MICROBIAL PARASITOIDS: 
GIANT VIRUSES AND TINY BACTERIA

by

CHRISTOPH MICHAEL DEEG

Diploma Biology, Albert Ludwig University of Freiburg, Germany, 2012

A thesis submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies
(Microbiology and Immunology)

The University of British Columbia
(Vancouver)

October 2018

© Christoph Michael Deeg, 2018
The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Microbial parasitoids: Giant viruses and tiny bacteria

submitted by Christoph Michael Deeg in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Microbiology and Immunology

Examining Committee:

Curtis Suttle; Microbiology & Immunology, Botany, Earth, Ocean and Atmospheric Sciences
Supervisor

Sean Crowe; Microbiology & Immunology; Earth, Ocean, and Atmospheric Sciences
Supervisory Committee Member

Patrick Keeling; Botany
Supervisory Committee Member

Julian Davies; Microbiology & Immunology
University Examiner

Rosemary Redfield; Zoology
University Examiner

Additional Supervisory Committee Members:

Laura Wegener-Parfrey, Botany, Zoology
Supervisory Committee Member
ABSTRACT

Microbial parasitoids that exploit other microbes are abundant, but remain a poorly explored frontier in microbiology. To study such pathogens, a high throughput screen was developed using ultrafiltration and flow cytometry, resulting in the isolation of five giant viruses and one bacterial pathogen infecting heterotrophic flagellates, as well as a bacterial predator of prokaryotes. Bodo saltans virus (BsV) is the first characterized representative of the most abundant group of giant viruses in oceans, so far only known from metagenomic data. Its 1.39 Mb genome encodes 1227 predicted ORFs; yet, much of its translational apparatus has been lost, including all tRNAs. Essential genes are invaded by homing endonuclease-encoding self-splicing introns that may defend against competing viruses. Ankyrin-repeat proteins that are putative anti-host factors show extensive gene duplication via a genomic accordion, indicating an ongoing evolutionary arms race and highlighting the rapid evolution and genomic plasticity leading to genome gigantism in giant viruses. Chromulinavorax destructans is an isolate from the TM6/Dependentiae phylum that infects and lyses the abundant heterotrophic flagellate Spumella elongata. Chromulinavorax destructans is characterized by a high degree of reduction and specialization. Its 1.2 Mb genome shows no metabolic potential, relying on an extensive transporter system to import nutrients and energy in the form of ATP from the host. It replicates by extensively reorganizing and expanding the host mitochondrion. Almost half of the inferred proteins contain signal sequences for secretion, which include many proteins of unknown function as well as 98 copies of ankyrin-repeat proteins, suggesting the presence of an extensive host-manipulation apparatus. Bdellovibrio salishius was found to exploit a beta-proteobacterium in an epibiotic manner. Despite this, B. salishius encodes a complex genomic complement more similar to periplasmic species as well as several biosynthesis pathways not previously found in epibiotic species. Bdellovibrio salishius is a representative of a widely distributed basal cluster within the genus Bdellovibrio, suggesting that epibiotic feeding might be a common predation type in nature and ancestral feature in the genus. The microorganisms described here broaden our understanding of microbial diversity and the unusual genomic functions associated with a parasitoid lifestyle amongst microbes.
Lay Summary

Pathogens of microscopic organisms in aquatic environments are poorly studied. They include giant viruses that rival cells in complexity and bacteria with reduced genomes. I characterize several isolates of previously unknown pathogens that infect and kill aquatic microbes. The giant virus, BsV, is representative of an abundant virus group in aquatic environments and shows signatures of genome expansion, providing insight into how giant viruses became so complex. A bacterial pathogen infecting another microzooplankton has a smaller genome than BsV and does not encode a metabolism, implying that it relies exclusively on its host. Another bacterial predator of other bacteria consumes its prey from the outside in a vampire-like manner, providing insights into the genes that are essential for such a predation strategy. Together, this work uncovers several pathogens that hitherto were unknown, and reveals a number of evolutionary innovations that allow these remarkable pathogens to exploit their hosts.
Chapter 1 was written by me.

Chapter 2 was conceptualized by Curtis A. Suttle and me and work was exclusively performed by me.

Chapter 3 was conceptualized by Curtis A. Suttle and me and all work was performed by me with the exception of library preparation for Illumina MiSeq sequencing that was assisted by and supervised by Cheryl-Emiliane T. Chow. An adapted version of chapter 3 has been published:


Chapter 4 was conceptualized by Curtis A. Suttle and me and all work was performed by me with the following exceptions: Matthias M. Zimmer performed particle purification and high molecular weight DNA extraction for PacBio sequencing under my supervision. Emma E. George and Filip Husnik provided valuable feedback on experimental design and interpretation.

Chapter 5 was conceptualized by Curtis A. Suttle and me and all work was performed by me with the following exceptions. Initial genome annotation was performed by Tan T. Le under my supervision. PCR analysis was performed by Tan T. Le and Matthias M. Zimmer under my supervision. Plaque assays were performed by Matthias M. Zimmer under my supervision.

Chapter 6 was written by me.
# TABLE OF CONTENTS

Abstract .............................................................................................................................. iii
Lay Summary .................................................................................................................... iv
Preface ............................................................................................................................... v
Table of Contents ............................................................................................................ vi
List of Tables .................................................................................................................... xii
List of Figures .................................................................................................................. xiii
List of Abbreviations ...................................................................................................... xvi
Acknowledgements ......................................................................................................... xxi
Dedication ........................................................................................................................ xxii

1 Introduction: Microbial parasitoids across the boundary of life ......................... 1

1.1 Trophic interactions in microbial ecosystems ....................................................... 1
    1.1.1 Microbial ecology ............................................................................................. 1
    1.1.2 Exploitative interactions in microbial communities: Predators, parasitoids and parasites ..................................................................................................................... 2
    1.1.3 Microbial predators .......................................................................................... 4
    1.1.4 Microbial parasitoids ....................................................................................... 4
    1.1.5 Microbial parasites .......................................................................................... 5

1.2 Implication of exploitative interactions in microbial ecosystems .................... 6
    1.2.1 Predators and parasitoids drive microbial community structures ................. 6
    1.2.2 Predators and parasitoids stimulate biogeochemical cycling ....................... 7

1.3 Parasitoids across the boundary of life ................................................................. 8
    1.3.1 The conundrum of complex viruses and simple bacteria ............................... 8
        1.3.1.1 Breaking the virus paradigm ................................................................. 8
        1.3.1.2 Giant viruses rival cellular life in many ways .................................... 9
        1.3.1.3 Highly host dependent bacteria show convergent traits to giant viruses .... 11
    1.3.2 The problems in defining the concept of “virus” ....................................... 12
    1.3.3 Viruses are of ancient origin and play a pivotal role in the evolution of cellular life ......................................................................................................................... 13
    1.3.4 The role of viruses in the definition of “life” .............................................. 13
1.4 Microbial predators, parasitoids, and parasites: The weird and the wonderful......14

2 Isolation and characterization of parasitoids of heterotrophic protists ...............16

2.1 Synopsis..............................................................................................................16

2.2 Introduction.......................................................................................................16

2.3 Materials and methods.......................................................................................17

2.3.1 Sampling and Host Isolation.........................................................................17

2.3.2 Virus concentration.......................................................................................21

2.3.3 Screen for lytic agents and replication kinetics ...........................................22

2.3.4 Electron microscopy.......................................................................................23

2.3.4.1 Negative staining transmission electron microscopy.................................23

2.3.4.2 Thin section microscopy.............................................................................23

2.4 Results ...............................................................................................................23

2.4.1 Virus concentrates .........................................................................................23

2.4.2 Isolates ...........................................................................................................25

2.4.2.1 Screen ........................................................................................................25

2.4.2.2 PspV, a putative viral lytic agent infecting *Procytophobia sp*.................26

2.4.2.3 PvV a putative non-icosahedral giant virus infecting *Paraphysomonas vestita* GflagA...............................................................26

2.4.2.4 HSV 2-3, a putative virus infecting the flagellated amoeba hs 2-3...........27

2.4.2.5 SspV, a viral lytic agent infecting *Spumella sp.*.......................................28

2.4.2.6 BsV NG1, a giant virus infecting *Bodo saltans* NG.................................29

2.4.2.7 Candidatus *Chromulinavorax destructans*, pathogenic bacterium infecting *Spumella elongata* CCAP 955/1...........................................................30

2.4.3 SspV, HsV and PspV lose infectivity during propagation...............................31

2.5 Discussion .........................................................................................................32

2.5.1 Pathogens of heterotrophic protists are readily isolated utilizing a screen employing ultrafiltration and flow cytometry....................................................32

2.5.2 Giant viruses might resist prolonged serial passage due to co-replicating virophages..................................................................................................................33

2.5.3 Isolation of parasitoids of heterotrophic protists – mining for microbial dark matter .........................................................................................................................34
3 The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea .................................................................35

3.1 Synopsis ........................................................................................................35
3.2 Introduction ..................................................................................................35
3.3 Materials and methods ...............................................................................37
  3.3.1 Sampling and Isolation: ..........................................................................37
  3.3.2 Transmission electron microscopy .........................................................38
    3.3.2.1 Negative staining ...............................................................................38
    3.3.2.2 High-pressure freezing and ultra-thin sectioning .............................38
  3.3.3 Virus concentration and sequencing ......................................................38
  3.3.4 Annotation ..............................................................................................40
  3.3.5 Phylogenetics .........................................................................................41
3.4 Results ........................................................................................................42
  3.4.1 Isolation and Infection Kinetics. ...............................................................42
  3.4.2 Virus morphology and Replication Kinetics. ..........................................42
  3.4.3 The Bodo saltans virus genome ...............................................................44
    3.4.3.1 Genome organization .......................................................................44
    3.4.3.2 Functional genome content ..............................................................46
    3.4.3.3 Repeat regions. ................................................................................50
    3.4.3.4 Genomic Mobilome ........................................................................50
  3.4.4 Phylogenetic placement and environmental representation of BsV. ..........52
3.5 Discussion .....................................................................................................56
  3.5.1 BsV represents the most abundant subfamily of the Mimiviridae in the TARA oceans data ........................................................................................................56
  3.5.2 BsV has acquired a host mechanism to facilitate membrane fusion, presumably employed during infection and virion morphogenesis ..........................56
  3.5.3 The BsV possesses an arsenal of mechanisms that could be involved in interference competition ..................................................................................57
  3.5.4 The contracted translation machinery in BsV suggest that this feature is a homoplastic trait of giant viruses ..............................................................58
An inflated genomic accordion due to evolutionary arms races is responsible for genome gigantism in BsV and sheds light on giant virus evolution

Summary

4 Chromulinavorax destructans, a pathogenic TM6 bacterium with an unusual replication strategy targeting protist mitochondrion

4.1 Synopsis

4.2 Introduction

4.3 Materials and methods

4.3.1 Sampling

4.3.2 Isolation

4.3.3 Transmission electron microscopy

4.3.3.1 Negative staining

4.3.3.2 Ultra-thin sectioning

4.3.4 Pathogen concentration and sequencing

4.3.5 Annotation

4.3.6 Phylogenetic analysis

4.4 Results

4.4.1 Chromulinavorax destructans is a lytic pathogen of Spumella elongata

4.4.2 Chromulinavorax destructans replicates in bodies nested in large mitochondrial invaginations

4.4.3 The Chromulinavorax destructans genome encodes a minimal set of core functions but an extensive assortment of genes involved in host modification

4.4.4 Phylogenetic position of Chromulinavorax destructans within the candidate phylum TM6/Dependentiae

4.5 Discussion

4.5.1 Chromulinavorax destructans is highly host-dependent

4.5.2 The extensive mitochondrial modifications caused by Chromulinavorax destructans might stem from limited host resources

4.5.3 Life history strategies and host-range might differ vastly amongst the TM6/Dependentiae and might promote diversity within the phylum

4.5.4 Summary
5 From the inside out: An epibiotic *Bdellovibrio* predator with an expanded genomic complement

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Synopsis</td>
<td>82</td>
</tr>
<tr>
<td>5.2</td>
<td>Introduction</td>
<td>82</td>
</tr>
<tr>
<td>5.3</td>
<td>Materials and methods</td>
<td>83</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Isolation and Culturing</td>
<td>83</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Pathogen sequencing</td>
<td>84</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Propagation and host range studies</td>
<td>84</td>
</tr>
<tr>
<td>5.3.4</td>
<td>Environmental sampling</td>
<td>85</td>
</tr>
<tr>
<td>5.3.5</td>
<td>Microscopy</td>
<td>85</td>
</tr>
<tr>
<td>5.3.5.1</td>
<td>Negative staining transmission electron microscopy</td>
<td>85</td>
</tr>
<tr>
<td>5.3.5.2</td>
<td>Ultra-thin sectioning transmission electron microscopy</td>
<td>86</td>
</tr>
<tr>
<td>5.3.6</td>
<td>Fluorescence In Situ Hybridization Epifluorescence Microscopy</td>
<td>86</td>
</tr>
<tr>
<td>5.3.7</td>
<td>Annotation</td>
<td>87</td>
</tr>
<tr>
<td>5.3.8</td>
<td>Phylogenetic analysis</td>
<td>87</td>
</tr>
<tr>
<td>5.4</td>
<td>Results</td>
<td>88</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Isolation, host range and distribution</td>
<td>88</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Morphology and replication cycle</td>
<td>90</td>
</tr>
<tr>
<td>5.4.3</td>
<td><em>Bdellovibrio salishius</em> has a complex genome for an epibiotic predator</td>
<td>91</td>
</tr>
<tr>
<td>5.4.3.1</td>
<td>Genome structure and content</td>
<td>91</td>
</tr>
<tr>
<td>5.4.3.2</td>
<td>Metabolism</td>
<td>92</td>
</tr>
<tr>
<td>5.4.3.3</td>
<td>Regulatory elements</td>
<td>93</td>
</tr>
<tr>
<td>5.4.3.4</td>
<td>Predatory arsenal</td>
<td>93</td>
</tr>
<tr>
<td>5.4.4</td>
<td><em>Bdellovibrio salishius</em> is a basal representative of its genus</td>
<td>94</td>
</tr>
<tr>
<td>5.5</td>
<td>Discussion</td>
<td>95</td>
</tr>
<tr>
<td>5.5.1</td>
<td><em>Bdellovibrio salishius</em> is an epibiotic predator of the beta-proteobacterium <em>Paraburkholderia fungorum</em></td>
<td>95</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Despite an epibiotic phenotype, <em>Bdellovibrio salishius</em> shares many genes with periplasmic predators</td>
<td>96</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Epibiotic predation is a common strategy of environmental <em>Bdellovibrio</em> species and could be the ancestral phenotype</td>
<td>97</td>
</tr>
</tbody>
</table>
5.5.4 Summary......................................................................................................................97

6 Concluding remarks ........................................................................................................98

6.1 Culturing still has a place in environmental microbiology in the age of metagenomics .........................................................................................................................98

6.2 Convergent evolution across the broad evolutionary distances.................................99

6.2.1 Evolutionary history is of limited relevance for the phenotype of parasitoids of heterotrophic protists ...........................................................................................................99

6.2.2 The bare necessities for bacterial pathogens and predators ....................................101

6.3 Limitations of the present study ..................................................................................102

6.4 Environmental microbiology in the 21st century .......................................................103

Bibliography ......................................................................................................................104

Appendices ........................................................................................................................116

A. Supplementary information to chapter 2 .......................................................................116

B. Supplementary information to chapter 3 .......................................................................117

C. Supplementary information to chapter 4 .......................................................................127

D. Supplementary information to chapter 5 .......................................................................128
LIST OF TABLES

Table 2-1 Sample locations ...........................................................................................................19
Table 2-2: Origin of cultures acquired by donation..................................................................21
Table 6-1: Convergent features of BsV and C. destructans ......................................................100

Supplementary Table 1 Comparison of translational machinery encoded by giant viruses based on phylogenetic analysis: .........................................................................................120
LIST OF FIGURES

Figure 1-1 : Exploitative interactions in microbial systems. .......................................................... 3
Figure 1-2 Impact of predators and parasitoids on bacterial size and motility. .......................... 7
Figure 1-3 Genome and particle size of microbial predators and parasitoids. .................................. 10
Figure 2-1. Sampled locations ........................................................................................................ 18
Figure 2-2: Concentration of pathogens of eukaryotes .................................................................... 22
Figure 2-3: Electron micrographs of large virus-like particles (VLPs) observed in the pathogen concentrates .................................................................................................................... 25
Figure 2-4: A lytic agent of Procrystobia sp. ..................................................................................... 26
Figure 2-5: A lytic agent of Paraphysomonas vestita GflagA ............................................................ 27
Figure 2-6: A lytic agent of the nanoflagellate isolate hs 2-3 ............................................................. 28
Figure 2-7: A lytic agent of Spumella sp. .......................................................................................... 29
Figure 2-8: A lytic agent of Bodo saltans NG ................................................................................... 30
Figure 2-9: A lytic agent of Spumella elongata CCAP 955/1 ............................................................ 31
Figure 2-10 Loss of infectivity with additional propagations of the putative viruses HSV 2-3, PspV and SspV ............................................................................................................................................. 31
Figure 2-11 Updated host-range of NCLDV and TM6/Dependentiae .................................................. 33
Figure 3-1: BsV induced lysis observed by flow cytometry. ............................................................... 42
Figure 3-2 Ultrastructure of BsV particles and replication. ............................................................... 43
Figure 3-3 Circularized genome plot of the linear BsV genome ....................................................... 46
Figure 3-4 BsV genome content ......................................................................................................... 47
Figure 3-5 Invasion of RNA polymerase genes by selfish genetic elements .................................... 52
Figure 3-6 BsV Phylogeny ................................................................................................................... 54
Figure 3-7 Intron/intein encoded endonuclease mediated interference completion between related viruses ............................................................................................................................................. 58
Figure 4-1: Micrographs of Chromulinavorax destructans ............................................................... 68
Figure 4-2: Effect of addition of Chromulinavorax destructans to a culture of Spumella elongata: ............................................................................................................................................. 68
Figure 4-3: Replication cycle of Chromulinavorax destructans .......................................................... 70
Figure 4-4: Replication cycle of Chromulinavorax destructans .......................................................... 71
Figure 4-5: Schematic representation of the inferred replication cycle of *Chromulinavorax destructans* .......................................................... 72

Figure 4-6 Genome content of *Chromulinavorax destructans* ........................................ 75

Figure 4-7: Phylogenetic placement of *Chromulinavorax destructans* .................................. 77

Figure 4-8: Schematic representation of inferred metabolic capabilities of *Chromulinavorax destructans* and its interactions with its host, *Spumella elongata* ........................................ 79

Figure 5-1 *Bdellovibrio salishius* host range and environmental distribution ......................... 89

Figure 5-2 Plaque assay ........................................................................................................... 90

Figure 5-3 *Bdellovibrio salishius* predation strategy and replication ........................................ 91

Figure 5-4: *Bdellovibrio salishius* genome ........................................................................... 92

Figure 5-5 *Bdellovibrio salishius* phylogenetic placement .................................................... 94

Figure 6-1: Hypothetical evolutionary scenario depicting the route of convergent evolutionary trajectories of BsV and *C. destructans* .......................................................... 101

Figure 6-2: Convergent complement of effector and transporter molecules in *B. salishius* and *C. destructans* ................................................................................................. 102

Supplementary Figure 1 Flow cytometry profiles of isolates .................................................. 116

Supplementary Figure 2 Ultrastructure of BsV particles and replication ............................ 117

Supplementary Figure 3: Ankyrin-repeat domain-containing proteins ............................... 118

Supplementary Figure 4: Evolutionary history of translational machinery found in giant viruses inferred by COUNT and abundance of ankyrin-repeat-domain genes ............................ 119

Supplementary Figure 5: Evolutionary history of gene family content found in giant viruses inferred by COUNT ............................................................................................................. 121

Supplementary Figure 6: Self-splicing group 1 introns found in BsV .................................... 122

Supplementary Figure 7: Phylogenetic analysis of five concatenated NCVOGs from selected NCLDVs ........................................................................................................... 124

Supplementary Figure 8: Phylogenetic analysis of D5-like helicase-primase (NCVOG0023;A-C) and DNA polymerase family B (NCVOG0038;D-F) from selected NCLDVs ......................................................................................................................... 125
Supplementary Figure 9: Phylogenetic analysis of DNA or RNA helicases of superfamily II (NCVOG0076; A-C), packaging ATPase (NCVOG0249; D-F), Poxvirus Late Transcription Factor VLTF3-like (NCVOG0262; G-I) from selected NCLDVs .............................................................. 126

Supplementary Figure 10: Pathway analysis of C. destructans according to pathway tools. 127

Supplementary Figure 11: Bdellovibrio salishius propagation in E. coli and continued presence in Nitobe Gardens pond. ........................................................................................................ 128

Supplementary Figure 12: Bdellovibrio salishius plaque assay. .................................................. 129
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>Amino acid</td>
</tr>
<tr>
<td>aaRS</td>
<td>Aminoacyl-tRNA synthetase</td>
</tr>
<tr>
<td>AaV</td>
<td>Aureococcus anophagefferens virus</td>
</tr>
<tr>
<td>AcV</td>
<td>Anomala cumrea entomopoxvirus</td>
</tr>
<tr>
<td>AMG</td>
<td>Auxiliary metabolic genes</td>
</tr>
<tr>
<td>AMaV</td>
<td>Acanthamoeba castellanii mamoavirus</td>
</tr>
<tr>
<td>AMgV</td>
<td>Mounoumavirus golette</td>
</tr>
<tr>
<td>AMoV</td>
<td>Acanthamoeba polyphaga mounoumavirus</td>
</tr>
<tr>
<td>ANK</td>
<td>Ankyrin repeat domain</td>
</tr>
<tr>
<td>ApMV</td>
<td>Acanthamoeba polyphaga mimivirus</td>
</tr>
<tr>
<td>ASFV</td>
<td>African swine fever virus BA71V</td>
</tr>
<tr>
<td>AtcV</td>
<td>Acanthocystis turfacea chlorella virus 1</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine-5'-triphosphate</td>
</tr>
<tr>
<td>BALO</td>
<td><em>Bdellovibrio</em> and like organisms</td>
</tr>
<tr>
<td>BER</td>
<td>Base excision repair</td>
</tr>
<tr>
<td>BI</td>
<td>Bayesian inference</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic local alignment search tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BpV</td>
<td>Bathycoccus sp. RCC1105 Virus</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>BsV</td>
<td>Bodo saltans virus NG1</td>
</tr>
<tr>
<td>CatV</td>
<td>Catovirus</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CDS</td>
<td>CoDing Sequence</td>
</tr>
<tr>
<td>CDD</td>
<td>Conserved domain database</td>
</tr>
<tr>
<td>CeV</td>
<td>Chrysochromulina ericina virus 1B</td>
</tr>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>COG</td>
<td>Clusters of orthologous groups of proteins</td>
</tr>
<tr>
<td>CroV</td>
<td>Cafeteria roenbergensis virus BV-PW1</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
</tbody>
</table>
dNTP ................................................................. Deoxyribonucleotide triphosphate
dsDNA ............................................................. Double-stranded deoxyribonucleic acid
eIF ................................................................. Eukaryotic translation initiation factor
EDTA ............................................................... Ethylenediaminetetraacetic acid
EhV ................................................................. Emiliana huxleyi virus 86
E-value ............................................................ Expectation value
FauV ................................................................. Faustovirus E12
GLIMMER ......................................................... Gene Locator and Interpolated Markov ModelER
HaRNAV .......................................................... Heterosigma akashiwo RNA virus
HAT ................................................................. Histone acetyl transferase
HGT ................................................................. Horizontal gene transfer
HokV ................................................................. Hokovirus
h.p.i. ................................................................. Hours post infection
HSV ................................................................. Howe Sound Virus 2-3
HvV ................................................................. Heliothis virescens ascovirus 3e
IiV ................................................................. Invertebrate iridescent virus
IndV ................................................................. Indivirus
ISKV ................................................................. Infectious spleen and kidney necrosis virus
KAAS .............................................................. KEGG Automatic Annotation Server
kb ........................................................................ Kilobase
KEGG .............................................................. Kyoto Encyclopedia of Genes and Genomes
KloV ................................................................. Klosneuvirus
LauV ................................................................. Lausannevirus
LUCA ................................................................. Last universal common ancestor
M ................................................................. Molar
MAG .............................................................. Metagenomically assembled genome
MarV ................................................................. Marseillevirus T19
Mb ....................................................................... megabase
MEME .......................................................... Multiple expectation maximation algorithm for motif elicitation
MCP ................................................................. Major capsid protein
ML ................................................................. Maximum likelihood
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOI</td>
<td>Multiplicity of infection</td>
</tr>
<tr>
<td>MoVs</td>
<td>Mollivirus sibericum</td>
</tr>
<tr>
<td>MpV</td>
<td>Micromonas pusillae Virus SP-1</td>
</tr>
<tr>
<td>MsV</td>
<td>Melanoplus sanguinipess entomopoxvirus</td>
</tr>
<tr>
<td>MUSCLE</td>
<td>Multiple sequence comparison by log-expectation</td>
</tr>
<tr>
<td>MVc</td>
<td>Megavirus chilensis</td>
</tr>
<tr>
<td>MW</td>
<td>Molecular weight</td>
</tr>
<tr>
<td>MyxV</td>
<td>Myxoma virus</td>
</tr>
<tr>
<td>NAD</td>
<td>Nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center for Biotechnology Information</td>
</tr>
<tr>
<td>NCLDV</td>
<td>Nucleo-cytoplasmic large DNA viruses</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding RNA</td>
</tr>
<tr>
<td>NCVOGS</td>
<td>Nucleo-cytoplasmic virus orthologous genes</td>
</tr>
<tr>
<td>NSF</td>
<td>N-ethylmaleimide-sensitive factor</td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>OLV1</td>
<td>Organic Lake Phycodnavirus 1</td>
</tr>
<tr>
<td>OLV2</td>
<td>Organic Lake Phycodnavirus 2</td>
</tr>
<tr>
<td>OtV</td>
<td>Ostreococcus tauri virus 1</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PbCV</td>
<td>Paramecium bursaria chlorella virus 1</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PES</td>
<td>Polyethersulfone</td>
</tr>
<tr>
<td>Pfam</td>
<td>Protein families</td>
</tr>
<tr>
<td>PFGE</td>
<td>Pulsed-field gel electrophoresis</td>
</tr>
<tr>
<td>PGAAIP</td>
<td>Prokaryotic genome annotation pipeline</td>
</tr>
<tr>
<td>PgV</td>
<td>Phaeocystis globosa virus 16T</td>
</tr>
<tr>
<td>PiVs</td>
<td>Pithovirus sibericum P1084-T</td>
</tr>
<tr>
<td>PolB</td>
<td>DNA-dependent DNA polymerase B</td>
</tr>
<tr>
<td>PoV</td>
<td>Pyramimonas orientalis virus</td>
</tr>
</tbody>
</table>
Phaeocystis pouchetii virus (PpV)
Procryptobia sp. Virus (PspV)
Pandoravirus dulcis (PVd)
Polyvinylidene fluoride (PVDF)
Pandoravirus salinus (PVs)
Paraphysomonas vestita virus (PvV)
Ribosomal DNA (rDNA)
Ribonuclease (RNase)
Ribonucleic acid (RNA)
Ribonucleotide reductase (RNR)
Reverse Position-Specific BLAST (RPS-BLAST)
Sodium dodecyl sulphate (SDS)
Spodoptera frugiperda ascovirus 1a (SfV)
Singapore grouper iridovirus (SGV)
N-ethylmaleimide-sensitive factor (NSF) attachment proteins (SNAP)
SNAP receptor (SNARE)
Single-stranded deoxyribonucleic acid (ssDNA)
Single-stranded ribonucleic acid (ssRNA)
Spumella sp. Virus (SspV)
Tris-borate-EDTA (TBE)
Tris-EDTA (TE)
Transmission electron microscopy (TEM)
Transcription factor (TF)
Tangential flow filtration (TFF)
Torf Mittelschich 6 (TM6)
Transfer-messenger RNA (tmRNA)
Trichoplusia ni ascovirus 2c (TnV)
Transfer RNA (tRNA)
Untranslated region (UTR)
Transfer ribonucleic acid (tRNA)
Vaccinia virus (VacV)
VF ................................................................. Virion factory
VLP ................................................................. Virus-like particles
WiV ................................................................. Wiseana iridescent virus
YLV1 ................................................................. Yellowstone lake phycodnavirus 1
ACKNOWLEDGEMENTS

I would thank members of the Suttle Lab, past and present, and especially Andrew Lang, Matthias Fischer, Cheryl Chow, Jan Finke, and Marli Vlok for helpful comments and guidance. As well, the assistance of Matthias Zimmer, Marie-Claire Veilleux-Foppiano, and Tan Le in the field and in the laboratory is greatly appreciated. Further, I would like to thank the following individuals for their contributions: Grieg Steward for insights into sample preparation and concentration. Denis Tikhonenkov for providing Bodo saltans HFCC12. John Archibald and Julius Lukeš for valuable feedback on bodonid biology. Thor Veen for assistance with R script writing. Patrick Keeling, Emma George, and Filip Husnik for their feedback and assistance on C. destructans biology. The UBC Bioimaging Facility for feedback and assistance with electron microscopy.

