage Era103
KM370384.1 Lelliottia phage phD2B
JQ837901.1 Pectobacterium phage PP1
KU946962.1 Proteus phage PM 116
KX279892.2 Escherichia phage AAPEc6
KM366097.1 Salmonella phage BP12B
KM247287.1 Enterobacteria phage J8 65
FR751545.1 Pantoea phage LIMEzero
KU310944.1 Pseudomonas phage YMC11 06 C171 PPU BP
KY953156.1 Pectobacterium phage vB PatP CB5
KX278418.2 Pectobacterium phage PP16
KT240186.1 Dickeya phage BF25 12
KY117485.1 Ralstonia phage phiAp1
AB854109.1 Ralstonia phage RSB3 DNA
FN594518.1 Pseudomonas phage phi 2
AM265639.1 Pseudomonas phage LKA1
MG845683.1 Pseudomonas phage phiNV3
LN610580.1 Pseudomonas phage vB PaeP PAO1 1 15pyo
KJ473423.1 Acinetobacter phage vB AbaP Acibel007
KF669656.1 Acinetobacter phage Petty
HQ186308.1 Acinetobacter phage phiAB1
MF754111.1 Vibrio phage vB VpaP KF1
MK387869.1 Providencia phage vB PstP PS3
KF319020.1 Proteus phage PM16
FR687252.1 Pantoea phage LIMElight
JQ267518.1 Enterobacter phage phiKDA1
MK562503.1 Shigella phage Buco
MF612071.1 Klebsiella phage AltoGao
MG878892.2 Salmonella phage vB SpuP Spp16
KY318515.1 Yersinia phage fHe Yen3 01
KT949345.1 Aeromonas phage Ahp1
JN651747.1 Aeromonas phage phiAS7
MK053931.3 Pectobacterium phage Arno160

Fig. S 14f

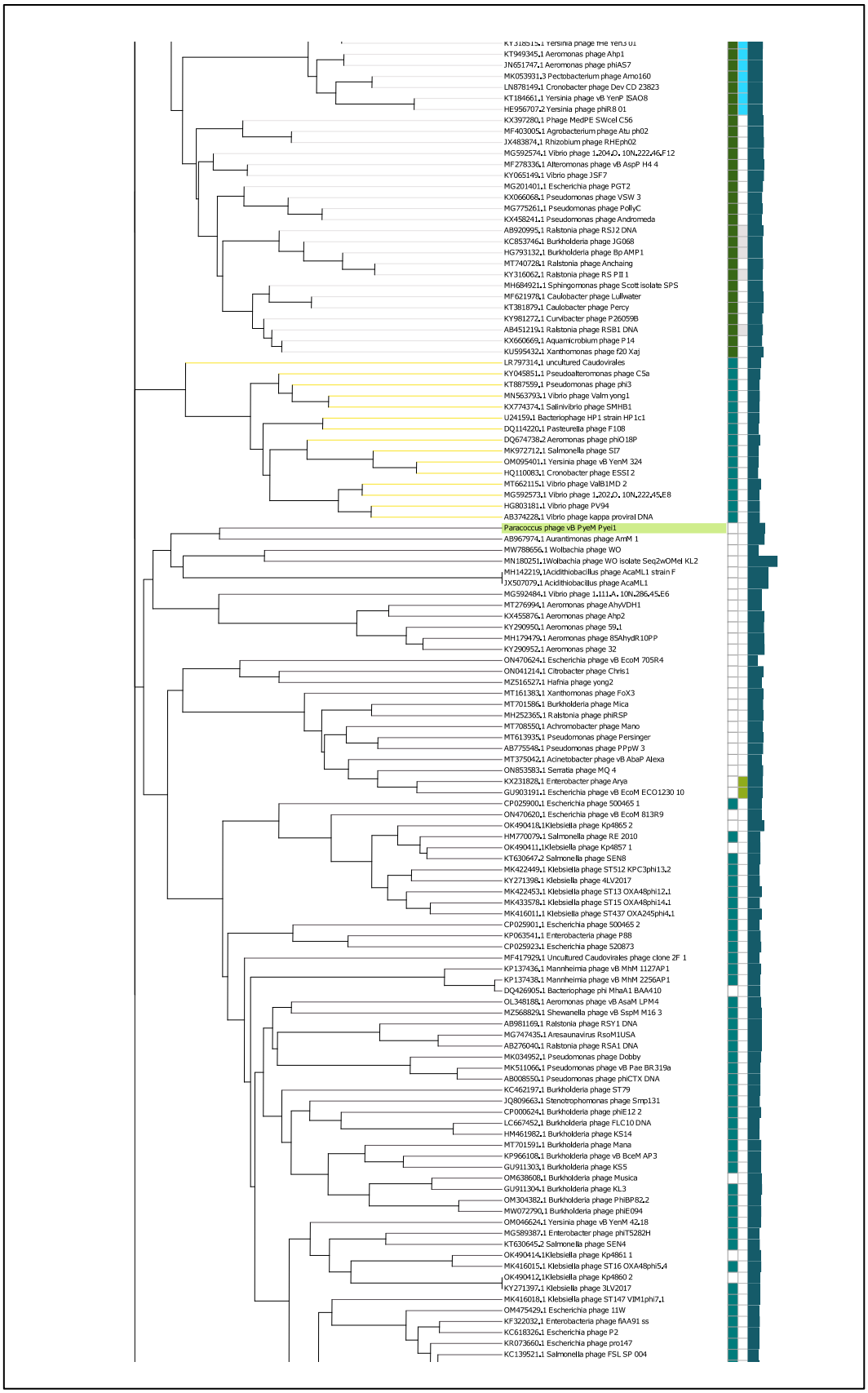

KT949345.1 Aeromonas phage Ahp1
JN651747.1 Aeromonas phage phiAS7
MK053931.1 Pectobacterium phage Arno160
LN878149.1 Cronobacter phage Dev CD 23823
KT184661.1 Yersinia phage vB YenP ISAO8
HE956707.2 Yersinia phage phiR8 01
KX397280.1 Phage MedPE SWcel C56
MF403005.1 Agrobacterium phage Atu ph02
JX483874.1 Rhizobium phage RHEph02
MG592574.1 Vibrio phage 1.204.O. 10N.222.46.F12
MF278336.1 Alteromonas phage vB AspP H4 4
KY065149.1 Vibrio phage JSF7
MG201401.1 Escherichia phage PGT2
KX066068.1 Pseudomonas phage VSW 3
MG775261.1 Pseudomonas phage PollyC
KX458241.1 Pseudomonas phage Andromeda
AB920995.1 Ralstonia phage RSJ2 DNA
KC853746.1 Burkholderia phage JG068
HG793132.1 Burkholderia phage Bp AMP1
MT740728.1 Ralstonia phage Anchaing
KY316062.1 Ralstonia phage RS PII 1
MH684921.1 Sphingomonas phage Scott isolate SPS
MF621978.1 Caulobacter phage Lullwater
KT381879.1 Caulobacter phage Percy
KY981272.1 Curvibacter phage P26059B
AB451219.1 Ralstonia phage RSB1 DNA
KX660669.1 Aquamicrobium phage P14
KU595432.1 Xanthomonas phage f20 Xaj
LR797314.1 uncultured Caudovirales
KY045851.1 Pseudoalteromonas phage CSa
KT887599.1 Pseudomonas phage phi3
MN563793.1 Vibrio phage Valm yong1
KX774374.1 Salinivibrio phage SMHB1
U24159.1 Bacteriophage HP1 strain HP1c1
DQ114220.1 Pasteurella phage F108
DQ674738.2 Aeromonas phage phiO18P
MK972712.1 Salmonella phage SI7
OM095401.1 Yersinia phage vB YenM 324
HQ110083.1 Cronobacter phage ESSI 2
MT662115.1 Vibrio phage ValB1MD 2
MG592573.1 Vibrio phage 1.202.O. 10N.222.45.E8
HG803181.1 Vibrio phage PV94
AB374228.1 Vibrio phage kappa proviral DNA
Paracoccus phage vB PyeM Pyei1
AB967974.1 Aurantimonas phage AmM 1
MW788656.1 Wolbachia phage WO
MN180251.1Wolbachia phage WO isolate Seq2wOMel KL2
MH142219.1Acidithiobacillus phage AcaML1 strain F
JX507079.1 Acidithiobacillus phage AcaML1
MG592484.1 Vibrio phage 1.111.A. 10N.286.45.E6
MT276994.1 Aeromonas phage AhyVDH1
KX455876.1 Aeromonas phage Ahp2
KY290950.1 Aeromonas phage 59.1
MH179479.1 Aeromonas phage 85AhydR10PP
KY290952.1 Aeromonas phage 32
ON470624.1 Escherichia phage vB EcoM 705R4
ON041214.1 Citrobacter phage Chris1
MZ516527.1 Hafnia phage yong2
MT161383.1 Xanthomonas phage FoX3
MT701586.1 Burkholderia phage Mica
MH252365.1 Ralstonia phage phiRSP
MT708550.1 Achromobacter phage Mano
MT613935.1 Pseudomonas phage Persinger
AB775548.1 Pseudomonas phage PPpW 3
MT375042.1 Acinetobacter phage vB AbaP Alexa
ON853583.1 Serratia phage MQ 4
KX231828.1 Enterobacter phage Arya
GU903191.1 Escherichia phage vB EcoM ECO1230 10
CP025900.1 Escherichia phage 500465 1
ON470620.1 Escherichia phage vB EcoM 813R9
OK490418.1Klebsiella phage Kp4865 2
HM770079.1 Salmonella phage RE 2010
OK490411.1Klebsiella phage Kp4857 1
KT630647.2 Salmonella phage SEN8
MK422449.1 Klebsiella phage ST512 KPC3phi13.2
KY271398.1 Klebsiella phage 4LV2017
MK422453.1 Klebsiella phage ST13 OXA48phi12.1
MK433578.1 Klebsiella phage ST15 OXA48phi14.1
MK416011.1 Klebsiella phage ST437 OXA245phi4.1
CP025901.1 Escherichia phage 500465 2
KP063541.1 Enterobacteria phage P88
CP025923.1 Escherichia phage 520873
MF417929.1 Uncultured Caudovirales phage clone 2F 1
KP137436.1 Mannheimia phage vB MhM 1127AP1
KP137438.1 Mannheimia phage vB MhM 2256AP1
DQ426905.1 Bacteriophage phi MhaA1 BAA410
OL348188.1 Aeromonas phage vB AsaM LPM4
MZ566820.1 Shewanella phage vB SspM M16 3
AB981169.1 Ralstonia phage RSY1 DNA
MG747435.1 Aresaunavirus RsoM1USA
AB276040.1 Ralstonia phage RSA1 DNA
MK034952.1 Pseudomonas phage Dobby
MK511066.1 Pseudomonas phage vB Pae BR319a
AB008550.1 Pseudomonas phage phiCTX DNA
KC462197.1 Burkholderia phage ST79
JQ809663.1 Stenotrophomonas phage Smp131
CP000624.1 Burkholderia phage phiE12 2
LC667452.1 Burkholderia phage FLC10 DNA
HM461982.1 Burkholderia phage KS14
MT701591.1 Burkholderia phage Mana
KP966108.1 Burkholderia phage vB BceM AP3
GU911303.1 Burkholderia phage KS5
OM638608.1 Burkholderia phage Musica
GU911304.1 Burkholderia phage KL3
OM304382.1 Burkholderia phage PhiBP82.2
MW072790.1 Burkholderia phage phiE094
OM046624.1 Yersinia phage vB YenM 42.18
MG589387.1 Enterobacter phage phiTS282H
KT630645.2 Salmonella phage SEN4
OK490414.1Klebsiella phage Kp4861 1
MK416015.1 Klebsiella phage ST16 OXA48phi5.4
OK490412.1Klebsiella phage Kp4860 2
KY271397.1 Klebsiella phage 3LV2017
MK416018.1 Klebsiella phage ST147 VIM1phi7.1
OM475429.1 Escherichia phage 11W
KF322032.1 Enterobacteria phage fAA91 ss
KC618326.1 Escherichia phage P2
KR073660.1 Escherichia phage pro147
KC139521.1 Salmonella phage FSL SP 004

Fig. S 14g

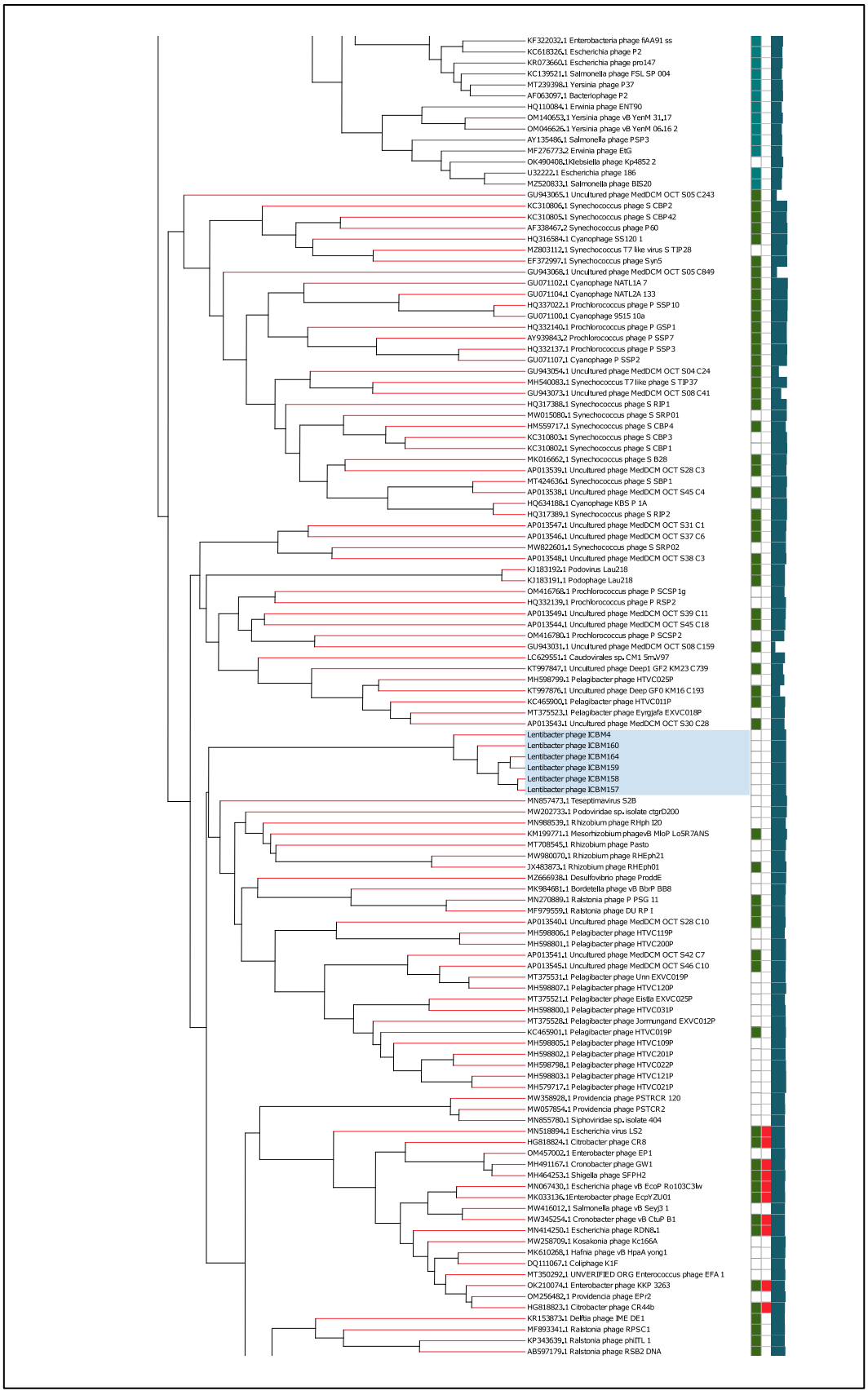

KF322032.1 Enterobacteria phage fIAA91 ss
KC618326.1 Escherichia phage P2
KR073660.1 Escherichia phage pro147
KC139521.1 Salmonella phage FSL SP 004
MT239398.1 Yersinia phage P37
AF063097.1 Bacteriophage P2
HQ110084.1 Erwinia phage ENT90
OM140653.1 Yersinia phage vB YenM 31.17
OM046626.1 Yersinia phage vB YenM 06.16 2
AY135486.1 Salmonella phage PSP3
MF276773.2 Erwinia phage EtG
OK490408.1Klebsiella phage Kp4852 2
U32222.1 Escherichia phage 186
MZ520833.1 Salmonella phage BIS20
GU943065.1 Uncultured phage MedDCM OCT S05 C243
KC310806.1 Synechococcus phage S CBP2
KC310805.1 Synechococcus phage S CBP42
AF338467.2 Synechococcus phage P60
HQ316584.1 Cyanophage SS120 1
MZ803112.1 Synechococcus T7 like virus S TIP28
EF372997.1 Synechococcus phage Syn5
GU943068.1 Uncultured phage MedDCM OCT S05 C849
GU071102.1 Cyanophage NATL1A 7
GU071104.1 Cyanophage NATL2A 133
HQ337022.1 Prochlorococcus phage P SSP10
GU071100.1 Cyanophage 9515 10a
HQ332140.1 Prochlorococcus phage P GSP1
AY939843.2 Prochlorococcus phage P SSP7
HQ332137.1 Prochlorococcus phage P SSP3
GU071107.1 Cyanophage P SSP2
GU943054.1 Uncultured phage MedDCM OCT S04 C24
MH540083.1 Synechococcus T7 like phage S TIP37
GU943073.1 Uncultured phage MedDCM OCT S08 C41
HQ317388.1 Synechococcus phage S RIP1
MW015080.1 Synechococcus phage S SRP01
HM559717.1 Synechococcus phage S CBP4
KC310803.1 Synechococcus phage S CBP3
KC310802.1 Synechococcus phage S CBP1
MK016662.1 Synechococcus phage S B28
AP013539.1 Uncultured phage MedDCM OCT S28 C3
MT424636.1 Synechococcus phage S SBP1
AP013538.1 Uncultured phage MedDCM OCT S45 C4
HQ634188.1 Cyanophage KBS P 1A
HQ317389.1 Synechococcus phage S RIP2
AP013547.1 Uncultured phage MedDCM OCT S31 C1
AP013546.1 Uncultured phage MedDCM OCT S37 C6
MW822601.1 Synechococcus phage S SRP02
AP013548.1 Uncultured phage MedDCM OCT S38 C3
KJ183192.1 Podovirus Lau218
KJ183191.1 Podophage Lau218
OM416768.1 Prochlorococcus phage P SCSP 1g
HQ332139.1 Prochlorococcus phage P RSP2
AP013549.1 Uncultured phage MedDCM OCT S39 C11
AP013544.1 Uncultured phage MedDCM OCT S45 C18
OM416780.1 Prochlorococcus phage P SCSP 2
GU943031.1 Uncultured phage MedDCM OCT S08 C159
LC629551.1 Caudovirales sp. CM1 5mV97
KT997847.1 Uncultured phage Deep1 GF2 KM23 C739
MH598799.1 Pelagibacter phage HTVC025P
KT997876.1 Uncultured phage Deep GF0 KM16 C193
KC465900.1 Pelagibacter phage HTVC011P
MT375523.1 Pelagibacter phage Eyrgjafa EXVC018P
AP013543.1 Uncultured phage MedDCM OCT S30 C28
Lentibacter phage ICBM4
Lentibacter phage ICBM160
Lentibacter phage ICBM164
Lentibacter phage ICBM159
Lentibacter phage ICBM158
Lentibacter phage ICBM157
MN857473.1 Teseptimavirus S2B
MW202733.1 Podoviridae sp. isolate ctgrD200
MN988539.1 Rhizobium phage RHph I20
KM199771.1 Mesorhizobium phagevB MloP Lo5R7ANS
MT708545.1 Rhizobium phage Pasto
MW980070.1 Rhizobium phage RHEph21
JX483873.1 Rhizobium phage RHEph01
MZ666938.1 Desulfovibrio phage ProddE
MK984681.1 Bordetella phage vB BbrP BB8
MN270889.1 Ralstonia phage P PSG 11
MF979559.1 Ralstonia phage DU RP I
AP013540.1 Uncultured phage MedDCM OCT S28 C10
MH598806.1 Pelagibacter phage HTVC119P
MH598801.1 Pelagibacter phage HTVC200P
AP013541.1 Uncultured phage MedDCM OCT S42 C7
AP013545.1 Uncultured phage MedDCM OCT S46 C10
MT375531.1 Pelagibacter phage Unn EXVC019P
MH598807.1 Pelagibacter phage HTVC120P
MT375521.1 Pelagibacter phage Eistla EXVC025P
MH598800.1 Pelagibacter phage HTVC031P
MT375528.1 Pelagibacter phage Jormungand EXVC012P
KC465901.1 Pelagibacter phage HTVC019P
MH598805.1 Pelagibacter phage HTVC109P
MH598802.1 Pelagibacter phage HTVC201P
MH598798.1 Pelagibacter phage HTVC022P
MH598803.1 Pelagibacter phage HTVC121P
MH579717.1 Pelagibacter phage HTVC021P
MW358928.1 Providencia phage PSTRCR 120
MW057854.1 Providencia phage PSTCR2
MN855780.1 Siphoviridae sp. isolate 404
MN518894.1 Escherichia virus LS2
HG818824.1 Citrobacter phage CR8
OM457002.1 Enterobacter phage EP1
MH491167.1 Cronobacter phage GW1
MH464253.1 Shigella phage SFPH2
MN067430.1 Escherichia phage vB EcoP Ro103C3lw
MK033136.1Enterobacter phage EcpYZU01
MW416012.1 Salmonella phage vB Seyj3 1
MW345254.1 Cronobacter phage vB CtuP B1
MN414250.1 Escherichia phage RDN8.1
MW258709.1 Kosakonia phage Kc166A
MK610268.1 Hafnia phage vB HpaA yong1
DQ111067.1 Coliphage K1F
MT350292.1 UNVERIFIED ORG Enterococcus phage EFA 1
OK210074.1 Enterobacter phage KKP 3263
OM256482.1 Providencia phage EPr2
HG818823.1 Citrobacter phage CR44b
KR153873.1 Delftia phage IME DE1
MF893341.1 Ralstonia phage RPSC1
KP343639.1 Ralstonia phage phITL 1
AB597179.1 Ralstonia phage RSB2 DNA

Fig. S 14h

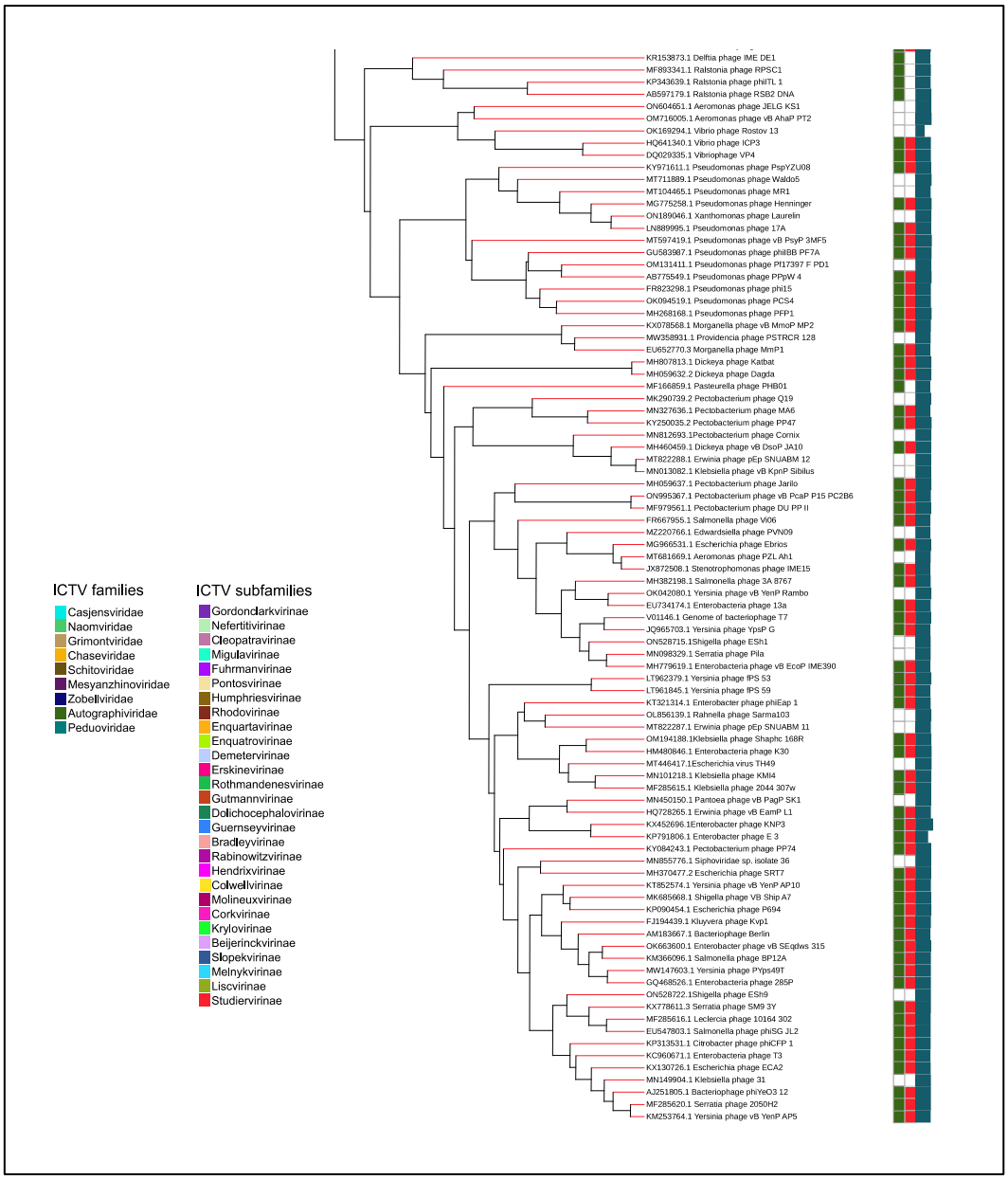


Fig. S 14i

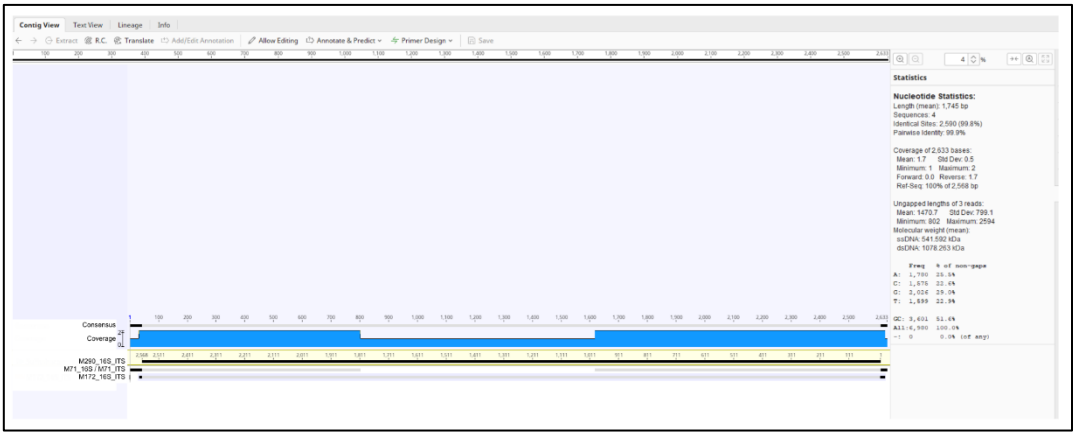


Fig. S 15: Sequence comparison of the 16S rRNA gene and the ITS region of strains M290, M71 and M172. Sequences of strains M71 and M172 were mapped against the sequence of M290 using Geneious Prime®.