The work was supported by grants to Curtis A. Suttle from the Natural Sciences and Engineering Research Council of Canada (NSERC; 05896), Canada Foundation for Innovation (25412), British Columbia Knowledge Development Fund, and the Canadian Institute for Advanced Research (IMB). Christoph Deeg was supported in part by a fellowship from the German Academic Exchange Service (DAAD).
DEDICATION

To Gaia, for she provides everything. This work is a celebration of her wonders.

To Cortney and Bonsai for enduring me through these years.
Chapter 1

1 Introduction: Microbial parasitoids across the boundary of life

1.1 Trophic interactions in microbial ecosystems

1.1.1 Microbial ecology

Microbiology is the study of a broad range of organisms that elude observation by the naked eye and can only be observed indirectly. Despite the monikers “microorganism” or “microbe”, they encompass the majority of phylogenetic diversity on earth. In fact, the term microbe is used to describe representatives of all three domains of cellular life, bacteria, archaea, and single-celled eukaryotes or protists. Further complicating the matter, viruses are also sometimes considered microbes and studied by microbiologists, despite their problematic placement on the “tree of life”. In addition to representing the majority of phylogenetic diversity, microbes also numerically dominate globally with approximately $10^{30}$ estimated prokaryotic cells [1]. This incredible number, larger than the number of stars in the known universe, translates to a global microbial biomass of more than 80 Gt of carbon, surpassed only by plants [2]. However, in contrast to plants, many of which have the majority of carbon deposited in biologically inert structural support features such as wood, most of the microbial carbon resides in physiologically active microorganisms [3]. Additionally, microorganisms, especially bacteria, are capable of a multitude of metabolic processes, that due to their immense scale, are essential for elemental cycling through ecosystems on a global scale, preventing the earth from reaching a thermodynamically inert equilibrium [4, 5]. The six major elemental constituents of living cells, H, C, N, O, S, and P, all depend on microbial activity to maintain global biogeochemical cycling [4]. This is especially relevant when considering nitrogen fixation, which before the invention of the Haber-Bosch process, supplied almost all bioavailable nitrogen on this planet [6]. How transformative microbially driven cycles can be on a global scale becomes apparent when considering the oxygenation of the atmosphere that culminated in the great oxidation event driven solely by the advent of microbial oxygenic-photosynthesis and changed the history of life on earth forever by altering its redox state [7]. Additionally, microbes play a crucial role in global biogeochemical cycles of trace elements such as iron [8]. Despite the fact that microbial activity is the single most important driver promoting global ecosystems, we know little about the intricate interactions that happen on community and individual levels within microbial populations that could influence the function of these ecosystem services that they provide.
1.1.2 Exploitative interactions in microbial communities: Predators, parasitoids and parasites

Due to their relevance in biogeochemistry, understanding the factors that shape microbial populations is paramount in tracing microbial community composition and function. While abiotic factors, such as salinity, influence microbial communities, biotic interactions, are arguably as important [9]. Similar to macrobial ecosystems, members of microbial ecosystems exhibit facilitative, neutral, amensal, and exploitative interactions with each other. Facilitative interactions are prevalent amongst microbes, as many metabolic pathways are split into functional steps that are performed by unrelated microbes enabling ecosystem functions that would be unattainable by a single cell [10]. On the other hand, successful systems attract parasites that gain from exploiting such productive systems. Accordingly, microbial systems are abound with parasitic, parasitoid, and predatory interactions, spanning across all domains of life from genetic elements to the organismal level (Figure 1-1).
Figure 1-1: Exploitative interactions in microbial systems. Blue arrows depict parasitic, yellow parasitoid and red arrows predatory interactions.
1.1.3 Microbial predators

Predators cause mortality and often consume their prey completely and take multiple prey individuals during their lifetime. The smallest microbial predators are social predatory bacteria such as swarming myxobacteria that hunt in “wolf packs”; they surround their prey cells, usually other bacteria, and cause lysis of the prey cell by secreting lytic enzymes [11]. Prey cell contents are consecutively consumed by members of the collaborating swarm [11]. Coordinating such complex behaviours necessitate some of the largest bacterial genomes [11]. Similar predatory strategies to myxobacteria are also deployed by members of the genera Lysobacter and Herpetosiphon. Heterotrophic nanoflagellates represent the next larger functional class of predators in aquatic microbial ecosystems. These usually rapidly swimming predators show low prey-specificity, with prey size being a major predator-specific determinant of predation rate, as well as prey behaviour and surface characteristics [12, 13]. Due to their high activity and abundance of up to $10^4$ cells ml$^{-1}$, predation by heterotrophic nanoflagellates is thought to be the mayor biotic factor regulating bacterial populations in aquatic systems, but some nanoflagellates also graze on smaller particles like viruses [14-17]. Larger protists, such as ciliates and amoeba show fewer restrictions in prey size range and therefore are even less discriminate in their predation on prokaryotes and smaller eukaryotes [18, 19]. Bridging the microbial and macrobial systems are the smallest metazoan predators, zooplankton such as copepods, rotifers or daphnia that prey on large prokaryotes and protists [20, 21].

1.1.4 Microbial parasitoids

In contrast to predators, parasitoids only require one prey or host individual to complete their lifecycle but nevertheless cause mortality in their host or prey in doing so. Viruses of single-celled organisms usually show parasitoid behavior as productive infection results in lysis of the single cell and therefore death. Bacteriophages, viruses infecting bacteria, are the most abundant viruses on earth, due to their host’s high abundance and productivity [22, 23]. Viruses are generally about an order of magnitude more abundant than their hosts, resulting in approximately $10^{31}$ virus particles in the ocean, where they play critical roles influencing biogeochemical cycles [24]. Bacteriophages are dominated by small to medium size double-strand DNA (dsDNA) viruses with other genetic systems (ssDNA or RNA) being comparatively rare [25]. Viruses of single-celled eukaryotes are also abundant, but are dominated by small picorna-like positive-sense single-strand RNA (+ssRNA) viruses, as well as very large viruses belonging to the nucleo-cytoplasmic large
DNA viruses (NCLDV) [26, 27]. The latter are remarkable for their genome and particle sizes, which overlap with those of cellular organisms, and curiously, some giant viruses are parasitized by smaller viruses, termed virophages [28, 29]. Prokaryotic parasitoids use both, other prokaryotes and eukaryotes as hosts. A diverse paraphyletic group termed *Bdellovibrio* and like organisms (BALOs), are parasitoids exploiting other prokaryotes [30]. Many of these BALOs, like the eponymous *Bdellovibrio bacteriovorus*, deploy free-living attack phase cells that attach to prey cells and enter their periplasm where the growth phase cells consume the prey cell and ultimately release several attack-phase progenies completing a parasitoid cycle that is similar to that of a virus [31]. Like bacteriophages, BALOs are thought to be broadly distributed and abundant in the environment [32]. Parasitoid prokaryotes of protists are rare, but examples exist, such as *Babela massiliensis* that infects and lyases amoeba; and metagenomic data suggest that relatives of such parasitoids are also widespread [33, 34]. In rare cases, protists can be parasitoids of metazoans as well [35].

### 1.1.5 Microbial parasites

In contrast to parasitoids, true parasites exploit but do not necessarily kill their host over several parasite generations. Parasitism is a widespread replication strategy in microbial systems that is deployed by a diversity of species. The definition of a parasitic lifestyle can also include parasitic genetic elements, such as self-splicing group I and II introns that are found in all life forms as well as viruses, but seem to be especially abundant in prokaryotes [36, 37]. Other widespread selfish genetic elements are transposons, which do not provide any benefit to their host while utilizing its resources [38]. While most known viruses of single-celled organisms are parasitoid, as they usually kill their hosts after successful replication, viruses infecting metazoa often show a more parasitic replication strategy, where death of the host is not a necessary outcome of successful replication. Accordingly, a substantial part of viral communities in the environment are metazoan infecting viruses [39]. Parasitic prokaryotes are also abundant in nature and infect protists and metazoan alike, while protists can be parasites of other protists and metazoan species [40].
1.2 Implication of exploitative interactions in microbial ecosystems

1.2.1 Predators and parasitoids drive microbial community structures

Exploitative interactions in microbial systems have a multitude of downstream effects on ecosystem structure and function. Primarily, predators and parasitoids force evolutionary arms races on their prey or hosts, altering system productivity by imposing costly defense mechanisms [41]. By preferentially infecting abundant and active host phenotypes, a phenomenon known as “killing the winner”, predators and parasitoids can induce diversification within populations and change community composition [41]. Since infection-resistant phenotypes often show lower fitness, a trade-off to acquire resistance, predators and parasitoids facilitate the coexistence of strains of different phenotypes by maintaining a level playing field among strains [41, 42]. Additionally, predators induce diversification as exemplified by the arms race between phage and their bacterial hosts. [43]. Therefore, viruses increase diversity at strain, species and community level in all ecosystems by imposing a selective pressure on their hosts.

Interestingly, different types of predators and parasitoids induce different selective pressures (Figure 1-1). When considering bacterial size and motility, viral parasitoids select for small immobile cells to reduce encounter rates [44]. Small cells would also be beneficial against BALOs because of reduced surface area for attachment, but the picture is different when considering grazing pressure by eukaryotic predators. Here, either very large or very small cell size, and mobility is advantageous since large cells can’t easily be phagocytized, small cells are “overlooked” and motile cells escape predation [12, 15, 45]. This suggests that different predators and parasitoids force bacterial cell size into a goldilocks zone, flanked on both sides by disadvantageous features exposing them to predators and parasitoids. This perfect size for predator and parasitoid avoidance, however, might not correlate with the optimal size for physiological processes, again proving a costly resistance mechanism.
1.2.2 Predators and parasitoids stimulate biogeochemical cycling.

Considerable evidence has demonstrated that in surface ocean waters viruses lyse about 20-40% of marine bacterial cells daily; thereby, promoting recycling of organic carbon, nutrients, and trace elements, and increasing overall productivity [24, 46]. Because of their abundance, most infection and lysis is of heterotrophic bacteria, which also benefit from the release of particulate and dissolved organic matter, as well as micronutrients, a phenomenon known as the viral shunt, promoting overall productivity [3, 5, 24, 46-48]. The viral shunt also promotes the retention of non-labile dissolved matter in the photic zone through the “Shunt and Pump” which fuels the biological carbon pump, and the regeneration of nutrients that limit primary productivity, such as N and P [24, 49-54]. These effects of viral lysis are especially prominent in systems with low primary production, such as deep sea sediments, where microbial populations are dominated by heterotrophic bacteria at the base of the food web [55]. The influences on the cycling of
micronutrients are assumed to be similarly beneficial, but data is limited beyond iron, which is a major limiting factor in the open ocean. Iron recycling through viral lysis is responsible for up to 90% of the requirement of primary production in high-nutrient, low-chlorophyll environments, due to the high bioavailability of siderophore-bound iron in viral lysates [47, 51, 56].

While BALOs are widely distributed in aquatic environments, there is little data on their influence on biogeochemistry [57-59]. Given that their abundance is presumed to be relatively low, their influence on nutrient cycling is minor. Predation on bacteria by heterotrophic flagellates, is similar to the mortality imposed by viruses, and therefore has significant consequences [60]. Under high C:N situations, protist grazing on bacteria can lead to nitrogen remineralization increasing primary production in the system [61, 62]. Similarly, “sloppy feeding” by metazoan predators like copepods promotes nutrient recycling and promotes the microbial loop [63, 64]. Collectively, it is apparent that exploitative interactions in microbial ecosystems increase turnover and nutrient recycling, therefore counterintuitively increasing ecosystem productivity.

1.3 Parasitoids across the boundary of life

1.3.1 The conundrum of complex viruses and simple bacteria

1.3.1.1 Breaking the virus paradigm

Viruses are typically defined as “… small infectious agents that replicate only inside the living cells of other organisms.” [https://en.wikipedia.org/wiki/Virus, retrieved: 24/06/18]. Historically, most studies on viruses focused on those affecting the health of humans, or economically important animals, plants and bacteria, creating a narrow view of viral diversity and a paradigm that viruses are small in size with relatively simple genomes. Forays into environmental virology, serendipitous discoveries of giant viruses, and the development and widespread adoption of culture-independent metagenomic sequencing has revealed the breadth of viral diversity, and that giant viruses with sizable and complex genomes are widespread and abundant in nature. There are many examples of transformative discoveries in virology, such as virus encoded auxiliary metabolic genes (AMGs), or viruses crossing the boundary between RNA and DNA, but it was the discovery of the giant Acanthamoeba-infecting Mimivirus that arguably had greatest impact on classical paradigms in virology [65-67]. This discovery suggested that very large viruses might also be in the environment, but had been missed in previous metagenomic surveys, which typically used 0.2-μm filtration to exclude cellular organisms before analyzing viral communities [26, 67].
After the discovery of Mimivirus, the door was opened to investigations of giant viruses and today, there is a plethora of diverse giant viruses known, both from culture based and metagenomic studies.

1.3.1.2 Giant viruses rival cellular life in many ways

The detailed study of giant viruses has transformed virology. The most obvious physical characteristics of giant viruses are their particle size that is well within the range of cellular organisms (Figure 1-3). The prototypical Mimivirus is approximately 750nm in diameter, including the fibers decorating its capsid, surpassing the size of many bacteria in oligotrophic environments, such as the cyanobacteria in the genus *Prochlorococcus* [67]. The non-icosahedral virus Pithovirus sibiricum is even larger at 1.6 μm, and surpasses the diameter of most aquatic prokaryotes [68]. The surprising finding that relatives of Mimivirus sport tail-like structures resulting in a length of 2.3μm emphasizes that selection can favour large particle size in viruses infecting *Acanthamoeba* [69]. Together, giant viruses demonstrate that particle size is not a good criterion to separate viruses from cells.

Genome size in giant viruses is equally hefty (Figure 1-3). The sequencing of Mimivirus revealed a 1.2 Mb genome with roughly a thousand protein-coding genes [28]. While this is impressive and similar in size to small bacteria, it is dwarfed by the largest known viral genome belonging to Pandoravirus dulcis, which at 2.5 Mb rivals that of some small eukaryotes [70]. Together, the large genomes and particle sizes demonstrate some of the difficulties in distinguishing viruses from cellular life.
Figure 1-3 Genome and particle size of microbial predators and parasitoids. Genetic elements: Self-splicing group I Intron. Viruses (green): Heterosigma akashiwo RNA virus (HaRNAV; Picronaviridae), T4 (Myoviridae), Paramecium bursaria Chlorella virus 1 (PBCV; Phycodnaviridae), Acanthamoeba polyphaga Mimivirus (Mimiviridae), Pandoravirus dulcis (Pandoraviridae), Pithovirus sibiricus (unassigned NCLDV). Bacteria (blue): Babela massiliensis (TM6/Dependentiae), Bdellovibrio bacteriovorus (delta-Proteobacteria), Myxococcus xanthus (delta-Proteobacteria). Eukarya (red): Bodo saltans (Euglenozoa), Tetrahymena thermophila (Alveolata), Tigriopus kingsejongensis (Opistokonta).
In addition to their large genome and particle sizes, giant viruses encode functions that were thought to be reserved for cellular life. The most surprising functions encoded by giant viruses are involved in translation. The first giant virus genome analyzed showed several genes that were thought to modify the host’s translation machinery; aminoacyl-tRNA synthetases (aaRS) were especially noteworthy, as they had never been seen encoded by viruses [28]. Consecutively, several members of the family Mimiviridae were found to encode a plethora of aaRS genes, some of which encompassed a complete set of aaRS for all amino acids [69, 71]. Virtually any compartment of translation, with the exception of ribosomes, seems to be encoded by some giant viruses. Nevertheless, the heterogeneity of the translational genome content suggests that it is not derived from an ancient shared origin, but was acquired more recently in response to selective pressure from the host. Other unexpected genes found in giant viruses are involved in sugar metabolism and fermentation, emphasizing that AMGs occur in giant viruses [72, 73].

Potentially the most surprising feature of giant viruses is that they can be “infected” by other viruses, termed virophages, in analogy to bacteriophages [29, 74, 75]. While different virophages deploy different strategies of entry, they depend on cytoplasmic viral factory that is established by their host giant virus during infection of the protist host. To this end, virophages either enter the protist host during infection by a giant virus, or enter independently and integrate into the host’s nuclear genome in anticipation of a future giant virus infection [29, 76]. In either case, virophage replication depends on the transcriptional machinery of its giant virus host and effectively is a hyperparasite.

1.3.1.3 Highly host dependent bacteria show convergent traits to giant viruses

While giant viruses exhibit many features of cellular organisms, there are many lineages of parasitic and symbiotic cellular life that have lost canonical features of free-living cells. For instance, extreme genomic reductions in symbionts of insects have resulted in bacterial genomes of only 160 kb, with a genomic content limited to DNA replication, transcription and translation [77, 78]. Such bacteria effectively have very limited or no metabolic capacity. Similarly, prokaryotic and eukaryotic pathogens show trends of metabolic reduction, and rely on their hosts to provide metabolites and energy, much like viral parasitoids [33, 79]. Another canonical feature often associated with cellular life is binary fission; however, many bacterial parasitoids do not replicate in this fashion and utilize “viral-like” replication strategies. Examples include the bdelloplast, where BALOs grow into a single or several large “cells” in the periplasm of the prey.
until the prey’s resources are exhausted, which induces septation into attack-phase cells [31]. Similarly, amoebal parasitoids like *Babela massiliensis* grow in disorganized large bodies before septating into daughter cells and lysing the host [34]. Finally, even eukaryotic parasites such as microsporidians forego binary fission in favor of rapid replication [80].

### 1.3.2 The problems in defining the concept of “virus”

As mentioned above, Wikipedia defines a virus as a “… small infectious agent that replicates only inside the living cells of other organisms.” [https://en.wikipedia.org/wiki/Virus, retrieved: 24/06/18]. Given that this definition includes small cellular pathogens and symbionts, as well as selfish genetic elements, it is apparent that a more comprehensive definition of a virus is needed.

Dimitry Ivanovski’s initial 1892 description of the causative agent of tobacco mosaic disease as an ultra-filterable, uncultivable and not visible agent started a string of negative definitions that defined a virus by characteristics it lacked rather than by some it possessed [81]. The term “virus” was first used in 1898 by Martinus Beijerinck, but it was not until viral particles were depicted in the first electron microscopes in the late 1930s that the scientific community became more attentive to the need to classify such infectious agents [82, 83]. The most consistently used definition of a virus arrived in the 1950ies with Andre Lwoff seminal papers defining viruses by, nevertheless, negative criteria to distinguish them from cellular microorganisms [84]: Viruses were to only contain one type of genetic material, should be able to reproduce from their nucleic acid only, never replicate by binary fission and never possess a Lipmann system. It is noteworthy that Lwoff did not include particle size in his discussion. The Lwoff-definition remained widely used in virology for the remainder of the 20th century until the discovery of giant viruses. In an extensive review and comment piece, Claverie and Abergel (2012) point out, as has been described above, that giant viruses violate most of Lwoff’s discrimination criteria with the exception of translation and binary fission [85]. However, several cellular parasites also forego binary fission, as has been discussed above, hence, blurring the line once more [31, 34]. At this point, one last divide between the viral domain and cellular life seemed to be the absence of a functional ribosome. Indeed, attempts have been made to unify “capsid” and “ribosome encoding lifeforms” in a single evolutionary theory, with one theory giving giant viruses the rank of a fourth domain of life [86, 87]. Given recent reports of ribosomal proteins found in several unrelated viral genomes, even the presence of a ribosome as a defining feature of cellular life could be threatened.
Together, these observations suggest that in the 21st century, well into the second century of virology, a functional working definition of a virus remains elusive.

### 1.3.3 Viruses are of ancient origin and play a pivotal role in the evolution of cellular life

Despite the absence of a clear definition of viruses, three classical theories have been postulated on their evolutionary origin [89]. According to the “virus first” hypothesis, virus-like entities originated in the primordial RNA world before the emergence of cellular life and established themselves as cellular parasites [90]. The “escape” hypothesis describes a scenario in which genomic fragments of cells became infectious and acquired capsids as a means of efficient cell-to-cell propagation [91]. Finally, the “regression” hypothesis proposes that viruses arose from cell-like precursors and were reduced to their present-day appearance due to evolutionary processes typical for intracellular parasites [92, 93]. There are strong supporting arguments and serious shortcomings when applying the three hypotheses to specific viral groups, but it is becoming increasingly clear that different viruses have different evolutionary histories, making it likely that all three proposed mechanisms have contributed to today’s diverse virophere [25]. Regardless, most authors agree that viruses are ancient, and predate the last universal cellular ancestor (LUCA) and have been active for at least 3.5 billion years [25, 89]. Since viruses deploy every conceivable genetic replication-expression strategy and have been accompanying cellular life throughout all its history, it follows that viruses were involved in key transitional events in the evolution of cellular life [25]. Their involvement in the transition from the RNA world to the DNA world is plausible and there are many widely accepted cases of the recruitment of viral innovation for cellular functions [89]. In light of such events, relaxed evolutionary constraints, shorter generation times, larger population sizes, and in many cases error-prone replication of their genetic material, makes it likely that many, if not most, genetic innovations originated inside the viral domain [25, 94]. Additionally, the classical arms race between viruses and hosts has long been appreciated as a key accelerator of cellular evolution and diversification [87].

### 1.3.4 The role of viruses in the definition of “life”

Since many giant viruses possess features associated with cellular life and many parasites of cellular origin lack features, the question arises whether some viruses should be considered “alive” [95]. Astonishingly, like the problematic definition of virus, the nature of the central theme in Biology, “life”, is poorly defined despite biology being literally the “study of life”.

[88]
Simple informational definitions require a replicator-like element able to reproduce itself. A popular definition, for instance, adapted by NASA, lists life as “… a self-sustained chemical system capable of undergoing Darwinian evolution” [96]. Most “biological” definitions of life include a basic checklist of six or seven features that must be satisfied to consider an entity “alive”. For instance, Koshland uses program, improvisation, compartmentalization, energy, regeneration, adaptability, and seclusion as his criteria [97]. Several other attempts use criteria that are surprisingly similar [98].

Irrespective of the definition used, giant viruses and small, highly host-dependent cellular parasites or symbionts are problematic, since they satisfy many of the criteria for life to the same degree, miss others, and could both be thought of as almost alive. Qualitatively, there are no hard delimitations separating complex giant viruses from simple cellular parasites, despite our innate desire to label cellular organisms as alive and viruses as not living.

Complicating these definitions is the idea that the evolving unit of a virus is the infected cell, or virocell and not the free virus particle. The virus particle can be thought of as the information package, or seed, that uses a suitable environment (the host cell) to propagate. This virocell is qualitatively different from the uninfected host since it possesses an array of novel genes and cellular processes are utilized to producing viral particles. This virus-virocell division recapitulates the scenario of cellular parasites forming inert spores in between infected hosts [99]. Such spores are functionally almost indistinguishable from virus particles until they infect their host forming what might be called a “sporocell”, again blurring the distinction between the two.

To summarize, no matter what definition for life is used, the result would mean that some entities termed viruses need to be considered alive, and that some cellular entities have to be considered non-living. Both of these eventualities leave us with an unsatisfying, counterintuitive outcome. Given the challenges of defining both viruses and life, the use of such definitions itself becomes questionable [100]. Ultimately however, not having clear definitions of viruses or life might not impede the advancement of science [101].

1.4 Microbial predators, parasitoids, and parasites: The weird and the wonderful

The scientific community over the last several decades has made major strides in understanding the key microbial players, and has laid the foundation for understanding global biodiversity and biogeochemical cycles. It is the unusual predators, parasitoids, and parasites that,
with a few notable exceptions, have eluded closer investigation, despite their importance to ecosystem processes and the evolution of life.

This thesis focuses on the isolation and characterization of novel viral and bacterial parasitoids recovered from local aquatic ecosystems. The insights from this study help inform our understanding of microbial biodiversity, and highlights overlooked and unusual players that influence microbial communities.
Chapter 2

2 Isolation and characterization of parasitoids of heterotrophic protists

2.1 Synopsis

Microbial dark matter, microorganisms known from metagenomic sequencing but evading culturing remain a poorly explored frontier in microbiology. Much of this microbial dark matter could be represented by the unusual niche of parasitoids of heterotrophic protists. Some occupants of this niche are bacterial pathogens, living an obligate intracellular replication cycle, and giant viruses of similar genomic complexity. Since very few of these representatives of microbial dark matter are available in culture, a high throughput screen was developed to isolate them from the environment using ultrafiltration and flow cytometry. This screen was used to establish five novel giant-virus systems in culture, as well as one novel bacterial pathogen of the TM6 candidate phylum. These isolates greatly expand the diversity of known giant viruses as well as the candidate phylum TM6.

2.2 Introduction

Metagenomic surveys regularly reveal sequences without recognizable similarity to known proteins or non-coding RNAs, suggesting that there is still unexplored sequence space encoded by uncultured microbes, a phenomenon often referred to as microbial dark matter. A substantial fraction of this dark matter is encoded by viruses and phage, potentially encoding most proteins on earth [102]. The discovery of giant viruses, members of a diverse class termed “nucleocytoplasmic large DNA viruses” (NCLDV) that infect heterotrophic protists provided an unexpected window into this unexplored genomic diversity [103]. Their genomes routinely exceed 1000 protein coding genes and rival the complexity of cellular life. Moreover, these genes are especially interesting given that at the time of their discovery merely 10% showed similarity to previously known protein functions [28]. After the first giant viruses were isolated, it became clear that sequences belonging to their relatives are well represented as parts of the microbial dark matter in bacterial metagenomes [104]. Indeed, analysis of metagenomic data from the TARA oceans global microbiome sampling suggested that relatives of giant viruses are present in the world’s oceans at levels similar to archaea [26]. Surprisingly, next to more moderately sized viruses in the
family *Phycodnaviridae*, which infect phytoplankton, members of the *Mimiviridae* comprise more than a third of all the sequences attributed to the NCLDV in these metagenomes. Although members of the genus *Acanthamoeba* serve as hosts to many mimivirus isolates, *Acanthamoeba* are not thought to be abundant in the open ocean, suggesting other hosts to these viruses exist. As giant viruses also infect heterotrophic and photosynthetic flagellates, which are abundant in the ocean, it suggests that these serve as hosts for most of the giant-virus diversity [105-107]. However, for most potential hosts the absence of a high-throughput screening method such as plaque assays makes the isolation of new giant viruses a tedious undertaking.

Like giant viruses, obligate intracellular bacterial pathogens and symbionts often encode unusual genes and, due to their strict host-dependence, resist culturing. Accordingly, these organisms also contribute to microbial dark matter. One phylum representing such organisms is the candidate phylum TM6, also known as *Dependentiae* [33, 108]. These bacterial pathogens and symbionts infect amoeba, but metagenomic sequencing suggests that more diverged relatives exhibit a host range beyond amoeba [34, 109].

To investigate pathogens of heterotrophic eukaryotes, a high throughput screen based on sampling, ultrafiltration and flow cytometry was established to isolate host-pathogen pairs for investigation in culture. In this screen five novel giant viruses as well as one pathogenic TM6 bacterium were isolated, thus expanding the known host ranges of NCLDV and TM6 bacteria.

### 2.3 Materials and methods

#### 2.3.1 Sampling and Host Isolation

Twenty-litre water samples were collected from eleven fresh, four brackish and four salt water environments in southwestern British Columbia, Canada using either a bucket or Niskin bottle (Table 1). Samples were stored at 4°C in the dark until virus concentration within 3 to 25 days.
Figure 2-1. Sampled locations: Locations sampled in 2014 for virus concentration and protist isolation
Table 2-1 Sample locations: Samples were pooled according to salinity in saltwater (dark blue), brackish water (light blue) and freshwater (light green)

<table>
<thead>
<tr>
<th>Location, #</th>
<th>Date</th>
<th>Type</th>
<th>Sampling</th>
<th>Depth</th>
<th>Salinity</th>
<th>Coordinates (N,W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saanich Inlet, 1</td>
<td>14.05.2014</td>
<td>Open Water</td>
<td>Niskin</td>
<td>6m</td>
<td>28ppt</td>
<td>48°35'18&quot;, 123°30'13&quot;</td>
</tr>
<tr>
<td>Saanich Inlet, 1</td>
<td>11.06.2014</td>
<td>Open Water</td>
<td>Niskin</td>
<td>6m</td>
<td>28ppt</td>
<td>48°35'18&quot;, 123°30'13&quot;</td>
</tr>
<tr>
<td>Jericho Pier, 2</td>
<td>16.07.2014</td>
<td>Near Shore</td>
<td>Bucket</td>
<td>Surface</td>
<td>13ppt</td>
<td>49°16'37&quot;, 123°12'5&quot;</td>
</tr>
<tr>
<td>Brohm Lake, 3</td>
<td>21.07.2014</td>
<td>Shallows</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°49'4&quot;, 123°7'46&quot;</td>
</tr>
<tr>
<td>Mamquam Blind Channel, 4</td>
<td>21.07.2014</td>
<td>Near Shore</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°42'5&quot;, 123°8'47&quot;</td>
</tr>
<tr>
<td>Britannia Beach, 5</td>
<td>21.07.2014</td>
<td>Shore</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°37'34&quot;, 123°12'27&quot;</td>
</tr>
<tr>
<td>Vedder River, 6</td>
<td>23.07.2014</td>
<td>River Bank</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°6'12&quot;, 122°4'38&quot;</td>
</tr>
<tr>
<td>Sumas River, 7</td>
<td>23.07.2014</td>
<td>River Bank</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°5'22&quot;, 122°7'1&quot;</td>
</tr>
<tr>
<td>Minnekhada Park, 8</td>
<td>23.07.2014</td>
<td>Swamp</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°18'10&quot;, 122°42'9&quot;</td>
</tr>
<tr>
<td>Howe Sound, 9</td>
<td>25.07.2014</td>
<td>Open Water</td>
<td>Niskin</td>
<td>10m</td>
<td>30ppt</td>
<td>49°27'17&quot;, 123°17'19&quot;</td>
</tr>
<tr>
<td>Howe Sound, 9</td>
<td>25.07.2014</td>
<td>Open Water</td>
<td>Bucket</td>
<td>Surface</td>
<td>13ppt</td>
<td>49°27'17&quot;, 123°17'19&quot;</td>
</tr>
<tr>
<td>Howe Sound, 10</td>
<td>25.07.2014</td>
<td>Near Shore</td>
<td>Bucket</td>
<td>Surface</td>
<td>8ppt</td>
<td>49°27'13&quot;, 123°19'36&quot;</td>
</tr>
<tr>
<td>Howe Sound, 11</td>
<td>25.07.2014</td>
<td>Open Water</td>
<td>Bucket</td>
<td>Surface</td>
<td>12ppt</td>
<td>49°18'38&quot;, 123°20'52&quot;</td>
</tr>
<tr>
<td>Howe Sound, 11</td>
<td>25.07.2014</td>
<td>Open Water</td>
<td>Niskin</td>
<td>10m</td>
<td>30ppt</td>
<td>49°18'38&quot;, 123°20'52&quot;</td>
</tr>
<tr>
<td>Fraser River Steveston, 12</td>
<td>30.07.2014</td>
<td>Swamp</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°8'27&quot;, 123°3'16&quot;</td>
</tr>
<tr>
<td>Fraser River Iona, 13</td>
<td>30.07.2014</td>
<td>Swamp</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°13'21&quot;, 123°12'43&quot;</td>
</tr>
<tr>
<td>Iona Lakes, 14</td>
<td>30.07.2014</td>
<td>Lakeshore</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°13'13&quot;, 123°12'41&quot;</td>
</tr>
<tr>
<td>Experimental Ponds UBC, 15</td>
<td>30.07.2014</td>
<td>Lakeshore</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°14'52&quot;, 123°13'59&quot;</td>
</tr>
<tr>
<td>Nitobe Memorial Garden, 16</td>
<td>30.07.2014</td>
<td>Lakeshore</td>
<td>Bucket</td>
<td>Surface</td>
<td>&lt;1ppt</td>
<td>49°15'58&quot;, 123°15'34&quot;</td>
</tr>
</tbody>
</table>
Protists for virus-isolation bioassays were either isolated from natural waters (Table 2-1) or obtained by donation form several researchers (Table 2-2). For the isolation of potential protist hosts of different size, 1 ml of water from each sampling location was passed through Durapore (Millipore) PVDF membranes with nominal pore sizes of 8-μm or 118 μm, respectively, or left unfiltered. The filtrate was added to 9 ml of appropriate liquid medium, as given below, supplemented with 0.01% yeast extract and 0.02% tryptone w/v and a wheat grain to enable the growth of heterotrophic bacteria as a food source for heterotrophic flagellates. Freshwater isolates were grown in modified DY-V artificial freshwater medium (2.03x10⁻⁷ M MgSO₄, 4.02x10⁻⁸ M KCl, 5.01x10⁻⁸ M NH₄Cl, 2.35x10⁻⁷ M NaNO₃, 1.00x10⁻⁸ M Na₂-ß-glycero-phosphate, 1.29x10⁻⁸ M H₃BO₃, 4.93x10⁻⁸ M Na₂SiO₃, 5.10x10⁻⁷ M CaCl₂, 3.70x10⁻⁹ M FeCl₃, 1.53x10⁻⁸ M Na₂EDTA, 3.84x10⁻¹⁰ M MnCl₂, 1.39x10⁻¹⁰ M ZnSO₄, 3.36x10⁻¹¹ M CoCl₂, 2.77x10⁻¹¹ M Na₂MoO₄, 9.2x10⁻¹² M Na₃VO₄, 3.10x10⁻¹¹ M H₂SeO₃, 9.56x10⁻⁷ M MOPS, 2.96x10⁻⁷ M thiamine, 2.05x10⁻⁹ M biotin, and 3.69x10⁻¹⁰ M cyanocobalamin at pH 6.8) [110]. Marine and brackish isolates were grow in f/2 enriched seawater medium at an appropriate salinity (8.25x10⁻⁶ M Na₂EDTA, 8.25x10⁻⁶ M FeCl₃-EDTA, 5.00x10⁻⁶ M K₂HPO₄, 5.50x10⁻⁴ M NaNO₃, 1.00x10⁻⁸ M Na₂SeO₃, 3.93x10⁻⁸ M CuSO₄, 5.87x10⁻⁹ M CuCl₂, 2.60x10⁻⁸ M Na₂MoO₄, 7.65x10⁻⁸ M ZnSO₄, 9.10x10⁻⁷ M MnCl₂, 2.96x10⁻⁷ M thiamine, 2.05x10⁻⁹ M biotin, and 3.69x10⁻¹⁰ M cyanocobalamin in 30 kDa MWCO ultrafiltrated seawater) [110]. Cultures were kept in the dark at room temperature and split once a month, where 9.5ml of media supplemented with yeast/tryptone and a wheat grain were inoculated with 500 μl of the previous month’s culture. Protist morphology was screened by phase contrast light microscopy (Olympus AX70) and heterogeneous cultures suggesting the presence of more than one species were subjected to several rounds of end-point dilutions until a homogeneous protist population was observed. Those isolates were presumed to be clonal. Additionally, large low-abundance protists were isolated by single-cell picking using a heat molded Pasteur pipette. Homogeneity and purity of the cultures was further confirmed by staining the nucleic acid with SYBR Green (Invitrogen Carlsbad, California, USA) or the food vacuole with LysoTracker Green (Molecular Probes, Eugene, OR, USA) and counting on a FACScalibur flow cytometer (Becton-Dickinson, Franklin Lakes, NJ, USA). A single population of large and bright green cells distinctly above the bacterial population (FSC vs. green fluorescence) was regarded a homogeneous population. Of 130 isolated and otherwise obtained protist cultures, 49 satisfied these criteria for
a clean culture and were screened. The identity of the isolates was confirmed by Sanger sequencing of the V1 to V3 variable region or full length 18s rDNA locus [111, 112].

<table>
<thead>
<tr>
<th>Species:</th>
<th>Donated by:</th>
<th>Culture collection # or reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neobodo designis</td>
<td>David Caron, USC</td>
<td>N.A.</td>
</tr>
<tr>
<td>Parauronema acutum</td>
<td>David Caron, USC</td>
<td>N.A.</td>
</tr>
<tr>
<td>Paraphysomonas bandaiensis</td>
<td>David Caron, USC</td>
<td>N.A.</td>
</tr>
<tr>
<td>Paraphysomonas imperforate</td>
<td>David Caron, USC</td>
<td>VS1: Lim et al. 1999 [113]</td>
</tr>
<tr>
<td>Uronema sp.</td>
<td>David Caron, USC</td>
<td>N.A.</td>
</tr>
<tr>
<td>Paraphysomonas vestita</td>
<td>David Caron, USC</td>
<td>GflagA: Macaluso et al. 2009 [114]</td>
</tr>
<tr>
<td>Spumella elongata</td>
<td>David Caron, USC</td>
<td>CCAP 955/1</td>
</tr>
<tr>
<td>Bicosoeca sp.</td>
<td>Naoji Yubuki; UBC</td>
<td>N.A.</td>
</tr>
<tr>
<td>Wobilna lunata</td>
<td>Naoji Yubuki; UBC</td>
<td>N.A.</td>
</tr>
<tr>
<td>Cafeteria roenbergensis</td>
<td>Matthias Fischer, MPI</td>
<td>Germany</td>
</tr>
</tbody>
</table>

2.3.2 Virus concentration

Water samples were pooled according to salinity into fresh (<1ppt), brackish (1-25ppt) and salt water (>25ppt). Protists were removed from pooled samples using pressure filtration through a 142mm GF/A filter (Millipore, Bedford, MA, USA; nominal pore size 1.1 μm) over a 0.8 μm pore-size PES membrane (Sterlitech, Kent, WA, USA) using a peristaltic pump [73]. Filtrates were concentrated using a 6m³ tangential flow filtration cartridge with a spiral wound 30kDa MW cut-off (Millipore, Bedford, MA, USA), down to a final volume of 250 ml and stored at 4°C in the dark [115].
2.3.3 Screen for lytic agents and replication kinetics

The 49 “clean” protist isolates were grown in triplicate cultures in appropriate media. Two 25 ml cultures were inoculated with 5 ml of the virus concentrate of appropriate salinity, the third 25-ml replicate served as a control. Cell numbers were screened by flow cytometry every other day for up to 20 days using LysoTracker Green (Molecular Probes) staining and FSC on a FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA) [116]. Once a lytic event was observed, as indicated by a drop in the cell density observed in flow cytometry, the lysate was vacuum-filtered through a 0.8 μm PES membrane (Sterlitech) to remove host cells. (Sterlitech). The lytic agents were propagated and clonal stocks were created by endpoint dilution and stored at 4°C in the dark. Stocks were tested by 16S rDNA PCR for the presence of bacterial pathogens [117].
2.3.4 Electron microscopy

2.3.4.1 Negative staining transmission electron microscopy

The efficacy of concentrating giant viruses and the nature of lytic agents in the supernatant of lysed host cultures were confirmed using negative-staining electron microscopy. Virus particles suspended in lysates were applied to the carbon side of a formvar carbon-coated 400 mesh copper grids (TedPella, CA, USA) and incubated at 4°C in the dark overnight under high humidity. Next, the lysate was removed and the grids were stained with 1% uranyl acetate for 30 s.

2.3.4.2 Thin section microscopy

Exponentially growing cultures were infected with 1 ml of lysate from a previous propagation. Cells were harvested from infected as well as from uninfected control cultures after 24 hours. Cells from 50 ml of control and experimental cultures were pelleted in two consecutive 10 min runs at 5000 xg in a fixed rotor tabletop centrifuge. The pellet was resuspended in 0.2 M Na cacodylate buffer, 0.2 M sucrose, 5% EM grade glutaraldehyde (pH 7.4) and incubated for 2 h on ice. After washing in 0.2 M Na cacodylate buffer, cells were postfixed with 1% osmium tetroxide. Samples were dehydrated through water/ethanol gradients and ethanol was substituted by acetone. Samples were embedded in an equal part mixture of Spurr's and Gembed embedding and the resin was polymerized at 60°C overnight. Fifty-nm thin sections were prepared using a Diatome ultra 45° knife (Diatome, Switzerland) on an ultra-microtome. The sections were collected on a 40x copper grid and stained for 10 min in 2% aqueous uranyl acetate and 5 min in Reynold's lead citrate.

Image data were recorded on a Hitachi H7600 transmission electron microscope at 80 kV. Image J (RRID:SCR_003070) was used to compile all the TEM images. Adjustments to contrast and brightness levels were applied equally to all parts of the image.

2.4 Results

2.4.1 Virus concentrates

Tangential flow filtration, an established method for concentrating viruses from natural waters [115], was adapted to concentrate micron size pathogens and bacteria (Fig. 2). Replacing the 0.22-μm prefiltration with 0.8 μm prefiltration allowed bacterial and viral populations to be concentrated up to 500-fold, while excluding flagellates (Fig 2, 3). Observation by TEM confirmed
abundant large icosahedral virus particles in all of the concentrates, which were pooled according to their salinity (freshwater, brackish-water and saltwater), as well as large numbers of bacteria and phage. The majority of VLPs showed simple architecture and small icosahedral capsids between 150 nm and 200 nm in diameter, which lacked detectable internal structures or external modifications, such as fibers (Fig 3 B-D), consistent with viruses belonging to the NCLDV, *Phycodnaviridae*, and the proposed Mesomimivirinae subfamily of the *Mimiviridae*, all of which predominantly infect eukaryotic picoplankton [26, 107, 118]. A smaller fraction of the particles were larger (Fig. 3 A,F), and with capsid sizes of ~0.5 μm resembled members of the proposed subfamily Megamimivirinae, within the *Mimiviridae*, albeit without the typical fibers decorating viruses infecting *Acanthamoeba* spp. Therefore, these viruses might be from a different viral subfamily, or the concentration may have removed the fibers. Further evidence that some of these particles are viruses in the *Mimiviridae* is the presence of a putative stargate (Fig 3F).
2.4.2 Isolates

2.4.2.1 Screen

Of the 49 cultures included in the initial screen, 14 cultures showed a lytic phenotype after inoculation with the mixed pathogen concentrate of the salinity appropriate to their culture condition. These initial lysates were filtered through a 0.8-µm PES or ATTP filter and used to inoculate the same host strain for a second propagation, while performing an end-point dilution to begin cloning the lytic agent. Of the 14 lysates subjected to a second-round propagation, six remained lytic and were subjected to closer interrogation using flow cytometry and electron microscopy.
2.4.2.2  PspV, a putative viral lytic agent infecting *Procryptobia sp.*

A kinetoplastid heterokont flagellate was isolated from the chlorophyll maximum in Saanich Inlet (48°35'18"N, 123°30'13"W) in May 2014. Sequencing of the 18S rDNA showed 99.8% identity to *Procryptobia sp.* (AY490216) over the 431 sequenced nucleotides. Inoculation with the mixed pathogen concentrate showed a decline in cell numbers in the treated samples compared to the untreated control, and could be propagated for one more round (Fig. 4A). Negative-staining electron microscopy showed virus-like particles of approximately 260 nm in diameter with an outer layer surrounding the capsid, and an interior electron-dense core (Fig 4B). A low density-population with a characteristic fluorescence and side-scatter profile was associated with cell lysis, suggesting the presence of low abundance large virus particles (Fig. 4C). The putative virus was provisionally named Procryptobia sp. Virus (PspV).

![Figure 2-4: A lytic agent of *Procryptobia sp.*:](image)

*Figure 2-4: A lytic agent of *Procryptobia sp.*: A: Cell numbers of *Procryptobia sp.* after inoculation with the lytic agent obtained from the initial screen (duplicate measurements, mean value shown). B: Negative-staining electron micrograph of a VLP in the lysate of *Procryptobia sp.* from the experiment in A, putatively representing PspV. Scale bar = 100nm. C: Flow cytometry profile of 0.8-μm filtered lysate stained with SYBR green. Population of the putative lytic agent highlighted by the black rectangle (see Supplementary Figure 1 for control).*

2.4.2.3  PvV a putative non-icosahedral giant virus infecting *Paraphysomonas vestita* GflagA

Cells of *Paraphysomonas vestita* GflagA declined sharply after inoculation with the fresh water mixed pathogen concentrate compared to the untreated control, and this phenotype could be propagated indefinitely (Fig. 5A). Thin-sectioning electron microscopy showed 600 by 300 nm amphora-shaped particles inside infected cells that were absent from control cells (Fig 5B). A low-density population with characteristic fluorescence and side-scatter occurred only in lysates, suggesting the presence of low abundance pathogens (Fig. 5C). The similarity with the amphora-
shaped viruses of the *Pandoraviridae* and pithovirus suggested the pathogen is viral, and thus was tentatively named *Paraphysomonas vestita virus* (PvV) [68, 70].

![Figure 2-5](image)

**Figure 2-5:** A lytic agent of *Paraphysomonas vestita* GflagA: A: Cell numbers of *Paraphysomonas vestita* after inoculation with the lytic agent from the initial screen (duplicate measurements, mean value shown). B: Electron micrograph of a VLP in *Paraphysomonas vestita* from the experiment in A, putatively representing PvV. Scale bar = 100nm. C: Flow cytometry profile of SYBR green stained 0.8-μm filtered lysate. Population of the putative lytic agent highlighted by the black rectangle (see Supplementary Figure 1 for control).

### 2.4.2.4 HSV 2-3, a putative virus infecting the flagellated amoeba hs 2-3

A flagellated amoeba was isolated from water near the sediment surface in Halkett Bay Marine Provincial Park, Howe Sound, BC (49°27'13"N, 123°19'36"W) in July 2014. Sequencing of the 18S rDNA showed 96% identity to *Cunea profundata* (KP862837) over the 405 sequenced nucleotides. Inoculation with the brackish water mixed pathogen concentrate showed a decline in cell numbers compared to the untreated control, and the lytic phenotype remained could be repeated for another propagation (Fig. 6A). Negative-staining electron microscopy showed virus-like particles, approximately 210-nm in diameter (Fig. 6B). A high-density population showing characteristic fluorescence was associated with cell lysis, suggesting high-abundance virus particles in the lysate (Fig 6C). As the host has not been with certainty be identified, the putative virus was provisionally named after the sampling location Howe Sound Virus 2-3 (HSV 2-3).
Figure 2-6: A lytic agent of the nanoflagellate isolate hs 2-3.: A: Cell numbers of hs 2-3 after inoculation with the lytic agent from the initial screen, and its subsequent propagation (duplicate measurements, mean value shown). B: Negative-staining electron micrograph of a HSV 2-3 VLP in the lysate of hs 2-3 from the experiment in A. Scale bar = 100nm. C: Flow cytometry profile of SYBR green stained 0.8-μm filtered lysate. Population of the putative lytic agent highlighted in a flow-cytometry plot (see Supplementary Figure 1 for control).

2.4.2.5 SspV, a viral lytic agent infecting Spumella sp.

A Spumella-like heterokont nanoflagellate was isolated from near the sediment surface of a freshwater bog in Minnekhada Park Regional Park (49°18'10"N, 122°42'9"W) in July 2014. Sequencing of the 18S rDNA showed 99.6 % identity to Spumella sp. (AB425951), and an environmental cercozoa sequence (EF024525) over the 523 sequenced residues. Inoculation with the mixed pathogen concentrate showed a stark decline in cell numbers compared to the untreated control. This lytic phenotype could be repeated for another propagation (Fig. 7A). Negative-staining electron microscopy showed virus-like particles approximately 230-nm in diameter (Fig 7B). A low density population with characteristic fluorescence and forward scatter was associated with cell lysis, suggesting low-abundance large viral particles in the lysate (Fig 7C). The lytic agent was provisionally named SspV1.
2.4.2.6 BsV NG1, a giant virus infecting Bodo saltans NG

A kinetoplastid heterokont flagellate was isolated from near sediment water of the pond in Nitobe Memorial Garden UBC (49°15'58"N, 123°15'34"W) in 2014. Sequencing of the 18S rDNA showed 99.8 % identity to Bodo saltans (AY490224) over the 2023 sequenced nucleotides [111]. The isolated strain (B. saltans NG) was deposited at the Canadian Center for the Culture of Microorganisms (reference number CCCM 6296). Inoculation with the mixed pathogen concentrate resulted in a stark decline in cell numbers compared to the untreated control, which could be propagated indefinitely (Fig. 8A). Negative-staining electron microscopy showed virus-like particles of approximately 230-nm in diameter (Fig 8B). A high density population with characteristic fluorescence and forward scatter was associated with cell lysis, suggesting the presence of high-abundance large viral particles (Fig 8C). Due to its viral appearance, the lytic agent was archived and classified as Bodo saltans virus NG1 (BsV NG1).
Figure 2-8: A lytic agent of Bodo saltans NG: A: Cell numbers of Bodo saltans after inoculation with the lytic agent from the initial screen, and its propagation (duplicate measurements, mean value shown). B: Negative-staining electron micrograph of a BsV VLP in the lysate of Bodo saltans from the experiment in A. Scale bar = 100nm. C: Flow cytometry profile of SYBR green stained 0.8-μm filtered lysate. Population of the putative lytic agent highlighted by the black rectangle (see Supplementary Figure 1 for control).

2.4.2.7 Candidatus Chromulinavorax destructans, pathogenic bacterium infecting Spumella elongata CCAP 955/1

Cultures of Spumella elongata CCAP 955/1 declined steeply when treated with the mixed pathogen concentrate compared to the untreated control, and the agent could be propagated indefinitely (Fig. 9A). Negative-staining electron microscopy showed abundant polymorphic particles of about 350-nm in diameter with an irregular depression (Fig 9B). A high density population with characteristic fluorescence and forward scatter was associated with cell lysis, suggesting the presence of high-abundance pathogens in the lysate (Fig 9C). The unusual particle size prompted the concentration and 16S rDNA sequencing of the concentrate revealing a 16S rDNA sequence with high similarity to a metagenomic sequence belonging to the candidate phylum TM6 (AY689681). Thus, the pathogen was termed candidatus Chromulinavorax destructans (after the host family Chromulinaeae and latin “vorare” meaning “to devour”).
Figure 2-9: A lytic agent of *Spumella elongata* CCAP 955/1: A: Cell numbers of *Spumella elongata* after inoculation with the lytic agent from the initial screen (duplicate measurements, mean value shown). B: Negative-staining electron micrograph of *Chromulinavorax destructans* in the lysate of *Spumella elongata* from the experiment in A. Scale bar = 100nm. C: Flow cytometry profile of the lysate after 0.8μm filtration stained with SYBR green. Population of the putative lytic agent highlighted by the black rectangle (see Supplementary Figure 1 for control).

2.4.3 SspV, HsV and PspV lose infectivity during propagation

Similar to six of the initial 14 lytic agents, SspV, HsV and PspV lost their lytic potential after second-round propagation. Together, 57% of lytic agents lost their lytic activity after the first propagation and 50% of the remaining lytic agents lost their potential after the second propagation. The lysates from the initial screen and second propagation maintained their lytic potential for four years, suggesting that the loss of lytic activity is not simply a function of viral half-life, but rather a biological phenomenon related to the number of propagations.

Figure 2-10 Loss of infectivity with additional propagations of the putative viruses HSV 2-3, PspV and SspV (duplicate measurements, mean value shown).
2.5 Discussion

2.5.1 Pathogens of heterotrophic protists are readily isolated utilizing a screen employing ultrafiltration and flow cytometry

In line with predictions from metagenomic studies, diverse large virus-like particles occurred in marine and freshwater pathogen concentrates [26]. The particles were icosahedral in structure, as are members of the *Phycodnaviridae* and *Mimiviridae*, suggesting these viruses are abundant in natural virus assemblages [26]. However, these observations might be biased because icosahedral viruses are easily recognized in negative-staining electron micrographs, while viruses of other shapes are easily confused with bacteria, or attributed to staining artifacts [68, 70, 119]. 30 kDa-cutoff tangential flow ultrafiltration of ~0.8-µm filtrates is an effective way to concentrate large viruses and bacteria to screen for lytic agents. Coupled with flow cytometry for rapid, high-throughput screening of abundances of protist isolates that are not amenable to culture on solid media that could allow plaque assay, this approach provides a robust way to screen water samples for pathogens of protists.

Using these approaches, bacterial and giant viral pathogens of heterotrophic protists were isolated. These included a bacterium of the candidate phylum, TM6, of which only two isolates have been described. As well, isolates and putative isolates of giant viruses infect a spectrum of protists that are widely dispersed across the eukaryotic tree of life (Fig. 11). Specifically, the inclusion of members of the super group Excavata as hosts for giant viruses, and Stramenopiles as hosts for a bacterium in the TM6 group is remarkable, as amoeba were the only known hosts for these bacterial pathogens and symbionts [34, 109].
Figure 2-11 Updated host-range of NCLDV and TM6/Dependentiae. Isolated from this study are highlighted in bright colors. Adapted from Burki et al. 2014 with permission from the author [120].

2.5.2 Giant viruses might resist prolonged serial passage due to co-replicating virophages

A striking number of lytic agents lost their lytic potential during propagation: 57% of the agents were not lytic after the first passage, and 50% of the remaining agents were not lytic after the second passage, suggesting that a biological mechanism was involved. All the lytic agents that did not propagate past the second round were viral in appearance, consistent with the possibility
that virophages were involved. Similarly, the giant virus, Cafeteria roenbergensis virus (CroV), developed resistance to propagation due to mavirus, a virophage that parasitizes the virus factory of the giant virus CroV and blocks its replication [72, 74, 106]. This effect is further amplified by the ability of mavirus to enter C. roenbergensis cells irrespective of CroV infection, followed by integration of the virophages genome in the flagellates genome [76]. These inserted viral sequences can be reactivated upon viral infection by CroV, effectively functioning as an adaptive immune system for C. roenbergensis populations. Furthermore, virophages can parasitize several strains of mimiviruses that infect Acanthamoeba spp., and traces of virophage integration have been found in several protist genomes [29, 121]. Additionally, several reports suggest that virophages are abundant and dynamic players in aquatic environments [75, 122]. Together, these observations suggest that virophages likely caused the loss of lytic activity in putative giant viruses isolated in this study. These virophages may have been present in the mixed pathogen concentrates used for the protist bioassays, or may have been reactivated from within the host genomes after infection by a giant virus. Likely, virophages explain the comparatively low number of giant viruses that have been isolated, despite their high abundance in the environment [26].