Fig. S 16a-e: Genome maps of all complete Sulfivirus genomes, ordered by species cluster.


Species 8
ICBM41
ICBM119
Species 9
ICBM12
ICBM13
ICBM22
ICBM49
ICBM57
ICBM58
ICBM59
ICBM65
ICBM70
ICBM95
ICBM133
ICBM134

Fig. S 16b

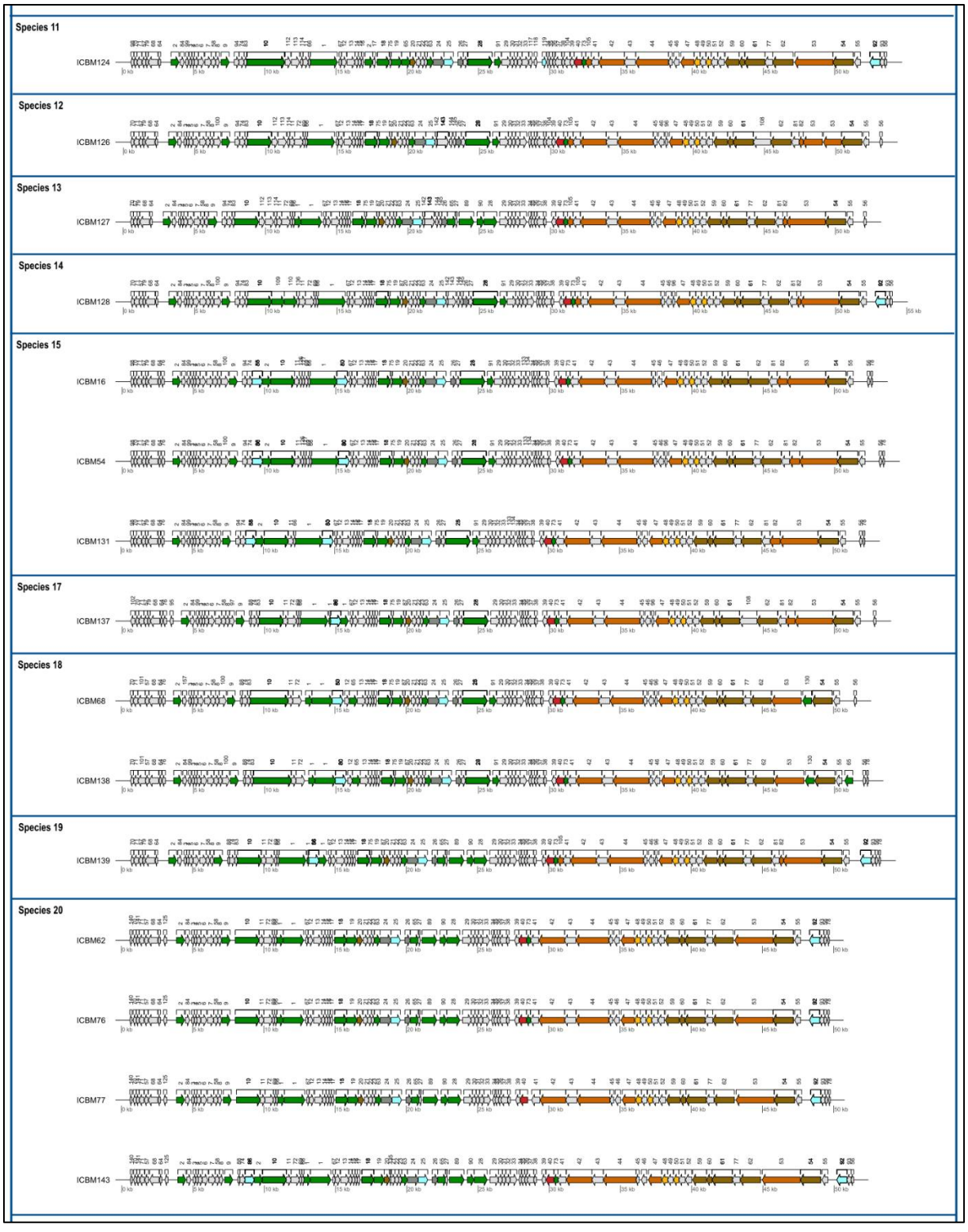

Species 11
ICBM124
Species 12
ICBM126
Species 13
ICBM127
Species 14
ICBM128
Species 15
ICBM16
ICBM54
ICBM131
Species 17
ICBM137
Species 18
ICBM68
ICBM138
Species 19
ICBM139
Species 20
ICBM62
ICBM76
ICBM77
ICBM143

Fig. S 16c

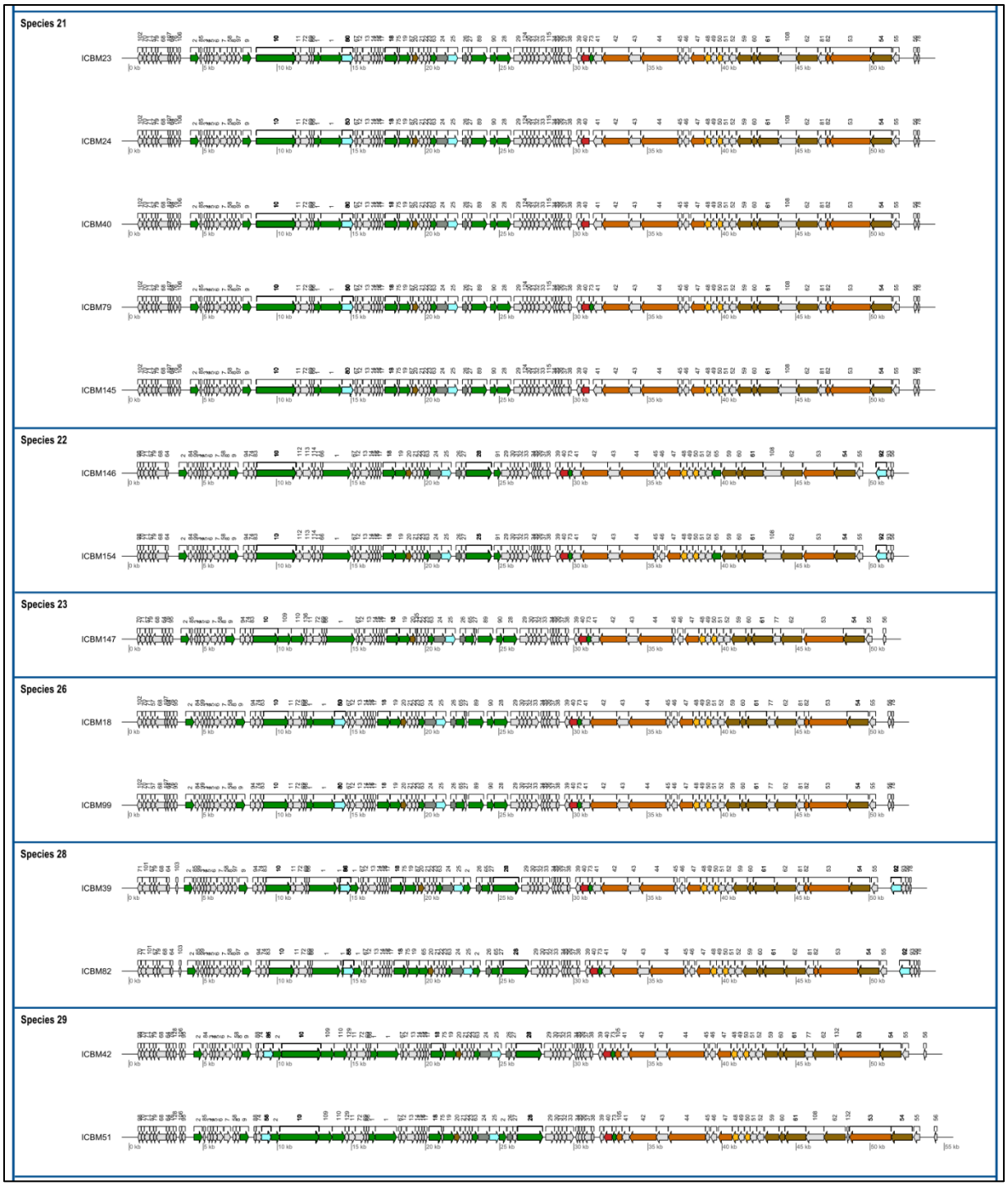

Species 21
ICBM23
ICBM24
ICBM40
ICBM79
ICBM145
Species 22
ICBM146
ICBM154
Species 23
ICBM147
Species 26
ICBM18
ICBM99
Species 28
ICBM39
ICBM82
Species 29
ICBM42
ICBM51

Fig. S 16d

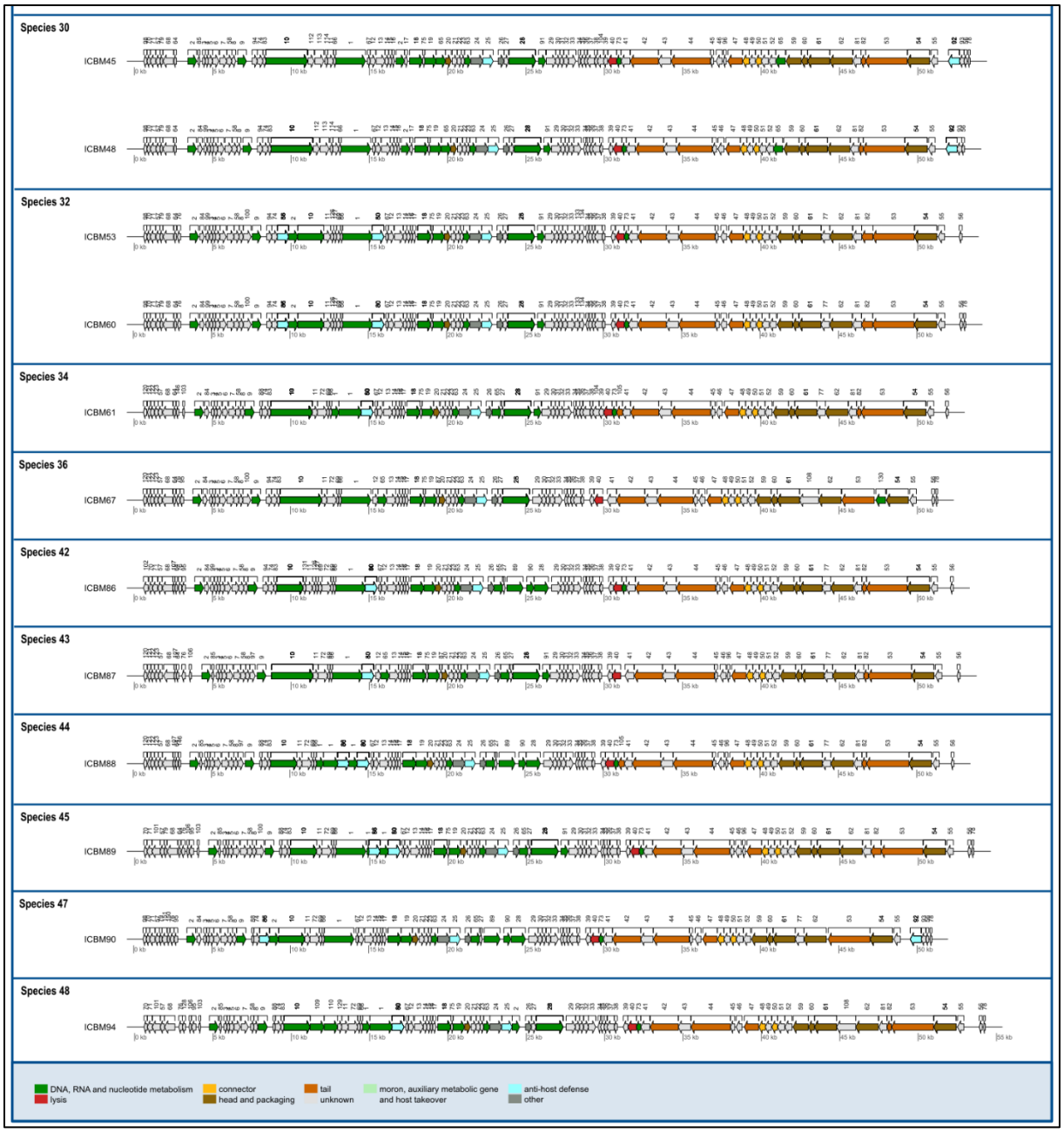

Species 30
ICBM45
ICBM48
Species 32
ICBM53
ICBM60
Species 34
ICBM61
Species 36
ICBM67
Species 42
ICBM86
Species 43
ICBM87
Species 44
ICBM88
Species 45
ICBM89
Species 47
ICBM90
Species 48
ICBM94
DNA, RNA and nucleotide metabolism
lysis
connector
head and packaging
tail
unknown
moron, auxiliary metabolic gene and host takeover
anti-host defense
other

Fig. S 16e

c) Supplementary text

Isolation of *Lentibacter* sp. MPI-62 and *Octadecabacter* sp. MM282 (by Anneke Heins, MPI, Bremen)

Lentibacter sp. MPI-62 was isolated from a seawater sample taken at the Helgoland Roads time series station (54°11’03”N, 7°54’00”E) in April 2017. *Octadecabacter* sp. MM282 was isolated from a seawater sample taken at high tide at the shore of Harlesiel (53°42'39"N 7°48'28"E) in October 2017. In both cases, one liter of seawater was transferred to a sedimentation cone (Imhoff, Brand, Wertheim, Germany). Sedimentation was allowed for 24 h at room temperature, and then the particle fraction was removed with the help of a stopcock at the bottom the cone. The particle fractions were homogenized and then diluted with sterile artificial seawater (following the recipe of Widdel and Bak (2013)). For the isolation of *Lentibacter* sp. MPI-62, the diluted fraction was plated with a sterile one-way inoculation loop on an MB-agar plate (recipe as described above). For the isolation of *Octadecabacter* sp. MM282 a Syl-agar plate was used (recipe after Hahnke and Harder (2013)). The plates were incubated for five weeks in the dark at 12 °C. Colonies were picked and transferred to new agar plates for strain purification.

Genomic information on “Ascunsovirus oldenburgi” ICBM5

>Ascunsovirus_oldenburgi_ICBM5
TTCAGACGAAAGGATGCGCGACGATAAAATTAGCTCTTGCAAACCCCGTCCAAATTCGACTAAGGT
CAATTTAGGGAATGGTTCCCGGCGCGGGTTTATCCCGTGGTGCTAACAAAGGAAAAACACATGGCT
GATAAAGCAACTTCCCCGACTACCGTCTCTAAAAAAGATCGTGATGATTTTGAGCTTCATGTTCGG
CTCTTGAGCCGGGCAGGTTACAAAACTGCCGAAGCTCGGACAATTGCATGGCTGCGCGGCCCTGCT
GGGCTGCAAGAAATGCTTGCGGGCGTCAAGGTCGCAAGCTAACATAGGTTGAGCGGTTCGCCGCTC
AACATTAACCGCAAAAAAGGAGAGAAATGTGAGTGCATTGAAACACTTTCTTACAAACCGTTTGAC
AGTCGTGGTTCTTACCTCGCTGGCAACGGCAGTAGGTACTGCGCTTGCGACGGAGTTTCCGTCCATT
TATAGCGCAGTCTGTGCCTAATGTCGGCTATCGCTGGTGCTCTTATCTCAGGTGGCGCAAGCCTCCT
AGGTGGTCTGTTCGGGCGTTCTTCTGCTAGCAAGCAGCAAGCCCGACAGAATGAGTACAACAAGCC
GATTAACATTCGCAAGCGGGCCGAGGAGGGGGGGTTCAACCCCCTGCTCTGGGCCGGTCAAGGCA
ACATCCAAATGCAGCCGGGTCCGTCCGGCATCATGGGTTCCGCTATTGCGAATGCTGGCCTAGCTC
TTGCCGATGGCATGAGCGAACAACGCCAGCTCGACCTTGAGCGTACCAAGCTCAAGCAAGATCAA
GAGCGTCTCGACGCTCTGATCGAAAAACAGACCATCCGGCCAAAGGTCGGCGGCATTTATGCCGG
GTCGCAACAAACGCCTTCTGTAGCGCGCGCTCCCGGTCGCCCGCTTATGAATGGCGCTCCTCAACC
CGGCTCTGCGCCGGTCTTTAACCCGCCAACGGAGTACAACCCAATCCCCGATGATGGCCCTCGCCT
GCAAACGAAAGTGATGCGTAGCGATGGCATGACCTCGGCGGATCCTGAAAATCCCGCCGAAATGG
AGGGCGATTGGTGGACGTGGGCCAGAGAGGGAACTTTCTGGCAAAACAACAACGAAATTCTGCGG
CGTAATACGCCGGAGACGTTGCACTACAAGGGTCGCGATGCCTTGTTCCCGAAAATGATTGACGGG
GCAAGGAAAGCGCATAAGAAGGCTCAAGAGGACTTCGAGAAAAACCCGCCAAAACGCCGCAAGC
TTAAAGGCGTCAACCCTAACCTAAACGACAAGAAATGGTAATGAACATGTCAAAGTATCAACGTC
CTACAAACACACGCCGCGAAAGCCGGACCATCGCTGGCCGGTTCCGTGGCGGCAAGTTGGCTCCTG
TTATGGCGTCCGCGTTCCGTGAGAGCGAAAGTGCAATCCTTTCGCAACAAGTTACCTATGAACTTG
ACCCAATCGCGGGCCGTATGATTACGCCGATCATGGCGGAACTTATCTCTGTATATGTTCCGGTCCA
AGCGATCGACGCCCTAAAAAACCCTGAGGAGGCTTATGCCGGTAACACTGAGGTTGTCCGTGACA

AGCTCCTCTCAGGAACGCCGCTGTTCGGCCTCGAAGACGAAAGCGAGATTTCGAAGCGTCTAGGCG
TTAACCCTATCTCTGTGGGTGGTGTCAAAAAAGTGAACGAGGCGGCACGCCTTGCGCATAACTGCG
CCGTGAACTTCCTACGTCAGCGCAAGTACGTCAACACGGTCAAGCTATTGGCCGATAATATGAACG
TAACGCCTGCGCTGATCTCTCAAACGGTTCTCGACCGGCTGAACGCCGTGCTCGATCCTGAGGATC
GTGTCAACGGTGCTGTGCAGCTTGACCTTGGGAACGTGCGTATGCCGGTCGAAGGCGTCGGCGTTG
AATACAATGCCGTTCCAGTGACTACCCCCGGTTGGAACGATACAGAGAATGACCGTGCTGCTGATC
CTCACTATTTCTCGGACCACGCCGCAGGCATTCGCCTTGCGATTGCCTCGGAGAATGGTCGTCCAGC
GGTCTACGCTGTTGGCGAGGGTCAAAATGCCGGCGGCATTTCCCTTACTGACTTCTACAACGCCGA
ACTGATGGACAGTCTTGTGCGCCAGATGCGGCAGATTGTCGATGATAACCCTGAGTATGGCGAAGA
AATGGTCACACGCTGGGCGCATGGTCTTTCTGTTGATAACGGCAAAACTCCTTGGATTTTGCACCA
AAGCCAGCAAATGTTCGGCAATCAATTCCGCCGTGCCATGGATGGTGCAAACCTCGATGTTGCTCA
ATCCGACATGATGCAAAGCCTTGAATTCACTGTCCCCGTGCCTCCAACGGAATTGGGCGGGGTCGT
GATTACCTTTGCATCTGTGAAGCCTGACGAAACACTAGGCTCTCAGCCGCACCCGTTCCTGTCGAC
AACATGGGAGGCGACTAACTACGTCTCCGACGAGTTGGCACGCGACCCTGAGGCCGTGACCATGC
GTCACCTGAACAGTGACGTTCCTATGGTCGATGAGGACACTCGTGCTCTCTACATCGGTCACAACG
GTCTGAAAAAAGCCTACATCTCATATGGCTTTAACCGGCATCTCGACCCTACTACGGTCGAAGCGA
AAACTGCCATCTGGCAGCTTGAGGTGCCAATGTCCGTAACGCCTGAAAGCGTTATCTATCCCGAGG
ACTTGGACCATTATCCGTTTGCGGATCAGCTTGCAGAGGTCTGCACCTATCAGGTTAGTTCGACCGC
TACAGTGCGGACGCCTATGGTCTTCGGTCCGACACCAGTTGAGGAACTGGCCCAGATTGAGACAGA
CAACGTCTTCGAAGACGTATAAAAAAAGGGGGGCGGTTATGCCGCCCCTCAATAAACCTTAAAAA
ACTGGAAAGATGAAATGCAAGTTAATGATGCAAAACGCTCCGTTCTTCTCGTCGGCCTTAATGGCG
GCGAAGTTTCTGTAATCTCCGCAGATGGTGAGATTATCGCGACCGAAGGTGTAACCGCTGGTCGGC
ATAAATGCTCCTCTTGGGTGCCATTTATGTCCAATGAGGGCGATGAATTGAGCTTCTCCGGCGATGT
GGTCCCAATGGTGCCAAATGGCGGTCGTGTCCGGCCTATGGCCTACGGCCCCGGTCAATTTGAAAG
CGGTGCCAATCCCGATTTCGTCGTTACTTCGGCGGATCGGATGGCTCGTGAGCTTGACCATAAAATT
AGGGGTCTTGACCAGACCGCTAAAAAAGTTGAGGCCCGTATTGCGCAGTTGAACAATCTAGCAGA
ACGCGCTAAAACACCTGTAGAAGTAGCAAAGGAGGAAAAAGTCGATGTTATTGATGATGACGATG
TTTCTGACGTTCGTAGCTCTGGCGACGATAGTGACGATACGGGACCTGATTTTGGCGCTGAAAAAG
TGGGGGCAGAATGAATGATGCCCCGAAAAAAAGCAATTCGGGTCGCAAACAAAGCTGGCTCCGCG
CGTATGCGCGGGGCCTCAAAGGCGGCACCCCGAAAACGCTCTGCGAAAGCCTGCGCCTGTGGTCA
ATCCCACAAGAGCAGGCGCAAGACGTGACAAAAAAAACTTACCTCAACGCCGCAGATGGCCCCAT
GATCTGGCTCGGTGAGCAGATTGTCAAACGCATGGAAAAGGCGGGTTATCCCTCCCGCATTTTCTG
CGGCTATCGTTCCCCAGAACAGCAAGACAAGGAGTTTGCAGAGGGCGACAGCAAGGCTAGGGCCT
ATCAAAGCCCACATCAGTTTTATGAAGCGGTAGACATAATTCATAAGACAAAAGCATGGAACGTCT
CTCAAGATTATTGGGATACACTCGCCGCAATTGTGCGAGTTGTAGAGCGTGAATATGCTATAGACC
TCGTGCATGGTTATGATTGGGGATGGGACAGTGCCCATATTGAGATTGCTTTGTGGCGTCAAGTTCC
AAAGCGTCAAATTGGCAAAACTGGCTCTAATTATCCCCCGTCACCTTGGGAACGTGAACAGCGTTT
CAAGGAGTTGCTTCCCGCTGTTTATGCGGCGCAACATCGGCCATAGACTTATGGCAATCTATGCAA
ACTAACAAGCGGGCGGCCTCTATGCGAGAGCCGCCCGTTTTGCATCCCTGCGGGCCAGTCGGAGAC
TGGCCAGCGGCAACTATATTCCGGGCCAAATCTTTTGGACGCGCCTTAGCGCCCCGGATGGCTACG
CAGCGTTAAGCTAGTGCCTAACAATGTGTCAGTAAGCTCGTCAACGTAAACCTCTCTCAGAATTGA
GGACGGTTTGCGTTGATCCCCCTAGCAGGCCCCCCTTGTTCCTGTATACACTTACTGACACAAAAAA
CAGGAGTTAAAGAATATGTGCAGTGATTTAATCCATATCGACGGGCAGCAATTTGCCTGTCGTAAG
TGTAATGAGTGCATCACCGCTAGGAAAAATGGTTGGGTTGCTCGTGCGATGGCTGAAAAGGCCGTC
ACTGCTGAAACTTTCAGCGTAACCCTAACTTATAATGATGCTACTCAGGAAAGCCGCGATGCGGCA
AAGACGTTCGAGTATCGACACGTCAAAAACTGGATTAAGAACCTTAGGCGTCAAATAGAGTACAC
CACCGGCCAGACTGGCCTTCTCCGCTACCTTGTAGCGGGTGAGCGCGGTTCCGATAAGGGCCGCTG
CCACTGGCATGTAATCTTGTTCTGTAACGCCGATATTTTAACGCTAGGAAAAATGACGCACTGGCC
CTCTGGCAAACTTGCTGCCGATAGGGCCGAAATTATAACTACAGGAAAAAGGAAAAAACGTATTA
ACTGGTCTCTCTGGCCGTATGGCTTCGTCAGCTTTCAAGAGCCTGACCAGTGGGGCATGGAGTACG
CCTTAGCGTATGCCCTCAAAGATCAGTTTAACATTGTCTCCGCTGCCGGTACGGCGCGGGAGGCTC
ACGTTTCCCGCACTTCTGCGGGTATGTTCCGCATGTCCAAAAAACCCCCAATCGGTTTCCCCTTTCT
TGAGCGCAAGCTCAACGCGCTAGATGCGCGTGGTCAGCTTCCAGTTGACCTAAAAATAAGGGTTCC
CGATTACAAAGGGTACTGGTATCCTACGGGAGCTATCCGCGAGTACATGCTCGACCGCTTGCGGAT
CAGCAATGAGCTTTACAAAGCTCAGCATGGGCGTAATGCGCCACAATGGACCTCGCTAACGCGAA
GCGTTGAGCAAAACGAAAAAGATTGGGAAAGGTTAATCCATGGCACCGAAGCGCAAGAAGAGGA
GCAAGTCGAAGACTTCGAGGAGTGGCAACGCTCAATCCTCCTCCGTACAAAAGAAATACGCCAAC

AACAGATCGACAGAGACACGCGAAAGCGATGTGGCGGGCTTTCTGCTTGCTCACGATGCCTTAACA
GCCTCACGCCGCAAGATTTCGACGCGGCGGCGAGATGGGCGGAACGCCAAGCTCGAAAACACGGC
GGCTATGACGCCGCCGAAAAACACTACCGCAGCGAAAACCGCTGCAATCCCTATTGTGGCTACAG
GGAACTACCAACCCAGAAAAGAGCCTTCAAAAAAGGAGCATGACGACAGGAA