2.5.3 Isolation of parasitoids of heterotrophic protists – mining for microbial dark matter

The presented method of isolating host-pathogen pairs has provided new insights into the diversity of giant viruses and lytic bacteria, and the range of hosts they can infect. Interrogation of the isolates would provide insights into the genomic complement carried by these unusual pathogens, and their host interactions. Further, the screening method could be optimized, as the main bottleneck is the isolation and maintenance of potential host organisms. Due to feasibility of maintaining a large number of protist cultures, the present study focused on easy-to-maintain heterotrophic flagellates. This biased the study by excluding many environmentally abundant and important heterotrophic protists, such as ciliates, amoebae and dinoflagellates. A more comprehensive approach for pathogen isolation would likely extend the phylogenetic space of protists that are can be infected by giant viruses and bacteria, as there are many sub-branches on the eukaryotic tree of life for which giant viruses and other lytic pathogens have yet to be isolated. Further optimization of screening methods to include a wider diversity of potential hosts will increase our understanding of the diversity of parasitoids of protists.
3 The kinetoplastid-infecting Bodo saltans virus (BsV), a window into the most abundant giant viruses in the sea

3.1 Synopsis

Giant viruses are ecologically important players in aquatic ecosystems that have challenged concepts of what constitutes a virus. The giant Bodo saltans virus (BsV) is the first characterized representative of the most abundant group of giant viruses in ocean metagenomes, and the first isolate of a klosneuvirus, a subgroup of the Mimiviridae proposed from metagenomic data. BsV infects an ecologically important microzooplankton, the kinetoplastid Bodo saltans. Its 1.39 Mb genome encodes 1227 predicted ORFs, including a complex replication machinery. Yet, much of its translational apparatus has been lost from the most recent predicted ancestor, including all tRNAs. Essential genes have been invaded by homing endonuclease-encoding self-splicing introns that may defend against competing viruses. Putative anti-host factors show extensive gene duplication via a genomic accordion indicating an ongoing evolutionary arms race and highlighting the rapid evolution and genomic plasticity that has led to genome gigantism and the enigma that is giant viruses.

3.2 Introduction

Viruses are the most abundant biological entities on the planet and there are typically millions of virus particles in each ml of marine or fresh waters that are estimated to kill about 20% of the living biomass each day in surface marine waters [24]. This has major consequences for global nutrient and carbon cycles, as well as for controlling the composition of the planktonic communities that are the base of aquatic foodwebs. Although the vast majority of aquatic viruses are less than 100 nm in diameter and primarily infect prokaryotes, it is increasingly clear that a subset of the viruses in aquatic ecosystems are comparative Leviathans that have been colloquially classified as giant viruses.

The first isolated giant virus in the family that later became known as the Mimiviridae infects a marine heterotrophic flagellate that was initially identified as Bodo sp. [106], and later shown to be Cafeteria roenbergensis [72]. Subsequently, the isolation and sequencing of mimivirus, a giant
virus infecting *Acanthamoeba polyphaga* [28, 67], transformed our appreciation of the biological and evolutionary novelty of giant viruses. This led to an explosion in the isolation of different groups of giant viruses infecting *Acanthamoeba spp.* including members of the genera Pandoravirus, Pithovirus, Mollivirus, Mimivirus and Marseillevirus [68, 70, 119, 123, 124]. Although each of these isolates expanded our understanding of the evolutionary history and biological complexity of giant viruses, all are pathogens of *Acanthamoeba spp.*, a widespread taxon that is representative of a single evolutionary branch of eukaryotes, and which is not a major component in the planktonic communities that dominate the world’s oceans and large lakes.

As knowledge of mimiviruses infecting *Acanthamoeba spp.* has expanded it has become evident based on analysis of metagenomic data that giant viruses and their relatives are widespread and abundant in aquatic systems [26, 71, 125]. However, except for Cafeteria roenbergensis Virus (CroV) that infects a microzooplankton [72], and the smaller phytoplankton-infected viruses Phaeocystis globosa virus PgV-16T [105, 107], and Aureococcus anophagefferens virus [118], the only members of the Mimiviridae that have been isolated and characterized infect *Acanthamoeba spp.*

Motivated by the lack of ecologically relevant giant-virus isolates, isolation and screening of representative microzooplankton was performed in order to isolate new giant-viruses that can serve as model systems for exploring their biology and function in aquatic ecosystems. In this screen, Bodo saltans virus (BsV) was isolated, a giant virus that infects the ecologically important kinetoplastid microzooplankter *Bodo saltans*, a member of the phylum Euglenazoa within the supergroup Excavata. This group of protists is well represented by bodonids in freshwater environments and by diplonemids in the oceans [126, 127]. Kinetoplastids are remarkable for their highly unusual RNA editing and having a single large mitochondrion, the kinetoplast, that contains circular concatenated DNA (kDNA) that comprises up to 25% of the total genomic content of the cell [127, 128], and are well known as causative agents of disease in humans (e.g. Leishmaniasis and sleeping sickness) and livestock [129, 130]. At 1.39 MB, BsV has the largest described complete genome of a cultured strain within the giant virus family Mimiviridae. Based on a recruitment analysis of metagenomic reads, BsV is representative of the most abundant group within the Mimiviridae in the ocean, and is the only isolate of the klosneuviruses, a group only known from metagenomic data [71]. The BsV genome exhibits evidence of significant genome rearrangements and recent adaptations to its host.
3.3 Materials and methods

3.3.1 Sampling and Isolation:

Virus concentrates were collected from 11 fresh water locations in southern British Columbia, Canada (49°49'4"N, 123° 7'6''W; 49°42'5"N, 123° 8'47''W; 49°37'34"N, 123°12'27''W; 49° 6'12"N, 122° 4'38''W; 49° 5'22"N, 122° 7'1''W; 49°18'10"N, 122°42'9''W; 49° 8'27"N; 123° 3'16''W; 49°13'21"N, 123°12'43''W; 49°13'13"N, 123°12'41''W; 49°14'52"N, 123°13'59''W; 49°15'58"N, 123°15'34''W). To concentrate giant viruses, 20 liter water samples were prefiltered with a GF-A filter (Millipore, Bedford, MA, USA; nominal pore size 1.1um) over a 0.8um PES membrane (Sterlitech, Kent, WA, USA). Filtrates from all locations were pooled and were concentrated using a 30kDa MW cut-off tangential flow filtration cartridge (Millipore, Bedford, MA, USA) [115].

*Bodo saltans* strain NG, the host of BsV, was isolated from a water sample collected near the sediment surface of the pond in Nitobe Memorial Garden, The University of British Columbia, Canada (49°15'58"N, 123°15'34''W). Clonal cultures were obtained by end-point dilution, and maintained in modified DY-V artificial fresh water media with yeast extract and a wheat grain [110]. The identity of the host organism was established by 18S sequencing and the strain was deposited at the Canadian Center for the Culture of Microorganisms (http://www3.botany.ubc.ca/cccm/) reference number CCCM 6296 [111]. Bodo saltans NG cultures were inoculated at approximately 2x10^5 cells/ml with the pooled giant virus concentrate from all 11 locations. Cell numbers were determined by flow cytometry and compared to a medium only mock-infected control culture (LysoTracker Green (Molecular Probes) vs. FSC on FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA))[116]. After a lytic event was observed, the lysate was filtered through a 0.8um PES membrane (Sterlitech) to remove host cells. The lytic agent was propagated and a monoclonal stock was created by three consecutive end point dilutions. The concentrations of the lytic agent were screened by flow cytometry using SYBR Green (Invitrogen Carlsbad, California, USA) nucleic acid stain after 2% glutaraldehyde fixation (vs SSC) [131]. The similarity to the flow cytometry profile of Cafeteria roenbergensis virus suggested that the lytic agent was a giant virus.
3.3.2 Transmission electron microscopy

3.3.2.1 Negative staining

*Bodo saltans* lysates after BsV infection were applied to the carbon side of a formvar-carbon coated 400 mesh copper grid (TedPella, CA, USA) and incubated at 4°C in the dark overnight in the presence of high humidity. Next, the lysate was removed and the grids were stained with 1% Uranyl acetate for 30 seconds before observation on a Hitachi H7600 transmission electron microscope at 80 kV.

3.3.2.2 High-pressure freezing and ultra-thin sectioning

Exponentially growing *B. saltans* cultures were infected at a concentration of 5x10^5 cells ml^-1 with BsV at a relative particle to cell ratio of ~5 to ensure synchronous infection. Cells were harvested from infected cultures at different time points (6, 12, 18, 24 h post infection) as well as from uninfected control cultures. Cells from 50 ml were pelleted in two consecutive 10 min at 5000 xg centrifugation runs in a Beckmann tabletop centrifuge. Pellets were resuspended in 10-15 µl DY-V culture medium with 20% (w/v) BSA and immediately place on ice. Cell suspensions were cryo-preserved using a Leica EM HPM100 high-pressure freezer. Vitrified samples were freeze-substituted in a Leica AFS system for 2 days at -85°C in a 0.5% glutaraldehyde 0.1% tannic acid solution in acetone, then rinsed ten times in 100% acetone at -85°C, and transferred to 1% osmium tetroxide, 0.1% uranyl acetate in acetone and stored for an additional 2 days at -85°C. The samples were then warmed to -20°C over 10 hours, held at -20°C for 6 hours to facilitate osmication, and then warmed to 4°C over 12 hours. The samples were then rinsed in 100% acetone 3X at room temperature and gradually infiltrated with an equal part mixture of Spurr’s and Gembed embedding media. Samples were polymerized in a 60°C oven overnight. 50 nm thin sections were prepared using a Diatome ultra 45° knife (Diatome, Switzerland) on an ultra-microtome. The sections were collected on a 40x copper grid and stained for 10 min in 2% aqueous uranyl acetate and 5 min in Reynold’s lead citrate. Image data were recorded on a Hitachi H7600 transmission electron microscope at 80 kV. Image J (RRID:SCR_003070) was used to compile all TEM images. Adjustments to contrast and brightness levels were applied equally to all parts of the image.

3.3.3 Virus concentration and sequencing

For Illumina sequencing, exponentially growing *B. saltans* cultures were infected at a concentration of approximately 5x10^5 cells ml^-1 with BsV lysate (10^7 VLP ml^-1) at a multiplicity
of infection (MOI) of ~0.5. After four days, when host cell densities had dropped below 30%, cultures were centrifuged in a Sorvall SLC-6000 for 20 min, 5000 rpm, 4°C to remove remaining host cells and the supernatant was consecutively subjected to tangential flow filtration with a 30kDa cut-off (Vivaflow PES) and concentrated approximately 100x. Viral concentrates were subjected to ultracentrifugation at 28,000 rpm, 15°C for 8h in a Ti90 fixed angle rotor (Beckman-Coulter, Brea, California, USA). Pellets were resuspended and virions lysed using laurosyl acid and proteinase K subjected to pulsed-field gel electrophoresis on a CHEF II pulse field gel electrophoresis aperture (BioRad) for 25h at 14°C in a 0.8% LMP agarose gel with 60-180S switchtimes and 16.170 ramping factor in 0.5 TBE under 5.5V/cm and 120°. Genomic DNA was visualized under UV light after 30min SYBR gold (Invitrogen Carlsbad, California, USA) staining. The dominant PFGE band belonging to genomic BsV DNA (1.35Mb) was cut and DNA was extracted using a GELase kit (Illumina, San Diego) and ethanol purified according to manufacturer’s protocol. Libraries were prepared using the Illumina Nextera XT kit (Illumina, San Diego, CA, USA) as per manufacturer’s recommendation and library quality and quantity were checked by Bioanalyzer 2100 with the HS DNA kit (Agilent Technology). 300bp paired-end sequencing was performed on an Illumina MiSeq platform by UCLA’s Genoseq center (Los Angeles, CA, USA) to a nominal sequencing depth of 800x. Sequence quality was examined using FastQC (RRID:SCR_014583 http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) and sequence reads were quality trimmed (parameters: minlen=50 qtrim=r1 trimq=15 ktrim=r k=21 mink=8 ref=$adapters hdist=2 hdist2=1 tbo=t tpe=t) and cleared of human (parameters: minid=0.95 maxindel=3 bw=0.16 bw=12 quickmatch fast minhits=2 qtrim=lr trimq=10) and PhiX (parameters: k=31 hdist=1) sequences against the whole respective genomes using BBMap v35 (http://sourceforge.net/projects/bbmap/).

For PacBio sequencing, BsV was concentrated using precentrifugation and TFF analogously to the Illumina sequencing step. Next, the concentrate was further concentrated by sedimenting it onto a 40% Optiprep 50 mM Tris-Cl, pH 8.0, 2mM MgCl₂ cushion for 30 min at 28,000 rpm, 15°C in a SW40Ti rotor in an ultracentrifuge (Beckman-Coulter, Brea, California, USA). An Optiprep (Sigma) gradient was created by underlaying a 10% Optiprep solution in 50 mM Tris-Cl, pH 8.0, 2mM MgCl₂ with a 30% solution followed by a 50% solution and was equilibrated over night at 4°C. One ml of viral concentrate from the 40% cushion was added atop the gradient and the concentrate was fractionated by ultracentrifugation in an SW40 rotor for 4h
at 25000rpm and 18°C. The viral fraction was extracted from the gradient with a syringe and washed twice with 50 mM Tris-Cl, pH 8.0, 2 mM MgCl\(_2\) followed by centrifugation in an SW40 rotor for 20 min at 7200rpm and 18°C and were finally collected by centrifugation in an SW40 rotor for 30 min at 7800rpm and 18°C. Purity of the concentrate was verified by flow cytometry (SYBR Green (Invitrogen Carlsbad, California, USA) vs SSC on a FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA). High molecular weight genomic DNA was extracted using phenol-chloroform-chloroform extraction. Length and purity were confirmed by gel electrophoresis and Bioanalyzer 2100 with the HS DNA kit (Agilent Technology). PacBio RSII 20kb sequencing was performed by the sequencing center of the University of Delaware. Reads were assembled using PacBio HGAP3 software with 20 kb seed reads and resulted in a single viral contig of 1,384,624bp, 286.1x coverage, 99.99% called bases and a consensus concordance of 99.9551% [132].

Cleaned up Illumina reads were mapped to the PacBio contig to confirm the PacBio assembly as well as extending the contig’s 5’ end by 1245bp to a total viral genome length of 1,385,869bp.

### 3.3.4 Annotation

Open reading frames were predicted using GLIMMER (RRID:SCR_011931, Dechler, 2007) with a custom start codon frequency of ATG, GTG, TTG, ATA, ATT at 0.8, 0.05, 0.05, 0.05, 0.05 as well as stop codons TAG, TGA, TAA, minimum length 100bp, max overlap 25, max threshold 30.

Promoter motives were analyzed by screening the 100bp upstream region of CDS using MEME (RRID:SCR_001783) . tRNAs were predicted with tRNAscanSE (RRID:SCR_010835)[133]. Group 1 introns were predicted by disruptions in coding sequences and secondary RNA structure was predicted using S-fold [134]. Intron splicing was confirmed using RT-PCR and Sanger sequencing with gene-specific primers designed to span the splice site predicted by S-fold. Functional analysis of CDS was performed after translation with BLASTp against the nr database with an e-value threshold of \(10^{-5}\) as well as rps conserved domain search against CDD v3.15. Coding sequences were manually assigned to functional classes based on predicted gene function using Geneious R9 [135]. The annotated genome of BsV-NG1 was deposited in GenBank under the accession number MF782455. The Bodo saltans NG 18S sequence was deposited in GenBank under the accession number MF962814.
3.3.5 Phylogenetics

Whole genome content phylogeny was performed by OrthoMCL (RRID:SCR_007839) [136]. Available whole genome sequences of NCLDV from NCBI were downloaded. First, gene clustering using OrthoMCL was performed with standard parameters (Blast E-value cutoff = 10−5 and mcl inflation factor = 1.5) on all protein-coding genes of length ≥ 100 aa. This resulted in the definition of 3,001 distinct clusters. A presence/absence matrix based on the gene clusters was computed and calculated a distance matrix using the according to Yutin et al. 2009 [137]. Gene clusters were used to infer ancestral gene content by posterior probabilities in a phylogenetic birth-and-death model in COUNT [138]. Additional gene substitutions were added to the model where phylogenetic analysis of individual genes strongly suggested accordingly (see next section).

Alignments of aa sequences were performed in Geneious R9 (RRID:SCR_010519) using MUSCLE with default parameters RRID:SCR_011812)[139]. Proteins used for the concatenated NCVOG tree were DNA polymerase elongation subunit family B (NCVOG0038), D5-like helicase-primase (NCVOG0023), packaging ATPase (NCVOG0249), Poxvirus Late Transcription Factor VLTF3-like (NCVOG0262), and DNA or RNA helicases of superfamily II (NCVOG0076) [140]. Residues not present in at least 2/3 of the sequences were trimmed and ProtTest 3.2 was used for amino-acid substitution model selection (RRID:SCR_014628). Maximum likelihood trees were constructed with RAxML rapid bootstrapping and ML search with 1000 Bootstraps utilizing the best fitting substitution matrixes determined by prottest (RRID:SCR_006086) [141]. Maximum likelihood trees of translational genes, based on alignments by Schulz et al. 2017 where available, were constructed using PhyML (RRID:SCR_014629) [71, 142]. Phylogenetic trees were computed with PhyloBayes-7MPI 1.4f in two Markov Chain Monte Carlo chains under the CAT-GTR model for 10,000 to 25,000 generations. The consensus tree was based on both chains, removing the first 1000 generations. Convergence was confirmed with bcomp and tracecomp (RRID:SCR_006402) [143] Trees were visualized in Figtree (A. Rambaut - http://tree.bio.ed.ac.uk/software/figtree/).

Translated environmental assemblies identified by Hingamp et al. as representing NCLDV DNA polymerase B family genes were mapped to a Mimiviridae DNA polymerase B family reference tree created as described above with pplacer (RRID:SCR_004737) [26, 144]. Environmental reads were aligned to the reference alignment using clustalw and were mapped under Bayesian
setting. The fat tree was visualized with Archaeopteryx (https://sites.google.com/site/cmzmasek/home/software/archaeopteryx). Of the 401 input sequences, 256 mapped within the Mimiviridae and are displayed in the figure.

3.4 Results

3.4.1 Isolation and Infection Kinetics.

In an effort to isolate giant viruses that infect ecologically relevant organisms, protistan microzooplankton from a variety of habitats were isolates and screened against their associated virus assemblages. One such screen using water collected from a temperate eutrophic pond in southern British Columbia, Canada, yielded a giant virus that was classified as Bodo saltans virus, Strain NG1 (BsV-NG1) that infects an isolate of the widely occurring kinetoplastid, *Bodo saltans* (Strain NG, CCCM6296). The addition of BsV to a culture of *Bodo saltans* (~2.5x10^5 cells ml⁻¹) at a virus particle to cell ratio of two, measured by flow cytometry, resulted in free virus particles 18 h later. Viral concentrations peaked at 2.5x10^7 particles ml⁻¹, while host cell density dropped to 25% of uninfected control cultures (Figure 3-1). The closely related strain *Bodo saltans* HFCC12 could not be infected by BsV-NG1, suggesting strain specificity.

![Figure 3-1: BsV induced lysis observed by flow cytometry. A: Flow cytometry profile of uninfected Bodo saltans stained with Lysotracker (blue arrow head) B: Flow cytometry profile of BsV (red arrow head) and bacteria (green arrow head) stained with SYBR Green. C: The abundances of B. saltans cells and BsV particles after infection at a particle to cell ratio of 2.](image)

3.4.2 Virus morphology and Replication Kinetics.

Transmission electron microscopy (TEM) revealed that BsV is an icosahedral particle approximately 300 nm in diameter (Figure 3-2A). The particle consists of at least six layers akin
to observations of Acanthamoeba polyphaga mimivirus (ApMV) [145]. The DNA-containing core of the virion was surrounded by a core wall and an inner membrane, and a putative membrane sitting under a double-capsid layer ([Figure 3-2A]. A halo of approximately 25 nm surrounds the virion. A possible stargate-like structure, as observed in ApMV, is associated with a depression of the virus core (Supplementary Figure 2B, D), which presumably releases the core from the capsid during infection [146, 147].

Figure 3-2 Ultrastructure of BsV particles and replication. A: Mature BsV virion: DNA containing core is surrounded by two putative membranous layers. The capsid consists of at least two proteinaceous layers. The bright halo hints to the presence of short (~40nm) fibers as observed in ApMV. The top vertex of the virion contains a possible stargate structure (See also Figure 3.3). (Scale bar =100nm) B: Healthy *Bodo saltans* cell: Nucleus with nucleolus and heterochromatin structures (Back arrow head) and kinetoplast genome (white arrow head) are clearly visible. (Scale bar = 500nm) C: Cell of *Bodo saltans* 24 h post BsV infection: Most subcellular compartments of healthy cells have been displaced by the virus factory now taking up a third of the cell. Virion production is directed towards the periphery of the cell (black arrow). Kinetoplast genome remains intact (white arrow head) while the nuclear genome is degraded (black arrow head; Scale bar = 500nm) D: BsV virion assembly and maturation: Lipid vesicles migrate through the virion factory where capsid proteins attach for the proteinaceous shell. Vesicles burst and accumulate at the virus
factory periphery where the capsid assembly completes (black arrow). Once the capsid is assembled, the virion is filled with the genome and detaches from the virus factory. Internal structures develop inside the virion in the cell’s periphery where mature virions accumulate until the host cell bursts (Scale bar = 500nm). See Supplementary Figure 2 for further information.

The healthy *Bodo saltans* cell presents intricate intracellular structures, including the characteristic kinetoplast and a pronounced cytostome and cytopharynx (Figure 3-2 B, Supplementary Figure 2A). In infected cells, virus factories were always observed in the cell’s posterior and particles always matured toward the posterior cell pole in a more spatially organized way compared to other *Mimiviridae* (Figure 3-2C, Supplementary Figure 2C) [148]. As infection progressed, the Golgi apparatus disappeared and the nucleus degraded, as evidenced by the loss of the nucleolus and heterochromatin (Figure 3-2B, C); yet, the kinetoplast remained intact, as indicated by the persistence of the characteristic kDNA structure (Figure 3-2B, C, Supplementary Figure 2A, C). Virus factories were first observed at 6h post-infection (p.i.) as electron-dense diffuse areas in the cytoplasm. By 12h p.i., the virus factory had expanded significantly and reached a maximum size of about one-third of the host cell, taking up most of the cytoplasm. The first capsid structures appeared at this time. At 18h p.i., the first mature virus particles were observed, coinciding with the first free virus particles observed by flow cytometry (Figure 3-1). By 24h p.i., most infected cells were at the late stage of infection with mature virus factories (Figure 3-2C, D). During virus replication, membrane vesicles were recruited through the virus factory where capsid proteins accumulated and disrupted the vesicles (Figure 3-2) [145]. The vesicle/capsid structures accumulated in the periphery of the virus factory where the capsid was formed (Figure 3-2C, D). Once the capsid was completed, the viral genome was packaged into the capsid at the vortex opposite to the putative stargate structure (Figure 3-2C, D). The internal structures of the virus particle matured in the cell periphery and accumulated below the host cytoplasmic membrane where they often remained for an extended period of time (Figure 3-2D, Supplementary Figure 2D). Besides being released during cell lysis, mature virus particles were observed budding in vesicles from the host membrane, reminiscent of a mechanism described for Marseillelevirus (Supplementary Figure 2 D) [149].

### 3.4.3 The *Bodo saltans* virus genome

#### 3.4.3.1 Genome organization.

Combined PacBio RSII and Illumina MiSeq sequencing resulted in the assembly of a 1,385,869bp linear double-stranded DNA genome (accession number MF782455), making the
BsV genome one of the largest complete viral genomes described to date, surpassing those of mimiviruses infecting Acanthamoeba spp. The GC content is 25.3% (Figure 3-3) and, as reported for other giant viruses [28], much lower than the ~50% observed for Bodo spp. [129]; this suggests the absence of large scale horizontal gene transfer with the host in recent evolutionary history. The genome encodes 1227 predicted open-reading frames (ORFs) with a coding density of 85%, with the ORFs distributed roughly equally between the two strands consistent with the constant GC-skew (Figure 3-3). Unlike ApMV, BsV does not display a central peak in GC skew and therefore does not have an organized bacterial like origin of replication [28]. The genomic periphery has a slightly skewed GC ratio due to the tandem orientation of repeated ORFs. Codon preference is highly biased towards A/T-rich codons and the amino acids Lysine, Asparagine, Isoleucine, and Leucine (10, 9.8, 9.6, 8%), which are preferentially encoded by A/T only triplets. The translation of the predicted ORFs resulted in proteins ranging from 43 to 4840 aa in length with an average length of 320 aa. Promotor analysis revealed a highly-conserved early promotor motif “AAAAATTGA” that is identical to that found in mimiviruses and CroV [72, 150]. A poorly-conserved late promotor motif “TGCG” surrounded by AT-rich regions was also observed. ORFs are followed by palindromic sequences, suggesting a hairpin-based transcription termination mechanism similar to ApMV [151]. Several non-coding stretches rich in repetitive sequences were observed, but no function could be attributed to them. Based on a BLASTp analysis, 40% of ORFs had no significant similarity to any other sequences and remained ORFans (Figure 3-4B). Most identified proteins (27%) matched sequences from eukaryotes; 2% of these matched best to sequences from isolates of B. saltans. The next largest fraction (22%) were most similar to viruses in the nucleocytoplasmic large DNA viruses (NCLDV) group, while the remaining ORFs were most similar to bacterial (9%) or archaeal (1%) sequences. In gene cluster analysis, only 45% of protein-coding gene clusters are shared with related viruses such as CroV and klosneuviruses, highlighting the low number of conserved core genes amongst these viruses (Supplementary Figure 5). Essential genes for replication, translation, DNA replication and virion structure are located in the central part of the genome, while the periphery is occupied by duplicated genes, including 148 copies of ankyrin-repeat-containing proteins (Figure 3-3).
3.4.3.2 Functional genome content.

While no function could be attributed to 54% of ORFs, the largest identifiable fraction of annotations are involved in DNA replication and repair (Figure 3-4A). Coding sequences for proteins associated with all classes of DNA repair mechanisms were identified including DNA mismatch repair (MutS and Uvr helicase/DDEDDh 3’-5’ exonucleases), nucleotide excision repair (family-2 AP endonucleases), damaged-base excision (uracil-DNA glycosylase and formamidopyrimidine-DNA glycosylase) and photoreactivation (deoxyribodipyrimidine photolyase). The repair pathways are completed by DNA polymerase family X and NAD-dependent DNA ligase. Sequences were also found that code for proteins involved in DNA
replication, including several primases, helicases, and an intein-containing family-B DNA polymerase, as well as replication factors A and C, a chromosome segregation ATPase, and topoisomerases 1 (two subunits) and 2. Sequences associated with proteins mediating recombination were also identified including endonucleases and resolvases, as well as the aforementioned DNA repair machinery.

There were 47 sequences identified that matched enzymes involved in protein and signal modification, with the majority being serine/threonine kinases/phosphatases. These are potentially involved in host cell takeover.