>ICBM5_gene_1
MADKATSPTTVSKKDRDDFELHVRLLSRAGYKTAEARTIAWLRGPAGLQEMLAGVKVAS

>ICBM5_gene_2
MSAIAGALISGGASLLGGLFGRSSASKQQARQNEYNKPINIRKRAEEGGFNPLLWAGQGNIQMQPGPSG
IMGSAIANAGLALADGMSEQRQLDLERTKLKQDQERLDALIEKQTIRPKVGGIYAGSQQTPSVARAPGR
PLMNGAPQPGSAPVFNPPTEYNPIPDDGPRLQTKVMRSDGMTSADPENPAEMEGDWWTWAREGTFW
QNNNEILRRNTPETLHYKGRDALFPKMIDGARKAHKKAQEDFEKNPPKRRKLKGVNPNLNDKKW

>ICBM5_gene_3
MVMNMSKYQRPTNTRRESRTIAGRFRGGKLAPVMASAFRESESAILSQQVTYELDPIAGRMITPIMAELI
SVYVPVQAIDALKNPEEAYAGNTEVVRDKLLSGTPLFGLEDESEISKRLGVNPISVGGVKKVNEAARLA
HNCAVNFLRQRKYVNTVKLLADNMNVTPALISQTVLDRLNAVLDPEDRVNGAVQLDLGNVRMPVEG
VGVEYNAVPVTTPGWNDTENDRAADPHYFSDHAAGIRLAIASENGRPAVYAVGEGQNAGGISLTDFY
NAELMDSLVRQMRQIVDDNPEYGEEMVTRWAHGLSVDNGKTPWILHQSQQMFGNQFRRAMDGANL
DVAQSDMMQSLEFTVPVPPTELGGVVITFASVKPDETLGSQPHPFLSTTWEATNYVSDELARDPEAVTM
RHLNSDVPMVDEDTRALYIGHNGLKKAYISYGFNRHLDPTTVEAKTAIWQLEVPMSVTPESVIYPEDLD
HYPFADQLAEVCTYQVSSTATVRTPMVFGPTPVEELAQIETDNVFEDV

>ICBM5_gene_4
MQVNDAKRSVLLVGLNGGEVSVISADGEIIATEGVTAGRHKCSSWVPFMSNEGDELSFSGDVVPMVPN
GGRVRPMAYGPGQFESGANPDFVVTSADRMARELDHKIRGLDQTAKKVEARIAQLNNLAERAKTPVE
VAKEEKVDVIDDDDVSDVRSSGDDSDDTGPDFGAEKVGAE

>ICBM5_gene_5
MNDAPKKSNSGRKQSWLRAYARGLKGGTPKTLCESLRLWSIPQEQAQDVTKKTYLNAADGPMIWLGE
QIVKRMEKAGYPSRIFCGYRSPEQQDKEFAEGDSKARAYQSPHQFYEAVDIIHKTKAWNVSQDYWDTL
AAIVRVVEREYAIDLVHGYDWGWDSAHIEIALWRQVPKRQIGKTGSNYPPSPWEREQRFKELLPAVYA
AQHRP

>ICBM5_gene_6
MCSDLIHIDGQQFACRKCNECITARKNGWVARAMAEKAVTAETFSVTLTYNDATQESRDAAKTFEYR
HVKNWIKNLRRQIEYTTGQTGLLRYLVAGERGSDKGRCHWHVILFCNADILTLGKMTHWPSGKLAAD
RAEIITTGKRKKRINWSLWPYGFVSFQEPDQWGMEYALAYALKDQFNIVSAAGTAREAHVSRTSAGMF
RMSKKPPIGFPFLERKLNALDARGQLPVDLKIRVPDYKGYWYPTGAIREYMLDRLRISNELYKAQHGR
NAPQWTSLTRSVEQNEKDWERLIHGTEAQEEEQVEDFEEWQRSILLRTKEIRQQQIDRDTRKRCGGLSA
CSRCLNSLTPQDFDAAARWAERQARKHGGYDAAEKHYRSENRCNPYCGYRELPTQKRAFKKGA

d) List of supplementary files

These files would have been too large to be included into the appendix, thus they can be found as supplementary files published either together with the corresponding paper (SI files S2-x) or in the Zenodo research data repository (SI files S4-x and S5-x):

https://doi.org/10.1038/s41396-019-0362-7

- SI_file_S2-2_Cobaviruses_annots.xlsx
- SI_file_S2-3_Cobaviruses_genomes.txt
- SI_file_S2-4_Metagenomes.xlsx
- SI_file_S2-5_Biogeographical_distribution.xlsx
- SI_file_S2-6_VICTOR_distances.xlsx
- SI_file_S2-7a_code.txt
- SI_file_S2-7b_code.txt
- SI_file_S2-7c_code.txt
- SI_file_S2-7d_code.txt

https://doi.org/10.5281/zenodo.11221490

- SI_file_S4-1_Host_strains_Alejandre-Colomo.xlsx
- SI_file_S4-2_RAPD_gel_all.emf
- SI_file_S4-3_Isolation_Phages_selected_RAPD-PCR.csv
- SI_file_S4-4_VIRIDIC_heatmap_143assem.pdf
- SI_file_S4-5_VIRIDIC_sim-dist_table_143_assem.tsv
- SI_file_S4-6_Isolation_VIRIDIC_heatmap_94unique_complete.pdf
- SI_file_S4-7_Proteomic_tree_965_input_phages.csv
- SI_file_S4-8_Whole_proteomic_tree_round.pdf
- SI_file_S4-9_Whole_proteomic_tree.pdf
- SI_file_S4-10_Gene_annotations_ICBM_phages.xlsx
- SI_file_S4-11_VGC_4_VIRIDIC_heatmap.pdf
- SI_file_S4-12_VGC_4_VIRIDIC_cluster_table.tsv
- SI_file_S4-13_VirClust_heatmap_VGC25-VGC9.PDF
- SI_file_S5-1_Sulfiviruses_hosts_16S-ITS_tree_identity_matrix.csv
- SI_file_S5-2_Sulfiviruses_hosts_16S_tree_identity_matrix.csv
- SI_file_S5-3_Sulfiviruses_hosts_genomes_distance_table_GGDC.csv
- SI_file_S5-4_Sulfiviruses_genome_annotations.csv

- SI_file_S5-5_Sulfiviruses_complete_genome_maps.pdf
- SI_file_S5-6_Sulfiviruses_hosts_plasmid_clustering_protein_heatmap_0.7_distance.PDF
- SI_file_S5-7a_Sulfiviruses_hosts_prophages_chromosomes_PHASTER.xlsx
- SI_file_S5-7b_Sulfiviruses_hosts_prophages_plasmids_PHASTER.xlsx
- SI_file_S5-7c_Sulfiviruses_hosts_prophages_plasmids_ProphageHunter.xlsx

Bibliography

Abbasifar, Reza; Griffiths, Mansel W.; Sabour, Parviz M.; Ackermann, Hans-Wolfgang; Vandersteegen, Katrien; Lavigne, Rob et al. (2014): Supersize me: Cronobacter sakazakii phage GAP32. In *Virology* 460-461, pp. 138–146. DOI: 10.1016/j.virol.2014.05.003.

Abedon, Stephen T. (2022): Bacteriophages, a Brief Introduction. In Stephen T. Abedon (Ed.): Bacteriophages as Drivers of Evolution. Cham: Springer International Publishing, pp. 3–14.

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.

Ackermann, H-W (2007): 5500 Phages examined in the electron microscope. In *Archives of virology* 152 (2), pp. 227–243. DOI: 10.1007/s00705-006-0849-1.

Adriaenssens, E; Tolstoy, Igor; Kropinski, Andrew; Moraru, Cristina; Wittman, J. (2020): ICTV Taxonomy Proposal 2020.133B: Create one new subfamily (Rhodovirinae) including seven genera (Caudovirales: Schitoviridae). Available online at https://ictv.global/ictv/proposals/2020.133B.R.Rhodovirinae.zip.

Adriaenssens, E. M.; am Kropinski; Turner, D; Krupovic, M.; Millard, A.; Dutilh, BE, Oksanen, HM et al. (2021): ICTV Taxonomy Proposal 2021.001B: Abolish the order Caudovirales and the families Myoviridae, Siphoviridae and Podoviridae (Caudoviricetes). Available online at https://ictv.global/ictv/proposals/2021.001B.R.abolish_Caudovirales.zip.

Aguilar, Alejandro; Peralta, Humberto; Mora, Yolanda; Díaz, Rafael; Vargas-Lagunas, Carmen; Girard, Lourdes; Mora, Jaime (2016): Genomic Comparison ofAgrobacterium pusenseStrains Isolated from Bean Nodules. In *Frontiers in microbiology* 7, p. 1720. DOI: 10.3389/fmicb.2016.01720.

Ahern, Stephen J.; Das, Mayukh; Bhowmick, Tushar Suvra; Young, Ry; Gonzalez, Carlos F. (2014): Characterization of novel virulent broad-host-range phages of Xylella fastidiosa and Xanthomonas. In *Journal of bacteriology* 196 (2), pp. 459–471. DOI: 10.1128/JB.01080-13.

Aiewsakun, Pakorn; Simmonds, Peter (2018): The genomic underpinnings of eukaryotic virus taxonomy: creating a sequence-based framework for family-level virus classification. In *Microbiome* 6 (1), p. 38. DOI: 10.1186/s40168-018-0422-7.

Aird, Daniel; Ross, Michael G.; Chen, Wei-Sheng; Danielsson, Maxwell; Fennell, Timothy; Russ, Carsten et al. (2011): Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. In *Genome biology* 12 (2), R18. DOI: 10.1186/gb-2011-12-2-r18.

Alanin, Katrine Wacenius Skov; Olsen, Nikoline S.; Djurhuus, Amaru Miranda; Carstens, Alexander Byth; Nielsen, Tue Kjærgaard; Wagner, Natalia et al. (2022): Three novel Erwinia billingiae phages represent three new genera isolated from organic waste. Available online at https://assets.researchsquare.com/files/rs-2102185/v1/fa01d72d-215c-4e8f-96d9-d24f3ba11d19.pdf?c=1664906155, checked on 11/11/2022.

Alberti, Adriana; Poulain, Julie; Engelen, Stefan; Labadie, Karine; Romac, Sarah; Ferrera, Isabel et al. (2017): Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. In *Scientific data* 4, p. 170093. DOI: 10.1038/sdata.2017.93.

Alejandre-Colomo, Carlota; Harder, Jens; Fuchs, Bernhard M.; Rosselló-Móra, Ramon; Amann, Rudolf (2020): High-throughput cultivation of heterotrophic bacteria during a spring phytoplankton bloom in the North Sea. In *Systematic and applied microbiology* 43 (2), p. 126066. DOI: 10.1016/j.syapm.2020.126066.

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.

Alonso, M. Carmen; Rodriguez, Jaime; Borrego, Juan J. (2000): Enumeration and isolation of viral particles from oligotrophic marine environments by tangential flow filtration. In *International microbiology : the official journal of the Spanish Society for Microbiology* (2), pp. 227–232. Available online at https://core.ac.uk/download/pdf/159084697.pdf, checked on 7/31/2023.

Amarillas, Luis; Estrada-Acosta, Mitzi; León-Chan, Ruben G.; López-Orona, Carlos; Lightbourn, Luis (2020): Complete genome sequence of Phobos: a novel bacteriophage with unusual genomic features that infects Pseudomonas syringae. In *Archives of virology*. DOI: 10.1007/s00705-020-04618-2.

Amgarten, Deyvid; Martins, Layla Farage; Lombardi, Karen Cristina; Antunes, Luciana Principal; Souza, Ana Paula Silva de; Nicastro, Gianlucca Gonçalves et al. (2017): Three novel Pseudomonas phages isolated from composting provide insights into the evolution and diversity of tailed phages. In *BMC Genomics* 18 (1), p. 346. DOI: 10.1186/s12864-017-3729-z.

Amin, S. A.; Hmelo, L. R.; van Tol, H. M.; Durham, B. P.; Carlson, L. T.; Heal, K. R. et al. (2015): Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. In *Nature* 522 (7554), pp. 98–101. DOI: 10.1038/nature14488.

Anderson, Donald M. (2009): Approaches to monitoring, control and management of harmful algal blooms (HABs). In *Ocean & Coastal Management* 52 (7), p. 342. DOI: 10.1016/j.ocecoaman.2009.04.006.

Andrade-Domínguez, Andrés; Kolter, Roberto (2016): Complete Genome Sequence of Pseudomonas aeruginosa Phage AAT-1. In *Genome Announc.* 4 (4). DOI: 10.1128/genomeA.00165-16.

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.

Ankrah, Nana Y. D.; Budinoff, Charles R.; Wilson, William H.; Wilhelm, Steven W.; Buchan, Alison (2014a): Genome sequence of the *Sulfitobacter* sp. strain 2047-infecting lytic phage {Phi}CB2047-B. In *Genome announcements* 2 (1), e00945-13. DOI: 10.1128/genomeA.00945-13.

Ankrah, Nana Y. D.; Budinoff, Charles R.; Wilson, William H.; Wilhelm, Steven W.; Buchan, Alison (2014b): Genome sequences of two temperate phages, ΦCB2047-A and ΦCB2047-C, infecting *Sulfitobacter* sp. Strain 2047. In *Genome announcements* 2 (3). DOI: 10.1128/genomeA.00108-14.

Aparna, S.; Parvathi, A.; Pradeep Ram, A. S.; Sime-Ngando, T. (2019): Genome analysis of Marinobacter phage AS1 suggests its close interactions with host Marinobacter sp. In *Aquat. Microb. Ecol.* 83 (2), pp. 119–129. DOI: 10.3354/ame01912.

Ariff, Amir; Wise, Michael J.; Kahler, Charlene M.; Tay, Chin Yen; Peters, Fanny; Perkins, Timothy T.; Chang, Barbara J. (2015): Novel Moraxella catarrhalis prophages display hyperconserved non-

structural genes despite their genomic diversity. In *BMC Genomics* 16, p. 860. DOI: 10.1186/s12864-015-2104-1.

Arndt, David; Grant, Jason R.; Marcu, Ana; Sajed, Tanvir; Pon, Allison; Liang, Yongjie; Wishart, David S. (2016): PHASTER: a better, faster version of the PHAST phage search tool. In *Nucleic acids research* 44 (W1), p. 21. DOI: 10.1093/nar/gkw387.

Atanasiu, C.; Su, T-J; Sturrock, S. S.; Dryden, D. T. F. (2002): Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. In *Nucl Acids Res* 30 (18), pp. 3936–3944. DOI: 10.1093/nar/gkf518.

Badawy, Shimaa; Pajunen, Maria I.; Haiko, Johanna; Baka, Zakaria A. M.; Abou-Dobara, Mohamed I.; El-Sayed, Ahmed K. A.; Skurnik, Mikael (2020): Identification and Functional Analysis of Temperate Siphoviridae Bacteriophages of Acinetobacter baumannii. In *Viruses* 12 (6). DOI: 10.3390/v12060604.

Bai, Yang; Müller, Daniel B.; Srinivas, Girish; Garrido-Oter, Ruben; Potthoff, Eva; Rott, Matthias et al. (2015): Functional overlap of the Arabidopsis leaf and root microbiota. In *Nature* 528 (7582), pp. 364–369. DOI: 10.1038/nature16192.

Bair, Catherine L.; Black, Lindsay W. (2007): A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. In *Journal of molecular biology* 366 (3), pp. 768–778. DOI: 10.1016/j.jmb.2006.11.051.

Bakenhus, Insa; Dlugosch, Leon; Billerbeck, Sara; Giebel, Helge-Ansgar; Milke, Felix; Simon, Meinhard (2017): Composition of total and cell-proliferating bacterioplankton community in early summer in the North Sea - Roseobacters are the most active component. In *Frontiers in microbiology* 8, p. 1771. DOI: 10.3389/fmicb.2017.01771.

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.

Bankevich, Anton; Nurk, Sergey; Antipov, Dmitry; Gurevich, Alexey A.; Dvorkin, Mikhail; Kulikov, Alexander S. et al. (2012): SPAdes: a new genome assembly algorithm and its applications to aingle-cell sequencing. In *Journal of Computational Biology* 19 (5), pp. 455–477. DOI: 10.1089/cmb.2012.0021.

Barr, Jeremy J.; Auro, Rita; Furlan, Mike; Whiteson, Katrine L.; Erb, Marcella L.; Pogliano, Joe et al. (2013): Bacteriophage adhering to mucus provide a non-host-derived immunity. In *PNAS* 110 (26), pp. 10771–10776. DOI: 10.1073/pnas.1305923110.

Barrero-Canosa, Jimena; Moraru, Cristina (2019): PhageFISH for monitoring phage infections at single cell level. In Martha R. J. Clokie, Andrew M. Kropinski, Rob Lavigne (Eds.): Bacteriophages. Methods and protocols. Methods in Molecular Biology., IV. New York: Humana Press; Springer (Springer protocols, volume 1898), pp. 1–26.

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.

Baum, Lisa; Nguyen, Mary T. H. D.; Jia, Yunke; Biazik, Joanna; Thomas, Torsten (2021): Characterization of a novel roseophage and the morphological and transcriptional response of the sponge

symbiont Ruegeria AU67 to infection. In *Environmental microbiology* 23 (5), pp. 2532–2549. DOI: 10.1111/1462-2920.15474.

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.

Belfort, Marlene (Ed.) (2005): Homing endonucleases and inteins. Berlin, Great Britain: Springer-Verlag (Nucleic Acids and Molecular Biology, vol. 16).

Bellas, Christopher M.; Schroeder, Declan C.; Edwards, Arwyn; Barker, Gary; Anesio, Alexandre M. (2020): Flexible genes establish widespread bacteriophage pan-genomes in cryoconite hole ecosystems. In *Nature communications* 11 (1), p. 4403. DOI: 10.1038/s41467-020-18236-8.

Belle, Archana; Landthaler, Markus; Shub, David A. (2002): Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. In *Genes & development* 16 (3), pp. 351–362. DOI: 10.1101/gad.960302.

Bench, Shellie R.; Hanson, Thomas E.; Williamson, Kurt E.; Ghosh, Dhritiman; Radosovich, Mark; Wang, Kui; Wommack, K. Eric (2007): Metagenomic characterization of Chesapeake Bay virioplankton. In *Applied and environmental microbiology* 73 (23), pp. 7629–7641. DOI: 10.1128/AEM.00938-07.

Benler, Sean; Koonin, Eugene V. (2021): Fishing for phages in metagenomes: what do we catch, what do we miss? In *Current opinion in virology* 49, pp. 142–150. DOI: 10.1016/j.coviro.2021.05.008.

Berge, Odile; Guinebretière, Marie-Hélène; Achouak, Wafa; Normand, Philippe; Heulin, Thierry (2002): Paenibacillus graminis sp. nov. and Paenibacillus odorifer sp. nov., isolated from plant roots, soil and food. In *International journal of systematic and evolutionary microbiology* 52 (Pt 2), pp. 607–616. DOI: 10.1099/00207713-52-2-607.

Bernheim, Aude; Millman, Adi; Ofir, Gal; Meitav, Gilad; Avraham, Carmel; Shomar, Helena et al. (2021): Prokaryotic viperins produce diverse antiviral molecules. In *Nature* 589 (7840), pp. 120–124. DOI: 10.1038/s41586-020-2762-2.

Bernheim, Aude; Sorek, Rotem (2020): The pan-immune system of bacteria: antiviral defence as a community resource. In *Nat Rev Micro* 18 (2), pp. 113–119. DOI: 10.1038/s41579-019-0278-2.

Berrios, Louis; Ely, Bert (2019): The Isolation and Characterization of Kronos, a Novel Caulobacter Rhizosphere Phage that is Similar to Lambdoid Phages. In *Curr Microbiol* 76 (5), pp. 558–565. DOI: 10.1007/s00284-019-01656-1.

Bertelli, Claire; Greub, Gilbert (2012): Lateral gene exchanges shape the genomes of amoeba-resisting microorganisms. In *Frontiers in cellular and infection microbiology* 2, p. 110. DOI: 10.3389/fcimb.2012.00110.

Bettarel, Yvan; Motegi, Chiaki; Weinbauer, Markus G.; Mari, Xavier (2016): Colonization and release processes of viruses and prokaryotes on artificial marine macroaggregates. In *FEMS Microbiology Letters* 363 (1), fnv216. DOI: 10.1093/femsle/fnv216.

Biddanda, Bopaiah; Benner, Ronald (1997): Carbon, nitrogen, and carbohydrate fluxes during the production of particulate and dissolved organic matter by marine phytoplankton. In *Limnol. Oceanogr.* 42 (3), pp. 506–518. DOI: 10.4319/lo.1997.42.3.0506.

Biggs, Tristan E. G.; Huisman, Jef; Brussaard, Corina P. D. (2021): Viral lysis modifies seasonal phytoplankton dynamics and carbon flow in the Southern Ocean. In *ISME J* 15 (12), pp. 3615–3622. DOI: 10.1038/s41396-021-01033-6.

Billerbeck, Sara; Orchard, Julia; Tindall, Brian J.; Giebel, Helge-Ansgar; Brinkhoff, Thorsten; Simon, Meinhard (2015): Description of Octadecabacter temperatus sp. nov., isolated from the southern North Sea, emended descriptions of the genus Octadecabacter and its species and reclassification of Octadecabacter jejudonensisPark and Yoon 2014 as Pseudooctadecabacter jejudonensis gen. nov., comb. nov. In *International journal of systematic and evolutionary microbiology* 65 (Pt 6), pp. 1967–1974. DOI: 10.1099/ijs.0.000205.

Bischoff, V; Adriaenssens, E M; Kropinski, A M; Duhaime, M; Moraru, C (2020): ICTV-proposal *Zobellviridae 2020.187B:* Create one new family (*Zobellviridae*) including one new subfamily (*Cobavirinae*), seven new genera and 12 new species (*Caudovirales*). Available online at https://ictv.global/ictv/proposals/2020.187B.R.Zobellviridae.zip.

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.

Bleriot, Ines; Trastoy, Rocío; Blasco, Lucia; Fernández-Cuenca, Felipe; Ambroa, Antón; Fernández-García, Laura et al. (2020): Genomic analysis of 40 prophages located in the genomes of 16 carbapenemase-producing clinical strains of Klebsiella pneumoniae. In *Microbial genomics* 6 (5). DOI: 10.1099/mgen.0.000369.

Bobay, Louis-Marie; Ochman, Howard (2018): Biological species in the viral world. In *Proceedings of the National Academy of Sciences of the United States of America* 115 (23), pp. 6040–6045. DOI: 10.1073/pnas.1717593115.

Boeckman, Justin; Korn, Abby; Yao, Guichun; Ravindran, Aravind; Gonzalez, Carlos; Gill, Jason (2022): Sheep in wolves' clothing: Temperate T7-like bacteriophages and the origins of the Autographiviridae. In *Virology* 568, pp. 86–100. DOI: 10.1016/j.virol.2022.01.013.

Bolduc, Benjamin; Jang, Ho Bin; Doulcier, Guilhem; You, Zhi-Qiang; Roux, Simon; Sullivan, Matthew B. (2017a): vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. In *PeerJ* 5, e3243. DOI: 10.7717/peerj.3243.

Bolduc, Benjamin; Youens-Clark, Ken; Roux, Simon; Hurwitz, Bonnie L.; Sullivan, Matthew B. (2017b): iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. In *The ISME journal* 11, 7-14. DOI: 10.1038/ismej.2016.89.

Bolger, Anthony M.; Lohse, Marc; Usadel, Bjoern (2014): Trimmomatic: a flexible trimmer for Illumina sequence data. In *Bioinformatics (Oxford, England)* 30 (15), pp. 2114–2120. DOI: 10.1093/bioinformatics/btu170.

Bollivar, David W.; Bernardoni, Brooke; Bockman, Matthew R.; Miller, Brenda M.; Russell, Daniel A.; Delesalle, Veronique A. et al. (2016): Complete Genome Sequences of Five Bacteriophages That Infect Rhodobacter capsulatus. In *Genome announcements* 4 (3). DOI: 10.1128/genomeA.00051-16.

Bondy-Denomy, Joe; Pawluk, April; Maxwell, Karen L.; Davidson, Alan R. (2013): Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. In *Nature* 493 (7432), pp. 429–432. DOI: 10.1038/nature11723.

Bondy-Denomy, Joseph; Qian, Jason; Westra, Edze R.; Buckling, Angus; Guttman, David S.; Davidson, Alan R.; Maxwell, Karen L. (2016): Prophages mediate defense against phage infection through diverse mechanisms. In *ISME J* 10 (12), pp. 2854–2866. DOI: 10.1038/ismej.2016.79.

Born, Yannick; Fieseler, Lars; Marazzi, Janine; Lurz, Rudi; Duffy, Brion; Loessner, Martin J. (2011): Novel virulent and broad-host-range Erwinia amylovora bacteriophages reveal a high degree of mosaicism and a relationship to Enterobacteriaceae phages. In *Applied and environmental microbiology* 77 (17), pp. 5945–5954. DOI: 10.1128/AEM.03022-10.

Bouzar, H.; Jones, J. B. (2001): Agrobacterium larrymoorei sp. nov., a pathogen isolated from aerial tumours of Ficus benjamina. In *International journal of systematic and evolutionary microbiology* 51 (Pt 3), pp. 1023–1026. DOI: 10.1099/00207713-51-3-1023.

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.

Breider, Sven; Sehar, Shama; Berger, Martine; Thomas, Torsten; Brinkhoff, Thorsten; Egan, Suhelen (2019): Genome sequence of Epibacterium ulvae strain DSM 24752T, an indigoidine-producing, macroalga-associated member of the marine Roseobacter group. In *Environmental Microbiome* 14 (1), p. 4. DOI: 10.1186/s40793-019-0343-5.

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.

Breitbart, Mya; Salamon, Peter; Andresen, Bjarne; Mahaffy, Joseph M.; Segall, Anca M.; Mead, David et al. (2002): Genomic analysis of uncultured marine viral communities. In *Proceedings of the National Academy of Sciences of the United States of America* 99 (22), pp. 14250–14255. DOI: 10.1073/pnas.202488399.

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.

Brum, Jennifer R.; Ignacio-Espinoza, J. Cesar; Roux, Simon; Doulcier, Guilhem; Acinas, Silvia G.; Alberti, Adriana et al. (2015): Patterns and ecological drivers of ocean viral communities. In *Science* 348 (6237), p. 1261498. DOI: 10.1126/science.1261498.

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.

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.

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.

Buchan, Alison; LeCleir, Gary R.; Gulvik, Christopher A.; González, José M. (2014): Master recyclers: features and functions of bacteria associated with phytoplankton blooms. In *Nature reviews. Microbiology* 12 (10), pp. 686–698. DOI: 10.1038/nrmicro3326.

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.

Bukhari, A. I.; Taylor, A. L. (1975): Influence of insertions on packaging of host sequences covalently linked to bacteriophage Mu DNA. In *Proceedings of the National Academy of Sciences of the United States of America* 72 (11), pp. 4399–4403. DOI: 10.1073/pnas.72.11.4399.