Figure 3-4 BsV genome content.: A: Functional assignment of BsV genome content based on BLASTp and CDD rps-BLAST B: Domain of best BLASTp hits C: Evolutionary history of translational machinery found in giant viruses inferred by COUNT. The size of the black circles mapped on a cladogram of the large members of the Mimiviridae (see Figure 3-6 for full phylogenetic analysis) represents the number of protein coding gene families involved in translation at each node or tip. Blue circles indicate the number of tRNA genes found in each genome. Gene gain and loss events are depicted along the branches. Genomes based on metagenomic assemblies are highlighted to indicate the possibility of incomplete representation of the translation machinery. See Figure 3.7 for the complete phylogenetic tree for members of the Mimiviridae. See Supplementary Table 1 for a table of all genes included in the analysis. See Supplementary Figure 5 for a cladogram depicting the inferred evolutionary history of all gene families in within the Mimiviridae. MVc: Megavirus chilensis, AMoV: Acanthamoeba polyphaga Moumouvirus, ApMV: Acanthamoeba polyphage Mimivirus, CatV: Catovirus, BsV: Bodo saltans virus, HokV: Hokovirus, KloV: Klosneuvirus (KlosnV), IndV: Indivirus, CroV: Cafeteria roenbergensis virus.
The genome of BsV is rich in coding sequences involved in transcription. An early transcription factor putatively recognizing the highly conserved AAAAAATTGA motif and a late transcription factor putatively targeting TGCG were identified; whereas, the target sequence of a third transcription factor is unknown. Further, a TATA-binding protein, a transcription initiation factor (TFIIB) and a transcription elongation factor (TFIIS) were identified that should aid transcription. As well, RNA polymerase subunits a,b,c,e,f,g and I were identified and are assisted by DNA topoisomerases Type 2 and 1B. BsV encodes a mRNA specific RNase III, a poly A polymerase, several 5’ capping enzymes and methyl transferases. Transcription is presumably terminated in a manner similar to that described in ApMV, since hairpin structures were detected in the 3’ UTR of most transcripts [150]. They are probably recognized and processed by the viral encoded RNase III in a manner similar to ApMV [151]. After hairpin loop cleavage, the poly-A tail is added by the virally encoded poly-A polymerase. The 5’ capping is accomplished by the virus-encoded mRNA capping enzyme, as well as several cap-specific methyltransferases. The extensive cap modification suggests that BsV is independent of the trans-splicing of splice-leader mRNA containing cap structures found in kinetoplastids [152].

BsV also encodes several enzymes associated with nucleic-acid transport and metabolism, including several AT-specific nucleic-acid synthesis pathway components. For instance, adenylosuccinate, thymidylate and pseudouridine synthetases and kinases, as well as ribonucleoside-diphosphate reductase were evident. Other ORFs were associated with nucleotide salvaging pathways, including nucleoside kinases, phosphoribosyl transferases, and cytidine and deoxycytidylate deaminase. A mitochondrial carrier protein was identified that, similar to ApMV, likely provides dATP and dTTP directly from the kinetoplast to the virus factory, as evident from electron microscopic observations [153].

Several genes were identified that are involved in membrane trafficking. A system based on soluble N-ethylmaleimide-sensitive factor (NSF) attachment proteins (SNAPs) and the SNAP receptors (SNAREs) appears to have been acquired from the host by horizontal gene transfer in the recent evolutionary past. In combination with several NSF homologues, including the vesicular-fusion ATPases that also seems to have been acquired from the host. Other proteins putatively involved in membrane trafficking are rab-domain containing proteins, ras-like GTPases, and kinesin motor proteins.
The BsV genome encodes four major capsid proteins. One of these proteins contains several large insertions between conserved domains shared among all four capsid proteins, and with 4194 aa boasts a size almost seven times that of its paralogs. This enlarged version of the major capsid protein might be responsible for creating the halo around the virus particles observed by TEM, by producing shortened fibers similar to those observed in ApMV (Figure 3-2A) [154]. Further, the genome contains two core proteins, several chaperones and glycosylation enzymes suggesting that proteins are highly modified before being incorporated into the virus particle. There were numerous ORFs that were similar to genes encoding metabolic proteins, like enzymes putatively involved in carbohydrate metabolism. However, no one continuous metabolic pathway could be assembled and therefore these enzymes likely complement host pathways. BsV also encodes coenzyme synthetases such as CoA and NADH and to meet the demand for amino acids that are rare in the host, BsV encodes the key steps in the synthesis pathways of glutamine, histidine, isoleucine, and asparagine.

Another group of genes putatively mediate competitive interactions, either directly with the host, or with other viruses or intracellular pathogens. These include genes involved in the production of several toxins such as a VIP2-like protein as well as antitoxins containing BRO domains. Further, a partial bleomycin detox pathway was found, as well as multidrug export pumps and partial restriction modification systems.

While BsV encodes a partial translation machinery, it differs markedly from those described in other members of the Mimiviridae and shows the largest turnover of these genes resulting in a net contraction (Fig 4C). Eukaryotic translation initiation factors include the commonly seen eIF-2a, eIF-2b, eIF-2g, eIF-4A-III and eIF-4E, as well as several pseudogenes related to eIFs. Eukaryotic elongation factor 1 is also present as is eukaryotic peptide chain release factor subunit 1. Notable is the absence of eIF-1; instead, BsV encodes a version of IF-2 that appears to have been acquired from the host and is functionally analogous to eIF-1 in kinetoplastids. The most striking difference to other NCLDV's is the absence of tRNAs. Uniquely among NCLDV's, BsV encodes several tRNA repair genes. These genes include putative RtcB-like RNA-splicing ligase, putative CAA-nucleotidyltransferase, tRNA 2'-phosphotransferase/Ap4A_hydrolase, putative methyltransferase, a TRM13-like protein, pseudouridine synthase and tRNA ligase/uridine kinase. Most of these genes appear to have been recently acquired from the host (Supplementary Table 1). Other translation modification enzymes
found in BsV and other NCLDVs include tRNA(Ile)-lysidine synthase, tRNA pseudouridine synthase B and tRNA 2'-phosphotransferase. Similar to the tRNAs, there are few aminoacyl-tRNA synthetases (aaRS) in BsV. Three of the recognizable aaRS are pseudogenes and show signs of recent nonsense mutations or ORF disruptions by genome rearrangements (aspRS, glnRS, and asnRS). The only complete aaRS proteins are isoleucine-tRNA synthetase, found in all members of the Mimiviridae, and a phenylalanyl-tRNA synthetase.

3.4.3.3 Repeat regions.

Genes in the genomic periphery have undergone massive duplication, with 148 copies of ankyrin-repeat proteins, mostly present in directional tandem orientation (Figure 3-3 Circularized genome plot of the linear BsV genome.). These sequences are quite variable and encode between four and 17 ankyrin-repeat domains. There is evidence of very recent sequence duplication resulting in direct or inverted repeat regions that contain complete ankyrin-repeat coding sequences and further expand the repeat clusters (Supplementary Figure 3A). Interestingly, the 5’ coding region of many ankyrin-repeat containing protein ORFs contain fragments of catalytic domains of essential viral genes such as DNA polymerases or the MutS repair protein (Supplementary Figure 3).

3.4.3.4 Genomic Mobilome.

In contrast to described giant virus genomic mobilomes consisting of virophages and transpovirons, the BsV genome is dominated by inteins, autocatalytic proteinases, and self-splicing group 1 introns [29, 105, 155, 156]. These mobile elements spread by targeting the DNA coding regions of essential genes for virus replication by deploying unrelated homing endonucleases encoded by internal ORFs nested within the elements (red ORFs in Figure 3-5). Inteins that are closely related to those in Mimiviridae and Phycodnaviridae are found in the BsV DNA polymerase family B gene, while other unrelated inteins are found in the DNA dependent RNA polymerase subunits A and B genes (polr2a and polr2b: Figure 3-5). The inteins in the RNA polymerase genes seem to be devoid of an active homing endonuclease, and are therefore fixed, suggesting an evolutionary ancient invasion. The intein in DNA pol B may be an exception, as an HNH endonuclease is located in close genomic proximity and could promote homing in a trans-acting fashion.

The group 1 self-splicing introns seem to have independently invaded the RNA polymerase subunit 1 and 2 genes, since these introns carry different homing endonucleases (HNH and GIY-YIG type) and their ribozymes differ in secondary structure (Figure 3-5,
Supplementary Figure 6B). Subsequent to the initial integration of introns containing endonucleases (e.g. polr2a-i1b in Figure 3-5), these homing endonucleases seeded “offspring” introns within the same gene (e.g. polr2a-i1a and –i1c in Figure 3-5). These secondary introns show conserved secondary RNA structure, but lack the homing endonuclease of their parental intron. Therefore, the secondary introns probably rely on the trans-homing of their parental introns’ endonucleases. The highly conserved sequence for some of the offspring introns (94.4% sequence identity for polr2a-i1a and -i1c, Supplementary Figure 6A) suggests that these have spread relatively recently, while other introns that only show conservation in their secondary structure probably represent older invasions. Besides proliferating introns, the BsV genome is also home to two distinct actively proliferating transposon classes.
**Figure 3-5 Invasion of RNA polymerase genes by selfish genetic elements.** Organization of self-splicing group 1 introns (grey) and inteins (black) within genes polr2a and polr2b (green). ORFs within the introns encode homing endonucleases (red). The precursor proteins (orange) are expressed after introns are excised from the pre-mRNA including the internal ORFs encoding the endonucleases. Consequently, the inteins (grey) excise themselves from the precursor protein resulting in the mature protein (yellow). Secondary structure of related self-splicing group 1 introns (“parental” intron polr2a-i1b and “offspring” polr2a-i1a and –i1c) is shown above the coding sequence with conserved self-splicing catalytic site highlighted by arrows. Additional secondary structure predictions and sequence alignment are available in Figure 3.10.

### 3.4.4 Phylogenetic placement and environmental representation of BsV.

Phylogenetic analysis of BsV places it within the Mimiviridae. Whole genome analysis based on gene cluster presence/absence of NCLDVs resulted in BsV clustering within the large mimiviruses (recently proposed “Megavirinae”) and separate from the small mimiviruses (“Mesomimivirinae”) (Figure 3-6A) [107]. BsV is closest affiliated with the genomes of the
Klosneuviruses, metagenomically assembled genomes (MAGs), and to a lesser degree with CroV. Phylogenetic analysis of five concatenated highly conserved NCLDV core genes reproduced this pattern within the Mimiviridae (Figure 3-6B). Within the “Megavirinae”, three clades emerged, the Acanthamoeba-infecting Mimiviruses, the klosneuvirus MAGs (Cato-, Hoko-, Klosneu-, and Indivirus) with BsV, and CroV as the sole member of an outgroup (Figure 3-6B). Phylogenies of the individual genes placed BsV within the klosneuviruses in three of five cases (Supplementary Figure 8, Supplementary Figure 9). When metagenomic reads of NCLDV DNA polymerase B sequences from the TARA oceans project were mapped to a maximum likelihood tree of DNA polymerase B sequences from the Mimiviridae, it was apparent that the “Megavirinae” were dominated by klosneuviruses, with BsV as their only isolate (Figure 3-6C). Thus, BsV is representative of the largest group of identifiable icosahedral giant viruses in the oceans with CroV being the sole representative of the second most abundant clade.
Figure 3-6 BsV Phylogeny. A: Phylogenetic distance matrix of NCLDV genomes based on whole genome content of gene family presence/absence. Both axes are identical and “heat” of the color increases with dissimilarity in genome content. The shaded “taxa” on the axes indicate viral sequences assembled from metagenomic data. B: Bayesian posterior consensus tree with Bayesian posterior probability of five concatenated Nucleo-Cytoplasmic Virus Orthologous Groups (NCVOGs) from selected NCLDVs based on two independent MCMC chains (16100 generations rel_diff: 0.104001 effsize: 297). The shaded labels at the end of the branches represent “taxa” based on sequences assembled from metagenomic data C: Maximum likelihood phylogenetic tree of DNA Polymerase family B of BsV within the Mimiviridae. Branch width correlated to the distribution of 256 metagenomic sequences identified as NCLDV DNA polymerase Family B genes from the TARA oceans project recruited to the tree with pplacer.

AaV: Aureococcus anophagefferens virus; AcV: Anomala cumrea entomopoxivirus; AMaV: Acanthamoeba castellanii mamavirus; AMgV: Mounouivirus goulette; AMoV: Acanthamoeba polyphaga moumouvir; ApMV: Acanthamoeba polyphaga mimivirus; ASFV: African swine fever virus BA71V; AtcV: Acanthocystis turfacea chlorella virus 1; BpV: Bathycoccus sp. RCC1105 Virus; BsV: Bodo saltans virus NG1; CatV: Catovirus; CeV: Chrysochromulina ericina virus 1B; CroV: Cafeteria roenbergensis virus BV-PW1; EhV: Emiliana huxleyi virus 86; FauV: Faustovirus E12; HokV: Hokovirus; HvV: Heliothis
virescens ascovirus 3e; IiV: Invertebrate iridescent virus; IndV: Indivirus; ISKV: Infectious spleen and kidney necrosis virus; KloV: Klosneuvirus; LauV: Lausannevirus; MarV: Marseillevirus T19; MoVs: Mollivirus sibericum; MpVS: Micromoas pusillae Virus SP-1; MsV: Melanoplus sanguinipess entomopoxvirus; MVc: Megavirus chilensis; MyxV: Myxoma virus; OLV1: Organic Lake Phycodnavirus 1; OLV2: Organic Lake Phycodnavirus 2; OtV: Ostreococcus tauri virus 1; PbCV: Paramecium bursaria chlorella virus 1; PgV: Phaeocystis globosa virus 16T; PiVs: Pithovirus sibericum P1084-T; PoV: Pyramimonas orientalis virus; PpV: Phaeocystis pouchetii virus; PVd: Pandoravirus dulcis; PVs: Pandoravirus salinus; SfV: Spodoptera frugipera ascovirus 1a; SGV: Singapore grouper iridovirus; TnV: Trichoplusia ni ascovirus 2c; VacV: Vaccinia virus; WiV: Wiseana iridescent virus; YLV1: Yellowstone lake phycodnavirus 1
3.5 Discussion

3.5.1 BsV represents the most abundant subfamily of the Mimiviridae in the TARA oceans data

Particle structure, functional features, like the transcription machinery, and phylogenetic analysis firmly place BsV within the Mimiviridae, making it the largest completely sequenced genome of the family. BsV groups with the klosneuviruses and is separate from CroV and the Acanthamoeba infecting mimiviruses, and falls within the proposed “Megavirinae” (Figure 3-6). A separate subfamily was proposed for the metagenomic klosneuviruses, the “Klosneuvirinae”, which would make BsV as the first isolate and the type species of this subfamily [71]. The high representation of the klosneuviruses and BsV-like viruses in metagenomic reads from the TARA oceans survey is consistent with BsV being representative of the most abundant group of Mimiviridae, and possibly all icosahedral giant viruses in the oceans [26]. The initial detection of the klosneuviruses in low complexity fresh water metagenomes further supports the global prevalence of this group.

3.5.2 BsV has acquired a host mechanism to facilitate membrane fusion, presumably employed during infection and virion morphogenesis

The SNAP/SNARE membrane fusion system found in BsV appears to have been recently acquired from the bodonid host via horizontal gene transfer. This system could mediate membrane fusion in a pH-dependent manner [157]. Accordingly, a phagocytosis based infection strategy is proposed for BsV: As described for ApMV, BsV is ingested through the cytostome and is phagocytosed in the cytopharynx before being transported in a phagosome towards the posterior of the cell [148]; here the viral SNAP/SNARE interacts with the host counterparts to initiate the fusion of the inner virus membrane with the phagolysosomal membrane upon phagolysosome acidification, releasing the viral genome into the cytoplasm. This scenario is supported by the localization of the virus factory at the posterior of the cell and virus particle structure (Figure 3-2A,C and Supplementary Figure 2B-C). According to this hypothesis, SNAP/SNARE proteins must be present in membranes of the mature virus particles and only get exposed after the stargate opens. The SNAP/SNARE system might also be involved in recruiting membrane vesicles from host organelles to the virus factory during maturation of the virus particle as has been described for pox viruses (Figure 3-2D).
3.5.3 The BsV possesses an arsenal of mechanisms that could be involved in interference competition

As a representative of environmentally highly abundant viruses, BsV might regularly experience competition for host resources. The putative toxin-antitoxin systems observed in BsV might be involved in competing with other parasites of viral or prokaryotic nature for these resources, by inhibiting their metabolism or damaging their genome as proposed for ApMV [158]. Most remarkable, however, are the site-specific homing endonucleases encoded by the self-splicing group 1 introns and inteins that have invaded several genes essential for BsV replication. These invasions also seem to be part of the competitive arsenal of BsV, fending off related virus strains competing for abundant and common hosts such as bodonids. During superinfection of two related viruses, having selfish elements encoding homing endonucleases targeting essential genes might be a competitive advantage. As the two competing virus factories are established in the cytoplasm, the endonucleases encoded within the intron or intein cleave the unoccupied locus in the genome of the intron/intein-free virus. The intron or intein containing virus’ genome stays intact, since the target sequences of the endonucleases within its genes are masked by the insertion of the intron or intein. Thus, the intron/intein-containing virus is reducing the ability of the competing virus to replicate (Figure 3-7). A similar mechanism has been described for competing phages in which an intron-encoded or derived homing endonuclease mediates marker exclusion during superinfection, causing selective sweeps of genes in the vicinity of the endonuclease through the phage population [36, 159, 160]. More credence to this hypothesis is given by the RNA polymerase sequences encoded by the proposed catovirus and klosneuvirus [71]. These genes (polr2a and polr2b) are fragmented in a manner similar to that observed in BsV, and also appear to encode homing endonucleases between gene fragments, suggesting the presence of self-splicing introns is common in the relatives of BsV such as the klosneuviruses. Since the hosts for klosneuviruses are unknown, it is possible that they compete with BsV for the same hosts, or at least experience similar competition. The presence of non-fixed inteins in other giant viruses hints to past invasions and selection for inteins in a manner resembling the hypothesis proposed here for introns [107, 161]. The retention of fixed inteins in BsV and other giant viruses suggests that there are still viruses in the environment that encode the relevant endonucleases that apply selective pressure to retain the inteins. A similar situation might explain the presence of an intein in the DNA polymerase of Pandoravirus salinus, but not in P. dulcis. Thus, pandoraviruses may be an
excellent model to experimentally explore the proposed mechanism of intron homing endonuclease mediated competition.

Figure 3-7 Intron/intein encoded endonuclease mediated interference completion between related viruses. 1) Two related viruses infect the same host cell. The green virus genome contains a selfish element encoding a homing endonuclease. 2) During initial replication, the endonuclease is expressed and cleaves the unoccupied locus on the yellow virus’ genome impairing its replication. 3) Due to suppressing its competitor’s replication, the majority of the viral progeny is of the green virus’ type. 4) The yellow virus can rescue its genome by using the green virus’ genome as a template. This creates a chimeric genome containing the selfish element and the endonuclease as well as adjacent sequences originating from the green virus’ genome.

3.5.4 The contracted translation machinery in BsV suggest that this feature is a homoplastic trait of giant viruses

The absence of tRNAs in the BsV genome is remarkable since tRNAs are found in all complete genomes of giant viruses and even in many moderately sized NCLDV genomes (Figure 3-4C, Supplementary Figure 4). This might be an adaptation to the unusual RNA modification found in kinetoplastids that also encompasses tRNA editing [152, 162]. Similarly, Trypanosoma
mitochondrial tRNAs are exclusively nuclear encoded [163]. BsV likely cannot replicate this unusual editing, and thus relies on using host tRNAs. Hence, BsV encodes tRNA repair genes to compensate for the lack of tRNA synthesis and to maintain the available tRNA pool in the host cell. Most of these genes appear to have been recently acquired from the host (Figure 3-4C, Supplementary Figure 4). Like the tRNAs, most virus-encoded aminoacyl-tRNA synthetases might not recognize the highly modified tRNAs present in the host and are therefore degrading in the absence of positive selective pressure (Figure 3-4C, Supplementary Figure 4, Supplementary Table 1). The presence of three recognizable pseudogenes of aminoacyl-tRNA synthetases is especially remarkable in this context and argues for an evolutionary recent degradation (Supplementary Table 1). This turnover in translational machinery components in BsV and the klosneuviruses, combined with the apparent diverse origin of these genes, suggests that the translation machinery found in giant viruses is the result of rapid adaptation by gene acquisition via horizontal gene transfer, as has been recently proposed by Schultz et al. [71]. BsV demonstrates that such genes can be readily purged from the virus genome if they are not required in a new host. Thus, BsV provides further evidence that the translation machinery encoded by NCLDVs is a homoplastic trait and need not be ancient in origin.

3.5.5 An inflated genomic accordion due to evolutionary arms races is responsible for genome gigantism in BsV and sheds light on giant virus evolution

The 148 copies of ankyrin-repeat domain proteins in the genomic periphery of BsV are telltale signs of an expanded genomic accordion (Figure 3-3) [164]. The observation of almost identical sequences in the very periphery of the genome is consistent with the genomic accordion hypothesis, in which the most recent duplications are closest to the genome ends (Figure 3-3, Supplementary Figure 3A,B). The genomic recombinations causing the gene duplications can also lead to the disruption of coding sequences that might explain the comparatively low coding density of BsV. Proteins with ankyrin-repeat domains are multifunctional attachment proteins that in pox viruses determine host range by inhibiting host innate immune system functions [165]. Further ankyrin-repeat domain proteins are used by bacterial intracellular pathogens like Legionella to manipulate eukaryotic host cells [166]. The presence of fragments of the catalytic domains of essential viral genes in many ankyrin-repeat containing genes is of further importance (Supplementary Figure 3C). This suggests a decoy defense mechanism, where these fusion proteins mimic the targets of host anti-viral defense systems disrupting essential viral functions.
By acting as decoy targets, they immobilize the proposed host factors upon binding via their ankyrin-repeat domains similar to what occurs in vaccinia virus [167]. The immobilized host factors might even be degraded in a ubiquitin dependent manner reminiscent of the situation in pox viruses as suggested by the presence of several ubiquitin conjugating enzymes encoded in the BsV genome [168]. An ankyrin-repeat based defense system might explain the observation of cells surviving or avoiding infection that can persist in the presence of the virus (Figure 3-1C). Alternatively, the protein-protein interaction mode of ankyrin-repeat proteins might aid attachment and induction of phagocytosis as the bodonid host cells have changing surface antigens [129]. Whatever the true function of the ankyrin-repeat proteins might be, they clearly highlight the importance of the genomic accordion in giant virus genome evolution driven by evolutionary arms races and complement previous observations of a contracting genomic accordion in ApMV [158].

3.5.6 Summary

Bodo saltans virus (BsV) has the largest sequenced genome for any isolate of the Mimiviridae, and is representative of the most abundant members of this family in aquatic ecosystems. BsV is also the first described DNA virus that infects kinetoplasts, or any member of the supergroup Excavata, a major evolutionary lineage of eukaryotes, and is the first isolate of a subfamily within the Mimiviridae that was previously based only on metagenomic data. BsV highlights the genomic plasticity of giant viruses via the genomic accordion, which allows for large-scale genome expansions and contractions via non-homologous recombination. The recent duplications in BsV demonstrate genome expansion in action and exemplifies the mechanisms leading to genome gigantism in the Mimiviridae. Further, the function of the expanding genes suggests that strong evolutionary pressure is placed on these viruses by a virus-host arms race that has driven genomic expansion. Genomic plasticity is further apparent in the translational machinery, which shows signs of recent gene loss and rapid adaptation to its bodonid host. This emphasizes that the translational machinery of giant viruses is an acquired homoplastic trait not derived from a common ancestor. An invasion of selfish elements in essential genes suggests interference competition among related viruses for shared hosts. Bodo saltans virus provides new insights into giant virus diversity and their biology.
Chapter 4

4 Chromulinavorax destructans, a pathogenic TM6 bacterium with an unusual replication strategy targeting protist mitochondrion

4.1 Synopsis

Most of the diversity of microbial life is not available in culture, and as such we lack even a fundamental knowledge of the biological diversity of several branches on the tree of life. One branch that is highly underrepresented is the bacterial candidate phylum TM6, also known as the Dependentiae. Their biology is known only from reduced genomes recovered from metagenomes around the world and two isolates infecting amoebae, all suggest that they live highly host-associated lifestyles as parasites or symbionts. Chromulinavorax destructans is an isolate from the TM6/Dependentiae that infects and lyses the abundant hetero-trophic flagellate, Spumella elongata. Chromulinavorax destructans is characterized by a high degree of reduction and specialization for infection, so much so it was discovered in a screen for giant viruses. Its 1.2 Mbp genome shows no metabolic potential and C. destructans instead relies on extensive transporter system to import nutrients, and even energy in the form of ATP, from the host. Accordingly, it replicates in a viral-like fashion, while extensively reorganizing and expanding the host mitochondrion. 44% of proteins contain signal sequences for secretion, which includes many proteins of unknown function as well as 98 copies of ankyrin-repeat domain proteins, known effectors of host modulation, suggesting the presence of an extensive host-manipulation apparatus.

4.2 Introduction

The vast majority of genetic diversity on earth is encoded by microorganisms, most of which are not represented in culture and are only known from barcoding or metagenomic studies [169]. As a result, there are many major branches in the tree of life where we know little or sometimes nothing at all about their biology. One such case is the candidate phylum TM6, also known as Dependentiae, which was originally described from a 16S environmental amplicon library prepared from a peat sample collected in northern Germany [170]. Later, low complexity metagenomic data from a hospital biofilm, soil, and waste water allowed for the assembly of almost complete genomes from representatives of the TM6 phylum [33, 108]. These genomes implied very limited metabolic capability, while encoding an extensive system of transporters,
including ATP transporters, suggesting that TM6 bacteria may be symbionts or parasites that rely extensively on their hosts for energy and metabolites [33, 108]. The genomic complement, enriched in conserved gene functions found in endosymbionts or parasites of eukaryotes such as Chlamydia, Wolbachia, and Rickettsia, further suggested that TM6 was associated with eukaryotic hosts [33, 108]. Indeed, two representatives of the TM6 phylum were since isolated in amoebae, competent hosts to a vast array of bacterial, eukaryotic, and viral symbionts and pathogens [34, 109]. The first isolate, Babela massiliensis, is an obligate pathogen that causes lysis of its host, Acanthamoeba castellanii, while the other, Vermiphilus pyriformis, maintains a stable relationship with its host Vermamoeba vermiformis, suggesting that different TM6 bacteria might employ different life history strategies [34, 109]. Since the genome of Babela massiliensis is the only available complete sequence, a detailed comparison of the genomic complement involved in the different lifestyles is not possible. Both isolates show unusual replication strategies within their hosts that delay cell fission and initially produce large replication bodies before dividing into progeny cells. Recent advances in metagenomics and single-cell genomics have added further almost-complete genomes to the proposed candidate phylum that seem to be distinct from the amoeba-infecting species; however, the interpretation of their genomic complement remains speculative with only amoeba-infecting isolates reported to be in culture [171-173].

Amoebae are probably the best-studied microbial eukaryotic hosts of bacterial symbionts and pathogens, but represent only a fraction of eukaryotic diversity. Another large and paraphyletic group of microbial eukaryotes that are abundant in most natural environments are heterotrophic nanoflagellates, defined by being between 2 and 20 μm in size, and by their lifestyle of preying on bacteria, viruses, and other microbial eukaryotes [16, 174]. They are important in aquatic ecosystems, as the component of the microbial loop that most directly links to higher trophic levels. Grazing by nanoflagellates, together with viral lysis, are the primary mortality agents of microbial populations in aquatic environments [175]. Chrysophytes are a diverse group of nanoflagellates within the stramenopiles, and include a number of mixotrophs in the genera Ochromonas and Chromulina, for example, as well as pure heterotrophs, such as members of the genus Spumella [12, 176]. Present in fresh and salt waters, Spumella is one of Earth’s most abundant phagotrophic nanoflagellates and is readily isolated from many environments [177].

Here, Chromulinavorax destructans, a novel isolate of the TM6/Dependentiae, is described which infects and lyses Spumella elongata. Its unusual replication cycle superficially resembles
the infection cycle of a virus, and includes the creation of a replication ‘factory’ and the
remodelling of the host mitochondrion within its host cell. The genome is highly reduced and lacks
nearly all genes related to energy metabolism. Moreover, like metagenomically assembled
genomes (MAGs) of other TM6, it encodes a large suite of transporters, including ATP
transporters.

4.3 Materials and methods

4.3.1 Sampling

Samples were collected from 11 freshwater locations in southern British Columbia, Canada
(49°49'4"N, 123° 7'46"W; 49°42'5"N, 123° 8'47"W; 49°37'34"N, 123°12'27"W; 49° 6'12"N, 122°
4'38"W; 49° 5'22"N, 122° 7'1"W; 49°18'10"N, 122°42'9"W; 49° 8'27"N; 123° 3'16"W;
49°13'21"N, 123°12'43"W; 49°13'13"N, 123°12'41"W; 49°14'52"N, 123°13'59"W; 49°15'58"N,
123°15'34"W. See Chapter 2). To concentrate pathogens, 20 liter water samples were prefiltered
with a GF-A filter (Millipore, Bedford, MA, USA; nominal pore size 1.1 μm) over a 0.8- μm pore-
size PES membrane (Sterlitech, Kent, WA, USA) [73]. Filtrates from all locations were pooled
and concentrated using a 30kDa MW cut-off tangential flow ultrafiltration cartridge (Millipore,
Bedford, MA, USA) [115].

4.3.2 Isolation

The host organism, Spumella elongata strain CCAP 955/1 was kindly provided by David
Caron (University of Southern California) and maintained in modified DY-V artificial fresh water
media with yeast extract and a wheat grain [110]. Cultures of S. elongata at approximately 2x10^5
cells/ml were inoculated with the pooled microbial concentrates from all 11 locations. Cell
numbers of the inoculated culture were monitored by flow cytometry and compared to a medium-
only mock-infected control culture using flow cytometry (LysoTracker Green (Molecular Probes)
vs. FSC on FACScalibur (Becton-Dickinson, Franklin Lakes, New Jersey, USA)) [116]. After cell
lysis, the lysate was filtered through a 0.8-μm pore-size PES membrane (Sterlitech) to remove
remaining host cells. The lytic agent was propagated and made clonal by three serial end-point
dilutions. The concentrations of the lytic agent were screened by flow cytometry using SYBR
Green (Invitrogen Carlsbad, California, USA) nucleic-acid stain after 2% glutaraldehyde fixation
(vs SSC). The flow cytometry profile presented as a population clearly distinct from heterotrophic
bacteria, phage, and eukaryotes; however, their larger size heterogeneity when compared to giant
virus isolates suggested a non-viral nature [178].
4.3.3 Transmission electron microscopy

4.3.3.1 Negative staining

Lysates of *Spumella elongata* after infection with *Chromulina vorax destructans* were applied to the carbon side of a formvar carbon-coated 400-mesh copper grids (TedPella, CA, USA) and incubated at 4°C in the dark overnight in the presence of high humidity. Next, the lysate was removed and the grids were stained with 1% Uranyl acetate for 30 s.

4.3.3.2 Ultra-thin sectioning

Exponentially growing cultures of *S. elongata* at a concentration of 5x10^5 cells ml^-1 were infected with *C. destructans* at a ratio of ~5 pathogen to host cells to ensure synchronous infection. Cells were harvested from infected cultures at 3, 6, 9, 12, 18, and 24 h post infection, as well as from uninfected control cultures. Cells from 50 ml were pelleted in two consecutive 10 min at 5000 xg centrifugation runs in a fixed angle Beckmann tabletop centrifuge.

For chemical fixation, the pellet was resuspended in 0.2 M Na cacodylate buffer, 0.2 M sucrose, 5% EM grade glutaraldehyde, pH 7.4 and incubated for 2 h on ice. After washing in 0.2 M Na cacodylate buffer, cells were postfixed with 1% osmium tetroxide. Samples were dehydrated through water/ethanol gradients and ethanol was substituted by acetone. An equal part mixture of Spurr’s and Gembed resin was used to embed the cells, which was polymerized at 60°C overnight. For high pressure freezing, cell pellets were resuspended in 10-15 µl of DY-V culture medium with 20% (w/v) BSA and immediately placed on ice. Cell suspensions were cryo-preserved using a Leica EM HPM100 high-pressure freezer. Vitrified samples were freeze-substituted in a Leica AFS system for 2 d at -85°C in a 0.5% glutaraldehyde / 0.1% tannic acid solution in acetone, then rinsed ten times in 100% acetone at -85°C, and transferred to 1% osmium tetroxide, 0.1% uranyl acetate in acetone and stored for an additional 2 d at -85°C. The samples were then warmed to -20°C for 10 h, held at -20°C for 6 h to facilitate osmication, and then warmed to 4°C for 12 h. The samples were then rinsed in 100% acetone 3X at room temperature and gradually infiltrated with an equal part mixture of Spurr’s and Gembed embedding media. Samples were polymerized in a 60°C oven overnight. Fifty nm thin sections were prepared using a Diatome ultra 45° knife (Diatome, Switzerland) on an ultra-microtome. The sections were collected on a 400x copper grid and stained for 10 min in 2% aqueous uranyl acetate and 5 min in Reynold’s lead citrate. Image data were recorded on a Hitachi H7600 transmission electron microscope at 80 kV. Image J
(RRID:SCR_003070) was used to compile all TEM images. Adjustments to contrast and brightness levels were applied equally to all parts of the image.

4.3.4 Pathogen concentration and sequencing

For PacBio sequencing, exponentially growing *S. elongata* cultures at a concentration of approximately 5x10^5 cells ml^-1 were infected with *C. destructans* lysate (~10^7 cells ml^-1) at a multiplicity of infection (MOI) of ~0.5. After four days, when host cells were undetectable, cultures were centrifuged in a Sorvall SLC-6000 centrifuge with fixed angle rotor for 20 min and 5000 rpm at 4°C to remove remaining host cells and the supernatant was subjected to tangential flow filtration with at 30kDa cut-off (Vivaflow PES) and concentrated approximately 100x. Concentrates were ultracentrifuged at 28,000 rpm, 15°C for 8h in a Ti90 fixed-angle rotor (Beckman-Coulter, Brea, California, USA). Next, the concentrate was further concentrated by sedimenting it onto a 40% Optiprep 50 mM Tris-Cl, pH 8.0, 2mM MgCl$_2$ cushion for 30 min at 28,000 rpm and 15°C in a SW40Ti swinging-bucket rotor in an ultracentrifuge (Beckman-Coulter, Brea, California, USA). An Optiprep (Sigma) gradient was created by underlaying a 10% Optiprep solution in 50 mM Tris-Cl, pH 8.0, 2 mM MgCl$_2$ with a 30% Optiprep solution followed by a 50% Optoprep solution and was equilibration overnight at 4°C. One ml of concentrate from the 40% cushion was added atop the gradient and the concentrate was fractionated by ultracentrifugation in an SW40 rotor for 4 h at 25000 rpm and 18°C. The fraction corresponding to the pathogen was extracted from the gradient with a syringe and washed twice with 50 mM Tris-Cl, pH 8.0, 2 mM MgCl$_2$ followed by centrifugation in an SW40 rotor for 20 min at 7200 rpm and 18°C and finally collected by centrifugation in an SW40 rotor for 30 min at 7800 rpm and 18°C. Purity of the concentrate was verified by flow cytometry (SYBR Green (Invitrogen Carlsbad, California, USA) vs SSC on a FACScalibur flow cytometer (Becton-Dickinson, Franklin Lakes, New Jersey, USA). High molecular weight genomic DNA was extracted using phenol-chloroform-chloroform extraction. Length and purity of the DNA were confirmed by gel electrophoresis and a Bioanalyzer 2100 with the HS DNA kit (Agilent Technology). PacBio RSII 20kb sequencing was performed by the sequencing center of the University of Delaware. Reads were assembled using PacBio HGAP3 software with 20 kb seed reads and resulted in a single contig of 1,228,924bp, 819.19x coverage, 100% called bases and a consensus concordance of 99.9839% [132].
4.3.5 Annotation

The genome was circularized, resulting in a final chromosome size of 1,174,272 bp. Genome annotation was performed using the automated NCBI Prokaryotic Genome Annotation Pipeline (PGAAP). In parallel, open reading frames were predicted using GLIMMER (RRID:SCR_011931) with default settings [179]. Translated proteins were analyzed using BLASTp, CDD RPS-BLAST and pfam HMMER. These results were used to refine the PGAAP annotation. Signal peptides and trans-membrane domains were predicted using Phobius [180]. The annotated genome is available under the accession number CP025544. Metabolic pathways were predicted using the Kyoto Encyclopedia of Genes and Genomes (KEGG RRID:SCR_012773) automatic annotation server KAAS and Pathway Tools (RRID:SCR_013786) [181, 182].

4.3.6 Phylogenetic analysis

Full length 16S rDNA sequences belonging to the candidate phylum TM6 were downloaded from NCBI. Alignments of rDNA sequences were done in Geneious R9 (RRID:SCR_010519) using MUSCLE with default parameters (RRID:SCR_011812)[139]. Maximum likelihood trees were constructed with RAxML ML search with 1000 rapid bootstraps using GTR+GAMMA [141].

Near-complete metagenomically assembled genomes (MAGs) and complete genomes of isolates were retrieved from NCBI. There were 3820 clusters of orthologous genes defined by OrthoMCL (RRID:SCR_007839) with standard parameters (Blast E-value cut-off = $10^{-5}$ and mcl inflation factor = 1.5) on all protein-coding genes of length ≥ 100 aa [136]. Overlap in gene content was defined using a custom R script (see Deeg et al. 2018 [178]). Genes encoding ribosomal proteins were identified by BLAST. Ribosomal protein L2, L3, L4, L5, L6, L14, L15, L18, L16, L22, L24, S3, S8, S10, S17, and S19 sequences were aligned using MUSCLE with default parameters (RRID:SCR_011812) [139]. Maximum-likelihood trees were constructed with RAxML rapid bootstrapping and ML search with 1000 bootstraps utilizing GTR+CAT [141].
4.4 Results

4.4.1 Chromulinavorax destructans is a lytic pathogen of Spumella elongata

In an effort to isolate pathogens that infect ecologically relevant protistan zooplankton, microbial assemblages smaller than 0.8 μm from a variety of freshwater habitats in southwestern British Columbia were concentrated (Chapter 2). The concentrates were used to inoculate cultures of protists isolated from the same environments, as well as from culture collections, and cell densities were monitored by flow cytometry. One such screen yielded a pathogen infecting Spumella elongata (CCAP 955/1) that caused complete lysis of the culture within 48 h post infection (hpi). The lytic agent was named Chromulinavorax destructans (subsequently referred to as C. destructans). Infection was strain-specific and a different isolate of Spumella sp. from one of the environments sampled for pathogen concentration did not cause lysis. The pathogen continued to cause lysis of S. elongata after storage at 4°C for up to four years. Further, pre-treatment of the pathogen with a bacteriocidal multi-antibiotic cocktail containing Ampicillin, Vancomycin, and Rifampicin also did not ablate the lytic potential upon inoculation of Spumella elongata cultures. Under epifluorescence microscopy, infected cells developed large DNA-containing intracytoplasmic compartments starting at 8 hpi, followed by lysis starting at 18 hpi that was associated with numerous DNA-containing particles of approximately 0.4 μm bursting out of the cytoplasm (Figure 4-1B). Chromulinavorax destructans cells were first observed by flow cytometry 12 hpi (Figure 4-2 B), coinciding with a decline in host-cell density (Figure 4-2A). The population of C. destructans showed a homogeneous fluorescence signature suggesting that DNA replication does not occur outside of the host (Figure 4-2B). Side-scatter heterogeneity suggests that the cells vary in size, or are clumped (Figure 4-2B). In negative-staining electron microscopy, C. destructans cells present as 350-400 nm cocci with one polymorphous depression, similar to Babela massiliensis (Figure 4-1C, Figure 4-2C). [34]. In thin-section electron microscopy, free C. destructans cells show a gram-negative-like coccoid morphology with a lipid double layer and electron-dense material in the periplasm (Figure 4-1D). The central part of the cytoplasm shows an electron-dense nucleoid. No depression similar to that seen in negative staining was observed in thin sections, suggesting that this is a preparation artefact of negative staining.
Figure 4-1: Micrographs of Chromulinavorax destructans. A: Epifluorescence micrograph of DAPI-stained Spumella elongata (prominently-stained nucleus). B: Epifluorescence micrograph of two S. elongata cells 19 h after exposure to C. destructans. Coccoid cells of C. destructans are seen bursting from a cell undergoing lysis. C: Purified C. destructans cell in negative-staining electron micrographs show a depression on the cell surface. D: Thin-sectioned, high-pressure frozen electron micrographs of C. destructans reveal 350- to 400-nm particles with inner and outer membranes, as well as dark-staining periplasmic material. The genome is condensed into a nucleoid during dormancy. Scale bars: A, B: 5 μm; C, D: 250 nm.

Figure 4-2: Effect of addition of Chromulinavorax destructans to a culture of Spumella elongata: A: Change in the cell number of S. elongate with (infected) and without (control) exposure to C. destructans is added (detection limit 10^5 cells ml^-1). The error bars represent the SEM of triplicate
cultures, and the asterisks, time points where the differences between the control and infected cultures were significantly different (p<0.005). B: Flow cytometric profile at 48hpi of *C. destructans* (red arrow) in a culture of S. elongata and stained with SYBR-Green.

4.4.2 *Chromulinavorax destructans* replicates in bodies nested in large mitochondrial invaginations

The life cycle of the lytic bacterium *Chromulinavorax destructans* putatively begins when ingested by the phagotrophic protist, *Spumella elongata*, a heterokont and slightly elongated flagellate (Figure 4-1A, Figure 4-3A, Figure 4-4C), although it can also form larger amoeboid cells. Once in the food vacuoles, the cells appear to secrete outer membrane vesicles (Figure 4-4A/B), and by 3 hpi they appear as a spherical mass in the host’s cytoplasm with the mitochondrion partially wrapped around the bacterial replication body by forming a deep invagination (Figure 4-3B). The invagination becomes more pronounced over time until the mitochondrion completely surrounds the replication body. During this process, the membrane integrity of the parasitoid and the mitochondrion, including its cristae, remain intact, suggesting that there is no invasion of the mitochondrial matrix (Figure 4-3C). At 9 hpi, the bacterial replication bodies are surrounded by the mitochondrion, and appear as several amorphous elongated bodies (Figure 4-4). This expansion phase culminates around 12 hpi, when the mitochondrion, now containing numerous invaginations and inclusions occupied by the replication bodies, takes up two thirds of the host cytoplasm (Figure 4-3D). Despite this extensive modification, the mitochondrion is still intact as membrane integrity is not compromised and cristae structures are preserved (Figure 4-3D). This integrity is contrasted by the degradation of the nucleus, as well as extensive membrane disarray inside the cell, including membrane vesicles budding from the cell (Figure 4-3D). The *C. destructans* replication bodies now show signs of regularly spaced invaginations, preceding division into the mature cocci (Figure 4-3D). The replication cycle completes at around 19 hpi, when *Spumella elongata* ghost cells with emptied cytoplasm appear (Figure 4-3E). The mature coccoid form of *C. destructans* are seen both intracellularly and extracellularly, presumably released by bursting of the compromised host cell membrane (Figure 4-3E, Figure 4-3E). Occasionally, replication bodies that did not complete the replication cycle, possibly due to prematurely exhausting the host cell’s resources, were observed within ghost cells (Figure 4-3E, Figure 4-3E).
Figure 4-3: Replication cycle of Chromulinavorax destructans. A: Electron micrograph of a healthy Spumella elongata cell showing the nucleus (black arrow) and the mitochondrion (white arrow). B: At 3 hpi, C. destructans (black arrow head) has invaded the host cytoplasm and the mitochondrion (white arrow) has wrapped around the replication body, while the nucleus stays intact (black arrow). C: Close-up of the replication body (black arrow head) surrounded by the host mitochondrion (white arrow) showing intact membranes of both the pathogen and the organelle (black arrow: nucleus). D: At 12 hpi, C. destructans shows long filaments and the beginning of septation (black arrowhead), surrounded by the highly invaginated mitochondrion (white arrow). The nucleus is present, but shows signs of degradation. E: Ghost cells of S. elongata with degraded cytoplasm and organelles (white arrow: mitochondrion) and C. destructans (black arrow heads) at 18 hpi. Scale bars: A: 500 nm; C: 250 nm; B,D,E: 1 μm
Figure 4-4: Replication cycle of Chromulinavorax destructans. A: Putative *C. destructans* (black arrow head) inside the food vacuole of *Spumella elongata* showing apparent secretion of outer membrane vesicles (white arrow head) and an intact *S. elongata* mitochondrion (white arrow). B: Putative *C. destructans* (black arrow head) inside the food vacuole of *Spumella elongata* showing liposomes and outer membrane vesicles (white arrow head). C: Extensive *C. destructans* replication inside *S. elongata* at 9 hpi. D: Mature cells of *C. destructans* being released from a ghost cell of *S. elongata* at 24 h post-infection. Scale bars: A,B: 250 nm; C: 500 nm; D: 1 μm
Figure 4-5: Schematic representation of the inferred replication cycle of Chromulinavorax destructans. The host S. elongata cell is shown in green, with the mitochondrion in brown, nucleus in dark green, Golgi in orange, and vacuoles in light orange. C. destructans is depicted in blue throughout its replication cycle.

4.4.3 The Chromulinavorax destructans genome encodes a minimal set of core functions but an extensive assortment of genes involved in host modification

The 1,174,272 bp circular ds DNA Chromulinavorax destructans genome has a GC content of 33.7%, similar to other intracellular parasitoids of eukaryotes [28, 34]. GC skew and the location of the presumptive DnaA box “TTATCCACA” suggest that the origin of replication lies at 651641-652233bp (Figure 4-6A). The genome encodes two typical complete rDNA loci and 35 tRNAs covering all twenty amino acids (Figure 4-6A). Other ncRNAs include 4.8S RNA (ffs ncRNA), RNase P RNA (rnpB ncRNA), as well as a tmRNA (ssrA) (Figure 4-6A). Of the 1,081 predicted open reading frames 55% had a functional prediction with the largest fractions involved in DNA replication, translation, trans-membrane transporters and host manipulation (Figure 4-6B).

The DNA replication and repair machinery of C. destructans consists of a simple origin binding complex (lacking dnaC) and a typical prokaryotic replicase complex. DNA repair pathways include nucleotide excision repair (uvr system), short- and long-patch base-excision repair and
mismatch repair (mutL/S/Y and recJ). Homologous recombination is supported by the rec and ruv pathways (with dpo), as well as site-specific recombinases (xerJ/D). The RNA polymerases assisted in transcription by several transcriptional regulators (merR, yebC, fmdB), transcription termination factors (nusA/B/G), and RNase H.

The ribosome, missing proteins 30S-22, 50S-8, 50S-25, 50S-26, 50S-30 and the rRNAs are edited by several ribosomal modification enzymes such as methyl and acetyl transferases. Similarly, the tRNAs are modified by a large number of modifying enzymes such as methyl transferases and dihydrouridine synthases and C. destructans possesses a full set of 21 aminoacyl-tRNA synthetases. Peptide chain release factors (prfA/B), ribonucleases (RNaseY/P), ribosome recycling factor (rrf), translation initiation and elongation factors promote translation (infA/B/C, fusA, lepA, elf, tsf, tufB). Post translational modifications are assisted by several chaperones and some proteins are subjected to transport to membranes.

Chromulinavorax destructans utilizes nucleoside salvage for both purine and pyrimidine biosynthesis and is able to phosphorylate the products into all required nucleotides. Therefore, precursor nucleosides must be imported from the host. Similarly, amino-acid biosynthesis is restricted to simple conversions between related amino acids (ser/cys, cys/ala, glu/gln, ser/gly), emphasizing that C. destructans is highly dependent on the host.

Chromulinavorax destructans encodes very rudimentary cell-division machinery, with only ftsA, K, L, W, and Z present, supported by zapA. Cell shape is determined by mreB, C, and RodA. Consistent with microscopic observations of a gram-negative-like phenotype, no LPS biosynthesis genes were observed. However, a complete mur pathway of peptidoglycan biosynthesis is likely responsible for generating the electron-dense material in the periplasm observed in electron micrographs (Figure 4-1D). Several surface antigens of unknown function are also encoded, as well as Type IV pilus assembly proteins, which might be assembled through a derived pulD channel. Several signaling trans-membrane receptors and signal-transduction proteins influencing the cell cycle are putatively involved, which could be involved in switching from dormancy to active replication.

As there was no evidence for complete metabolic pathways encoded by the genome of C. destructans, it implies that the cells must rely on extensive transmembrane transport systems to import metabolites and other resources from its host (Supplementary Figure 10). A large array of ABC transporters involved in importing oligopeptides and amino acids through the pot and
opp/dpp systems, presumably provide amino acids for protein synthesis. Phospholipids and lipoproteins are imported through the mal and lol ABC transporter systems respectively. ABC transporter systems fep and tro import trace elements such as iron, zinc and manganese (znu). Other ABC transporter systems of uncharacterized specificity are present, including putative multidrug exporters. Besides the ABC transporter systems, several specialized symporters, antiporters, and pumps are predicted to import potassium (trk), sodium (nha/als/put), inorganic ions (mgt), and nucleosides (nup). Further, specialized multidrug exporters are also present (rhat/eama). Central to energy requirements, there are two copies of a tlcc ATP/ADP antiporter that allows for the exchange of ADP for ATP from the host, which seems to be the only source of ATP. The membrane potential, essential for many transporters and antiporters, appears to be maintained by an ATP-synthase running “in reverse”, which may be regulated by passive proton channels. Large biomolecules can be imported by mechanosensitive channels (msc) and a biopolymer transporter (exb).

A substantial proportion of the genome encodes proteases, nuclease, and hydrolases, which are putatively involved in processing and breaking down host-cell structures and biopolymers that can be imported and reused. These effectors can be classified as secreted, membrane bound, and cytoplasmic factors. Amongst the secreted hydrolases is a close relative of chitinases and several HAD superfamily hydrolases (yigB, mphC, glcD). Secreted proteinases include M3, M15, M23, M41, M48, C11, and C65 peptidases. C65, M23, and M50 peptidases are predicted to be membrane bound. The final step in host biopolymer degradation is accomplished by cytoplasmic endonucleases (endoU), hydrolases (NUDIX and HIT-type), and proteases (lon, M16, M17). As well, a putative superoxide dismutase is encoded, which could help the cells cope with stressors such as hydrogen peroxide from close proximity to the mitochondrion.

More than 10% of the proteins encoded by C. destructans are predicted to be involved in modifying and influencing the host cell. Most prominent are 98 copies of ankyrin-repeat domain proteins. This concurs with observations of other intracellular parasitoids among the TM6 candidate phylum, such as Babela massiliensis, but also includes unrelated intracellular parasites such as Legionella spp., and giant viruses [34, 178, 183, 184]. The exact function of ankyrin-repeat domain proteins is unknown, but they have been implicated in membrane modification and counteracting the host immune system [165, 166, 183]. A CDS for a protein distantly resembling
mitofilin was also found, and is an intriguing candidate for the extensive manipulation of the mitochondrion.

A sec secretion system is likely used to export proteins in the periplasm and to secrete proteins. Interestingly, 44% of all CDS possess a putative signal peptide that targets them to the secretion system, either as membrane proteins or as secreted proteins. This subset of secreted proteins is enriched in ORFans and proteins of unknown function compared to the complete set of CDS (Figure 4-6B). However, other overrepresented fractions include putative CDS for proteins that break down or alter the host, such as proteinases and nucleases, and most prominently the ankyrin-repeat domain proteins, 78% of which contain a signal protein.

Figure 4-6 Genome content of Chromulinavorax destructans. A: Genome map showing open reading frames encoded by the plus and negative strand (dark and light blue respectively), tRNAs (green), rRNAs (red) and other ncRNAs (black). Inner circles depict GC content (red/green) and GC skew (teal/purple). B: Functional groups of putative protein-coding genes (Proteome) and the fraction coding a signal sequence (Secretome).

4.4.4 Phylogenetic position of Chromulinavorax destructans within the candidate phylum TM6/Dependentiae

Full length 16S rDNA maximum likelihood (ML) phylogenetic analysis of the TM6/Dependentiae places C. destructans into the proposed Babeliales (ord. nov.), within the Babeliae (class nov.) [33], but within a sister family to the Babeliacea (fam. nov.), for which the
name *Chromulinavoracea* (fam. nov.) is proposed. This family also contains the original MAG of JCVI_TM6SC1 and might also harbour *Vermiphilus pyriformis*, based on a partial 16S sequence [109]. The basal nodes in the 16S rDNA ML tree are poorly supported, likely due to undersampling, and similar to previous analysis (Figure 4-7A) [33]. The phylogenetic position of *C. destructans* was also explored by ML analysis of concatenated ribosomal proteins from MAGs that confidently contained all of the ribosomal CDSs investigated. The resulting tree reflects the architecture of the 16S rDNA tree and supports the proposed sister families *Babeliacea* and *Chromulinavoracea* within the order *Babeliales*, which is distinct from a second well-supported order of environmental sequences (Figure 4-7B). Genome content analysis of representatives of the two families within the *Babeliales* supports the separation of the families *Babeliacea* and *Chromulinavoracea*. The proposed members of the *Chromulinavoraceae* share more clusters of orthologous genes with each other than with *Babela masiliensis* of the *Babeliacea*, despite the *Chromulinavoraceae* including potentially incomplete MAGs (UASB293, SOIL82, and JCVI TM6SCI) and the *Babeliacea* being represented by a complete genome (Figure 4-7C).
**Figure 4.7**: Phylogenetic placement of *Chromulinavorax destructans*. A) 16S full length rDNA ML tree (1000 BS replicates) showing the TM6/Dependentiae phylum. “Substantially complete” MAGs highlighted in blue, isolates in red. Solid circles show node support of >90 and open circle >50. B) ML tree of ribosomal proteins (L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, S19; BS support based on 1000BS replicates). C) Shared gene clusters between *C. destructans* and members of its family *Chromulinavoracea* (UASB293, SOIL82, and JCVI TM6SCI) and *Babela massiliensis* as a representative of the Babeliaeaea.