Bukhari, A. I.; Zipser, D. (1972): Random Insertion of Mu-1 DNA within a Single Gene. In *Nature New Biology* 236 (69), pp. 240–243. DOI: 10.1038/newbio236240a0.

Cady, Kyle C.; Bondy-Denomy, Joe; Heussler, Gary E.; Davidson, Alan R.; O'Toole, George A. (2012): The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages. In *Journal of bacteriology* 194 (21), pp. 5728–5738. DOI: 10.1128/JB.01184-12.

Cahill, Jesse; Young, Ry (2019): Phage Lysis: Multiple Genes for Multiple Barriers. In *Advances in virus research* 103, pp. 33–70. DOI: 10.1016/bs.aivir.2018.09.003.

Cai, Lanlan; Chen, Yue; Xiao, Shiwei; Liu, Riyue; He, Maoqiu; Zhang, Rui; Zeng, Qinglu (2023): Abundant and cosmopolitan lineage of cyanopodoviruses lacking a DNA polymerase gene. In *ISME J* 17 (2), pp. 252–262. DOI: 10.1038/s41396-022-01340-6.

Cai, Lanlan; Ma, Ruijie; Chen, Hong; Yang, Yunlan; Jiao, Nianzhi; Zhang, Rui (2019): A newly isolated roseophage represents a distinct member of Siphoviridae family. In *Virology journal* 16 (1), p. 128. DOI: 10.1186/s12985-019-1241-6.

Cai, Lanlan; Yang, Yunlan; Jiao, Nianzhi; Zhang, Rui (2015): Complete genome sequence of vB_DshP-R2C, a N4-like lytic roseophage. In *Marine genomics* 22, pp. 15–17. DOI: 10.1016/j.margen.2015.03.005.

Calendar, R; Abedon, S T (Eds.) (2006): The Bacteriophages. Second edition: Oxford University Press.

Camacho, Christiam; Coulouris, George; Avagyan, Vahram; Ma, Ning; Papadopoulos, Jason; Bealer, Kevin; Madden, Thomas L. (2009): BLAST+: architecture and applications. In *BMC bioinformatics* 10, p. 421. DOI: 10.1186/1471-2105-10-421.

Carl, Gaby; Jäckel, Claudia; Grützke, Josephine; Hertwig, Stefan; Grobbel, Mirjam; Malorny, Burkhard et al. (2017): Complete Genome Sequence of the Temperate Klebsiella pneumoniae Phage KPP5665-2. In *Genome Announc.* 5 (43). DOI: 10.1128/genomeA.01118-17.

Casjens, Sherwood R.; Gilcrease, Eddie B. (2009): Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. In *Methods in molecular biology (Clifton, N.J.)* 502, pp. 91–111. DOI: 10.1007/978-1-60327-565-1_7.

Castro-Mejía, Josué L.; Muhammed, Musemma K.; Kot, Witold; Neve, Horst; Franz, Charles M. A. P.; Hansen, Lars H. et al. (2015): Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. In *Microbiome* 3 (1), p. 64. DOI: 10.1186/s40168-015-0131-4.

Chafee, Meghan; Fernàndez-Guerra, Antonio; Buttigieg, Pier Luigi; Gerdts, Gunnar; Eren, A. Murat; Teeling, Hanno; Amann, Rudolf I. (2018): Recurrent patterns of microdiversity in a temperate coastal marine environment. In *The ISME journal* 12 (1), pp. 237–252. DOI: 10.1038/ismej.2017.165.

Chan, Jacqueline Z-M; Millard, Andrew D.; Mann, Nicholas H.; Schäfer, Hendrik (2014): Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like Roseophage specific genes. In *Frontiers in microbiology* 5, p. 506. DOI: 10.3389/fmicb.2014.00506.

Charif, D.; Lobry, J. R. (2007): SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In U. Bastolla, M. Porto, H.E. Roman, M. Vendruscolo (Eds.): Structural approaches to sequence evolution: Molecules, networks, populations. New York: Springer Verlag (Biological and Medical Physics, Biomedical Engineering), pp. 207–232.

Chassard, Christophe; Delmas, Eve; Lawson, Paul A.; Bernalier-Donadille, Annick (2008): Bacteroides xylanisolvens sp. nov., a xylan-degrading bacterium isolated from human faeces. In *International journal of systematic and evolutionary microbiology* 58 (Pt 4), pp. 1008–1013. DOI: 10.1099/ijs.0.65504-0.

Chen, Rujia; Xu, Xuling; Han, Pengjun; Hu, Yunjia; Qin, Hongbo; Li, Mengzhe et al. (2022): Complete Genome Sequence of Aeromonas hydrophila Bacteriophage BUCT552. In *Microbiol Resour Announc* 11 (2), e0117221. DOI: 10.1128/mra.01172-21.

Chen, Zhangran; Zhang, Jingyan; Lei, Xueqian; Zhang, Bangzhou; Cai, Guanjing; Zhang, Huajun et al. (2014): Influence of plaque-forming bacterium, *Rhodobacteraceae* sp. on the growth of *Chlorella vulgaris*. In *Bioresource technology* 169, pp. 784–788. DOI: 10.1016/j.biortech.2014.07.021.

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.

Choi, In Young; Lee, Ju-Hoon; Kim, Hye-Jin; Park, Mi-Kyung (2017): Isolation and Characterization of a Novel Broad-host-range Bacteriophage Infecting Salmonella enterica subsp. enterica for Biocontrol and Rapid Detection. In *Journal of microbiology and biotechnology* 27 (12), pp. 2151–2155. DOI: 10.4014/jmb.1711.11017.

Chow, L. T.; Bukhari, A. I. (1976): The invertible DNA segments of coliphages Mu and P1 are identical. In *Virology* 74 (1), pp. 242–248. DOI: 10.1016/0042-6822(76)90148-3.

Clark, James; Awah, Adey; Moreland, Russell; Liu, Mei; Gill, Jason J.; Ramsey, Jolene (2019): Complete Genome Sequence of Vibrio natriegens Phage Phriendly. In *Microbiol Resour Announc* 8 (40). DOI: 10.1128/MRA.01096-19.

Clokie, Martha R. J.; Mann, Nicholas H. (2006): Marine cyanophages and light. In *Environmental microbiology* 8 (12), pp. 2074–2082. DOI: 10.1111/j.1462-2920.2006.01171.x.

Cobbley, Hunter K.; Evans, Seth I.; Brown, Hannah M. F.; Eberhard, Braden; Eberhard, Nathaniel; Kim, Minji et al. (2022): Complete Genome Sequences of Six Chi-Like Bacteriophages That Infect Proteus and Klebsiella. In *Microbiol Resour Announc* 11 (4), e0121521. DOI: 10.1128/mra.01215-21.

Cochlan, William P.; Wikner, Johan; Steward, Grieg F.; Smith, David C.; Azam, Farroq (1993): Spatial distribution of viruses, bacteria and chlorophyll a in neritic, oceanic and estuarine environments. In *Marine ecology progress series* (92), pp. 77–87. Available online at https://www.int-res.com/articles/meps/92/m092p077.pdf, checked on 6/15/2023.

Coclet, Clément; Roux, Simon (2021): Global overview and major challenges of host prediction methods for uncultivated phages. In *Current opinion in virology* 49, pp. 117–126. DOI: 10.1016/j.coviro.2021.05.003.

Comeau, André M.; Short, Steven; Suttle, Curtis A. (2004): The use of degenerate-primed random amplification of polymorphic DNA (DP-RAPD) for strain-typing and inferring the genetic similarity among closely related viruses. In *Journal of virological methods* 118 (2), pp. 95–100. DOI: 10.1016/j.jviromet.2004.01.020.

Conn, H. J. (1942): Validity of the Genus Alcaligenes. In *Journal of bacteriology* 44 (3), pp. 353–360. DOI: 10.1128/jb.44.3.353-360.1942.

Coutinho, Felipe H.; Silveira, Cynthia B.; Gregoracci, Gustavo B.; Thompson, Cristiane C.; Edwards, Robert A.; Brussaard, Corina P. D. et al. (2017): Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. In *Nature communications* 8, p. 15955. DOI: 10.1038/ncomms15955.

Crona, Mikael; Avesson, Lotta; Sahlin, Margareta; Lundin, Daniel; Hinas, Andrea; Klose, Ralph et al. (2013): A rare combination of ribonucleotide reductases in the social amoeba *Dictyostelium discoideum*. In *The Journal of biological chemistry* 288 (12), pp. 8198–8208. DOI: 10.1074/jbc.M112.442434.

Cruz-López, Ricardo; Maske, Helmut (2016): The vitamin B1 and B12 required by the marine dinoflagellate *Lingulodinium polyedrum* can be provided by its associated bacterial community in culture. In *Frontiers in microbiology* 7, p. 560. DOI: 10.3389/fmicb.2016.00560.

Cuervo, Ana; Chagoyen, Mónica; Pulido-Cid, Mar; Camacho, Ana; Carrascosa, José L. (2013): Structural characterization of T7 tail machinery reveals a conserved tubular structure among other *Podoviridae* family members and suggests a common mechanism for DNA delivery. In *Bacteriophage* 3 (4), e27011. DOI: 10.4161/bact.27011.

Culot, A.; Grosset, N.; Gautier, M. (2019): Overcoming the challenges of phage therapy for industrial aquaculture: A review. In *Aquaculture* 513, p. 734423. DOI: 10.1016/j.aquaculture.2019.734423.

Dai, Yanhui; Yang, Shangbo; Zhao, Dan; Hu, Chuanmin; Xu, Wang; Anderson, Donald M. et al. (2023): Coastal phytoplankton blooms expand and intensify in the 21st century. In *Nature* 615 (7951), pp. 280–284. DOI: 10.1038/s41586-023-05760-y.

de Jonge, Patrick A.; Nobrega, Franklin L.; Brouns, Stan J. J.; Dutilh, Bas E. (2019): Molecular and Evolutionary Determinants of Bacteriophage Host Range. In *Trends in microbiology* 27 (1), pp. 51–63. DOI: 10.1016/j.tim.2018.08.006.

Decewicz, Przemyslaw; Dziewit, Lukasz; Golec, Piotr; Kozlowska, Patrycja; Bartosik, Dariusz; Radlinska, Monika (2019): Characterization of the virome of Paracoccus spp. (Alphaproteobacteria) by combined in silico and in vivo approaches. In *Scientific reports* 9 (1), p. 7899. DOI: 10.1038/s41598-019-44460-4.

Decewicz, Przemyslaw; Golec, Piotr; Szymczak, Mateusz; Radlinska, Monika; Dziewit, Lukasz (2020): Identification and Characterization of the First Virulent Phages, Including a Novel Jumbo Virus, Infecting Ochrobactrum spp. In *International journal of molecular sciences* 21 (6). DOI: 10.3390/ijms21062096.

Del Dot, Therese; Osawa, Ro; Stackebrandt, Erko (1993): Phascolarctobacterium faecium gen. nov, spec. nov., a Novel Taxon of the Sporomusa Group of Bacteria. In *Systematic and applied microbiology* 16 (3), pp. 380–384. DOI: 10.1016/S0723-2020(11)80269-9.

Delafont, Vincent; Brouke, Amélie; Bouchon, Didier; Moulin, Laurent; Héchard, Yann (2013): Microbiome of free-living amoebae isolated from drinking water. In *Water Research* 47 (19), pp. 6958–6965. DOI: 10.1016/j.watres.2013.07.047.

Delesalle, Veronique A.; Kuhn, Jens; Kropinski, Andrew; Adriaenssens, E M; Bollivar, David W. (2016a): ICTV Taxonomy Proposal 2016.050a-dB: To create one (1) new genus, Titanvirus, including two (2) new species in the family Siphoviridae. Available online at https://ictv.global/ictv/proposals/2016.050a-dB.A.v1.Titanvirus.pdf.

Delesalle, Veronique A.; Kuhn, Jens; Kropinski, Andrew; Adriaenssens, E M; Bollivar, David W. (2016b): ICTV Taxonomy Proposal 2016.056a-dB: To create one (1) new genus, Cronusvirus, including one (1) new species in the family Siphoviridae. Available online at https://ictv.global/ictv/proposals/2016.056a-dB.A.v1.Cronusvirus.pdf.

DeLong, John P.; van Etten, James L.; Al-Ameeli, Zeina; Agarkova, Irina V.; Dunigan, David D. (2023): The consumption of viruses returns energy to food chains. In *PNAS* 120 (1), e2215000120. DOI: 10.1073/pnas.2215000120.

Demina, Tatiana A.; Luhtanen, Anne-Mari; Roux, Simon; Oksanen, Hanna M. (2021): Virus-host interactions and genetic diversity of Antarctic sea ice bacteriophages. DOI: 10.1101/2021.05.28.446129.

Deng, Li; Ignacio-Espinoza, J. Cesar; Gregory, Ann C.; Poulos, Bonnie T.; Weitz, Joshua S.; Hugenholtz, Philip; Sullivan, Matthew B. (2014): Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. In *Nature* 513 (7517), pp. 242–245. DOI: 10.1038/nature13459.

Dhillon, E. K.; Dhillon, T. S.; Lai, A. N.; Linn, S. (1980a): Host range, immunity and antigenic properties of lambdoid coliphage HK97. In *The Journal of general virology* 50 (1), pp. 217–220. DOI: 10.1099/0022-1317-50-1-217.

Dhillon, E. K.; Dhillon, T. S.; Lam, Y. Y.; Tsang, A. H. (1980b): Temperate coliphages: classification and correlation with habitats. In *Applied and environmental microbiology* 39 (5), pp. 1046–1053. DOI: 10.1128/aem.39.5.1046-1053.1980.

Dhillon, T. S.; Dhillon, E. K.; Chau, H. C.; Li, W. K.; Tsang, A. H. (1976): Studies on bacteriophage distribution: virulent and temperate bacteriophage content of mammalian feces. In *Applied and environmental microbiology* 32 (1), pp. 68–74. DOI: 10.1128/aem.32.1.68-74.1976.

Didelot, Xavier; Wilson, Daniel J. (2015): ClonalFrameML: efficient inference of recombination in whole bacterial genomes. In *PLoS computational biology* 11 (2), e1004041. DOI: 10.1371/journal.pcbi.1004041.

Ding, Hao; Moksa, Michelle M.; Hirst, Martin; Beatty, J. Thomas (2014): Draft Genome Sequences of Six Rhodobacter capsulatus Strains, YW1, YW2, B6, Y262, R121, and DE442. In *Genome announcements* 2 (1). DOI: 10.1128/genomeA.00050-14.

Dion, Moïra B.; Oechslin, Frank; Moineau, Sylvain (2020): Phage diversity, genomics and phylogeny. In *Nat Rev Micro* 18 (3), pp. 125–138. DOI: 10.1038/s41579-019-0311-5.

Dlugosch, Leon; Bunse, Carina; Bunk, Boyke; Böttcher, Lea; Tran, Den Quoc; Dittmar, Thorsten et al. (2023): Naturally induced biphasic phytoplankton spring bloom reveals rapid and distinct substrate and bacterial community dynamics. In *FEMS microbiology ecology* 99 (8). DOI: 10.1093/femsec/fiad078.

Dogs, Marco; Wemheuer, Bernd; Wolter, Laura; Bergen, Nils; Daniel, Rolf; Simon, Meinhard; Brinkhoff, Thorsten (2017): Rhodobacteraceae on the marine brown alga Fucus spiralis are abundant and show physiological adaptation to an epiphytic lifestyle. In *Systematic and applied microbiology* 40 (6), pp. 370–382. DOI: 10.1016/j.syapm.2017.05.006.

Dong, Zhaoxia; Xing, Shaozhen; Liu, Jin; Tang, Xizhe; Ruan, Lifang; Sun, Ming et al. (2018): Isolation and characterization of a novel phage Xoo-sp2 that infects Xanthomonas oryzae pv. oryzae. In *Journal of General Virology* 99 (10), pp. 1453–1462. DOI: 10.1099/jgv.0.001133.

Doron, Shany; Melamed, Sarah; Ofir, Gal; Leavitt, Azita; Lopatina, Anna; Keren, Mai et al. (2018): Systematic discovery of antiphage defense systems in the microbial pangenome. In *Science (New York, N.Y.)* 359 (6379). DOI: 10.1126/science.aar4120.

Downes, Julia; Sutcliffe, Iain C.; Hofstad, Tor; Wade, William G. (2006): Prevotella bergensis sp. nov., isolated from human infections. In *International journal of systematic and evolutionary microbiology* 56 (Pt 3), pp. 609–612. DOI: 10.1099/ijs.0.63888-0.

Dreiseikelmann, Brigitte; Bunk, Boyke; Spröer, Cathrin; Rohde, Manfred; Nimtz, Manfred; Wittmann, Johannes (2017): Characterization and genome comparisons of three Achromobacter phages of the family Siphoviridae. In *Arch Virol* 162 (8), pp. 2191–2201. DOI: 10.1007/s00705-017-3347-8.

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.

Duhaime, Melissa B.; Solonenko, Natalie; Roux, Simon; Verberkmoes, Nathan C.; Wichels, Antje; Sullivan, Matthew B. (2017): Comparative omics and trait analyses of marine Pseudoalteromonas phages advance the phage OTU concept. In *Frontiers in microbiology* 8, p. 1241. DOI: 10.3389/fmicb.2017.01241.

Dujon, B. (1989): Group I introns as mobile genetic elements: facts and mechanistic speculations--a review. In *Gene* 82 (1), pp. 91–114. DOI: 10.1016/0378-1119(89)90034-6.

Dwivedi, Bhakti; Xue, Bingjie; Lundin, Daniel; Edwards, Robert A.; Breitbart, Mya (2013): A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. In *BMC evolutionary biology* 13, p. 33. DOI: 10.1186/1471-2148-13-33.

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

Eggerth, A. H.; Gagnon, B. H. (1933): The Bacteroides of Human Feces. In *Journal of bacteriology* 25 (4), pp. 389–413. DOI: 10.1128/jb.25.4.389-413.1933.

Ehrenreich, A.; Widdel, F. (1994): Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. In *Applied and environmental microbiology* 60 (12), pp. 4517–4526.

Eilers, H.; Pernthaler, J.; Amann, R. (2000a): Succession of pelagic marine bacteria during enrichment: a close look at cultivation-induced shifts. In *Applied and environmental microbiology* 66 (11), pp. 4634–4640. DOI: 10.1128/AEM.66.11.4634-4640.2000.

Eilers, H.; Pernthaler, J.; Glöckner, F. O.; Amann, R. (2000b): Culturability and In situ abundance of pelagic bacteria from the North Sea. In *Applied and environmental microbiology* 66 (7), pp. 3044–3051. DOI: 10.1128/AEM.66.7.3044-3051.2000.

Eilers, H.; Pernthaler, J.; Peplies, J.; Glöckner, F. O.; Gerdts, G.; Amann, R. (2001): Isolation of novel pelagic bacteria from the German bight and their seasonal contributions to surface picoplankton. In *Applied and environmental microbiology* 67 (11), pp. 5134–5142. DOI: 10.1128/AEM.67.11.5134-5142.2001.

El Yacoubi, Basma; Bailly, Marc; Crécy-Lagard, Valérie de (2012): Biosynthesis and function of posttranscriptional modifications of transfer RNAs. In *Annual review of genetics* 46, pp. 69–95. DOI: 10.1146/annurev-genet-110711-155641.

Eller, Monique R.; Salgado, Rafael L.; Vidigal, Pedro M. P.; Alves, Maura P.; Dias, Roberto S.; Oliveira, Leandro L. de et al. (2013): Complete Genome Sequence of the Pseudomonas fluorescens Bacteriophage UFV-P2. In *Genome Announc.* 1 (1). DOI: 10.1128/genomeA.00006-12.

Ely, Bert; Berrios, Louis; Thomas, Quill (2022): S2B, a Temperate Bacteriophage That Infects Caulobacter Crescentus Strain CB15. In *Curr Microbiol* 79 (4), p. 98. DOI: 10.1007/s00284-022-02799-4.

Enav, Hagay; Béjà, Oded; Mandel-Gutfreund, Yael (2012): Cyanophage tRNAs may have a role in cross-infectivity of oceanic Prochlorococcus and Synechococcus hosts. In *ISME J* 6 (3), pp. 619–628. DOI: 10.1038/ismej.2011.146.

Engelhardt, Tim; Sahlberg, Monika; Cypionka, Heribert; Engelen, Bert (2011): Induction of prophages from deep-subseafloor bacteria. In *Environmental microbiology reports* 3 (4), pp. 459–465. DOI: 10.1111/j.1758-2229.2010.00232.x.

Erdrich, Sebastian H.; Sharma, Vikas; Schurr, Ulrich; Arsova, Borjana; Frunzke, Julia (2022): Isolation of Novel Xanthomonas Phages Infecting the Plant Pathogens X. translucens and X. campestris. In *Viruses* 14 (7). DOI: 10.3390/v14071449.

Esakkiraj, Palanichamy; Usha, Rajamony; Palavesam, Arunachalam; Immanuel, Grasian (2012): Solid-state production of esterase using fish processing wastes by Bacillus altitudinis AP-MSU. In *Food and Bioproducts Processing* 90 (3), pp. 370–376. DOI: 10.1016/j.fbp.2011.12.008.

Escarmis, Cristina; Garcia, Pedro; Mendez, Enrique; López, Rubens; Salas, Margarita; García, Ernesto (1985): Inverted terminal repeats and terminal proteins of the genomes of pneumococcal phages. In *Gene* 36 (3), pp. 341–348. DOI: 10.1016/0378-1119(85)90189-1.

Essoh, Christiane; Vernadet, Jean-Philippe; Vergnaud, Gilles; Coulibaly, Adama; Kakou-N'Douba, Adele; N'Guetta, Assavo S-P et al. (2020): Characterization of sixteen Achromobacter xylosoxidans phages from Abidjan, Cote d'Ivoire, isolated on a single clinical strain. In *Archives of virology* 165 (3), pp. 725–730. DOI: 10.1007/s00705-019-04511-7.

Fancello, Laura; Trape, Sébatien; Robert, Catherine; Boyer, Mickaël; Popgeorgiev, Nikolay; Raoult, Didier; Desnues, Christelle (2013): Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. In *ISME J* 7 (2), pp. 359–369. DOI: 10.1038/ismej.2012.101.

Farlow, Jason; Freyberger, Helen R.; He, Yunxiu; Ward, Amanda M.; Rutvisuttinunt, Wiriya; Li, Tao et al. (2020): Complete Genome Sequences of 10 Phages Lytic against Multidrug-Resistant Pseudomonas aeruginosa. In *Microbiol Resour Announc* 9 (29). DOI: 10.1128/MRA.00503-20.

Farris, James S. (1972): Estimating phylogenetic trees from distance matrices. In *The American Naturalist* 106 (951), pp. 645–668. DOI: 10.1086/282802.

Feiner, Ron; Argov, Tal; Rabinovich, Lev; Sigal, Nadejda; Borovok, Ilya; Herskovits, Anat A. (2015): A new perspective on lysogeny: prophages as active regulatory switches of bacteria. In *Nat Rev Micro* 13 (10), pp. 641–650. DOI: 10.1038/nrmicro3527.

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.

Fiebig, Anne; Pradella, Silke; Petersen, Jörn; Päuker, Orsola; Michael, Victoria; Lünsdorf, Heinrich et al. (2013): Genome of the R-body producing marine alphaproteobacterium *Labrenzia alexandrii* type strain (DFL-11(T)). In *Standards in genomic sciences* 7 (3), pp. 413–426. DOI: 10.4056/sigs.3456959.

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.

Fogg, Paul C. M.; Hynes, Alexander P.; Digby, Elizabeth; Lang, Andrew S.; Beatty, J. Thomas (2011): Characterization of a newly discovered Mu-like bacteriophage, RcapMu, in Rhodobacter capsulatus strain SB1003. In *Virology* 421 (2), pp. 211–221. DOI: 10.1016/j.virol.2011.09.028.

Ford, Brian E.; Sun, Bruce; Carpino, James; Chapler, Elizabeth S.; Ching, Jane; Choi, Yoon et al. (2014): Frequency and fitness consequences of bacteriophage φ6 host range mutations. In *PloS one* 9 (11), e113078. DOI: 10.1371/journal.pone.0113078.

Fu, Chaoqun; Zhao, Qin; Li, Zhiying; Wang, Yongxia; Zhang, Shiying; Lai, Yonghong et al. (2017): Complete genome sequence of Halomonas ventosae virulent halovirus QHHSV-1. In *Arch Virol* 162 (10), pp. 3215–3219. DOI: 10.1007/s00705-017-3415-0.

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

Fukui, Youhei; Abe, Mahiko; Kobayashi, Masahiro; Satomi, Masataka (2015): Sulfitobacter pacificus sp. nov., isolated from the red alga Pyropia yezoensis. In *Antonie van Leeuwenhoek* 107 (5), pp. 1155–1163. DOI: 10.1007/s10482-015-0407-5.

Gantzer, Christophe; Henny, Joseph; Schwartzbrod, Louis (2002): Bacteroides fragilis and Escherichia coli bacteriophages in human faeces. In *International journal of hygiene and environmental health* 205 (4), pp. 325–328. DOI: 10.1078/1438-4639-00152.

Gao, Linyi; Altae-Tran, Han; Böhning, Francisca; Makarova, Kira S.; Segel, Michael; Schmid-Burgk, Jonathan L. et al. (2020): Diverse enzymatic activities mediate antiviral immunity in prokaryotes. In *Science (New York, N.Y.)* 369 (6507), pp. 1077–1084. DOI: 10.1126/science.aba0372.

García-Alcalde, Fernando; Okonechnikov, Konstantin; Carbonell, José; Cruz, Luis M.; Götz, Stefan; Tarazona, Sonia et al. (2012): Qualimap: evaluating next-generation sequencing alignment data. In *Bioinformatics (Oxford, England)* 28 (20), pp. 2678–2679. DOI: 10.1093/bioinformatics/bts503.

Garneau, Julian R.; Depardieu, Florence; Fortier, Louis-Charles; Bikard, David; Monot, Marc (2017): PhageTerm: a tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. In *Scientific reports* 7 (1), p. 8292. DOI: 10.1038/s41598-017-07910-5.

Gerdts, Gunnar; Wichels, Antje; Dpke, Hilke; Klings, Karl-Walter; Gunkel, Wilfried; Schtt, Christian (2004): 40-year long-term study of microbial parameters near Helgoland (German Bight, North Sea): historical view and future perspectives. In *Helgol Mar Res* 58 (4), pp. 230–242. DOI: 10.1007/s10152-004-0189-z.

Giebel, Helge-Ansgar; Kalhoefer, Daniela; Lemke, Andreas; Thole, Sebastian; Gahl-Janssen, Renate; Simon, Meinhard; Brinkhoff, Thorsten (2011): Distribution of *Roseobacter* RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. In *The ISME journal* 5 (1), pp. 8–19. DOI: 10.1038/ismej.2010.87.