### 4.5 Discussion

#### 4.5.1 *Chromulinavorax destructans* is highly host-dependent

*Chromulinavorax destructans* is an intracellular pathogen of *S. elongata* that appears to be dependent on its host for replication, to the extent that it does not appear to encode any complete metabolic pathways (Figure 4-6, Supplementary Figure 10). Many putative effector molecules, such as proteases, nucleases and ankyrin–repeat domain proteins, that could manipulate and break down host structures, contain signal peptides and appear to be secreted from the bacteria (Figure 4-4A). Similarly, putative outer membrane vesicles that could also contain such effector molecules
bud during the early stages of infection and might serve a similar function (Figure 4-4A). The reorganization and breakdown of host structures provide resources, possibly due to aforementioned effector molecules, that can be imported into the \textit{C. destructans} replication bodies (Figure 4-8). With no evidence of lipid biosynthesis, in combination with the observed host membrane disarray during late infection, suggests that host lipids are used for the bacterial cell membrane. Remodeling of the host mitochondrion is one of the most unusual features of the replication process. The expansion of the mitochondrion while maintaining membrane integrity contrasts with observations from other mitochondrion-invading pathogens and symbionts that actively disrupt the mitochondrial integrity [185, 186]. The proximity of bacterial replication to the expanded mitochondrion would allow for a steady supply of ATP mediated by the encoded ADP/ATP antiporter, but it would also require free radicals to be neutralized, a function likely performed by bacterial encoded superoxide dismutase. This high host-dependence is consistent with the pathogen being dormant until taken up by a host cell, and the lack of evidence of genome replication in free bacteria. It may also be the reason that the cells remained infectious after four years at 4°C. Complete host dependence and extracellular dormancy are life-cycle characteristics that highlight traits of convergent evolution found in many obligate pathogens, including viruses and some of the most reduced eukaryotic parasites. For example, Microsporidia completely lack energy generation pathways, while ATP transporters are widespread [99, 187, 188]. The high similarity in life style and genome content between TM6/Dependentiae and giant viruses, provides an intriguing example of converging evolutionary trajectories from vastly different evolutionary backgrounds.
4.5.2 The extensive mitochondrial modifications caused by *Chromulinavorax destructans* might stem from limited host resources

Rudimentary cell-division machinery and the lack of many hallmark genes has been highlighted for the TM6/Dependentia group, specifically based on MAGs [33, 108]. Similarly, *Babela massiliensis* showed delayed cell division after initial growth in amorphous bodies (i.e. replication bodies), and, while not directly described, *Vermiphilus pyriformis* likely uses a similar replication strategy [34, 109]. Although replication of *C. destructans* in replication bodies is much
like *B. massiliensis*, the close association and modulation of the mitochondrion was unique to *C. destructans*. In part, this difference might reflect the availability of cytoplasmic resources, such as metabolites and ATP, that presumable are more scarce in the flagellate cell compared to the much larger amoeba; Accordingly, *C. destructans* causes severe deformation of the cellular architecture of the nanoflagellate that was not seen in amoeba hosts (Figure 4-3, Figure 4-4). The scarcity of cytoplasmic ATP could lead to the close association of *C. destructans* with the host mitochondrion and initiate the mitochondrion’s expansion and rearrangement. A limitation of energy could have led to the incomplete replication bodies that were frequently found inside ghost cells, suggesting that ATP availability might restrain *C. destructans* replication. The differences between amoeba-infecting and nanoflagellate-infecting TM6/Dependentiae resemble observations made for giant viruses infecting different hosts. Nanoflagellate-infecting giant viruses show highly spatially oriented replication and preserve the mitochondrion; whereas, amoeba-infecting viruses seem to be unrestricted by the host-cell architecture [67, 178]. Additionally, a similarly high percentage of the genome of the lytic TM6/Dependentiae representatives is dedicated to ankyrin-repeat domain proteins, suggesting that these are crucial effectors for intra-eukaryotic replication and host cell take-over [34, 178]. Although their specific function of these proteins within these host-pathogen systems remain unknown, their nature as protein-protein interaction domains suggests they might be involved in host modification strategies [183]. Where ankyrin-repeat domain proteins have been explored in other host-pathogen systems, their functions include blocking of the host immune system, the induction of phagocytosis, the facilitation of cytoplasmic invasion, and the self-protection from potentially damaging effector molecules [166, 168, 183, 189, 190]. Any one of these functions could benefit *C. destructans* during its replication and therefore might be deployed.

**4.5.3 Life history strategies and host-range might differ vastly amongst the TM6/Dependentiae and might promote diversity within the phylum**

Phylogenetic analysis confidently places *C. destructans* within the candidate phylum TM6/Dependentiae, making it the first reported isolate of the phylum that does not infect an amoeba, and the first to infect a representative of the stramenopile supergroup of eukaryotes. Other studies have reported a correlation of TM6/Dependentiae with ciliates, suggesting that they infect or are symbionts of other eukaryotic lineages [33]. Life-history strategy seems to be independent of the host taxonomic group, given that *C. destructans* and *B. massiliensis* are lytic, but infect distantly related hosts, while *V. pyriformis* is symbiotic, and like *B. massiliensis* infects an amoeba.
This suggests that life-history strategy, and resulting pathogenicity, is a product of coevolution within a specific host-pathogen pair. The restricted host ranges of *C. destructans* and *V. pyriformis* are consistent with members of the TM6/Dependentiae being highly adapted to their hosts. The highly restricted host range of members of the TM6/Dependentiae also suggests that they are extraordinarily diverse, as is reflected by the great diversity of related 16S rRNA sequences from environmental surveys[33]. Given their potentially broad host range and distribution, TM6/Dependentiae could play a significant role in controlling the abundance and diversity of aquatic autotrophic and heterotrophic protists.

### 4.5.4 Summary

*Chromulinavorax destructans* is the first described isolate of the candidate phylum TM6/Dependentiae that does not infect *Acanthamoeba* and is the first isolate of a novel family so far only represented by MAGs. It causes extensive reorganization of the host cell, most notably the mitochondrion wrapping around the bacterium’s replication bodies. The genomic complement of *C. destructans* is highly reduced in metabolic capabilities, encoding not a single complete metabolic pathway and instead relies extensively on host resources such as metabolites and even ATP for energy supply. In contrast, the *C. destructans* is rich in putative effector molecules, putatively breaking down and reorganizing the host cells, as well as a large number of uncharacterized proteins and proteins not represented in databases. *C. destructans*’ narrow host range provides hints to a regulatory role of parasitic bacteria in the diversity and abundance of heterotrophic protists.
Chapter 5

5 From the inside out: An epibiotic *Bdellovibrio* predator with an expanded genomic complement

5.1 Synopsis

Bdellovibrio and like organisms are abundant environmental predators of prokaryotes that show a diversity of predation strategies, ranging from intra-periplasmic to epibiotic predation. The novel epibiotic predator *Bdellovibrio salishius* was isolated from a eutrophic freshwater pond in British Columbia, where it was a continual part of the microbial community. *Bdellovibrio salishius* was found to preferentially prey on the beta-proteobacterium *Paraburkholderia fungorum*. Despite its epibiotic replication strategy, *B. salishius* encodes a complex genomic complement more similar to periplasmic predators as well as several biosynthesis pathways not previously found in epibiotic predators. *Bdellovibrio salishius* is representative of a widely distributed basal cluster within the genus *Bdellovibrio*, suggesting that epibiotic predation might be a common predation type in nature and ancestral to the genus.

5.2 Introduction

Biotic factors regulating bacterial populations in nature are often assumed to be viral lysis and zooplankton grazing [175]; however, an underappreciated cause of mortality is predation by other prokaryotes. Such predators, collectively referred to as Bdellovibrio and like organisms (BALOs), have evolved several times independently and deploy a variety of “hunting strategies”. Many facultative predators with broad host ranges, such as *Ensifer adhaerens* and *Myxococcus xanthus*, deploy a “wolfpack strategy” where a prey cell is surrounded by several predators and lysed [11, 191]. Other, more specialized obligate predators have a narrower host range and specific predation strategies; for example, *Bdellovibrio spp.* enters the periplasm of the prey cell to consume the prey’s cytoplasm [31, 192].

*Bdellovibrio spp.* are delta-proteobacteria predators that use a biphasic lifestyle comprising an attack phase, in which a small, highly motile flagellated cell seeks out prey, and a growth phase, characterized by the predator penetrating the outer membrane of the prey cell and consuming its cytoplasm [31]. During the growth phase, the predator forms a characteristic structure in the prey’s periplasm known as the bdelloplast, which consists of a rounded, osmotically stable outer
membrane of the prey cell and several replicating *Bdellovibrio* cells. The bdelloplast continues to grow until the resources of the prey cell are exhausted and culminates in the septation and release of several to dozens of new attack-phase cells. This dichotic lifestyle switch is mediated by a highly expressed riboswitch in *B. bacteriovorus* [193]. The related genera *Bacteriovorax* and *Predibacter* are in the family *Bacteriovoracea*, which is a sister family to the prototypical *Bdellovibrionacea* within the order *Bdellovibrionales* [30].

Curiously, the alpha-proteobacteria genus *Micavibrio*, which is unrelated to the *Bdellovibrionales* leads a remarkably similar lifestyle to *Bdellovibrio* species, with high prey specificity. However, these bacteria prey in an epibiotic fashion on the outside of the prey cell instead of penetrating into the periplasm [194]. Due to their similar lifestyles, *Micavibrio* spp. are included into the BALOs.

Recently, an isolate of a newly described species, *Bdellovibrio exovorus*, in the family *Bdellovibrionacea* that is closely related to periplasmic bdelloplast-forming *Bdellovibrio* species, was shown to have an extremely narrow host range, and employ a different epibiotic replication strategy [195]. In the attack-phase, cells of *Bdellovibrio exovorus* resemble those of other *Bdellovibrio* isolates; whereas, in growth-phase the cells do not penetrate into the cytoplasm, but stay attached to the outside of the prey, strongly resembling *Micavibrio* species. Further, in growth-phase, *B. exovorus* does not induce a bdelloplast and seems to extract the cytoplasmic contents of the prey across both membranes. Once the resources of the prey are exhausted, growth-phase result in binary fission releasing two progeny attack-phase cells. The comparatively small genome of *B. exovorus* has been linked to its epibiotic predation strategy and reductionist evolution from an ancestor capable of intra-periplasmic replication [195-197]. *Bdellovibrio salishius* is the second epibiotic predator within the genus *Bdellovibrio*. Its genomic complement, phylogenetic placement and environmental distribution broaden our understanding of the ecology and evolution of this genus.

### 5.3 Materials and methods

#### 5.3.1 Isolation and Culturing

An isolate of a lytic bacterium, here named *Bdellovibrio salishius sp. nov.*, was obtained from a water sample collected near the sediment surface of a eutrophic pond in Nitobe Memorial Garden at the University of British Columbia, Canada (49°15′58″N, 123°15′34″W). As part of a
bioassay for pathogens infecting heterotrophic protists a subsample of the water was inoculated into modified DY-V artificial freshwater medium with yeast extract and a wheat grain [110].

5.3.2 Pathogen sequencing

For PacBio sequencing, exponentially growing mixed cultures containing *B. salishius* as well as *Bodo saltans* NG1 and its virus BsV were centrifuged in a Sorvall SLC-6000 for 20 min and 5000 rpm at 4°C to remove eukaryotic cells [178]. Particles in the supernatant concentrated approximately 100-fold by tangential flow ultrafiltration with at 30kDa cut-off (Vivaflow 200, PES) cartridge. To concentrate the cells further they were centrifuged at 28,000 rpm, 15°C for 8 h in a Beckman ultracentrifuge using a Ti90 fixed-angle rotor (Beckman-Coulter, Brea, California, USA), and then sedimented onto a 40% Optiprep 50 mM Tris-Cl, pH 8.0, 2mM MgCl$_2$ cushion for 30 min at 28,000 rpm, and 15°C in a SW40Ti swing-out rotor. The Optiprep gradient was created by underlaying a 10% Optiprep solution in 50 mM Tris-Cl, pH 8.0, 2 mM MgCl$_2$ with a 30% solution followed by a 50% solution and equilibration overnight at 4°C. One ml of concentrate from the 40% cushion was added atop the gradient and the concentrate was fractionated by centrifugation in an SW40 rotor for 4 h at 25000 rpm and 18°C. The fraction corresponding to the pathogen was extracted from the gradient with a syringe and washed twice with 50 mM Tris-Cl, pH 8.0, 2 mM MgCl$_2$ followed by centrifugation in an SW40 rotor for 20 min at 7200 rpm and 18°C and were finally collected by centrifugation in an SW40 rotor for 30 min at 7800 rpm and 18°C. Purity of the concentrate was verified by fluorescence vs SSC of SYBR-Green stained samples (Invitrogen Carlsbad, California, USA) on a FACScalibur flow cytometer (Becton-Dickinson, Franklin Lakes, New Jersey, USA). High molecular weight genomic DNA was extracted using phenol-chloroform-chloroform extraction. Length and purity were confirmed by gel electrophoresis and by using a Bioanalyzer 2100 with the HS DNA kit (Agilent Technology). PacBio RSII 20kb sequencing was performed by the sequencing center of the University of Delaware. Reads were assembled using PacBio HGAP3 software with 20 kb seed reads resulting in a single contig of 3,376,027 bp, 97.08 x coverage, 99.92% called bases and a consensus concordance of 99.9954 % [132].

5.3.3 Propagation and host range studies

Plaque assays were performed by mixing 0.5 ml putative host cultures in logarithmic growth stage and 10μl of Bdellovibrio salishius stock culture with 4.5 ml molten 0.5% DY-V agar and incubation for 48h. Propagation of *Bdellovibrio salishius* in liquid culture was monitored by
PCR with custom primers set specific to *B. salishius* 16S rDNA (Forward-5’- AGTCGAACGGGTAGCAATAC-3’, Reverse-5’-CTGACTTAGAAGCCCACCTAC-3’) as well as a BALO-specific primer set by Davidov et al. [30]. To obtain a pure isolate of prey cells present in the mixed microbial assemblage, culture samples were streaked onto a DY-V agar plate and incubated at room temperature. Distinct colonies were picked and propagated in liquid DY-V medium. Propagation of *Bdellovibrio salishius* using these cultures as hosts was confirmed by PCR. The identity of the prey cell cultures was confirmed by universal 16S rDNA Sanger sequencing (515F-5’-GTGYCAGCMGCCGCGGTAA-3’, 926R-5’-CCGYCAATTYMTTTRAGTTT-3’) [117]. To clean up the predator culture, *E. coli* (Thermo Fisher) cells were grown in LB medium and pelleted at 3,900 x g (4500 rpm) for 10 min, washed with 10 ml of HEPES/CaCl2 buffer (25 mM HEPES, 2 mM CaCl2), centrifuged in a fixed angle rotor centrifuge at 3,900 x g for 5 min, and re-suspended in 19 ml of HEPES/CaCl2 buffer. This cell suspension was then inoculated with 1 ml of 0.8-μm PVDF membrane filtered lysate of the *Bdellovibrio* containing culture and *B. salishius* propagation was monitored by PCR. Bdellovibrio remained viable at 4°C storage for up to two years and glycerol stocks of the native community containing *Bdellovibrio* as well as an inoculated E.coli TOP10 culture was stored at -80°C for archival purposes.

5.3.4 Environmental sampling

The presence of *B. salishius* in the Nitobe-Garden pond was determined in 20-L water samples that were taken bimonthly during spring and summer 2017 filtered through GF-A filters (Millipore, Bedford, MA, USA; nominal pore size 1.1 μm) laid over a 0.8-μm pore-size PES membrane (Sterlitech, Kent, WA, USA). The remaining particulate material was concentrated into 250 ml using a 30-kDa MW cut-off tangential flow filtration cartridge (Millipore, Bedford, MA, USA). DNA from these concentrates was extracted using phenol-chlorophorm extraction and subjected to PCR using *Bdellovibrio salishius* specific 16S rDNA primers to confirm its presence.

5.3.5 Microscopy

5.3.5.1 Negative staining transmission electron microscopy

Cultures of *Escherichia coli* TOP10 were inoculated with *B. salishius* at two hour time intervals and infected cultures, as well as an uninfected control were diluted tenfold and fixed in 4% glutaraldehyde. Next, the samples were applied to the carbon side of formvar carbon-coated
400-mesh copper grids (TedPella, CA, USA) and incubated at 4°C in the dark overnight under high humidity. The liquid was then removed and the grids stained with 1% uranyl acetate for 30 s.

5.3.5.2 Ultra-thin sectioning transmission electron microscopy

For higher resolution images, cells of *E. coli* infected with *Bdellovibrio salishius* were harvested at 4h intervals, as well as from uninfected control cultures. Cells from 10 ml of culture were pelleted at 5000 xg in a Beckmann tabletop centrifuge using a fixed angle rotor. The pellet was resuspended in 0.2 M Na-cacodylate buffer, 0.2 M sucrose, 5% EM-grade glutaraldehyde, pH 7.4 and incubated for 2 h on ice. After washing in 0.2 M Na-cacodylate buffer, cells were post-fixed with 1% Osmium tetroxide. Samples were dehydrated through water/ethanol gradients and ethanol was substituted by acetone. Samples were embedded in an equal part mixture of Spurr’s and Gembed embedding and the resin was polymerized at 60°C overnight. Fifty-nm thin sections were prepared using a Diatome ultra 45° knife (Diatome, Switzerland) on an ultra-microtome. The sections were collected on a 400x copper grid and stained for 10 min in 2% aqueous uranyl acetate and 5 min in Reynold’s lead citrate. Image data were recorded on a Hitachi H7600 transmission electron microscope at 80 kV. Image J (RRID:SCR_003070) was used to compile all TEM images. Adjustments to contrast and brightness levels were applied equally to all parts of the image.

5.3.6 Fluorescence In Situ Hybridization Epifluorescence Microscopy

To confirm epibiotic predation, cultures for fluorescence in-situ hybridization (FISH) were prepared as outlined below. Two 10-ml volumes of *E. coli* TOP10 were centrifuged at 3900 xg in a Beckman tabletop fixed angle centrifuge (4500 rpm) for 10 min, washed with 5 ml of HEPES/CaCl2 buffer, centrifuged at 3900 g for 5 min, and re-suspended in 9 ml of HEPES/CaCl2 buffer. One ml of *B. salishius* containing culture was added to the resuspended *E. coli* while another served as a control. Both cultures were incubated at room temperature for 24 h and were centrifuged again at 3900 x g (4500 rpm) for 10 min, washed with 10 ml of PBS, centrifuged at 3,900 x g for 5 min, and re-suspended in 5 ml of PBS. Two ml of the cultures were fixed in a 1:3 dilution of 10% buffered formalin (pH 7.0; 10 ml of 37% formaldehyde, 0.65 g Na₂HPO₄, 0.4 g NaH₂PO₄, 90 ml of Milli-QTM H2O) at 4°C for 3 h. Cells were then centrifuged again at 3,900 x g, washed twice in 10 ml of PBS, re-suspended in 10 ml of a mixture of PBS and 96% EtOH (1:1), and vortexed. In order to localize the predator an Alexa-488 tagged probe specific to *Bdellovibrio* 16S rDNA was designed (5’-/5Alexa488NTGCTGCCTCCCCGTAGGAGT-3’) based on Mahmoud et al. which also served as a template for the incubation protocol [198]. Ten μl of sample
was spotted onto a 70% EtOH-cleaned slide, dried at room temperature, and then taken through a dehydration series of 50%, 80%, and 95% EtOH. 25 μl of the hybridization master mix (20 mM Tris-HCl [pH 7.4], 0.1 % SDS, 5 mM EDTA, 0.8 M NaCl, 37% formalin, 1 ng/μl of final probe concentration) was added onto the sample. A cover glass was placed onto each sample and the slides incubated for 2 h at 46°C. With the cover slip removed, the slides were subsequently submerged into a bath of wash buffer and incubated at 48°C for 30 min. Slides were rinsed with sterile deionized H2O and dried at room temperature. A drop of ProLongTM Diamond Antifade Mountant with DAPI (4,6-diamidine-2-phenylindole) was spotted onto a new cover glass and placed on the sample. Finally, the slides were incubated at room temperature in the dark for 24 h prior to observation on an Olympus FV 1000 system.

5.3.7 Annotation

The genome was circular and 3,348,710 bp in length. Genome annotation was performed using the automated NCBI Prokaryotic Genome Annotation Pipeline (PGAAP). In parallel, open reading frames were predicted using GLIMMER (RRID:SCR_011931) with default settings [179]. Translated proteins were analyzed using BLASTp, CDD RPS-BLAST and pfam HMMER. These results were used to refine the PGAAP annotation. Signal peptides and trans-membrane domains were predicted using Phobius [180]. The annotated genome is available under the accession number CP025544. Metabolic pathways were predicted using the Kyoto Encyclopedia of Genes and Genomes (KEGG RRID:SCR_012773) automatic annotation server KAAS and Pathway Tools (RRID:SCR_013786) [181, 182].

5.3.8 Phylogenetic analysis

Full length 16S rDNA sequences of completely sequenced isoaltes of *Bdellovibrio* spp., as well as full-length uncultured top BLAST hits were downloaded from NCBI. Alignments of rDNA sequences were performed in Geneious R9 (RRID:SCR_010519) using MUSCLE with default parameters (RRID:SCR_011812)[139]. Maximum likelihood trees were constructed with RAxML ML search with 1000 rapid bootstraps using GTR+GAMMA [141].

Phylogenetic analysis of the genome content by orthologous gene clusters was performed by OrthoMCL (RRID:SCR_007839) [136] using whole genome sequences downloaded from NCBI. OrthoMCL was run with standard parameters (Blast E-value cutoff = 10−5 and mcl inflation factor = 1.5) on all protein-coding genes of length ≥ 100 aa. This resulted in the definition of 4242 distinct gene clusters.
5.4 Results

5.4.1 Isolation, host range and distribution

A lytic pathogen of bacteria was collected in a mixed microbial assemblage from a temperate eutrophic pond in southwestern British Columbia, Canada. Based on full-genome sequencing and electron microscopy of the infection cycle the pathogen was determined to be a new species of bacterium, here named *Bdellovibrio salishius*, after the Coast Salish, the indigenous peoples who populate southwestern British Columbia. *B. salishius* propagated in a mixed microbial assemblage from the sample site (Figure 5-1). Additionally, *B. salishius* could also be propagated on specific isolates from this assemblage that were identified as the gamma-proteobacterium *Pseudomonas fluorescence* as well as the beta-proteobacterium *Paraburkholderia fungorum* (Figure 5-1). Further, *B. salishius* was also able to propagate in a liquid culture of *E.coli* TOP10 (Supplementary Figure 11A). However, only inoculation of *Paraburkholderia fungorum* resulted in the observation of high numbers of putative attack-phase cells under phase-contrast microscopy. Similarly, small clear plaques were only observed on plates of *P. fungorum*, but not on plates of *P. fluorescence* or *E. coli* (Figure 5-2). *Bdellovibrio salishius* was detected at several time points in DNA extracted from water concentrates from Nitobe Gardens UBC in 2017, four years after the initial isolation, indicating the population persists in the pond (Supplementary Figure 12).
Figure 5-1 Bdellovibrio salishius host range and environmental distribution. A: B. salishius propagation assessed by PCR using strain-specific primers: = water control, DY-V=medium control, += Positive control from mixed Nitobe garden pond culture, B= *Pseudomonas fluorescense* culture, B+= *Pseudomonas fluorescense* culture after inoculation and two propagation cycles of B. salishius, T= *Paraburkholderia fungorum* culture, T+ = *Paraburkholderia fungorum* culture after inoculation and two times propagation of B. salishius
5.4.2 Morphology and replication cycle

*Bdellovibrio salishius* attack-phase cells are free swimming highly motile flagellated rods that are about 1 μm by 0.4 μm in size and exhibit a sheathed flagellum (Figure 5-3 A). Once attack-phase cells contact a prey cell, they attach irreversibly and form a broad predatory synapse and discard the flagellum (Figure 5-3 B). No invasion of the prey cell was observed during growth-phase, nor were bdelloplasts, implying that *Bdellovibrio salishius* is an epibiotic predator (Figure 5-3). The growth-phase cell attached to the prey cell empties the cytoplasm of the host cell, leaving behind an empty ghost cell (Figure 5-3 C,E,G). Simultaneously, the growth-phase *Bdellovibrio* cell grows in size and once the resources of the prey cell are exhausted, the growth-phase cumulates in binary fission and the production of two offspring attack-phase cells that repeat actively searching for new prey cells by rapid locomotion. Throughout the growth-phase, the cell membranes of the prey as well as the predator remain intact and instead of periplasmic invasion, an electron dense layer is observed on both the prey’s and predator’s membranes suggesting that a high concentration of effector molecules like transmembrane transporters are likely recruited to these sited to facilitate predation (Figure 5-3 F). Predation was dependent on the growth-phase of the host cell with cells in logarithmic growth supporting the highest *B. salishius* concentrations.
5.4.3 *Bdellovibrio salishius* has a complex genome for an epibiotic predator

5.4.3.1 Genome structure and content

The 3,348,710-bp *B. salishius* genome is similar in size to periplasmic *Bdellovibrio* spp., but considerably larger than that of *Bdellovibrio exovorus*, another epibiotic predator with a 2.66 Mb genome and with 38.9% also has the lowest GC content of any species within the genus. The GC content is relatively constant and exhibits a dichotomy in GC-skew that is typical of a circular...
bacterial genome (Figure 5-4A). Furthermore, the *B. salishius* genome contains one complete rDNA operon, similar to other epibiotic predators, and 31 tRNAs, three non-coding RNAs (ssrS, rnpB and ffs), and three putative riboswitches (Figure 5-4B). A total of 3166 protein coding genes were identified and are distributed equally between the plus and minus strands(Figure 5-4A). These proteins represent 22 different functional clusters of orthologous genes (Figure 5-4B).

**Figure 5-4:** Bdellovibrio salishius genome. A: Genomic map of *B. salishius*. From outside to inwards: Plus-strand CDS (light blue), Minus-strand CDS (dark blue), tRNAs (black), rRNAs (red), and non-coding RNAs (pink), GC-content (purple/mustard), and GC-skew (red/green). B: Abundance of 866 identified functional clusters of orthologous genes in the *B. salishius* genome.

### 5.4.3.2 Metabolism

The *B. salishius* genome encodes a metabolism typical for a predatory bacterium. Glycolysis and the complete TCA cycle, as well as a core set of pentose phosphate pathway genes are present, suggesting *B. salishius* is capable of several sugar conversions, and is able to provide the precursors for riboflavin biosynthesis. Pyruvate metabolism is coded for, but propanoate metabolism is only partially possible. A vitamin B6 biosynthesis pathway is encoded and acetyl-CoA biosynthesis is possible via pantoate. Nicotinamide metabolism and biosynthesis pathways also exist. Oxidative phosphorylation is encoded with the exception of cytochrome C reductase. The presence of core mevalonate pathway enzymes suggests this pathway is functional.

Based on inferred CDS, *B. salishius* can synthesize pyrimidines and purines *de novo*, and in contrast to other epibiotic predators, produce inosine. A complete DNA polymerase complex
facilitates DNA replication. As well, all types of DNA repair pathways are present, including base excision, nucleotide excision, mismatch repair and homologous recombination, the latter being limited to single-stand-break repair.

*Bdellovibrio salishius* encodes a complete ribosome except for the non-essential protein L28, and there is a complete set of tRNAs loaded by aminoacyl-tRNA-synthetases for every amino acid. The genome encodes a core RNA degradasome for post-transcriptional regulation and nucleotide recycling. Amino-acid biosynthesis pathways are limited and only cysteine, methionine, glutamate, lysine, proline and threonine can be completely synthesized *de novo*. Glycine and serine can be synthesized via the one-carbon pool pathway using tetrahydrofolate. Aspartate, alanine leucine, isoleucine, valine phenylalanine and tyrosine can be converted from their direct precursors, which are presumably acquired from the prey.

Complete fatty-acid degradation pathways are coded for, but fatty-acid elongation seems limited. Also encoded are complete sec and gsp pathways for protein secretion, as well as a partial tat pathway (subunits tatA, tatC, tatD). Extensive peptidoglycan production, as well as partial lipopolysaccharide biosynthesis pathways putatively decorate the periplasmic space and cell surface.

### 5.4.3.3 Regulatory elements

Master regulators, such as sigma factor 28 / FliA, which is proposed to enable the switch between attack and growth-phase modes in other *Bdellovibrio* species, is present in *B. salishius* and might work alongside putative riboswitch elements similar to those found in *B. bacteriovorus* [193]. No homologue of the host interaction (“hit”) locus protein bd0108, of *B. bacteriovorus*, was found in *B. salishius*, suggesting that it may deploy a different pilus regulation mechanism, and therefore might not be able to switch between facultative and obligate predation.