Glukhov, Anatoly S.; Krutilina, Antonina I.; Shlyapnikov, Michael G.; Severinov, Konstantin; Lavysh, Daria; Kochetkov, Vladimir V. et al. (2012): Genomic analysis of Pseudomonas putida phage tf with localized single-strand DNA interruptions. In *PloS one* 7 (12), e51163. DOI: 10.1371/journal.pone.0051163.

Gobler, Christopher J. (2020): Climate Change and Harmful Algal Blooms: Insights and perspective. In *Harmful algae* 91, p. 101731. DOI: 10.1016/j.hal.2019.101731.

Göker, Markus; García-Blázquez, Gema; Voglmayr, Hermann; Tellería, M. Teresa; Martín, María P. (2009): Molecular taxonomy of phytopathogenic fungi: a case study in *Peronospora*. In *PloS one* 4 (7), e6319. DOI: 10.1371/journal.pone.0006319.

Göller, Pauline C.; Elsener, Tabea; Lorgé, Dominic; Radulovic, Natasa; Bernardi, Viona; Naumann, Annika et al. (2021): Multi-species host range of staphylococcal phages isolated from wastewater. In *Nat Commun* 12 (1), p. 6965. DOI: 10.1038/s41467-021-27037-6.

Golomidova, Alla K.; Kulikov, Eugene E.; Prokhorov, Nikolai S.; Guerrero-Ferreira, Ricardo C.; Knirel, Yuriy A.; Kostryukova, Elena S. et al. (2016): Branched lateral tail fiber organization in T5-like bacteriophages DT57C and DT571/2 is revealed by genetic and functional analysis. In *Viruses* 8 (1). DOI: 10.3390/v8010026.

Goodridge, Lawrence D.; Bisha, Bledar (2011): Phage-based biocontrol strategies to reduce foodborne pathogens in foods. In *Bacteriophage* 1 (3), pp. 130–137. DOI: 10.4161/bact.1.3.17629.

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.

Gosink, J. J.; Herwig, R. P.; Staley, J. T. (1997): Octadecabacter arcticus gen. nov., sp. nov., and O. antarcticus, sp. nov., Nonpigmented, Psychrophilic Gas Vacuolate Bacteria from Polar Sea Ice and Water. In *Systematic and applied microbiology* 20 (3), pp. 356–365. DOI: 10.1016/S0723-2020(97)80003-3.

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.

Green, D. H.; Hart, M. C.; Blackburn, S. I.; Bolch, C. J.S. (2010): Bacterial diversity of *Gymnodinium catenatum* and its relationship to dinoflagellate toxicity. In *Aquat. Microb. Ecol.* 61 (1), pp. 73–87. DOI: 10.3354/ame01437.

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.

Gross, Miryam; Marianovsky, Irina; Glaser, Gad (2006): MazG -- a regulator of programmed cell death in *Escherichia coli*. In *Molecular microbiology* 59 (2), pp. 590–601. DOI: 10.1111/j.1365-2958.2005.04956.x.

Guannel, M. L.; Horner-Devine, M. C.; Rocap, G. (2011): Bacterial community composition differs with species and toxigenicity of the diatom *Pseudo-nitzschia*. In *Aquat. Microb. Ecol.* 64 (2), pp. 117–133. DOI: 10.3354/ame01513.

Gunathilake, K. M. Damitha; Tremblay, Denise M.; Plante, Pier-Luc; Jensen, Ellen C.; Nickerson, Kenneth W.; Moineau, Sylvain (2022): Genome Sequence of SN1, a Bacteriophage That Infects

Sphaerotilus natans and Pseudomonas aeruginosa. In *Microbiol Resour Announc* 11 (9), e0047822. DOI: 10.1128/mra.00478-22.

Guo, Fei; Liu, Zheng; Fang, Ping-An; Zhang, Qinfen; Wright, Elena T.; Wu, Weimin et al. (2014): Capsid expansion mechanism of bacteriophage T7 revealed by multistate atomic models derived from cryo-EM reconstructions. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (43), E4606-14. DOI: 10.1073/pnas.1407020111.

Gutiérrez, Diana; Martín-Platero, Antonio M.; Rodríguez, Ana; Martínez-Bueno, Manuel; García, Pilar; Martínez, Beatriz (2011): Typing of bacteriophages by randomly amplified polymorphic DNA (RAPD)-PCR to assess genetic diversity. In *FEMS Microbiology Letters* 322. DOI: 10.1111/j.1574-6968.2011.02342.x.

Guy, Lionel; Roat Kultima, Jens; Andersson, Siv G. E. (2010): genoPlotR: comparative gene and genome visualization in R. In *Bioinformatics* 26 (18), pp. 2334–2335. DOI: 10.1093/bioinformatics/btq413.

Haaber, Jakob; Rousseau, Geneviève M.; Hammer, Karin; Moineau, Sylvain (2009): Identification and characterization of the phage gene sav, involved in sensitivity to the lactococcal abortive infection mechanism AbiV. In *Appl. Environ. Microbiol.* 75 (8), pp. 2484–2494. DOI: 10.1128/AEM.02093-08.

Hahnke, Richard L.; Bennke, Christin M.; Fuchs, Bernhard M.; Mann, Alexander J.; Rhiel, Erhard; Teeling, Hanno et al. (2015): Dilution cultivation of marine heterotrophic bacteria abundant after a spring phytoplankton bloom in the North Sea. In *Environmental microbiology* 17 (10), pp. 3515–3526. DOI: 10.1111/1462-2920.12479.

Hahnke, Richard L.; Harder, Jens (2013): Phylogenetic diversity of Flavobacteria isolated from the North Sea on solid media. In *Systematic and applied microbiology* 36 (7), pp. 497–504. DOI: 10.1016/j.syapm.2013.06.006.

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.

Hammerl, Jens A.; Göllner, Cornelia; Al Dahouk, Sascha; Nöckler, Karsten; Reetz, Jochen; Hertwig, Stefan (2016): Analysis of the First Temperate Broad Host Range Brucellaphage (BiPBO1) Isolated from B. inopinata. In *Frontiers in microbiology* 7, p. 24. DOI: 10.3389/fmicb.2016.00024.

Hampton, Hannah G.; Watson, Bridget N. J.; Fineran, Peter C. (2020): The arms race between bacteria and their phage foes. In *Nature* 577 (7790), pp. 327–336. DOI: 10.1038/s41586-019-1894-8.

Han, Pengjun; Hu, Yunjia; An, Xiaoping; Song, Lihua; Fan, Huahao; Tong, Yigang (2021): Biochemical and genomic characterization of a novel bacteriophage BUCT555 lysing Stenotrophomonas maltophilia. In *Virus research* 301, p. 198465. DOI: 10.1016/j.virusres.2021.198465.

Han, Ying; Zhang, Yan; Zhao, Junyi; Cheng, Kai (2017): Cyanosiphovirus S-ESS1 Infecting Marine Synechococcus (Chroococcales) Almost Shows No Genetic Relationship to Known Cyanosiphoviruses. In *gab*. DOI: 10.5376/gab.2017.08.0002.

Hansen, T. A.; Veldkamp, H. (1973): Rhodopseudomonas sulfidophila, nov. spec., a new species of the purple nonsulfur bacteria. In *Archiv fur Mikrobiologie* 92 (1), pp. 45–58. DOI: 10.1007/BF00409510.

Hardies, Stephen C.; Comeau, André M.; Serwer, Philip; Suttle, Curtis A. (2003): The complete sequence of marine bacteriophage VpV262 infecting vibrio parahaemolyticus indicates that an ancestral

component of a T7 viral supergroup is widespread in the marine environment. In *Virology* 310 (2), pp. 359–371. DOI: 10.1016/S0042-6822(03)00172-7.

Hardies, Stephen C.; Hwang, Yeon J.; Hwang, Chung Y.; Jang, Gwang I.; Cho, Byung C. (2013): Morphology, physiological characteristics, and complete sequence of marine bacteriophage ϕRIO-1 infecting *Pseudoalteromonas marina*. In *Journal of virology* 87 (16), pp. 9189–9198. DOI: 10.1128/JVI.01521-13.

Hardies, Stephen C.; Thomas, Julie A.; Black, Lindsay; Weintraub, Susan T.; Hwang, Chung Y.; Cho, Byung C. (2016): Identification of structural and morphogenesis genes of *Pseudoalteromonas* phage φRIO-1 and placement within the evolutionary history of *Podoviridae*. In *Virology* 489, pp. 116–127. DOI: 10.1016/j.virol.2015.12.005.

Hardy, Aël; Kever, Larissa; Frunzke, Julia (2023): Antiphage small molecules produced by bacteria - beyond protein-mediated defenses. In *Trends in microbiology* 31 (1), pp. 92–106. DOI: 10.1016/j.tim.2022.08.001.

Hassan, Yousef I.; Lepp, Dion; Li, Xiu-Zhen; Zhou, Ting (2015): Insights into the Hydrocarbon Tolerance of Two Devosia Isolates, D. chinhatensis Strain IPL18T and D. geojensis Strain BD-c194T, via Whole-Genome Sequence Analysis. In *Genome announcements* 3 (4). DOI: 10.1128/genomeA.00890-15.

Hayashi, Wataru; Iimura, Masaki; Horiuchi, Kazuki; Arai, Eriko; Natori, Tatsuya; Suzuki, Shin et al. (2021): Occurrence of blaNDM-1 in a Clinical Isolate of Acinetobacter lwoffii in Japan: Comparison of blaNDM-1-Harboring Plasmids between A. lwoffii and A. pittii Originated from a Hospital Sink. In *Japanese journal of infectious diseases* 74 (3), pp. 252–254. DOI: 10.7883/yoken.JJID.2020.806.

Heinrichs, Mara E.; Tebbe, Dennis A.; Wemheuer, Bernd; Niggemann, Jutta; Engelen, Bert (2020): Impact of Viral Lysis on the Composition of Bacterial Communities and Dissolved Organic Matter in Deep-Sea Sediments. In *Viruses* 12 (9). DOI: 10.3390/v12090922.

Heins, Anneke; Amann, Rudolf I.; Harder, Jens (2021): Cultivation of particle-associated heterotrophic bacteria during a spring phytoplankton bloom in the North Sea. In *Systematic and applied microbiology* 44 (5), p. 126232. DOI: 10.1016/j.syapm.2021.126232.

Heins, Anneke; Harder, Jens (2022): Particle-associated bacteria in seawater dominate the colony-forming microbiome on ZoBell marine agar. In *FEMS microbiology ecology* 99 (1). DOI: 10.1093/femsec/fiac151.

Herlemann, Daniel Pr; Labrenz, Matthias; Jürgens, Klaus; Bertilsson, Stefan; Waniek, Joanna J.; Andersson, Anders F. (2011): Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. In *ISME J* 5 (10), pp. 1571–1579. DOI: 10.1038/ismej.2011.41.

Heyerhoff, Benedikt; Heinrichs, Mara Elena; Bunse, Carina; Dlugosch, Leon; Böttcher, Lea; Niggemann, Jutta et al. (in prep.): Bacterioplankton-associated viruses respond in switching infection strategies during a spring bloom mesocosm experiment.

Hille, Frank; Richter, Hagen; Wong, Shi Pey; Bratovič, Majda; Ressel, Sarah; Charpentier, Emmanuelle (2018): The Biology of CRISPR-Cas: Backward and Forward. In *Cell* 172 (6), pp. 1239–1259. DOI: 10.1016/j.cell.2017.11.032.

Hiraishi, A.; Ueda, Y. (1995): Isolation and characterization of Rhodovulum strictum sp. nov. and some other purple nonsulfur bacteria from colored blooms in tidal and seawater pools. In *International journal of systematic bacteriology* 45 (2), pp. 319–326. DOI: 10.1099/00207713-45-2-319.

Holmfeldt, Karin; Middelboe, Mathias; Nybroe, Ole; Riemann, Lasse (2007): Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. In *Applied and environmental microbiology* 73 (21), pp. 6730–6739. DOI: 10.1128/AEM.01399-07.

Holmfeldt, Karin; Solonenko, Natalie; Shah, Manesh; Corrier, Kristen; Riemann, Lasse; Verberkmoes, Nathan C.; Sullivan, Matthew B. (2013): Twelve previously unknown phage genera are ubiquitous in global oceans. In *Proceedings of the National Academy of Sciences of the United States of America* 110 (31), pp. 12798–12803. DOI: 10.1073/pnas.1305956110.

Hoshina, Ryo; Shimizu, Mayumi; Makino, Yoichi; Haruyama, Yoshihiro; Ueda, Shin-ichiro; Kato, Yutaka et al. (2010): Isolation and characterization of a virus (CvV-BW1) that infects symbiotic algae of *Paramecium bursaria* in Lake Biwa, Japan. In *Virology journal* 7, p. 222. DOI: 10.1186/1743-422X-7-222.

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.

Hu, Bo; Margolin, William; Molineux, Ian J.; Liu, Jun (2013): The bacteriophage T7 virion undergoes extensive structural remodeling during infection. In *Science (New York, N.Y.)* 339 (6119), pp. 576–579. DOI: 10.1126/science.1231887.

Huang, Sijun; Zhang, Yongyu; Chen, Feng; Jiao, Nianzhi (2011): Complete genome sequence of a marine roseophage provides evidence into the evolution of gene transfer agents in alphaproteobacteria. In *Virology journal* 8, p. 124. DOI: 10.1186/1743-422X-8-124.

Hunt, Dana E.; Klepac-Ceraj, Vanja; Acinas, Silvia G.; Gautier, Clement; Bertilsson, Stefan; Polz, Martin F. (2006): Evaluation of 23S rRNA PCR primers for use in phylogenetic studies of bacterial diversity. In *Applied and environmental microbiology* 72 (3), pp. 2221–2225. DOI: 10.1128/AEM.72.3.2221-2225.2006.

Hurwitz, Bonnie L.; Hallam, Steven J.; Sullivan, Matthew B. (2013): Metabolic reprogramming by viruses in the sunlit and dark ocean. In *Genome biology* 14 (11), R123. DOI: 10.1186/gb-2013-14-11-r123.

Hussain, Fatima Aysha; Dubert, Javier; Elsherbini, Joseph; Murphy, Mikayla; VanInsberghe, David; Arevalo, Philip et al. (2021): Rapid evolutionary turnover of mobile genetic elements drives bacterial resistance to phages. In *Science (New York, N.Y.)* 374 (6566), pp. 488–492. DOI: 10.1126/science.abb1083.

Hutinet, Geoffrey; Kot, Witold; Cui, Liang; Hillebrand, Roman; Balamkundu, Seetharamsingh; Gnanakalai, Shanmugavel et al. (2019): 7-Deazaguanine modifications protect phage DNA from host restriction systems. In *Nature communications* 10 (1), p. 5442. DOI: 10.1038/s41467-019-13384-y.

Hutinet, Geoffrey; Swarjo, Manal A.; Crécy-Lagard, Valérie de (2017): Deazaguanine derivatives, examples of crosstalk between RNA and DNA modification pathways. In *RNA biology* 14 (9), pp. 1175–1184. DOI: 10.1080/15476286.2016.1265200.

Hwang, Chung Y.; Cho, Yirang; Jang, Gwang I.; Cho, Byung C.; Hardies, Stephen C. (2020): Complete Genome Sequence of Sulfitobacter Phage ϕGT1, Isolated from a Tidal Flat. In *Microbiol Resour Announc* 9 (33). DOI: 10.1128/MRA.00779-20.

Hwang, Jinik; Park, So Yun; Park, Mirye; Lee, Sukchan; Jo, Yeonhwa; Cho, Won Kyong; Lee, Taek-Kyun (2016): Metagenomic characterization of viral communities in Goseong Bay, Korea. In *Ocean Sci. J.* 51 (4), pp. 599–612. DOI: 10.1007/s12601-016-0051-7.

Hwang, Jinik; Park, So Yun; Park, Mirye; Lee, Sukchan; Lee, Taek-Kyun (2017): Seasonal dynamics and metagenomic characterization of marine viruses in Goseong Bay, Korea. In *PloS one* 12 (1), pp. e0169841. DOI: 10.1371/journal.pone.0169841.

Hyman, Paul; Abedon, Stephen T. (2010): Chapter 7 - Bacteriophage Host Range and Bacterial Resistance. In : Advances in Applied Microbiology, vol. 70: Academic Press, pp. 217–248. Available online at https://www.sciencedirect.com/science/article/pii/S0065216410700071.

Hynes, Alexander (2014): The Phages and Phage-like Elements of *Rhodobacter capsulatus*. [doctoral thesis]. Available online at https://research.library.mun.ca/6296/1/Hynes_Alexander_Pierre_2013_112013_PhD.pdf.pdf, checked on 11/17/2022.

Ibrahim, Mohamed A.; Griko, Natalya; Junker, Matthew; Bulla, Lee A. (2010): Bacillus thuringiensis: a genomics and proteomics perspective. In *Bioengineered bugs* 1 (1), pp. 31–50. DOI: 10.4161/bbug.1.1.10519.

Illumina, Inc (2015): Nextera XT Library Prep: Tips and Troubleshooting. Available online at https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-troubleshooting-guide.pdf.

Inoue, K.; Sugiyama, K.; Kosako, Y.; Sakazaki, R.; Yamai, S. (2000): Enterobacter cowaniisp. nov., a new species of the family Enterobacteriaceae. In *Curr Microbiol* 41 (6), pp. 417–420. DOI: 10.1007/s002840010160.

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.

Ivanova, Elena P.; Gorshkova, Nataliya M.; Sawabe, Tomoo; Zhukova, Natalia V.; Hayashi, Karin; Kurilenko, Valerie V. et al. (2004): Sulfitobacter delicatus sp. nov. and Sulfitobacter dubius sp. nov., respectively from a starfish (Stellaster equestris) and sea grass (Zostera marina). In *International journal of systematic and evolutionary microbiology* 54 (Pt 2), pp. 475–480. DOI: 10.1099/ijs.0.02654-0.

Jackson, C. A.; Couger, M. B.; Prabhakaran, M.; Ramachandriya, K. D.; Canaan, P.; Fathepure, B. Z. (2017): Isolation and characterization of Rhizobium sp. strain YS-1r that degrades lignin in plant biomass. In *J Appl Microbiol* 122 (4), pp. 940–952. DOI: 10.1111/jam.13401.

Jain, Chirag; Rodriguez, Luis M.; Phillippy, Adam M.; Konstantinos, T.; Aluru, Srinivas (2018): High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.

Jeong, Hae Jin; Du Yoo, Yeong; Kim, Jae Seong; Seong, Kyeong Ah; Kang, Nam Seon; Kim, Tae Hoon (2010): Growth, feeding and ecological roles of the mixotrophic and heterotrophic dinoflagellates in marine planktonic food webs. In *Ocean Science Journal* 45 (2), pp. 65–91. DOI: 10.1007/s12601-010-0007-2.

Ji, Jianda; Zhang, Rui; Jiao, Nianzhi (2015): Complete genome sequence of Roseophage vB_DshP-R1, which infects Dinoroseobacter shibae DFL12. In *Stand in Genomic Sci* 10, p. 6. DOI: 10.1186/1944-3277-10-6.

Jiang, Jie; Lan, Guanda; Li, Jinghua; Yu, Jun; Huang, Honglan; Sun, Yanbo et al. (2022): Characterization and Genomic Analysis of JC01, a Novel unclassified Bacteriophage Infecting Cronobacter sakazakii. Available online at https://assets.researchsquare.com/files/rs-1876797/v1/e2ef476d-f988-4e69-9f04-82b9202c78e4.pdf?c=1659021933, checked on 11/24/2022.

Jin, Myeong Seo; Kim, Kyung Hyun; Baek, Ju Hye; Kim, Jeong Min; Jeon, Che Ok (2023): Octadecabacter algicola sp. nov. and Octadecabacter dasysiphoniae sp. nov., isolated from a marine red alga and emended description of the genus Octadecabacter. In *International journal of systematic and evolutionary microbiology* 73 (1). DOI: 10.1099/ijsem.0.005664.

Johnson, Matthew C.; Sena-Velez, Marta; Washburn, Brian K.; Platt, Georgia N.; Lu, Stephen; Brewer, Tess E. et al. (2017): Structure, proteome and genome of Sinorhizobium meliloti phage ΦM5: A virus with LUZ24-like morphology and a highly mosaic genome. In *Journal of Structural Biology* 200 (3), pp. 343–359. DOI: 10.1016/j.jsb.2017.08.005.

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.

Joung, Yochan; Kim, Haneul; Jang, Taeyong; Ahn, Tae-Seok; Joh, Kiseong (2011): Gramella jeungdoensis sp. nov., isolated from a solar saltern in Korea. In *J Microbiol.* 49 (6), pp. 1022–1026. DOI: 10.1007/s12275-011-1192-0.

Jourda, Cyril; Santini, Sébastien; Rocher, Caroline; Le Bivic, André; Claverie, Jean-Michel (2015): Draft Genome Sequence of an Alphaproteobacterium Associated with the Mediterranean Sponge Oscarella lobularis. In *Genome announcements* 3 (5). DOI: 10.1128/genomeA.00977-15.

Juhala, R. J.; Ford, M. E.; Duda, R. L.; Youlton, A.; Hatfull, G. F.; Hendrix, R. W. (2000): Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic mosaicism in the lambdoid bacteriophages. In *Journal of molecular biology* 299 (1), pp. 27–51. DOI: 10.1006/jmbi.2000.3729.

Kabwe, Mwila; Brown, Teagan; Speirs, Lachlan; Ku, Heng; Leach, Michael; Chan, Hiu Tat et al. (2020): Novel Bacteriophages Capable of Disrupting Biofilms From Clinical Strains of Aeromonas hydrophila. In *Frontiers in microbiology* 11, p. 194. DOI: 10.3389/fmicb.2020.00194.

Kala, Smriti; Cumby, Nichole; Sadowski, Paul D.; Hyder, Batool Zafar; Kanelis, Voula; Davidson, Alan R.; Maxwell, Karen L. (2014): HNH proteins are a widespread component of phage DNA packaging machines. In *PNAS* 111 (16), pp. 6022–6027. DOI: 10.1073/pnas.1320952111.

Kaliniene, Laura; Noreika, Algirdas; Kaupinis, Algirdas; Valius, Mindaugas; Jurgelaitis, Edvinas; Lazutka, Justas et al. (2021): Analysis of a Novel Bacteriophage vB_AchrS_AchV4 Highlights the Diversity of Achromobacter Viruses. In *Viruses* 13 (3). DOI: 10.3390/v13030374.

Kang, Ilnam; Jang, Hani; Oh, Hyun-Myung; Cho, Jang-Cheon (2012): Complete genome sequence of *Celeribacter* bacteriophage P12053L. In *Journal of virology* 86 (15), pp. 8339–8340. DOI: 10.1128/JVI.01153-12.

Kang, Ilnam; Oh, Hyun-Myung; Kang, Dongmin; Cho, Jang-Cheon (2013): Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. In *Proceedings of the National Academy of Sciences* 110. DOI: 10.1073/pnas.1219930110.

Kanukollu, Saranya; Wemheuer, Bernd; Herber, Janina; Billerbeck, Sara; Lucas, Judith; Daniel, Rolf et al. (2016): Distinct compositions of free-living, particle-associated and benthic communities of the

Roseobacter group in the North Sea. In *FEMS microbiology ecology* 92 (1). DOI: 10.1093/femsec/fiv145.

Kar, Priyanka; Das, Tridip Kr; Ghosh, Smita; Pradhan, Shrabani; Chakrabarti, Sudipta; Mondal, Keshab Ch; Ghosh, Kuntal (2022): Characterization of a Vibrio-infecting bacteriophage, VPMCC5, and proposal of its incorporation as a new genus in the Zobellviridae family. In *Virus research* 321, p. 198904. DOI: 10.1016/j.virusres.2022.198904.

Katoh, Kazutaka; Standley, Daron M. (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. In *Mol Biol Evol* 30 (4), pp. 772–780. DOI: 10.1093/molbev/mst010.

Kauffman, Kathryn M.; Chang, William K.; Brown, Julia M.; Hussain, Fatima A.; Yang, Joy Y.; Polz, Martin F.; Kelly, Libusha (2021): Resolving the structure of phage-bacteria interactions in the context of natural diversity. DOI: 10.1101/2021.06.27.449121.

Kauffman, Kathryn M.; Hussain, Fatima A.; Yang, Joy; Arevalo, Philip; Brown, Julia M.; Chang, William K. et al. (2018): A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. In *Nature*. DOI: 10.1038/nature25474.

Kauffman, Kathryn M.; Polz, Martin F. (2018): Streamlining standard bacteriophage methods for higher throughput. In *MethodsX* 5, pp. 159–172. DOI: 10.1016/j.mex.2018.01.007.

Kazaks, Andris; Dislers, Andris; Lipowsky, Gerd; Nikolajeva, Vizma; Tars, Kaspars (2012): Complete genome sequence of the Enterobacter cancerogenus bacteriophage Enc34. In *J. Virol.* 86 (20), pp. 11403–11404. DOI: 10.1128/JVI.01954-12.

Kazantseva, Olesya A.; Buzikov, Rustam M.; Pilipchuk, Tatsiana A.; Valentovich, Leonid N.; Kazantsev, Andrey N.; Kalamiyets, Emilia I.; Shadrin, Andrey M. (2021): The Bacteriophage Pf-10-A Component of the Biopesticide "Multiphage" Used to Control Agricultural Crop Diseases Caused by Pseudomonas syringae. In *Viruses* 14 (1). DOI: 10.3390/v14010042.

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.

Keen, Eric C. (2014): Tradeoffs in bacteriophage life histories. In *Bacteriophage* 4 (1), e28365. DOI: 10.4161/bact.28365.

Kemp, Priscilla; Garcia, L. René; Molineux, Ian J. (2005): Changes in bacteriophage T7 virion structure at the initiation of infection. In *Virology* 340 (2), pp. 307–317. DOI: 10.1016/j.virol.2005.06.039.

Kerimoglu, Onur; Hintz, Nils H.; Lücken, Leonhard; Blasius, Bernd; Böttcher, Lea; Bunse, Carina et al. (2022): Growth, organic matter release, aggregation and recycling during a diatom bloom: A model-based analysis of a mesocosm experiment.

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.

Kim, Sang Guen; Roh, Eunjung; Park, Jungkum; Giri, Sib Sankar; Kwon, Jun; Kim, Sang Wha et al. (2021): The Bacteriophage pEp_SNUABM_08 Is a Novel Singleton Siphovirus with High Host Specificity for Erwinia pyrifoliae. In *Viruses* 13 (7). DOI: 10.3390/v13071231.

Kim, Yiseul; Aw, Tiong Gim; Rose, Joan B. (2016): Transporting ocean viromes: Invasion of the aquatic biosphere. In *PloS one* 11 (4), e0152671. DOI: 10.1371/journal.pone.0152671.

Kloos, W. E.; Schleifer, K. H.; Smith, R. F. (1976): Characterization of Staphylococcus sciuri sp.nov. and Its Subspecies. In *International journal of systematic bacteriology* 26 (1), pp. 22–37. DOI: 10.1099/00207713-26-1-22.

Klumpp, Jochen; Dunne, Matthew; Loessner, Martin J. (2023): A perfect fit: Bacteriophage receptor-binding proteins for diagnostic and therapeutic applications. In *Current opinion in microbiology* 71, p. 102240. DOI: 10.1016/j.mib.2022.102240.

Knowles, B.; Silveira, C. B.; Bailey, B. A.; Barott, K.; Cantu, V. A.; Cobián-Güemes, A. G. et al. (2016): Lytic to temperate switching of viral communities. In *Nature* 531 (7595), pp. 466–470. DOI: 10.1038/nature17193.

Kokjohn, T. A.; Sayler, G. S. (1991): Attachment and replication of Pseudomonas aeruginosa bacteriophages under conditions simulating aquatic environments. In *Journal of general microbiology* 137 (3), pp. 661–666. DOI: 10.1099/00221287-137-3-661.

Koonin, Eugene V.; Dolja, Valerian V.; Krupovic, Mart; Varsani, Arvind; Wolf, Yuri I.; Yutin, Natalya et al. (2020a): 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.; Makarova, Kira S.; Wolf, Yuri I.; Krupovic, Mart (2020b): Evolutionary entanglement of mobile genetic elements and host defence systems: guns for hire. In *Nat Rev Genet* 21 (2), pp. 119–131. DOI: 10.1038/s41576-019-0172-9.

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.

Kropinski, Andrew; Adriaenssens, E; Tong, Yigang (2018): ICTV Taxnomy Proposal 2018.095B: To create one (1) new genus, Nanhaivirus, containing a single species in the family Siphoviridae. Available online at https://ictv.global/ictv/proposals/2018.095B.A.v1.Nanhaivirus.zip.

Kropinski, Andrew; Kuhn, Jens, H.; Adriaenssens, Evelien, M. (2016): ICTV Taxonomy proposal 2016.040a-dB: To create one (1) new genus, Rdjlvirus, including two (2) new species in the family Siphoviridae. Available online at https://ictv.global/ictv/proposals/2016.040a-dB.A.v3.Rdjlvirus.pdf.

Krüger, Karen; Chafee, Meghan; Ben Francis, T.; Glavina Del Rio, Tijana; Becher, Dörte; Schweder, Thomas et al. (2019): In marine Bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. In *ISME J* 13 (11), pp. 2800–2816. DOI: 10.1038/s41396-019-0476-y.

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.

Kuhlisch, Constanze; Schleyer, Guy; Shahaf, Nir; Vincent, Flora; Schatz, Daniella; Vardi, Assaf (2021): Viral infection of algal blooms leaves a unique metabolic footprint on the dissolved organic matter in the ocean. In *Science advances* 7 (25). DOI: 10.1126/sciadv.abf4680.

Kulczyk, A. W.; Richardson, C. C. (2016): The Replication System of Bacteriophage T7. In *The Enzymes* 39, pp. 89–136. DOI: 10.1016/bs.enz.2016.02.001.

Kulikov, Eugene E.; Golomidova, Alla K.; Letarova, Maria A.; Kostryukova, Elena S.; Zelenin, Alexandr S.; Prokhorov, Nikolai S.; Letarov, Andrey V. (2014): Genomic sequencing and biological characteristics of a novel Escherichia coli bacteriophage 9g, a putative representative of a new Siphoviridae genus. In *Viruses* 6 (12), pp. 5077–5092. DOI: 10.3390/v6125077.

Kutser, Tiit (2009): Passive optical remote sensing of cyanobacteria and other intense phytoplankton blooms in coastal and inland waters. In *International Journal of Remote Sensing* 30 (17), pp. 4401–4425. DOI: 10.1080/01431160802562305.

Kutter, Elizabeth; Vos, Daniel de; Gvasalia, Guram; Alavidze, Zemphira; Gogokhia, Lasha; Kuhl, Sarah; Abedon, Stephen T. (2010): Phage therapy in clinical practice: treatment of human infections. In *Current pharmaceutical biotechnology* 11 (1), pp. 69–86. DOI: 10.2174/138920110790725401.

Kwan, Tony; Liu, Jing; DuBow, Michael; Gros, Philippe; Pelletier, Jerry (2006): Comparative Genomic Analysis of 18 Pseudomonas aeruginosa Bacteriophages†. In *Journal of bacteriology* 188 (3), pp. 1184–1187. DOI: 10.1128/JB.188.3.1184-1187.2006.

Kwon, Young-Kyung; Kim, Jennifer Jooyoun; Kim, Ji Hyung; Jeon, Seon-Mi; Ye, Bo-Ram; Jang, Jiyi et al. (2012): Draft genome sequence of the xylan-degrading marine bacterium strain S124, representing a novel species of the genus Oceanicola. In *Journal of bacteriology* 194 (22), p. 6325. DOI: 10.1128/JB.01614-12.

Labrie, Simon J.; Samson, Julie E.; Moineau, Sylvain (2010): Bacteriophage resistance mechanisms. In *Nat Rev Micro* 8 (5), pp. 317–327. DOI: 10.1038/nrmicro2315.

Lamy, D.; Obernosterer, I.; Laghdass, M.; Artigas, F.; Breton, E.; Grattepanche, J. D. et al. (2009): Temporal changes of major bacterial groups and bacterial heterotrophic activity during a *Phaeocystis globosa* bloom in the eastern English Channel. In *Aquat. Microb. Ecol.* 58, pp. 95–107. DOI: 10.3354/ame01359.

Lawson, Paul A.; Citron, Diane M.; Tyrrell, Kerin L.; Finegold, Sydney M. (2016): Reclassification of Clostridium difficile as Clostridioides difficile (Hall and O'Toole 1935) Prévot 1938. In *Anaerobe* 40, pp. 95–99. DOI: 10.1016/j.anaerobe.2016.06.008.

Le, Shuai; He, Xuesong; Tan, Yinling; Huang, Guangtao; Zhang, Lin; Lux, Renate et al. (2013): Mapping the tail fiber as the receptor binding protein responsible for differential host specificity of Pseudomonas aeruginosa bacteriophages PaP1 and JG004. In *PloS one* 8 (7), e68562. DOI: 10.1371/journal.pone.0068562.

Leduc, D.; Graziani, S.; Meslet-Cladiere, L.; Sodolescu, A.; Liebl, U.; Myllykallio, H. (2004): Two distinct pathways for thymidylate (dTMP) synthesis in (hyper)thermophilic Bacteria and Archaea. In *Biochm. Soc. Trans.* 32 (2), pp. 231–235. DOI: 10.1042/bst0320231.

Lefort, Vincent; Desper, Richard; Gascuel, Olivier (2015): FastME 2.0: A Comprehensive, accurate, and fast distance-based phylogeny inference program. In *Molecular biology and evolution* 32 (10), pp. 2798–2800. DOI: 10.1093/molbev/msv150.

Legrand, Pierre; Collins, Barry; Blangy, Stéphanie; Murphy, James; Spinelli, Silvia; Gutierrez, Carlos et al. (2016): The Atomic Structure of the Phage Tuc2009 Baseplate Tripod Suggests that Host Recognition Involves Two Different Carbohydrate Binding Modules. In *mBio* 7 (1), e01781-15. DOI: 10.1128/mBio.01781-15.

Leiman, Petr G.; Arisaka, Fumio; van Raaij, Mark J.; Kostyuchenko, Victor A.; Aksyuk, Anastasia A.; Kanamaru, Shuji; Rossmann, Michael G. (2010): Morphogenesis of the T4 tail and tail fibers. In *Virology journal* 7, p. 355. DOI: 10.1186/1743-422X-7-355.

Lenk, Sabine; Moraru, Cristina; Hahnke, Sarah; Arnds, Julia; Richter, Michael; Kube, Michael et al. (2012): *Roseobacter* clade bacteria are abundant in coastal sediments and encode a novel combination of sulfur oxidation genes. In *The ISME journal* 6 (12), pp. 2178–2187. DOI: 10.1038/ismej.2012.66.

Letunic, Ivica; Bork, Peer (2016): Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. In *Nucleic acids research* 44 (W1), W242-5. DOI: 10.1093/nar/gkw290.

Letunic, Ivica; Bork, Peer (2021): Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. In *Nucleic acids research* 49 (W1), W293-W296. DOI: 10.1093/nar/gkab301.

Li, Baolian; Zhang, Si; Long, Lijuan; Huang, Sijun (2016a): Characterization and complete genome sequences of three N4-Like Roseobacter phages isolated from the South China Sea. In *Current microbiology* 73 (3), pp. 409–418. DOI: 10.1007/s00284-016-1071-3.

Li, Erna; Yin, Zhe; Ma, Yanyan; Li, Huan; Lin, Weishi; Wei, Xiao et al. (2016b): Identification and molecular characterization of bacteriophage phiAxp-2 of Achromobacter xylosoxidans. In *Scientific reports* 6 (1), p. 34300. DOI: 10.1038/srep34300.

Li, Mengzhe; Jin, Yanqiu; Lin, Hong; Wang, Jingxue; Jiang, Xiuping (2018): Complete Genome of a Novel Lytic Vibrio parahaemolyticus Phage VPp1 and Characterization of Its Endolysin for Antibacterial Activities. In *Journal of food protection* 81 (7), pp. 1117–1125. DOI: 10.4315/0362-028X.JFP-17-278.

Li, Zhao; Qu, Zhe; Zhang, Xiuming; Zhang, Xiao-Hua (2012): *Lentibacter algarum* gen. nov., sp. nov., isolated from coastal water during a massive green algae bloom. In *International journal of systematic and evolutionary microbiology* 62 (Pt 5), pp. 1042–1047. DOI: 10.1099/ijs.0.029868-0.

Liang, Kevin Y. H.; Orata, Fabini D.; Boucher, Yann F.; Case, Rebecca J. (2021): Roseobacters in a Sea of Poly- and Paraphyly: Whole Genome-Based Taxonomy of the Family Rhodobacteraceae and the Proposal for the Split of the "Roseobacter Clade" Into a Novel Family, Roseobacteraceae fam. nov. In *Frontiers in microbiology* 12, p. 683109. DOI: 10.3389/fmicb.2021.683109.

Liang, Yantao; Zhang, Yongyu; Zhou, Chao; Chen, Zhenghao; Yang, Suping; Yan, Changzhou; Jiao, Nianzhi (2016): Complete genome sequence of the siphovirus Roseophage RDJLΦ 2 infecting *Roseobacter denitrificans* OCh114. In *Marine genomics* 25, pp. 17–19. DOI: 10.1016/j.margen.2015.10.009.

Lin, Dan; Tang, Kai; Han, Yu; Li, Chenlan; Chen, Xiaofeng (2016): Genome sequence of an inducible phage in Rhodovulum sp. P5 isolated from the shallow-sea hydrothermal system. In *Marine genomics* 30, pp. 93–95. DOI: 10.1016/j.margen.2016.10.002.

Lin, Shih-Yao; Hameed, Asif; Hsieh, Yu-Ting; Young, Chiu-Chung (2019): Mesorhizobium composti sp. nov., isolated from compost. In *Antonie van Leeuwenhoek* 112 (9), pp. 1387–1398. DOI: 10.1007/s10482-019-01270-y.

Lipski, A.; Reichert, K.; Reuter, B.; Spröer, C.; Altendorf, K. (1998): Identification of bacterial isolates from biofilters as Paracoccus alkenifer sp. nov. and Paracoccus solventivorans with emended description of Paracoccus solventivorans. In *International journal of systematic bacteriology* 48 Pt 2, pp. 529–536. DOI: 10.1099/00207713-48-2-529.

Lisitskaya, Lidiya; Aravin, Alexei A.; Kulbachinskiy, Andrey (2018): DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. In *Nat Commun* 9 (1), p. 5165. DOI: 10.1038/s41467-018-07449-7.

Little, J. W. (2005): Lysogeny, Prophage Induction, and Lysogenic Conversion. In Matthew K. Waldor, David I. Friedman, Sankar L. Adhya (Eds.): Phages: Their Role in Bacterial Pathogenesis and Biotechnology. Washington, D.C. Available online at https://onlinelibrary.wiley.com/doi/abs/10.1128/9781555816506.ch3.

Liu, Minghsun; Deora, Rajendar; Doulatov, Sergei R.; Gingery, Mari; Eiserling, Frederick A.; Preston, Andrew et al. (2002): Reverse transcriptase-mediated tropism switching in Bordetella bacteriophage. In *Science (New York, N.Y.)* 295 (5562), pp. 2091–2094. DOI: 10.1126/science.1067467.

Liu, Qingqing; Belle, Archana; Shub, David A.; Belfort, Marlene; Edgell, David R. (2003): SegG endonuclease promotes marker exclusion and mediates co-conversion from a distant cleavage site. In *Journal of molecular biology* 334 (1), pp. 13–23. DOI: 10.1016/j.jmb.2003.09.027.

Liu, Yang; Du, Juan; Lai, Qiliang; Dong, Chunming; Shao, Zongze (2017): *Nioella sediminis* sp. nov., isolated from surface sediment and emended description of the genus *Nioella*. In *International journal of systematic and evolutionary microbiology* 67 (5), pp. 1271–1274. DOI: 10.1099/ijsem.0.001798.

Lohr, Jayme E.; Chen, Feng; Hill, Russell T. (2005): Genomic analysis of bacteriophage PhiJL001: insights into its interaction with a sponge-associated alpha-proteobacterium. In *Applied and environmental microbiology* 71 (3), pp. 1598–1609. DOI: 10.1128/AEM.71.3.1598-1609.2005.

Lopatina, Anna; Tal, Nitzan; Sorek, Rotem (2020): Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. In *Annual review of virology* 7 (1), pp. 371–384. DOI: 10.1146/annurev-virology-011620-040628.

Lopes, Anne; Tavares, Paulo; Petit, Marie-Agnès; Guérois, Raphaël; Zinn-Justin, Sophie (2014): Automated classification of tailed bacteriophages according to their neck organization. In *BMC Genomics* 15 (1), p. 1027. DOI: 10.1186/1471-2164-15-1027.

Lopez-Diaz, Maria; Bleriot, Ines; Pacios, Olga; Fernandez-Garcia, Laura; Blasco, Lucia; Ambroa, Anton et al. (2022): Molecular characteristics of phages located in Carbapenemase-Producing Escherichia coli clinical isolates: New Phage-Like Plasmids. DOI: 10.1101/2022.06.29.498070.

Lowe, Todd M.; Chan, Patricia P. (2016): tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. In *Nucleic acids research* 44 (W1), W54-7. DOI: 10.1093/nar/gkw413.

Ludwig, Wolfgang; Strunk, Oliver; Westram, Ralf; Richter, Lothar; Meier, Harald; Yadhukumar et al. (2004): ARB: a software environment for sequence data. In *Nucleic acids research* 32 (4), pp. 1363–1371. DOI: 10.1093/nar/gkh293.

Luhtanen, Anne-Mari; Eronen-Rasimus, Eeva; Oksanen, Hanna M.; Tison, Jean-Louis; Delille, Bruno; Dieckmann, Gerhard S. et al. (2018): The first known virus isolates from Antarctic sea ice have complex infection patterns. In *FEMS microbiology ecology* 94 (4). DOI: 10.1093/femsec/fiy028.

Lundström, K. H.; Bamford, D. H.; Palva, E. T.; Lounatmaa, K. (1979): Lipid-containing bacteriophage PR4: structure and life cycle. In *The Journal of general virology* 43 (3), pp. 583–592. DOI: 10.1099/0022-1317-43-3-583.

Lynch, Karlene H.; Stothard, Paul; Dennis, Jonathan J. (2012): Comparative analysis of two phenotypically-similar but genomically-distinct Burkholderia cenocepacia-specific bacteriophages. In *BMC Genomics* 13 (1), p. 223. DOI: 10.1186/1471-2164-13-223.

Makarova, Kira S.; Wolf, Yuri I.; Snir, Sagi; Koonin, Eugene V. (2011): Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. In *Journal of bacteriology* 193 (21), pp. 6039–6056. DOI: 10.1128/JB.05535-11.

Malone, Lucia M.; Warring, Suzanne L.; Jackson, Simon A.; Warnecke, Carolin; Gardner, Paul P.; Gumy, Laura F.; Fineran, Peter C. (2020): A jumbo phage that forms a nucleus-like structure evades CRISPR-Cas DNA targeting but is vulnerable to type III RNA-based immunity. In *Nature microbiology* 5 (1), pp. 48–55. DOI: 10.1038/s41564-019-0612-5.

Maniatis, T.; Ptashne, M. (1973): Multiple repressor binding at the operators in bacteriophage lambda. In *Proceedings of the National Academy of Sciences of the United States of America* 70 (5), pp. 1531–1535. DOI: 10.1073/pnas.70.5.1531.

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.

Manuel, Nikhil; Rushing, Leika; Ravindran, Aravind; Newkirk, Heather; Burrowes, Ben; Young, Ry; Gonzalez, Carlos (2021): Complete Genome Sequence of Rhizobium japonicum Podophage Pasto. In *Microbiol Resour Announc* 10 (3). DOI: 10.1128/MRA.01442-20.

Mari, Xavier; Kerros, Marie-Emmanuelle; Weinbauer, Markus G. (2007): Virus attachment to transparent exopolymeric particles along trophic gradients in the southwestern lagoon of New Caledonia. In *Applied and environmental microbiology* 73 (16), pp. 5245–5252. DOI: 10.1128/AEM.00762-07.

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.

Mascolo, Elia; Adhikari, Satish; Caruso, Steven M.; deCarvalho, Tagide; Folch Salvador, Anna; Serra-Sagristà, Joan et al. (2022): The transcriptional regulator CtrA controls gene expression in Alphaproteobacteria phages: Evidence for a lytic deferment pathway. In *Frontiers in microbiology* 13, p. 918015. DOI: 10.3389/fmicb.2022.918015.

Mavrich, Travis N.; Hatfull, Graham F. (2019): Evolution of Superinfection Immunity in Cluster A Mycobacteriophages. In *mBio* 10 (3). DOI: 10.1128/mBio.00971-19.

McCutcheon, Jaclyn G.; Lin, Andrea; Dennis, Jonathan J. (2022): Characterization of Stenotrophomonas maltophilia phage AXL1 as a member of the genus Pamexvirus encoding resistance

to trimethoprim-sulfamethoxazole. In *Scientific reports* 12 (1), p. 10299. DOI: 10.1038/s41598-022-14025-z.

McKenna, R.; Xia, D.; Willingmann, P.; Ilag, L. L.; Krishnaswamy, S.; Rossmann, M. G. et al. (1992): Atomic structure of single-stranded DNA bacteriophage phi X174 and its functional implications. In *Nature* 355 (6356), pp. 137–143. DOI: 10.1038/355137a0.

Meier-Kolthoff, Jan P.; Auch, Alexander F.; Klenk, Hans-Peter; Göker, Markus (2013): Genome sequence-based species delimitation with confidence intervals and improved distance functions. In *BMC bioinformatics* 14, p. 60. DOI: 10.1186/1471-2105-14-60.

Meier-Kolthoff, Jan P.; Carbasse, Joaquim Sardà; Peinado-Olarte, Rosa L.; Göker, Markus (2021): TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. In *Nucl Acids Res*. DOI: 10.1093/nar/gkab902.

Meier-Kolthoff, Jan P.; Carbasse, Joaquim Sardà; Peinado-Olarte, Rosa L.; Göker, Markus (2022): TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. In *Nucleic acids research* 50 (D1), D801-D807. DOI: 10.1093/nar/gkab902.

Meier-Kolthoff, Jan P.; Göker, Markus (2017): VICTOR: genome-based phylogeny and classification of prokaryotic viruses. In *Bioinformatics* 33 (21), pp. 3396–3404. DOI: 10.1093/bioinformatics/btx440.

Meier-Kolthoff, Jan P.; Göker, Markus (2019): TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. In *Nature communications* 10 (1), p. 2182. DOI: 10.1038/s41467-019-10210-3.

Meier-Kolthoff, Jan P.; Hahnke, Richard L.; Petersen, Jörn; Scheuner, Carmen; Michael, Victoria; Fiebig, Anne et al. (2014): Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. In *Standards in genomic sciences* 9, p. 2. DOI: 10.1186/1944-3277-9-2.

Meiring, Tracy L.; Tuffin, I. Marla; Cary, Craig; Cowan, Don A. (2012): Genome sequence of temperate bacteriophage Psymv2 from Antarctic Dry Valley soil isolate Psychrobacter sp. MV2. In *Extremophiles* 16 (5), pp. 715–726. DOI: 10.1007/s00792-012-0467-7.

Merchant, Nirav; Lyons, Eric; Goff, Stephen; Vaughn, Matthew; Ware, Doreen; Micklos, David; Antin, Parker (2016): The iPlant Collaborative: Cyberinfrastructure for Enabling Data to Discovery for the Life Sciences. In *PLoS biology* 14 (1), e1002342. DOI: 10.1371/journal.pbio.1002342.

Merrill, Bryan D.; Ward, Andy T.; Grose, Julianne H.; Hope, Sandra (2016): Software-based analysis of bacteriophage genomes, physical ends, and packaging strategies. In *BMC genomics* 17, p. 679. DOI: 10.1186/s12864-016-3018-2.

Meyer, Justin R.; Dobias, Devin T.; Medina, Sarah J.; Servilio, Lisa; Gupta, Animesh; Lenski, Richard E. (2016): Ecological speciation of bacteriophage lambda in allopatry and sympatry. In *Science (New York, N.Y.)* 354 (6317), pp. 1301–1304. DOI: 10.1126/science.aai8446.

Middelboe (2000): Bacterial Growth Rate and Marine Virus-Host Dynamics. In *Microb Ecol* 40 (2), pp. 114–124. DOI: 10.1007/s002480000050.

Miller, Robert V.; Day, Martin J. (2008): Contribution of lysogeny, pseudolysogeny, and starvation to phage ecology. In : Bacteriophage Ecology. 1st ed.: Cambridge University Press, pp. 114–144.

Miller, Robert V.; Ripp, Steven A. (2002): Pseudolysogeny. In : Horizontal gene transfer. 2nd ed. San Diego: Academic Press, pp. 81–91.

Millman, Adi; Bernheim, Aude; Stokar-Avihail, Avigail; Fedorenko, Taya; Voichek, Maya; Leavitt, Azita et al. (2020): Bacterial Retrons Function In Anti-Phage Defense. In *Cell* 183 (6), 1551-1561.e12. DOI: 10.1016/j.cell.2020.09.065.

Millman, Adi; Melamed, Sarah; Leavitt, Azita; Doron, Shany; Bernheim, Aude; Hör, Jens et al. (2022): An expanding arsenal of immune systems that protect bacteria from phages. DOI: 10.1101/2022.05.11.491447.

Mizuno, Carolina Megumi; Rodriguez-Valera, Francisco; Kimes, Nikole E.; Ghai, Rohit (2013): Expanding the Marine Virosphere Using Metagenomics. In *PLoS genetics* 9 (12). DOI: 10.1371/journal.pgen.1003987.

Moebus, K.; Nattkemper, H. (1981): Bacteriophage sensitivity patterns among bacteria isolated from marine waters. In *Helgolander Meeresunters* 34 (3), pp. 375–385. DOI: 10.1007/BF02074130.

Moraru, Cristina (2023): VirClust-A Tool for Hierarchical Clustering, Core Protein Detection and Annotation of (Prokaryotic) Viruses. In *Viruses* 15 (4). DOI: 10.3390/v15041007.

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.

Mori, Corinna; Beck, Melanie; Hintz, Nils H.; Merder, Julian; Bunse, Carina; Dittmar, Thorsten et al. (2021): Biogeochemical thallium cycling during a mesocosm phytoplankton spring bloom: Biotic versus abiotic drivers. In *Geochimica et Cosmochimica Acta* 313, pp. 257–276. DOI: 10.1016/j.gca.2021.08.002.

Morrow, Kathleen M.; Tedford, Abbey Rose; Pankey, M. Sabrina; Lesser, Michael P. (2018): A member of the Roseobacter clade, Octadecabacter sp., is the dominant symbiont in the brittle star Amphipholis squamata. In *FEMS microbiology ecology* 94 (4). DOI: 10.1093/femsec/fiy030.

Mruk, Iwona; Kobayashi, Ichizo (2014): To be or not to be: regulation of restriction-modification systems and other toxin-antitoxin systems. In *Nucl Acids Res* 42 (1), pp. 70–86. DOI: 10.1093/nar/gkt711.

Murphy, James; Mahony, Jennifer; Ainsworth, Stuart; Nauta, Arjen; van Sinderen, Douwe (2013): Bacteriophage orphan DNA methyltransferases: insights from their bacterial origin, function, and occurrence. In *Applied and environmental microbiology* 79 (24), pp. 7547–7555. DOI: 10.1128/AEM.02229-13.

Muyzer, G.; Teske, A.; Wirsen, C. O.; Jannasch, H. W. (1995): Phylogenetic relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. In *Arch. Microbiol.* 164 (3), pp. 165–172. DOI: 10.1007/BF02529967.

Naser, Iftekhar Bin; Hoque, M. Mozammel; Nahid, M. Ausrafuggaman; Tareq, Tokee M.; Rocky, M. Kamruzzaman; Faruque, Shah M. (2017): Analysis of the CRISPR-Cas system in bacteriophages active on epidemic strains of Vibrio cholerae in Bangladesh. In *Scientific reports* 7 (1), p. 14880. DOI: 10.1038/s41598-017-14839-2.

Newton, Ryan J.; Griffin, Laura E.; Bowles, Kathy M.; Meile, Christof; Gifford, Scott; Givens, Carrie E. et al. (2010): Genome characteristics of a generalist marine bacterial lineage. In *The ISME journal* 4 (6), pp. 784–798. DOI: 10.1038/ismej.2009.150.

Nguyen, James; Harb, Laith; Moreland, Russell; Liu, Mei; Gill, Jason J.; Ramsey, Jolene (2019): Complete Genome Sequence of Proteus mirabilis Siphophage Saba. In *Microbiol Resour Announc* 8 (41). DOI: 10.1128/MRA.01094-19.

Nguyen, Lam-Tung; Schmidt, Heiko A.; Haeseler, Arndt von; Minh, Bui Quang (2015): IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. In *Mol Biol Evol* 32 (1), pp. 268–274. DOI: 10.1093/molbev/msu300.

Nicholson, Ainsley C.; Humrighouse, Ben W.; Graziano, James C.; Emery, Brian; McQuiston, John R. (2016): Draft Genome Sequences of Strains Representing Each of the Elizabethkingia Genomospecies Previously Determined by DNA-DNA Hybridization. In *Genome Announc.* 4 (2). DOI: 10.1128/genomeA.00045-16.

Nishimura, Yosuke; Watai, Hiroyasu; Honda, Takashi; Mihara, Tomoko; Omae, Kimiho; Roux, Simon et al. (2017a): Environmental viral genomes shed new light on virus-host interactions in the ocean. In *mSphere* 2 (2). DOI: 10.1128/mSphere.00359-16.

Nishimura, Yosuke; Yoshida, Takashi; Kuronishi, Megumi; Uehara, Hideya; Ogata, Hiroyuki; Goto, Susumu (2017b): ViPTree: the viral proteomic tree server. In *Bioinformatics (Oxford, England)* 33 (15), pp. 2379–2380. DOI: 10.1093/bioinformatics/btx157.

Nissimov, Jozef I.; Vandzura, Rebecca; Johns, Christopher T.; Natale, Frank; Haramaty, Liti; Bidle, Kay D. (2018): Dynamics of transparent exopolymer particle production and aggregation during viral infection of the coccolithophore, Emiliania huxleyi. In *Environ Microbiol* 20 (8), pp. 2880–2897. DOI: 10.1111/1462-2920.14261.

Nobrega, Franklin L.; Vlot, Marnix; Jonge, Patrick A. de; Dreesens, Lisa L.; Beaumont, Hubertus J. E.; Lavigne, Rob et al. (2018): Targeting mechanisms of tailed bacteriophages. In *Nat Rev Micro* 16 (12), pp. 760–773. DOI: 10.1038/s41579-018-0070-8.

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.

Novikova, Olga; Jayachandran, Pradeepa; Kelley, Danielle S.; Morton, Zachary; Merwin, Samantha; Topilina, Natalya I.; Belfort, Marlene (2016): Intein Clustering Suggests Functional Importance in Different Domains of Life. In *Mol Biol Evol* 33 (3), pp. 783–799. DOI: 10.1093/molbev/msv271.

Numberger, Daniela; Ganzert, Lars; Zoccarato, Luca; Mühldorfer, Kristin; Sauer, Sascha; Grossart, Hans-Peter; Greenwood, Alex D. (2019): Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. In *Scientific reports* 9 (1), p. 9673. DOI: 10.1038/s41598-019-46015-z.

Ohtsubo, M.; Okazaki, H.; Nishimoto, T. (1989): The RCC1 protein, a regulator for the onset of chromosome condensation locates in the nucleus and binds to DNA. In *The Journal of cell biology* 109 (4 Pt 1), pp. 1389–1397. DOI: 10.1083/jcb.109.4.1389.

Oliveira, Pedro H.; Touchon, Marie; Rocha, Eduardo P. C. (2014): The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. In *Nucl Acids Res* 42 (16), pp. 10618–10631. DOI: 10.1093/nar/gku734.

Oliveira, Vanessa; Polónia, Ana R. M.; Cleary, Daniel F. R.; Huang, Yusheng M.; Voogd, Nicole J. de; Keller-Costa, Tina et al. (2022): Assessing the genomic composition, putative ecological relevance and

biotechnological potential of plasmids from sponge bacterial symbionts. In *Microbiological research* 265, p. 127183. DOI: 10.1016/j.micres.2022.127183.

Ortĭn, J.; Viñuela, E.; Salas, M.; Vasquez, C. (1971): DNA-protein complex in circular DNA from phage phi-29. In *Nature: New biology* 234 (52), pp. 275–277. DOI: 10.1038/newbio234275a0.

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.

Pagnier, Isabelle; Raoult, Didier; La Scola, Bernard (2008): Isolation and identification of amoeba-resisting bacteria from water in human environment by using an *Acanthamoeba polyphaga* co-culture procedure. In *Environmental microbiology* 10 (5), pp. 1135–1144. DOI: 10.1111/j.1462-2920.2007.01530.x.

Pagnier, Isabelle; Valles, Camille; Raoult, Didier; La Scola, Bernard (2015): Isolation of *Vermamoeba vermiformis* and associated bacteria in hospital water. In *Microbial pathogenesis* 80, pp. 14–20. DOI: 10.1016/j.micpath.2015.02.006.

Paquet, Valérie E.; Charette, Steve J. (2016): Amoeba-resisting bacteria found in multilamellar bodies secreted by *Dictyostelium discoideum*: social amoebae can also package bacteria. In *FEMS microbiology ecology* 92 (3). DOI: 10.1093/femsec/fiw025.

Park, Sooyeon; Yoon, Sun Young; Jung, Yong-Taek; Yoon, Jung-Hoon (2016): Octadecabacterponticola sp. nov., isolated from seawater. In *International journal of systematic and evolutionary microbiology* 66 (10), pp. 4179–4184. DOI: 10.1099/ijsem.0.001332.

Parte, Aidan C.; Sardà Carbasse, Joaquim; Meier-Kolthoff, Jan P.; Reimer, Lorenz C.; Göker, Markus (2020): List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. In *International journal of systematic and evolutionary microbiology* 70 (11), pp. 5607–5612. DOI: 10.1099/ijsem.0.004332.

Peters, Danielle L.; McCutcheon, Jaclyn G.; Stothard, Paul; Dennis, Jonathan J. (2019): Novel Stenotrophomonas maltophilia temperate phage DLP4 is capable of lysogenic conversion. In *BMC Genomics* 20 (1), p. 300. DOI: 10.1186/s12864-019-5674-5.

Petrovic, Aleksandra; Kostanjsek, Rok; Rakhely, Gabor; Knezevic, Petar (2017): The First Siphoviridae Family Bacteriophages Infecting Bordetella bronchiseptica Isolated from Environment. In *Microb Ecol* 73 (2), pp. 368–377. DOI: 10.1007/s00248-016-0847-0.

Piel, Damien; Bruto, Maxime; Labreuche, Yannick; Blanquart, François; Goudenège, David; Barcia-Cruz, Rubén et al. (2022): Phage-host coevolution in natural populations. In *Nat Microbiol* 7 (7), pp. 1075–1086. DOI: 10.1038/s41564-022-01157-1.

Pinhassi, Jarone; Sala, Maria Montserrat; Havskum, Harry; Peters, Francesc; Guadayol, Oscar; Malits, Andrea; Marrasé, Cèlia (2004): Changes in bacterioplankton composition under different phytoplankton regimens. In *Applied and environmental microbiology* 70 (11), pp. 6753–6766. DOI: 10.1128/AEM.70.11.6753-6766.2004.

Piya, Denish; Vara, Leonardo; Russell, William K.; Young, Ry; Gill, Jason J. (2017): The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. In *Molecular microbiology* 105 (3), pp. 399–412. DOI: 10.1111/mmi.13705.

Pleška, Maroš; Guet, Călin C. (2017): Effects of mutations in phage restriction sites during escape from restriction–modification. In *Biol. Lett.* 13 (12), p. 20170646. DOI: 10.1098/rsbl.2017.0646.

Potter, Simon C.; Luciani, Aurélien; Eddy, Sean R.; Park, Youngmi; Lopez, Rodrigo; Finn, Robert D. (2018): HMMER web server: 2018 update. In *Nucleic acids research* 46 (W1), W200-W204. DOI: 10.1093/nar/gky448.

Price, Morgan N.; Dehal, Paramvir S.; Arkin, Adam P. (2010): FastTree 2--approximately maximum-likelihood trees for large alignments. In *PloS one* 5 (3), e9490. DOI: 10.1371/journal.pone.0009490.

Qin, Fang; Du, Sen; Zhang, Zefeng; Ying, Hanqi; Wu, Ying; Zhao, Guiyuan et al. (2022): Newly identified HMO-2011-type phages reveal genomic diversity and biogeographic distributions of this marine viral group. In *The ISME journal*. DOI: 10.1038/s41396-021-01183-7.

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.

Quispe, Cristian F.; Esmael, Ahmed; Sonderman, Olivia; McQuinn, Michelle; Agarkova, Irina; Battah, Mohammed et al. (2017): Characterization of a new chlorovirus type with permissive and non-permissive features on phylogenetically related algal strains. In *Virology* 500, pp. 103–113. DOI: 10.1016/j.virol.2016.10.013.

R Core Team (2020): R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available online at https://www.R-project.org/.

Rambaut, A. (2006): FigTree 1.4.3 – a graphical viewer of phylogenetic trees and a program for producing publication-ready figures. Available online at http://tree.bio.ed.ac.uk/software/figtree/.

Ramos-Vivas, José; Superio, Joshua; Galindo-Villegas, Jorge; Acosta, Félix (2021): Phage Therapy as a Focused Management Strategy in Aquaculture. In *International journal of molecular sciences* 22 (19). DOI: 10.3390/ijms221910436.

Rapala, Jackson; Miller, Brenda; Garcia, Maximiliano; Dolan, Megan; Bockman, Matthew; Hansson, Mats et al. (2021): Genomic diversity of bacteriophages infecting Rhodobacter capsulatus and their relatedness to its gene transfer agent RcGTA. In *PloS one* 16 (11), e0255262. DOI: 10.1371/journal.pone.0255262.

Rathor, Nisha; Sachdeva, Pooja; Chaudhry, Rama (2022): Genome Sequence of Bacteriophage Infecting a Rare Pathogen, Pseudomonas luteola. In *Microbiol Resour Announc* 11 (2), Article e01113-21, e0111321. DOI: 10.1128/mra.01113-21.

Reyes-Robles, Tamara; Dillard, Rebecca S.; Cairns, Lynne S.; Silva-Valenzuela, Cecilia A.; Housman, Max; Ali, Afsar et al. (2018): Vibrio cholerae Outer Membrane Vesicles Inhibit Bacteriophage Infection. In *Journal of bacteriology* 200 (15). DOI: 10.1128/JB.00792-17.

Richlen, Mindy L.; Morton, Steve L.; Jamali, Ebrahim A.; Rajan, Anbiah; Anderson, Donald M. (2010): The catastrophic 2008–2009 red tide in the Arabian gulf region, with observations on the identification and phylogeny of the fish-killing dinoflagellate Cochlodinium polykrikoides. In *Harmful algae* 9 (2), pp. 163–172. DOI: 10.1016/j.hal.2009.08.013.

Riemann, Lasse; Grossart, Hans-Peter (2008): Elevated lytic phage production as a consequence of particle colonization by a marine Flavobacterium (Cellulophaga sp.). In *Microb Ecol* 56 (3), pp. 505–512. DOI: 10.1007/s00248-008-9369-8.

Rihtman, Branko; Puxty, Richard J.; Hapeshi, Alexia; Lee, Yan-Jiun; Zhan, Yuanchao; Michniewski, Slawomir et al. (2021): A new family of globally distributed lytic roseophages with unusual deoxythymidine to deoxyuridine substitution. In *Current biology : CB*. DOI: 10.1016/j.cub.2021.05.014.

Robinson, James T.; Thorvaldsdóttir, Helga; Winckler, Wendy; Guttman, Mitchell; Lander, Eric S.; Getz, Gad; Mesirov, Jill P. (2011): Integrative genomics viewer. In *Nature Biotechnology* 29. DOI: 10.1038/nbt.1754.

Rodwell, Ella V.; Wenner, Nicolas; Pulford, Caisey V.; Cai, Yueyi; Bowers-Barnard, Arthur; Beckett, Alison et al. (2021): Isolation and Characterisation of Bacteriophages with Activity against Invasive Non-Typhoidal Salmonella Causing Bloodstream Infection in Malawi. In *Viruses* 13 (3). DOI: 10.3390/v13030478.

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.

Rousset, François; Depardieu, Florence; Miele, Solange; Dowding, Julien; Laval, Anne-Laure; Lieberman, Erica et al. (2022): Phages and their satellites encode hotspots of antiviral systems. In *Cell host & microbe* 30 (5), 740-753.e5. DOI: 10.1016/j.chom.2022.02.018.

Roux, Simon; Adriaenssens, Evelien M.; Dutilh, Bas E.; Koonin, Eugene V.; Kropinski, Andrew M.; Krupovic, Mart et al. (2018): Minimum Information about an Uncultivated Virus Genome (MIUViG). In *Nature Biotechnology* 37. DOI: 10.1038/nbt.4306.

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

Roux, Simon; Emerson, Joanne B.; Eloe-Fadrosh, Emiley A.; Sullivan, Matthew B. (2017): Benchmarking viromics. An in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. In *PeerJ* 5, e3817. DOI: 10.7717/peerj.3817.

Roux, Simon; Enault, Francois; Hurwitz, Bonnie L.; Sullivan, Matthew B. (2015): VirSorter: mining viral signal from microbial genomic data. In *PeerJ* 3, pp. e985. DOI: 10.7717/peerj.985.

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.

Sabri, Mourad; Häuser, Roman; Ouellette, Marc; Liu, Jing; Dehbi, Mohammed; Moeck, Greg et al. (2011): Genome annotation and intraviral interactome for the Streptococcus pneumoniae virulent phage Dp-1. In *Journal of bacteriology* 193 (2), pp. 551–562. DOI: 10.1128/JB.01117-10.

Sakamoto, Mitsuo; Benno, Yoshimi (2006): Reclassification of Bacteroides distasonis, Bacteroides goldsteinii and Bacteroides merdae as Parabacteroides distasonis gen. nov., comb. nov., Parabacteroides goldsteinii comb. nov. and Parabacteroides merdae comb. nov. In *International journal of systematic and evolutionary microbiology* 56 (Pt 7), pp. 1599–1605. DOI: 10.1099/ijs.0.64192-0.

Sakamoto, Mitsuo; Suzuki, Masahito; Huang, Yi; Umeda, Makoto; Ishikawa, Isao; Benno, Yoshimi (2004): Prevotella shahii sp. nov. and Prevotella salivae sp. nov., isolated from the human oral cavity. In *International journal of systematic and evolutionary microbiology* 54 (Pt 3), pp. 877–883. DOI: 10.1099/ijs.0.02876-0.

Sakowski, Eric G.; Munsell, Erik V.; Hyatt, Mara; Kress, William; Williamson, Shannon J.; Nasko, Daniel J. et al. (2014): Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. In *Proceedings of the National Academy of Sciences of the United States of America* 111 (44), pp. 15786–15791. DOI: 10.1073/pnas.1401322111.

Salazar, Adam J.; Lessor, Lauren; O'Leary, Chandler; Gill, Jason; Liu, Mei (2019): Complete Genome Sequence of Klebsiella pneumoniae Siphophage Seifer. In *Microbiol Resour Announc* 8 (46). DOI: 10.1128/MRA.01289-19.

Santamaría, Rosa Isela; Bustos, Patricia; Sepúlveda-Robles, Omar; Lozano, Luis; Rodríguez, César; Fernández, José Luis et al. (2014): Narrow-host-range bacteriophages that infect Rhizobium etli associate with distinct genomic types. In *Applied and environmental microbiology* 80 (2), pp. 446–454. DOI: 10.1128/AEM.02256-13.

Sañudo-Wilhelmy, Sergio A.; Gómez-Consarnau, Laura; Suffridge, Christopher; Webb, Eric A. (2014): The role of B vitamins in marine biogeochemistry. In *Annual review of marine science* 6, pp. 339–367. DOI: 10.1146/annurev-marine-120710-100912.

Savilahti, H.; Bamford, D. H. (1993): Protein-primed DNA replication: role of inverted terminal repeats in the *Escherichia coli* bacteriophage PRD1 life cycle. In *J. Virol.* 67 (8), pp. 4696–4703.

Schmitz-Esser, Stephan; Toenshoff, Elena R.; Haider, Susanne; Heinz, Eva; Hoenninger, Verena M.; Wagner, Michael; Horn, Matthias (2008): Diversity of bacterial endosymbionts of environmental acanthamoeba isolates. In *Applied and environmental microbiology* 74 (18), pp. 5822–5831. DOI: 10.1128/AEM.01093-08.

Schulz, Frederik; Tyml, Tomáš; Pizzetti, Ilaria; Dyková, Iva; Fazi, Stefano; Kostka, Martin; Horn, Matthias (2015): Marine amoebae with cytoplasmic and perinuclear symbionts deeply branching in the *Gammaproteobacteria*. In *Scientific reports* 5, p. 13381. DOI: 10.1038/srep13381.

Seco, Elena M.; Zinder, John C.; Manhart, Carol M.; Lo Piano, Ambra; McHenry, Charles S.; Ayora, Silvia (2013): Bacteriophage SPP1 DNA replication strategies promote viral and disable host replication in vitro. In *Nucl Acids Res* 41 (3), pp. 1711–1721. DOI: 10.1093/nar/gks1290.

Seed, Kimberley D.; Bodi, Kip L.; Kropinski, Andrew M.; Ackermann, Hans-Wolfgang; Calderwood, Stephen B.; Qadri, Firdausi; Camilli, Andrew (2011): Evidence of a dominant lineage of *Vibrio cholerae*-specific lytic bacteriophages shed by cholera patients over a 10-year period in Dhaka, Bangladesh. In *mBio* 2 (1), e00334-10. DOI: 10.1128/mBio.00334-10.

Seed, Kimberley D.; Dennis, Jonathan J. (2005): Isolation and characterization of bacteriophages of the Burkholderia cepacia complex. In *FEMS Microbiology Letters* 251 (2), pp. 273–280. DOI: 10.1016/j.femsle.2005.08.011.

Seed, Kimberley D.; Yen, Minmin; Shapiro, B. Jesse; Hilaire, Isabelle J.; Charles, Richelle C.; Teng, Jessica E. et al. (2014): Evolutionary consequences of intra-patient phage predation on microbial populations. In *eLife* 3, e03497. DOI: 10.7554/eLife.03497.

Segev, Einat; Wyche, Thomas P.; Kim, Ki Hyun; Petersen, Jörn; Ellebrandt, Claire; Vlamakis, Hera et al. (2016): Dynamic metabolic exchange governs a marine algal-bacterial interaction. In *eLife* 5. DOI: 10.7554/eLife.17473.

Selje, Natascha; Simon, Meinhard; Brinkhoff, Thorsten (2004): A newly discovered *Roseobacter* cluster in temperate and polar oceans. In *Nature* 427 (6973), pp. 445–448. DOI: 10.1038/nature02272.

Sengupta, Rajib; Holmgren, Arne (2014): Thioredoxin and glutaredoxin-mediated redox regulation of ribonucleotide reductase. In *World journal of biological chemistry* 5 (1), pp. 68–74. DOI: 10.4331/wjbc.v5.i1.68.

Sepúlveda-Robles, Omar; Kameyama, Luis; Guarneros, Gabriel (2012): High diversity and novel species of Pseudomonas aeruginosa bacteriophages. In *Appl. Environ. Microbiol.* 78 (12), pp. 4510–4515. DOI: 10.1128/AEM.00065-12.

Serwer, Philip (2005): T3/T7 DNA Packaging. In Carlos Enrique Catalano (Ed.): Viral genome packaging machines. Genetics, structure, and mechanism. Georgetown, Tex.: Landes Bioscience/Eurekah.com; New York (Molecular Biology Intelligence Unit), pp. 59–79.

Seshadri, Rekha; Leahy, Sinead C.; Attwood, Graeme T.; Teh, Koon Hoong; Lambie, Suzanne C.; Cookson, Adrian L. et al. (2018): Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. In *Nat Biotechnol* 36 (4), pp. 359–367. DOI: 10.1038/nbt.4110.

Shahin, Khashayar; Soleimani-Delfan, Abbas; Zhang, Lili; Hedayatkhah, Abolghasem; Mansoorianfar, Mojtaba; Bao, Hongduo et al. (2022): Genome Sequence of Salmonella enterica Serovar Typhimurium Phage SAP12. In *Microbiol Resour Announc* 11 (6), e0108621. DOI: 10.1128/mra.01086-21.

Shen, Peter S.; Domek, Matthew J.; Sanz-García, Eduardo; Makaju, Aman; Taylor, Ryan M.; Hoggan, Ryan et al. (2012): Sequence and structural characterization of great salt lake bacteriophage CW02, a member of the T7-like supergroup. In *Journal of virology* 86 (15), pp. 7907–7917. DOI: 10.1128/JVI.00407-12.

Shimada, Kazunori; Weisberg, Robert A.; Gottesman, Max E. (1972): Prophage lambda at unusual chromosomal locations. In *Journal of molecular biology* 63 (3), pp. 483–503. DOI: 10.1016/0022-2836(72)90443-3.

Shkoporov, Andrei N.; Khokhlova, Ekaterina V.; Chaplin, Andrei V.; Kafarskaia, Lyudmila I.; Nikolin, Alexei A.; Polyakov, Vladimir Yu et al. (2013): Coprobacter fastidiosus gen. nov., sp. nov., a novel member of the family Porphyromonadaceae isolated from infant faeces. In *International journal of systematic and evolutionary microbiology* 63 (Pt 11), pp. 4181–4188. DOI: 10.1099/ijs.0.052126-0.

Silveira, Cynthia B.; Luque, Antoni; Rohwer, Forest (2021): The landscape of lysogeny across microbial community density, diversity and energetics. In *Environ Microbiol* 23 (8), pp. 4098–4111. DOI: 10.1111/1462-2920.15640.

Simmonds, Peter; Adriaenssens, Evelien M.; Zerbini, F. Murilo; Abrescia, Nicola G. A.; Aiewsakun, Pakorn; Alfenas-Zerbini, Poliane et al. (2023): Four principles to establish a universal virus taxonomy. In *PLoS biology* 21 (2), e3001922. DOI: 10.1371/journal.pbio.3001922.

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.

Skennerton, Connor T.; Angly, Florent E.; Breitbart, Mya; Bragg, Lauren; He, Shaomei; McMahon, Katherine D. et al. (2011): Phage encoded H-NS: a potential achilles heel in the bacterial defence system. In *PloS one* 6 (5), e20095. DOI: 10.1371/journal.pone.0020095.

Smet, Jeroen De; Zimmermann, Michael; Kogadeeva, Maria; Ceyssens, Pieter-Jan; Vermaelen, Wesley; Blasdel, Bob et al. (2016): High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. In *The ISME journal* 10 (8), p. 1823. DOI: 10.1038/ismej.2016.3.

Smith, Margaret C. M.; Brown, William R. A.; McEwan, Andrew R.; Rowley, Paul A. (2010): Site-specific recombination by phiC31 integrase and other large serine recombinases. In *Biochm. Soc. Trans.* 38 (2), pp. 388–394. DOI: 10.1042/BST0380388.

Song, Wenchen; Sun, Hai-Xi; Zhang, Carolyn; Cheng, Li; Peng, Ye; Deng, Ziqing et al. (2019): Prophage Hunter: an integrative hunting tool for active prophages. In *Nucleic acids research* 47 (W1), W74-W80. DOI: 10.1093/nar/gkz380.

Song, Yuli; Könönen, Eija; Rautio, Merja; Liu, Chengxu; Bryk, Anne; Eerola, Erkki; Finegold, Sydney M. (2006): Alistipes onderdonkii sp. nov. and Alistipes shahii sp. nov., of human origin. In *International journal of systematic and evolutionary microbiology* 56 (Pt 8), pp. 1985–1990. DOI: 10.1099/ijs.0.64318-0.

Sonnenschein, Eva C.; Nielsen, Kristian F.; D'Alvise, Paul; Porsby, Cisse H.; Melchiorsen, Jette; Heilmann, Jens et al. (2017): Global occurrence and heterogeneity of the Roseobacter-clade species Ruegeria mobilis. In *The ISME journal* 11 (2), pp. 569–583. DOI: 10.1038/ismej.2016.111.

Souza, Rocheli de; Sant'Anna, Fernando Hayashi; Ambrosini, Adriana; Tadra-Sfeir, Michele; Faoro, Helisson; Pedrosa, Fabio Oliveira et al. (2015): Genome of Rhizobium sp. UR51a, Isolated from Rice Cropped in Southern Brazilian Fields. In *Genome announcements* 3 (2). DOI: 10.1128/genomeA.00249-15.

Stepanova, O. A.; Boĭko, A. L.; Shcherbatenko, I. S. (2013): Computational genome analysis of three marine algoviruses. In *Mikrobiolohichnyi zhurnal (Kiev, Ukraine : 1993)* 75 (5), pp. 76–81.

Stoddard, Barry L. (2011): Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. In *Structure (London, England : 1993)* 19 (1), pp. 7–15. DOI: 10.1016/j.str.2010.12.003.

Sullivan, Matthew B.; Huang, Katherine H.; Ignacio-Espinoza, Julio C.; Berlin, Aaron M.; Kelly, Libusha; Weigele, Peter R. et al. (2010): Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. In *Environmental microbiology* 12 (11), pp. 3035–3056. DOI: 10.1111/j.1462-2920.2010.02280.x.

Sullivan, Matthew B.; Weitz, Joshua S.; Wilhelm, Steven (2017): Viral ecology comes of age. In *Environmental microbiology reports* 9 (1), pp. 33–35. DOI: 10.1111/1758-2229.12504.

Sullivan, Teresa; Manuel, Nikhil; Clark, James; Liu, Mei; Burrowes, Ben (2022): Complete Genome Sequence of Stenotrophomonas maltophilia Siphophage Suzuki. In *Microbiol Resour Announc* 11 (4), e0013622. DOI: 10.1128/mra.00136-22.

Summer, Elizabeth J.; Berry, Joel; Tran, Tram Anh T.; Niu, Lili; Struck, Douglas K.; Young, Ry (2007a): Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. In *Journal of molecular biology* 373 (5), pp. 1098–1112. DOI: 10.1016/j.jmb.2007.08.045.

Summer, Elizabeth J.; Gill, Jason J.; Upton, Chris; Gonzalez, Carlos F.; Young, Ry (2007b): Role of phages in the pathogenesis of Burkholderia, or 'Where are the toxin genes in Burkholderia phages?'. In *Current opinion in microbiology* 10 (4), pp. 410–417. DOI: 10.1016/j.mib.2007.05.016.

Summer, Elizabeth J.; Gonzalez, Carlos F.; Bomer, Morgan; Carlile, Thomas; Embry, Addie; Kucherka, Amalie M. et al. (2006): Divergence and mosaicism among virulent soil phages of the Burkholderia cepacia complex. In *Journal of bacteriology* 188 (1), pp. 255–268. DOI: 10.1128/JB.188.1.255-268.2006.

Sutorius, Mara; Mori, Corinna; Greskowiak, Janek; Boettcher, Lea; Bunse, Carina; Dittmar, Thorsten et al. (2022): Rare earth element behaviour in seawater under the influence of organic matter cycling during a phytoplankton spring bloom – A mesocosm study. In *Front. Mar. Sci.* 9, Article 895723. DOI: 10.3389/fmars.2022.895723.

Suttle, C. A. (1994): The significance of viruses to mortality in aquatic microbial communities. In *Microbial Ecology* 28 (2), pp. 237–243. DOI: 10.1007/BF00166813.

Suttle, Curtis A. (2005): Viruses in the sea. In *Nature* 437 (7057), pp. 356–361. DOI: 10.1038/nature04160.

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.

Suzuki, M. T.; Giovannoni, S. J. (1996): Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. In *Applied and environmental microbiology* 62 (2), pp. 625–630. DOI: 10.1128/aem.62.2.625-630.1996.

Synnott, Aidan J.; Kuang, Ying; Kurimoto, Miki; Yamamichi, Keiko; Iwano, Hidetomo; Tanji, Yasunori (2009): Isolation from sewage influent and characterization of novel Staphylococcus aureus bacteriophages with wide host ranges and potent lytic capabilities. In *Appl. Environ. Microbiol.* 75 (13), pp. 4483–4490. DOI: 10.1128/AEM.02641-08.

Szymczak, Mateusz; Grygorcewicz, Bartłomiej; Karczewska-Golec, Joanna; Decewicz, Przemysław; Pankowski, Jarosław Adam; Országh-Szturo, Hanna et al. (2020): Characterization of a Unique Bordetella bronchiseptica vB_BbrP_BB8 Bacteriophage and Its Application as an Antibacterial Agent. In *International journal of molecular sciences* 21 (4). DOI: 10.3390/ijms21041403.

Tal, Nitzan; Millman, Adi; Stokar-Avihail, Avigail; Fedorenko, Taya; Leavitt, Azita; Melamed, Sarah et al. (2022): Bacteria deplete deoxynucleotides to defend against bacteriophage infection. In *Nat Microbiol* 7 (8), pp. 1200–1209. DOI: 10.1038/s41564-022-01158-0.

Tal, Nitzan; Sorek, Rotem (2022): SnapShot: Bacterial immunity. In *Cell* 185 (3), 578-578.e1. DOI: 10.1016/j.cell.2021.12.029.

Tang, Chaofei; Deng, Chuanjiang; Zhang, Yi; Xiao, Cong; Wang, Jing; Rao, Xiancai et al. (2018): Characterization and Genomic Analyses of Pseudomonas aeruginosa Podovirus TC6: Establishment of Genus Pa11virus. In *Frontiers in microbiology* 9, p. 2561. DOI: 10.3389/fmicb.2018.02561.

Tang, Kai; Lin, Dan; Zheng, Qiang; Liu, Keshao; Yang, Yujie; Han, Yu; Jiao, Nianzhi (2017): Genomic, proteomic and bioinformatic analysis of two temperate phages in *Roseobacter* clade bacteria isolated from the deep-sea water. In *BMC genomics* 18 (1), p. 485. DOI: 10.1186/s12864-017-3886-0.

Tang, Kai; Yang, Yujie; Lin, Dan; Li, Shuhui; Zhou, Wenchu; Han, Yu et al. (2016): Genomic, physiologic, and proteomic insights into metabolic versatility in Roseobacter clade bacteria isolated from deep-sea water. In *Scientific reports* 6 (1), p. 35528. DOI: 10.1038/srep35528.

Tang, Ying Zhong; Koch, Florian; Gobler, Christopher J. (2010): Most harmful algal bloom species are vitamin B1 and B12 auxotrophs. In *Proceedings of the National Academy of Sciences of the United States of America* 107 (48), pp. 20756–20761. DOI: 10.1073/pnas.1009566107.

Tariq, Mohammad A.; Everest, Francesca L. C.; Cowley, Lauren A.; Wright, Rosanna; Holt, Giles S.; Ingram, Hazel et al. (2019): Temperate Bacteriophages from Chronic Pseudomonas aeruginosa Lung Infections Show Disease-Specific Changes in Host Range and Modulate Antimicrobial Susceptibility. In *mSystems* 4 (4). DOI: 10.1128/mSystems.00191-18.

Taylor, Véronique L.; Fitzpatrick, Alexa D.; Islam, Zafrin; Maxwell, Karen L. (2019): The Diverse Impacts of Phage Morons on Bacterial Fitness and Virulence 103, pp. 1–31. DOI: 10.1016/bs.aivir.2018.08.001.

Taylor-Brown, Alyce; Vaughan, Lloyd; Greub, Gilbert; Timms, Peter; Polkinghorne, Adam (2015): Twenty years of research into *Chlamydia*-like organisms: a revolution in our understanding of the biology and pathogenicity of members of the phylum *Chlamydiae*. In *Pathogens and disease* 73 (1), pp. 1–15. DOI: 10.1093/femspd/ftu009.

Teeling, Hanno; Fuchs, Bernhard M.; Becher, Dörte; Klockow, Christine; Gardebrecht, Antje; Bennke, Christin M. et al. (2012): Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. In *Science (New York, N.Y.)* 336 (6081), pp. 608–611. DOI: 10.1126/science.1218344.

Teeling, Hanno; Fuchs, Bernhard M.; Bennke, Christin M.; Krüger, Karen; Chafee, Meghan; Kappelmann, Lennart et al. (2016): Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. In *eLife* 5, e11888. DOI: 10.7554/eLife.11888.

Terzian, Paul; Olo Ndela, Eric; Galiez, Clovis; Lossouarn, Julien; Pérez Bucio, Rubén Enrique; Mom, Robin et al. (2021): PHROG: families of prokaryotic virus proteins clustered using remote homology. In *NAR genomics and bioinformatics* 3 (3), lqab067. DOI: 10.1093/nargab/lqab067.

Teve, Michael; Clark, James; Le, Tram; Burrowes, Ben; Liu, Mei (2022): Complete Genome Sequence of Stenotrophomonas maltophilia Siphophage Sonora. In *Microbiol Resour Announc* 11 (4), e0016722. DOI: 10.1128/mra.00167-22.

Thiaville, Jennifer J.; Kellner, Stefanie M.; Yuan, Yifeng; Hutinet, Geoffrey; Thiaville, Patrick C.; Jumpathong, Watthanachai et al. (2016): Novel genomic island modifies DNA with 7-deazaguanine derivatives. In *PNAS* 113 (11), E1452-9. DOI: 10.1073/pnas.1518570113.

Thingstad, T. Frede (2000): Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. In *Limnol. Oceanogr*. 45 (6), pp. 1320–1328. DOI: 10.4319/lo.2000.45.6.1320.

Thornton, Daniel C.O. (2014): Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. In *European Journal of Phycology* 49 (1), pp. 20–46. DOI: 10.1080/09670262.2013.875596.

Tikhe, Chinmay Vijay; Martin, Thomas M.; Gissendanner, Chris R.; Husseneder, Claudia (2015): Complete Genome Sequence of Citrobacter Phage CVT22 Isolated from the Gut of the Formosan Subterranean Termite, Coptotermes formosanus Shiraki. In *Genome announcements* 3 (4). DOI: 10.1128/genomeA.00408-15.

Tolstoy, I.; Turner, D.; Moraru, C.; Adriaenssens, EM.; am Kropinski (2021): ICTV Taxonomy Proposal 2021.015B: Create one new family (Casjensviridae) including 20 new genera and four existing genera (Caudoviricetes). Available online at https://ictv.global/ictv/proposals/2021.015B.R.Casjensviridae.zip.

Torres-Acosta, Mario A.; Clavijo, Viviana; Vaglio, Christopher; González-Barrios, Andrés F.; Vives-Flórez, Martha J.; Rito-Palomares, Marco (2019): Economic evaluation of the development of a phage therapy product for the control of Salmonella in poultry. In *Biotechnology progress* 35 (5), e2852. DOI: 10.1002/btpr.2852.

Tremblay, Denise M.; Tegoni, Mariella; Spinelli, Silvia; Campanacci, Valérie; Blangy, Stéphanie; Huyghe, Céline et al. (2006): Receptor-binding protein of Lactococcus lactis phages: identification and characterization of the saccharide receptor-binding site. In *Journal of bacteriology* 188 (7), pp. 2400–2410. DOI: 10.1128/JB.188.7.2400-2410.2006.

Turner, D.; Kropinski, Andrew, M.; Adriaenssens Evelien, M. (2020): ICTV Taxonomy proposal 2020.010B: Rename 2,532 bacterial virus species to the binomial format. Available online at https://ictv.global/ictv/proposals/2021.010B.R.Binomial_names.zip.

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

Uchiyama, Jumpei; Rashel, Mohammad; Matsumoto, Tetsuya; Sumiyama, Yoshinobu; Wakiguchi, Hiroshi; Matsuzaki, Shigenobu (2009): Characteristics of a novel Pseudomonas aeruginosa bacteriophage, PAJU2, which is genetically related to bacteriophage D3. In *Virus research* 139 (1), pp. 131–134. DOI: 10.1016/j.virusres.2008.10.005.

Urtecho, Guillaume; Campbell, Danielle E.; Hershey, David M.; Hussain, Fatima A.; Whitaker, Rachel J.; O'Toole, George A. (2020): Discovering the Molecular Determinants of Phaeobacter inhibens Susceptibility to Phaeobacter Phage MD18. In *mSphere* 5 (6). DOI: 10.1128/mSphere.00898-20.

Vacheron, Jordan; Kupferschmied, Peter; Resch, Grégory; Keel, Christoph (2018): Genome Sequence of the Pseudomonas protegens Phage ΦGP100. In *Genome Announc.* 6 (25). DOI: 10.1128/genomeA.00261-18.

van Zyl, Leonardo Joaquim; Nemavhulani, Shonisani; Cass, James; Cowan, Donald Arthur; Trindade, Marla (2016): Three novel bacteriophages isolated from the East African Rift Valley soda lakes. In *Virology journal* 13 (1), p. 204. DOI: 10.1186/s12985-016-0656-6.

van Zyl, Lonnie; Adriaenssens, E. M.; Kropinski, Andrew (2018): ICTV Taxonomy Proposal 2018.124B: To create one (1) new genus, Vhulanivirus, containing one species in the family Siphoviridae. Available online at https://ictv.global/ictv/proposals/2018.124B.A.v1.Vhulanivirus.zip.

Vandamme, Erick J.; Mortelmans, Kristien (2019): A century of bacteriophage research and applications: impacts on biotechnology, health, ecology and the economy! In *J. Chem. Technol. Biotechnol.* 94 (2), pp. 323–342. DOI: 10.1002/jctb.5810.

Vargas, Colomban de; Audic, Stéphane; Henry, Nicolas; Decelle, Johan; Mahé, Frédéric; Logares, Ramiro et al. (2015): Eukaryotic plankton diversity in the sunlit ocean. In *Science* 348. DOI: 10.1126/science.1261605.

Vila-Nistal, Marina; Maestre-Carballa, Lucia; Martinez-Hernández, Francisco; Martinez-Garcia, Manuel (2023): Novel RNA viruses from the Atlantic Ocean: Ecogenomics, biogeography, and total virioplankton mass contribution from surface to the deep ocean. In *Environ Microbiol*. DOI: 10.1111/1462-2920.16502.

Villamor, Judith; Ramos-Barbero, María Dolores; González-Torres, Pedro; Gabaldón, Toni; Rosselló-Móra, Ramón; Meseguer, Inmaculada et al. (2018): Characterization of ecologically diverse viruses infecting co-occurring strains of cosmopolitan hyperhalophilic *Bacteroidetes*. In *The ISME journal* 12 (2), pp. 424–437. DOI: 10.1038/ismej.2017.175.

Villarroel, Julia; Kleinheinz, Kortine Annina; Jurtz, Vanessa Isabell; Zschach, Henrike; Lund, Ole; Nielsen, Morten; Larsen, Mette Voldby (2016): HostPhinder. A Phage Host Prediction Tool. In *Viruses* 8 (5). DOI: 10.3390/v8050116.

Vlot, Marnix; Houkes, Joep; Lochs, Silke J. A.; Swarts, Daan C.; Zheng, Peiyuan; Kunne, Tim et al. (2018): Bacteriophage DNA glucosylation impairs target DNA binding by type I and II but not by type V CRISPR-Cas effector complexes. In *Nucl Acids Res* 46 (2), pp. 873–885. DOI: 10.1093/nar/gkx1264.

Wagemans, Jeroen; Tsonos, Jessica; Holtappels, Dominique; Fortuna, Kiandro; Hernalsteens, Jean-Pierre; Greve, Henri de et al. (2020): Structural Analysis of Jumbo Coliphage phAPEC6. In *International journal of molecular sciences* 21 (9). DOI: 10.3390/ijms21093119.

Wagner-Döbler, Irene; Ballhausen, Britta; Berger, Martine; Brinkhoff, Thorsten; Buchholz, Ina; Bunk, Boyke et al. (2010): The complete genome sequence of the algal symbiont *Dinoroseobacter shibae*: a hitchhiker's guide to life in the sea. In *The ISME journal* 4 (1), pp. 61–77. DOI: 10.1038/ismej.2009.94.

Waldor, M. K.; Mekalanos, J. J. (1996): Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin. In *Science* 272 (5270), pp. 1910–1914. DOI: 10.1126/science.272.5270.1910.

Walker, Peter J.; Siddell, Stuart G.; Lefkowitz, Elliot J.; Mushegian, Arcady R.; Adriaenssens, Evelien M.; Alfenas-Zerbini, Poliane et al. (2022): Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). In *Arch Virol* 167 (11), pp. 2429–2440. DOI: 10.1007/s00705-022-05516-5.

Wallace, James C.; Youngblood, Jessica E.; Port, Jesse A.; Cullen, Alison C.; Smith, Marissa N.; Workman, Tomomi; Faustman, Elaine M. (2018): Variability in metagenomic samples from the Puget Sound: Relationship to temporal and anthropogenic impacts. In *PloS one* 13. DOI: 10.1371/journal.pone.0192412.

Wang, Chao-Nan; Liu, Yuan; Wang, Jing; Du, Zong-Jun; Wang, Ming-Yi (2021): Sulfitobacter algicola sp. nov., isolated from green algae. In *Arch Microbiol* 203 (5), pp. 2351–2356. DOI: 10.1007/s00203-021-02213-w.

Wang, Hui; Tomasch, Jürgen; Jarek, Michael; Wagner-Döbler, Irene (2014): A dual-species co-cultivation system to study the interactions between *Roseobacters* and dinoflagellates. In *Frontiers in microbiology* 5, p. 311. DOI: 10.3389/fmicb.2014.00311.

Wang, Lianrong; Chen, Shi; Xu, Tiegang; Taghizadeh, Koli; Wishnok, John S.; Zhou, Xiufen et al. (2007): Phosphorothioation of DNA in bacteria by dnd genes. In *Nat Chem Biol* 3 (11), pp. 709–710. DOI: 10.1038/nchembio.2007.39.

Wang, Qiong; Cai, Lanlan; Zhang, Rui; Wei, Shuzhen; Li, Fang; Liu, Yuanfang; Xu, Yongle (2022): A Unique Set of Auxiliary Metabolic Genes Found in an Isolated Cyanophage Sheds New Light on Marine Phage-Host Interactions. In *Microbiology spectrum* 10 (5), e0236722. DOI: 10.1128/spectrum.02367-22.

Wang, Xiaoxue; Kim, Younghoon; Ma, Qun; Hong, Seok Hoon; Pokusaeva, Karina; Sturino, Joseph M.; Wood, Thomas K. (2010): Cryptic prophages help bacteria cope with adverse environments. In *Nat Commun* 1 (1), p. 147. DOI: 10.1038/ncomms1146.

Weaver, P. F.; Wall, J. D.; Gest, H. (1975): Characterization of Rhodopseudomonas capsulata. In *Arch. Microbiol.* 105 (3), pp. 207–216. DOI: 10.1007/BF00447139.

Wei, Cuihua; Liu, Junli; Maina, Alice Nyambura; Mwaura, Francis B.; Yu, Junping; Yan, Chenghui et al. (2017): Developing a bacteriophage cocktail for biocontrol of potato bacterial wilt. In *Virologica Sinica* 32 (6), pp. 476–484. DOI: 10.1007/s12250-017-3987-6.

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.

Wemheuer, Bernd; Wemheuer, Franziska; Hollensteiner, Jacqueline; Meyer, Frauke-Dorothee; Voget, Sonja; Daniel, Rolf (2015): The green impact: bacterioplankton response toward a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches. In *Frontiers in microbiology* 6, p. 805. DOI: 10.3389/fmicb.2015.00805.

Wichels, A.; Biel, S. S.; Gelderblom, H. R.; Brinkhoff, T.; Muyzer, G.; Schütt, C. (1998): Bacteriophage diversity in the North Sea. In *Applied and environmental microbiology* 64 (11), pp. 4128–4133. DOI: 10.1128/AEM.64.11.4128-4133.1998.

Wick, Ryan R.; Schultz, Mark B.; Zobel, Justin; Holt, Kathryn E. (2015): Bandage: interactive visualization of de novo genome assemblies. In *Bioinformatics (Oxford, England)* 31 (20), pp. 3350–3352. DOI: 10.1093/bioinformatics/btv383.

Widdel, Friedrich; Bak, Friedhelm (2013): Gram-Negative Mesophilic Sulfate-Reducing Bacteria. In Albert Balows, Hans G. Trüper, Martin Dworkin, Wim Harder, Karl-Heinz Schleifer (Eds.): The Prokaryotes. A Handbook on the Biology of Bacteria: Ecophysiology, Isolation, Identification, Applications. 2nd ed. New York, NY: Springer, pp. 3352–3378.

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.

Wiltshire, Karen Helen; Kraberg, Alexandra; Bartsch, Inka; Boersma, Maarten; Franke, Heinz-Dieter; Freund, Jan et al. (2010): Helgoland Roads, North Sea: 45 Years of Change. In *Estuaries and Coasts* 33 (2), pp. 295–310. DOI: 10.1007/s12237-009-9228-y.

Winget, Danielle M.; Wommack, K. Eric (2008): Randomly amplified polymorphic DNA PCR as a tool for assessment of marine viral richness. In *Applied and environmental microbiology* 74 (9), pp. 2612–2618. DOI: 10.1128/AEM.02829-07.

Wittmann, Johannes; Turner, Dann; Millard, Andrew D.; Mahadevan, Padmanabhan; Kropinski, Andrew M.; Adriaenssens, Evelien M. (2020): From Orphan Phage to a Proposed New Family-the Diversity of N4-Like Viruses. In *Antibiotics (Basel, Switzerland)* 9 (10). DOI: 10.3390/antibiotics9100663.

Wolf, Arite; Fritze, Antje; Hagemann, Martin; Berg, Gabriele (2002): Stenotrophomonas rhizophila sp. nov., a novel plant-associated bacterium with antifungal properties. In *International journal of systematic and evolutionary microbiology* 52 (Pt 6), pp. 1937–1944. DOI: 10.1099/00207713-52-6-1937.

Woods, Donald E.; Jeddeloh, Jeffrey A.; Fritz, David L.; DeShazer, David (2002): Burkholderia thailandensis E125 harbors a temperate bacteriophage specific for Burkholderia mallei. In *Journal of bacteriology* 184 (14), pp. 4003–4017. DOI: 10.1128/JB.184.14.4003-4017.2002.

Xu, Yongle; Zhang, Rui; Jiao, Nianzhi (2015): Complete genome sequence of Paracoccus marcusii phage vB_PmaS-R3 isolated from the South China Sea. In *Standards in genomic sciences* 10, p. 94. DOI: 10.1186/s40793-015-0089-7.

Xu, Yongle; Zhang, Rui; Wang, Nannan; Cai, Lanlan; Tong, Yigang; Sun, Qiang et al. (2018): Novel phage-host interactions and evolution as revealed by a cyanomyovirus isolated from an estuarine environment. In *Environmental microbiology* 20 (8), pp. 2974–2989. DOI: 10.1111/1462-2920.14326.

Yang, Jhung-Ahn; Kang, Ilnam; Moon, Mikyung; Ryu, Uh-Chan; Kwon, Kae Kyoung; Cho, Jang-Cheon; Oh, Hyun-Myung (2016): Complete genome sequence of Celeribacter marinus IMCC12053(T), the host strain of marine bacteriophage P12053L. In *Marine genomics* 26, pp. 5–7. DOI: 10.1016/j.margen.2015.11.012.

Yang, Qiao; Ge, Ya-ming; Iqbal, Nurhezreen Md; Yang, Xi; Zhang, Xiao-ling (2021): Sulfitobacter alexandrii sp. nov., a new microalgae growth-promoting bacterium with exopolysaccharides bioflocculanting potential isolated from marine phycosphere. In *Antonie van Leeuwenhoek* 114 (7), pp. 1091–1106. DOI: 10.1007/s10482-021-01580-0.

Yang, Ying-Jie; Zhang, Ning; Ji, Shi-Qi; Lan, Xin; Zhang, Kun-di; Shen, Yu-Long et al. (2014): Dysgonomonas macrotermitis sp. nov., isolated from the hindgut of a fungus-growing termite. In *International journal of systematic and evolutionary microbiology* 64 (Pt 9), pp. 2956–2961. DOI: 10.1099/ijs.0.061739-0.

Yang, Yunlan; Cai, Lanlan; Ma, Ruijie; Xu, Yongle; Tong, Yigang; Huang, Yong et al. (2017): A novel Roseosiphophage isolated from the oligotrophic South China Sea. In *Viruses* 9 (5). DOI: 10.3390/v9050109.

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.

Zaczek-Moczydłowska, Maja A.; Young, Gillian K.; Trudgett, James; Fleming, Colin C.; Campbell, Katrina; O'Hanlon, Richard (2020): Genomic Characterization, Formulation and Efficacy in Planta of a Siphoviridae and Podoviridae Protection Cocktail against the Bacterial Plant Pathogens Pectobacterium spp. In *Viruses* 12 (2). DOI: 10.3390/v12020150.

Zeng, Jian; Wang, Yan; Zhang, Ju; Yang, Shixing; Zhang, Wen (2021): Multiple novel filamentous phages detected in the cloacal swab samples of birds using viral metagenomics approach. In *Virol J* 18 (1), p. 240. DOI: 10.1186/s12985-021-01710-0.

Zhai, Zhiqiang; Zhang, Zefeng; Zhao, Guiyuan; Liu, Xinxin; Qin, Fang; Zhao, Yanlin (2021): Genomic Characterization of Two Novel RCA Phages Reveals New Insights into the Diversity and Evolution of Marine Viruses. In *Microbiology spectrum* 9 (2), e0123921. DOI: 10.1128/Spectrum.01239-21.

Zhan, Yuanchao; Chen, Feng (2019a): Bacteriophages that infect marine roseobacters: genomics and ecology. In *Environmental microbiology* 21 (6), pp. 1885–1895. DOI: 10.1111/1462-2920.14504.

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

Zhan, Yuanchao; Huang, Sijun; Chen, Feng (2018): Genome Sequences of Five Bacteriophages Infecting the Marine Roseobacter Bacterium Ruegeria pomeroyi DSS-3. In *Microbiol Resour Announc* 7 (12), p. 910. DOI: 10.1128/MRA.00959-18.

Zhan, Yuanchao; Huang, Sijun; Voget, Sonja; Simon, Meinhard; Chen, Feng (2016): A novel roseobacter phage possesses features of podoviruses, siphoviruses, prophages and gene transfer agents. In *Scientific reports* 6, p. 30372. DOI: 10.1038/srep30372.

Zhang, Carolyn; Song, Wenchen; Ma, Helena R.; Peng, Xiao; Anderson, Deverick J.; Fowler, Vance G. et al. (2020a): Temporal encoding of bacterial identity and traits in growth dynamics. In *PNAS* 117 (33), pp. 20202–20210. DOI: 10.1073/pnas.2008807117.

Zhang, Fuxing; Fan, Yongxiang; Zhang, Danyang; Chen, Shuangshuang; Bai, Xue; Ma, Xiaohong et al. (2020b): Effect and mechanism of the algicidal bacterium Sulfitobacter porphyrae ZFX1 on the mitigation of harmful algal blooms caused by Prorocentrum donghaiense. In *Environmental pollution (Barking, Essex : 1987)* 263 (Pt A), p. 114475. DOI: 10.1016/j.envpol.2020.114475.

Zhang, Yongyu; Jiao, Nianzhi (2009): Roseophage RDJL Phi1, infecting the aerobic anoxygenic phototrophic bacterium *Roseobacter denitrificans* OCh114. In *Applied and environmental microbiology* 75 (6), pp. 1745–1749. DOI: 10.1128/AEM.02131-08.

Zhang, Zefeng; Chen, Feng; Chu, Xiao; Zhang, Hao; Luo, Haiwei; Qin, Fang et al. (2019a): Diverse, abundant and novel viruses infecting “unculturable” but abundant marine bacteria 42. DOI: 10.1101/699256.

Zhang, Zefeng; Chen, Feng; Chu, Xiao; Zhang, Hao; Luo, Haiwei; Qin, Fang et al. (2019b): Diverse, Abundant, and Novel Viruses Infecting the Marine Roseobacter RCA Lineage. In *mSystems* 4 (6). DOI: 10.1128/mSystems.00494-19.

Zhang, Zefeng; Qin, Fang; Chen, Feng; Chu, Xiao; Luo, Haiwei; Zhang, Rui et al. (2021): Culturing novel and abundant pelagiphages in the ocean. In *Environmental microbiology* 23 (2), pp. 1145–1161. DOI: 10.1111/1462-2920.15272.

Zhao, Yanlin; Qin, Fang; Zhang, Rui; Giovannoni, Stephen J.; Zhang, Zefeng; Sun, Jing et al. (2018): Pelagiphages in the Podoviridae family integrate into host genomes. In *Environmental microbiology*. DOI: 10.1111/1462-2920.14487.

Zhao, Yanlin; Temperton, Ben; Thrash, J. Cameron; Schwalbach, Michael S.; Vergin, Kevin L.; Landry, Zachary C. et al. (2013): Abundant SAR11 viruses in the ocean. In *Nature* 494 (7437), pp. 357–360. DOI: 10.1038/nature11921.

Zhao, Yanlin; Wang, Kui; Jiao, Nianzhi; Chen, Feng (2009): Genome sequences of two novel phages infecting marine roseobacters. In *Environmental microbiology* 11 (8), pp. 2055–2064. DOI: 10.1111/j.1462-2920.2009.01927.x.

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.

Acknowledgements

In 2015, during my very first practical course within the Master´s program in Microbiology, I joined the group supervised by Dr. Cristina Moraru working on phages - and already I was infected. Since then, I have been on the hunt for new roseophages, starting as a Hiwi, then for my Master´s thesis and finally as a PhD student. This journey would not have been as successful and beautiful without the support of many people I want to acknowledge here.

First, I would like to thank Prof. Meinhard Simon for providing me the opportunity to join his working group at the ICBM and the Roseobacter SFB. Thank you for your support throughout the years, for reviewing my thesis and for being my first examiner.

I am deeply grateful to Dr. Cristina Moraru, who has been a constant source of support throughout these years and whose enthusiasm for phage research has been so inspiring to me. Thank you very much for being not only a great supervisor but also a friend!

Furthermore, I would like to thank Prof. Karin Holmfeldt and Dr. Bert Engelen for joining my thesis committee and being part of my examination board as second and third reviewer.

Many thanks are owed to the entire AG Simon and AG Cypionka for cultivating such a warm and welcoming working environment, especially to the "TA-Backbone" of Andrea, Mathias, Birgit, Rolf, and Frank, who were always there with their helping hands and expertise.

I am also immensely grateful to the "ICBM phagers" family including Cristina, Falk, Ismail, Aaron, Ben, and Andrea, as well as Anne, Martin, and Eileen. Working in a team with you has been such a great experience and without your dedicated support, the multitude of phage isolations, DNA extractions and host range tests, involving an insane number of agar plates, would have been impossible. Thank you so much!

Thanks also to Carlota, Anneke, and the Planktotron crew for providing the bacterial host strains and the seawater samples that laid the foundation for so many successful roseophage isolations.

Further, I am very thankful to all my fellow PhD colleagues and friends who made this journey unforgettable. A big thank you to Falk, Mara, and Melissa for proofreading my thesis!

Finally, I would like to thank my family - my parents and my brother, who have made this path possible for me and who have always believed in me. - Vielen Dank für eure Unterstützung! ♥

And to Viktor, who suffered most during the last years, guiding me through the many ups and downs and always taking care of me. - DANKE! ♥

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Zusätzlich erkläre ich, dass diese Dissertation weder in ihrer Gesamtheit noch in Teilen einer anderen wissenschaftlichen Hochschule zur Begutachtung in einem Promotionsverfahren vorliegt oder vorgelegen hat. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

_____ _____

Ort, Datum Unterschrift