### 5.4.3.4 Predatory arsenal

The *B. salishius* genome shows many adaptations to a predatory lifestyle. A complete biosynthesis pathway for flagellar assembly and regulation provides locomotion in the attack-phase. Attack phase cells are likely guided by a canonical, almost, complete chemotaxis pathway that is only missing cheY. Substrate recognition is putatively mediated by tatC and two copies of von Willebrand factors [199]. A type IV pilus appears to be present that is putatively involved in prey-cell attachment through pilZ. To access resources within the prey, an array of transporters are coded for, many of which show signal peptides that facilitate export, and might insert into the
prey-cell membrane. ABC-type transporters likely import phosphate (pst), phosphonate (phn), and lipopolysaccharides (lpt), while there appears to be partial ABC transporter systems for lipoproteins, thiamine, branched-chain amino acids, oligo and dipeptides, microcin, phospholipids, biotin, daunorubicine, alkylphosphosphate, methionine, iron and siderophores, cobalt, sugar and organic solvents. Non-ABC transporter system CDS are present for potassium (kdp), biopolymers (exbD and tol), iron (ofeT), heavy metals (cusA), biotin (bioY), threonine (rhtB), as well as for several multidrug exporters (bcr, cflA, arcB). CDS for low-specificity transporters include MFS and EamE transporters, as well as transporter for ions and cations, macrolide, chromate, as well as sodium-dependent transporters and others of uncharacterized specificity.

5.4.4 *Bdellovibrio salishius* is a basal representative of its genus

In phylogenetic analysis of the 16S rDNA locus, *B. salishius* occupies a well-supported basal branch within the genus *Bdellovibrio*; the closest relatives are found in environmental amplicon sequences, as well as in poorly characterized isolates (Figure 5-5A). Notably, these strains from a tight cluster basal to *B. exovorus* and *B. bacterivorax* strains and show rather low boot-strap support within their clade despite being closely related. This cluster appears to be equivalent to the “cluster 2” described by Davidov et al. [30].

![Phylogenetic tree and gene cluster analysis](Figure 5-5 Bdellovibrio salishius phylogenetic placement. A: 16S phylogenetic tree showing the completely sequenced *Bdellovibrio* species as well as the top BLAST hits to *Bdellovibrio salishius* from uncharacterized isolates, as well as from metagenomic data. B: Shared gene cluster analysis of complete *Bdellovibrio* genomes comparing epibiotic and periplasmic species.)
Shared gene cluster analysis was congruent with the 16S phylogeny with the majority of gene clusters in *B. salishius* being shared with other members of the genus *Bdellovibrio* (Figure 5-5B). Surprisingly, *B. salishius* shares more than 300 gene clusters with periplasmic predators, which are not found in the epibiotic predator *B. exovorus*, despite their similar predation strategy. On the other hand, the epibiotic predators *B. salishius* and *B. exovorus* shared more than 130 genes that are not found in periplasmic predators, and might be involved in epibiotic predation. These exclusively epibiotic genes within members of the genus *Bdellovibrio* include CDS for proteases, peroxiredoxin, glutathione-dependent formaldehyde-activating enzyme, nucleases, hydrolases thioesterase, polysaccharide deacetylase, an amino-acid ABC transporter, as well as many poorly characterized proteins. Gene clusters exclusively shared with *B. exovorus*, as well as the phylogenetically distant *M. aeruginosavorus* that also deploys an epibiotic predation strategy, are remarkable as they might highlight a common set of genes are required for epibiotic predation. Six such genes that were identified include anhydro-N-acetylmuramic acid kinase, peptidoglycan translocase, a FAD/NAD binding protein, a cation-transporting P-type ATPase and a pseudouridine synthase.

### 5.5 Discussion

#### 5.5.1 *Bdellovibrio salishius* is an epibiotic predator of the beta-proteobacterium *Paraburkholderia fungorum*

While propagation of *B. salishius* was observed in both, liquid cultures of *Pseudomonas fluorescence* as well as *E. coli*, plaque assay failed to confirm lysis of these potential host. Accordingly, detection of continued *B. salishius* replication in such cultures was probably an artefact of carry-over of *Paraburkholderia fungorum* from the *B. salishius* stock culture, as plaque assay provides a more specific readout than PCR detection in liquid culture. This is in line with the higher density of putative attack-phase cells observed in *P. fungorum* cultures compared to the other putative hosts. Similarly, the low number of *B. salishius* cells and their slow replication dynamics observed in microscopy might be a function of the low number of *P. fungorum* present in these systems as a result of carry-over.
5.5.2 Despite an epibiotic phenotype, *Bdellovibrio salishius* shares many genes with periplasmic predators.

Microscopic analysis clearly shows *B. salishius* deploying an epibiotic predation strategy, which is reflected in its genome that shares several features previously identified to be involved in epibiotic predators such as *B. exovorus* and *Micavibrio aeruginosavorus*. These include physical features such as the number of rDNA loci, as well as metabolic capabilities based on gene content that suggests limited fatty-acid elongation and the absence of polyhydroxyalkanoate depolymerase and the siderophore aerobactin, all present in periplasmic predators [197]. In contrast, *B. salishius* also has coding sequences for the biosynthesis of isoleucine and tyrosine, as well as for riboflavin and vitamin B6, which had been found in periplasmic but not epibiotic predators reflecting its comparatively large and complex genome [196]. The linkage of these genes with periplasmic replication was by association and not for functional reasons; hence, the presence of several of these genes in *B. salishius* may simply reflect its relatively larger genome size. Cluster analysis of orthologous genes in *B. salishius*, other *Bdellovibrio* spp. as well as the unrelated BALOs *Micavibrio aeruginosavorus* and *Halobacteriovorax marinus* reveals just six gene clusters associated with epibiotic predation. Strikingly, these genes suggest that the prey peptidoglycan is salvaged by N-acetylmuramic-acid kinase as well as a peptidoglycan translocase that is specific to epibiotic predators. Since this limited complement of genes was found between distantly related taxa, there might not be a clear functional separation between epibiotic and periplasmic predators. This is consistent with epibiotic predation evolving independently within the genera *Bdellovibrio* and *Micavibrio*. Therefore, conclusions regarding function based on gene-cluster analysis should be interpreted with caution, especially since functionally equivalent proteins can group into different gene clusters, and thus may escape such analysis. Accordingly, gene-cluster comparison among species in the same genus that deploy different predation strategies is more informative and revealed a specialized complement of proteases nucleases, hydrolases and detoxifying enzymes in epibiotic *Bdellovibrio* species (Figure 5-5 B). Notably, the addition of *B. salishius* as a second epibiotic *Bdellovibrio* species greatly decreased the number of genes associated with epibiotic predation, as its large genome has greater overlap with periplasmic *Bdellovibrio* species. Further, this suggests that different mechanisms can be deployed in both, epibiotic and periplasmic predation.
5.5.3 Epibiotic predation is a common strategy of environmental *Bdellovibrio* species and could be the ancestral phenotype

The closest known relatives to *B. salishius* are poorly characterized isolates, and environmental 16S rRNA sequences from diverse environments. These environments include “commercial aquaculture preparations”, soils, waste-water activated sludge, and iron-oxidizing freshwater environments [30, 200-202]. This broad diversity of habitats suggests that *Bdellovibrio* species that are closely related to *B. salishius* are widely distributed in freshwaters, and could be a major contributor to global BALO diversity. Because of the phylogenetic placement and the broad distribution of related isolates, epibiotic predation might be common among BALOs, despite being underrepresented in isolates. The recurrent detection of populations of *B. salishius* in the pond from which it was isolated confirms that is part of the natural community and therefore supports the idea that epibiotic predation might be common in nature. Additionally, both *B. salishius* and *B. exovorus* are epibiotic and branch basal within the genus, compared to the periplasmic predators. This suggests that epibiotic predation might be the ancestral predation type in the genus.

5.5.4 Summary

Detailed knowledge of environmental BALO diversity is still missing. The discovery and analysis of *Bdellovibrio salishius* provides new insights into this fascinating group. Its intermediate genomic complement blurs the lines between what is required for epibiotic and periplasmic replication. Moreover, as a representative of a basal and widespread member within the genus, *Bdellovibrio*, it suggests that epibiotic predation is common in the environment, and might be the ancestral form of predation in this genus.
Chapter 6

6 Concluding remarks

6.1 Culturing still has a place in environmental microbiology in the age of metagenomics

Since the turn of the century, the advent of high throughput next generation sequencing technology has led to an unprecedented influx of genomic information. Environmental microbiology has not been an exception from this trend and the new sequencing methods have enabled the study of microbial communities and their members without the restraints of culturing. While this field, known as metagenomics, has provided insights into the composition of microbial life on earth, it also has created novel questions. While it has been established early on that the majority of microbes cannot easily be cultured and therefore evade investigation, a phenomenon termed the “great plate count anomaly”, metagenomics have opened our eyes to the true diversity of life. An example of this is the phenomenon is SAR11, one of the most abundant bacteria in the oceans, that was only recognized due to metagenomic studies [203]. Similarly, the “candidate phyla radiation” suggests that there is a large group of basal phyla comprising up to 15% of the total bacterial diversity, that at the time of discovery did not have any cultured representatives at all [173]. While metabolic capabilities encoded by many of these genomes could be derived by sequence analogies, other MAGs remain unexplored [204]. Accordingly, the biology of such organisms remains largely enigmatic.

Classical culturing techniques have been on the sideline of environmental microbiology over the last decade, but still provide the confidence of actual in vitro experiments that no in-silico predictions can produce. Since the uncultured majority does exist in the environment, there is no prohibitive reason why such organisms can’t be cultured in general. Rather the conditions in culture so far are not suitable for the organisms in question. The SAR11 again serves as a prime example, since targeted culturing approaches, informed by metagenomic information, was ultimately successful in establishing SAR11 in culture [205].

Similarly, the present study utilized the information of metagenomic studies to specifically target pathogens of nanoflagellates for establishment in culture [26, 122]. Using a novel combination of particle concentration and high throughput screening, the present study was able to isolate a broad diversity of microbial parasitoids, confirming that they are common and abundant members of microbial ecosystems. All three isolates studied in detail showed surprising biology
and a large number of novel genes (ORFans) reflecting the high degree of bias in the databases towards cultured organisms.

6.2 Convergent evolution across the broad evolutionary distances

A surprising feature of the presented work is the apparent convergence of features shown by the studied isolates. The parasitoids of nanoflagellates, BsV and C. destructans possess an array of shared featured despite their drastically different evolutionary backgrounds. Chromulinavorax destructans and B. salishius also show a number of convergently evolved mechanisms of extracting resources from their host/prey cells.

6.2.1 Evolutionary history is of limited relevance for the phenotype of parasitoids of heterotrophic protists

Bodo saltans virus and Chromulinavorax destructans both infect heterotrophic nanoflagellates. While their host are phylogenetically unrelated, the pathogens do share features in their biology, so far so, that C. destructans was initially thought to be a giant virus (Table 6-1). Both are rather similar in size and while the 400nm C. destructans particle is relatively small for a bacterium, BsV is a very large icosahedral virus. This size might be an adaptation to the prey size preference of their host cells, since phagocytosis is hypothesized to be the mode of entry [14]. Both pathogens initially expand in cytoplasmic replication factories that reorganize the host cell and show intimate relationships with the mitochondrion. In the case of BsV, the mitochondrion is the only organelle that remains intact until late stages of infection, whereas C. destructans effectively is surrounded by the mitochondrion, highlighting the high energy demands of both pathogens. Similarly, the two genomes are similar in size, and several shared genes. Remarkable is the abundance of ankyrin-repeat domain proteins in both genomes that make up approximately 10% of the total number of protein coding genes in either genome. Similarly, both genomes also exhibit a high number of ORFans highlighting the gaps in out understanding of pathogens of protists.
Table 6-1: Convergent features of BsV and *C. destructans*

<table>
<thead>
<tr>
<th>Feature</th>
<th>Bodo saltans virus</th>
<th>Chromulavorax destructans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle size</td>
<td>300nm</td>
<td>400nm</td>
</tr>
<tr>
<td>Genome size</td>
<td>1.4 Mb</td>
<td>1.2 Mb</td>
</tr>
<tr>
<td>Replication</td>
<td>Cytoplasmic near Mitochondrion</td>
<td>Cytoplasmic nested in Mitochondrion</td>
</tr>
<tr>
<td># of ANKs</td>
<td>141</td>
<td>93</td>
</tr>
<tr>
<td>% ORFans</td>
<td>47%</td>
<td>46%</td>
</tr>
<tr>
<td>Metabolic activity</td>
<td>None predicted</td>
<td>None predicted</td>
</tr>
<tr>
<td>DNA replication</td>
<td>Complete</td>
<td>Complete</td>
</tr>
<tr>
<td>Transcription</td>
<td>Complete</td>
<td>Complete</td>
</tr>
<tr>
<td>Translation</td>
<td>tRNA editing and accessory proteins</td>
<td>Complete</td>
</tr>
<tr>
<td>Particle Permanence</td>
<td>Partial</td>
<td>Continuous</td>
</tr>
</tbody>
</table>

These analogies are surprising given the evolutionary history of both pathogens. While *C. destructans* presumably evolved from a more complex free living bacterium, BsV and other giant viruses appear to be derived from smaller viruses [140]. Selective pressure seems to have opposite effects on the genomes of the two pathogens, causing continued expansion in the virus and reduction in the bacterium, ultimately resulting in a larger genome of BsV than that of *C. destructans* (Figure 6-1). This suggests that there is an optimal genome size for parasitoids of heterotrophic protists that allows for both, the coding of a sufficiently complex replication machinery, as well as a sufficiently small genome to effectively use limited resources and facilitate rapid genome replication. Interestingly, *C. destructans* faces several unique evolutionary constraints, such as having to maintain a functional and complete translational machinery, compared to BsV, which resorts to modifying and reusing the host’s machinery. Further, *C. destructans* faces the additional burden of a continuously maintained cell membrane that all effectors and host resources have to be transported across, utilizing a complex secretion system and an array of different transporter. BsV on the other hand can access the host resources more easily since the viral factory, still clearly distinct from the host cytoplasm, is not physically separated from the host cytoplasm. Therefore, *C. destructans* appears to encode a significantly larger number of essential genes to maintain its cellular integrity throughout the infection cycle. Accordingly, a substantial proportion of BsV genes may be accessory in nature, as has been demonstrated for Mimivirus [158]. Contrarily, the overlap in gene content between both pathogens could be a function of their lifestyle, as the cytoplasmic milieu of heterotrophic protists is an
environment thought to be highly conducive for horizontal gene transfer [206]. An alternative explanation to the observed convergences would be the assumption that giant viruses, such as BsV are descendants of more complex cell like organisms, as has been proposed by several authors [207, 208].

![Figure 6-1: Hypothetical evolutionary scenario depicting the route of convergent evolutionary trajectories of BsV and C. destructans](image)

6.2.2 The bare necessities for bacterial pathogens and predators

While *B. salishius* and *C. destructans* live markedly different lifestyles, the former being an active predator, the latter an infective “spore”, they still face a similar problem: How to get effector molecules into the host/prey cell, and later, how to import resources into their own cell. Accordingly, these two bacteria have acquired a convergent assortment effector molecules, transporters, and secretion systems (Figure 6-2): Both organisms encode a large number of proteases, nucleases and hydrolases to break down the host/prey cell, many of these effector molecules possess signal peptides for export using a sec secretion system. In the other direction, both organisms deploy a large number of transporters such as extensive ABC transporter systems as well as major facilitator superfamily transporters (MFS) and specific antiporters. The identifiable sections of the *C. destructans* genome are limited to DNA replication, transcription,
translation and host modification/exploitation. The *B. salishius* genome is almost three time the size, as it needs to accommodate all these functions on top of the necessary genomic information to maintain a metabolism, locomotion, and chemotaxis. Nevertheless, the convergence in a “core-parasitoidome” of prokaryotes is remarkable across this broad phylogenetic gap.

![Diagram](image)

*Figure 6-2: Convergent complement of effector and transporter molecules in B. salishius and C. destructans*

### 6.3 Limitations of the present study

The present study on the characterization of parasitoids of protists has several limitations that are based on biological and technical reasons.

For feasibility reasons, the screening was limited to heterotrophic nanoflagellates that could be readily isolated and maintained. Further, only flagellates that showed a distinct flow cytometry profile could be included in the screen, narrowing down the range of potential host organisms further. Accordingly, the host and pathogen pairs discovered here are highly biased compared to the true diversity of heterotrophic protists found in the environment. Unfortunately, only alternative isolation and culturing techniques, most of them considerably more laborious, could allow for a less biased analysis.
The analysis of the genome content is hampered by the high number of protein coding genes that do not show similarities to proteins represented in any available database, as well as the absence of functional information for many genes that do show homologies to known genes. Accordingly, almost half the genome content of the two parasitoids of flagellates sequenced here are of unknown function. In the absence of readily available genetic manipulation systems, the functional analysis of such genes is a tedious undertaking. One possibility would be the utilization of CRISPR directed mutagenesis, but in the presence of multiple genome copies within each infected host cell, the practicability of such approaches is questionable and would require extensive screening for mutants. On top, the host genomes are not, or not well studied, making direct functional inferences problematic. Finally, the heterotrophic phagotrophic hosts necessitate the presence of a diverse bacterial community with their own parasites, phage and genetic elements that could drastically influence the genome of the parasitoids studied here [158].

6.4 Environmental microbiology in the 21st century

The 21st century is poised to be a pivotal time for environmental microbiology. For the first time, we have the necessary methods to truly access the global biodiversity. Metagenomic surveys by themselves, as well as by informing targeted isolation and mesocosm experiments can provide unprecedented insights into life on our planet. The current bottle neck to unveiling this diversity and its function on a global scale is the knowledge of the basic biology to interpret metagenomic information. This is where targeted isolation and culturing experiments, like the ones presented in this study, fill the gap in providing insights in the “biology” that goes with this “information”. However, given the sporadic and short sighted nature currently exhibited by funding agencies, such basic research is often difficult to conduct. Accordingly, it is in the hand of funding agencies to determine how soon we will break through into a new area of “informed metagenomics”. Since these agencies are largely driven by politics and therefore ultimately the general public, the fate of this research is in the hands of the research community themselves, as they must engage the public through effective science communication. A scientific mindset, fact-based decisions making, is something that the public at large, but especially politics could greatly benefit from in such pivotal times for the future of our planet and species.


81. Lecoq, H., *Discovery of the first virus, the tobacco mosaic virus: 1892 or 1898?* 2001.
100. Van Regenmortel, M., *The metaphor that viruses are living is alive and well, but it is no more than a metaphor*. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 2016. 59: p. 117-124.


A. Supplementary information to chapter 2

Supplementary Figure 1 Flow cytometry profiles of isolates. Lysates of cultures 72hpi with the appropriate lytic agent compared to uninfected cultures.
**B. Supplementary information to chapter 3**

*Supplementary Figure 2* **Ultrastructure of BsV particles and replication.** (A) Healthy *Bodo saltans* cell: Visible structures include the cytostome (black arrow head), the Golgi, mitochondrial arms protruding from the kinetoplast center with the kinetoplast genome (white arrow head), the flagellar root of both flagella as well as several vacuoles (back arrow: food vacuole containing partially digested bacterial prey) are visible traveling from the cytopharynx (white arrow) to the posterior cell pole (Scale bar = 500 nm) (B) Negative staining of a BsV particle with a blossom like opened stargate (C) *Bodo saltans* cell 24 hr post-BsV infection showing degraded intracellular structures and an extensive BsV virion factory (black arrow head). (D) BsV virions inside a vesicle. A closed stargate is visible at the apex of the bottom left virion (black arrow head).
Supplementary Figure 3: Ankyrin-repeat domain-containing proteins. A: Location of recently duplicated region in the 3' end of the genome (highlighted in black bars, duplicated CDS in red) B: Alignment of recently duplicated sequence from A, encoding several Ankyrin repeat domain-containing proteins. A region containing four CDS was duplicated in a reverse complement orientation maintaining the integrity and sequence identity of the original sequence (% DNA ID in the central region is 99.7%) C: ID, annotation and functional class of ankyrin-repeat domain-containing proteins with a recognizable domain fragment in the N-terminal region of the protein.
Supplementary Figure 4: Evolutionary history of translational machinery found in giant viruses inferred by COUNT and abundance of ankyrin-repeat-domain genes. The size of the black circles mapped on a cladogram of members of the Mimiviridae (see Figure 3-6 for full phylogenetic analysis) represents the number of gene families involved in translation at each node or tip. Blue circles indicate the number of ankyrin-repeat-domain encoding genes found in each genome. Gene gain and loss events are depicted along the branches. MVC: Megavirus chilensis, AMoV: Acanthamoeba polyphaga Mounovirus, ApMV: Acanthamoeba polyphage Mimivirus, CatV: Catovirus, BsV: Bodo saltans virus, HokV: Hokovirus, KloV: Klosneuvirus (KlosnV), IndV: Indivirus, CroV: Cafeteria roenbergensis virus, OLV1: Organic Lake Phycodnavirus 1, OLV2: Organic Lake Phycodnavirus 2, CeV: Chrysochromulina Ericina Virus, PgV: Phaeocystis globosa virus PgV-16T, AaV: Aureococcus anophagefferens virus
**Supplementary Table 1** Comparison of translational machinery encoded by giant viruses based on phylogenetic analysis: BsV Translational machinery compared to other NCLDV: blue: monophyletic group, red: recently host acquired, green: recently acquired from bodonid host, ψ: pseudogene MimV: Mimiviruses, , CatV: Catovirus, BsV: Bodo saltans virus, HokV: Hokovirus, KloV: Klosneuvirus (KlosnV), IndV: Indivirus, CroV: Cafeteria roenbergensis virus, PhycV: Phycodnaviridae, PanV: Pandoraviridae.

<table>
<thead>
<tr>
<th>Name</th>
<th>Abbr.</th>
<th>Class</th>
<th>BsV</th>
<th>Klov</th>
<th>IndV</th>
<th>CatV</th>
<th>HokV</th>
<th>MimV</th>
<th>CroV</th>
<th>PhycV</th>
<th>PanV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoleucine-tRNA synthetase</td>
<td>IleRS</td>
<td>aRS</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methionine-tRNA synthetase</td>
<td>MetRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamyl-tRNA synthetase</td>
<td>GluRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aspartyl-tRNA synthetase</td>
<td>AspRS</td>
<td>aRS</td>
<td>ψ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alanine-tRNA synthetase</td>
<td>AlaRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threonine-tRNA synthetase</td>
<td>ThrRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cysteinyl-tRNA synthetase</td>
<td>CysRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arginine-tRNA synthetase</td>
<td>ArgRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histidine-tRNA synthetase</td>
<td>HisRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proline-tRNA synthetase</td>
<td>ProRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lysine-tRNA synthetase</td>
<td>LysRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serine-tRNA synthetase</td>
<td>SerRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyrosine-tRNA synthetase</td>
<td>TyrRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glycine-tRNA synthetase</td>
<td>GlyRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leucine-tRNA synthetase</td>
<td>LeuRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tryptophan-tRNA synthetase</td>
<td>TrpRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenylalanine-tRNA synthetase</td>
<td>PheRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valyl-tRNA</td>
<td>ValRS</td>
<td>aRS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asparagine-tRNA synthetase</td>
<td>AsnRS</td>
<td>aRS</td>
<td>ψ</td>
<td>xx</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glutamine-tRNA synthetase, cytoplasmic</td>
<td>GluRS</td>
<td>aRS</td>
<td>ψ</td>
<td>xx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation initiation factor IF-2</td>
<td>IF-2</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 1</td>
<td>eIF-1</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 1A</td>
<td>eIF-1A</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 2 subunit alpha</td>
<td>eIF-2a</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 2 subunit beta 1</td>
<td>eIF-2b</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 2 subunit beta 2</td>
<td>eIF-2b2</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 2 subunit gamma</td>
<td>eIF-2g</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 4A-III</td>
<td>eIF-4A-III</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 4E</td>
<td>eIF-4E</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle domain of eukaryotic initiation factor 4G</td>
<td>eIF-4G</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 5A</td>
<td>eIF-5A</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic initiation factor 5B</td>
<td>eIF-5B</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic elongation factor 1</td>
<td>eEF-1</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eukaryotic peptide chain release factor subunit 1</td>
<td>eRF-1</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GTP-binding elongation/initiation factor</td>
<td>TF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA Z'-phosphotransferase 1 phospho</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA pseudouridine synthase B.pseudo</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peptidyl-tRNA hydrolase peptidyl</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA(Ile)-lysidine synthase lysidine</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA(Asp)-guanylyl transferase guany</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative tRNA ligase/uridine kinase</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative RtcB-like tRNA-splicing ligase</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative CAA-nucleotidyltransferase</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative methyltransferase TRM13-like</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHH family phosphohydrolase-like protein</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudouridine synthase</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DEXD helicases</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA specific mase3</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Putative ATP-dependent RNA helicase</td>
<td>TM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Figure 5: Evolutionary history of gene family content found in giant viruses inferred by COUNT. The size of the black circles mapped on a cladogram of the Mimiviridae (see Figure 3-6 for full phylogenetic analysis) represents the number of gene families at each node or tip. Gene gain and loss events are depicted on top of the branches in red and green, gene duplication and contraction in blue and purple below. MVc: Megavirus chilensis, AMoV: Acanthamoeba polyphaga Mounouvirus, ApMV: Acanthamoeba polyphage Mimivirus, CatV: Catovirus, BsV: Bodo saltans virus, HokV: Hokovirus, KloV: Klosneuvirus (KlosnV), IndV: Indivirus, CroV: Cafeteria roenbergensis virus, OLV1: Organic Lake Phycodnavirus 1, OLV2: Organic Lake Phycodnavirus 2, CeV: Chrysochromulina Ericina Virus, PgV: Phaeocystis globosa virus PgV-16T, AaV: Aureococcus anophagefferens virus
Supplementary Figure 6: Self-splicing group 1 introns found in BsV. A: DNA Sequence alignment of recently duplicated self-splicing group 1 introns polr2a-i1a and polr2a-i1c (94% DNA sequence ID). B: Secondary structure and conserved autocatalytic ribozyme centers (blue arrows) of other polr encoded introns.
Supplementary Figure 7: Phylogenetic analysis of five concatenated NCVOGs from selected NCLDV.

A: Maximum likelihood phylogenetic tree based on 1000 bootstrap replicates as calculated by RAxML.
B: Maximum likelihood phylogenetic tree based on 1000 bootstrap replicates as calculated by PhyML.
C: Bayesian posterior consensus tree with Bayesian posterior probability on two independent MCMC chains (16100 generations).

Supplementary Figure 8: Phylogenetic analysis of D5-like helicase-primase (NCVOG0023; A-C) and DNA polymerase family B (NCVOG0038; D-F) from selected NCLDV. . A/D: Maximum likelihood phylogenetic tree based on 1000 bootstrap replicates as calculated by RAxML. B/E: Maximum likelihood phylogenetic tree based on 1000 bootstrap replicates as calculated by PhyML. C/E: Bayesian posterior consensus tree with Bayesian posterior probability on two independent MCMC chains. MVC: Megavirus chilensis, AMoV: Acanthamoeba polyphaga Mounouvirus, ApMV: Acanthamoeba polyphage Mimivirus, CatV: Catovirus, BsV: Bodo saltans virus, HokV: Hokovirus, KloV: Klosneuvirus (KlosnV), IndV: Indivirus, CroV: Cafeteria roenbergensis virus, OLVI: Organic Lake Phycodnavirus 1, OLV2: Organic Lake Phycodnavirus 2, CeV: Chrysochromulina Ericina Virus, PgV: Phaeocystis globosa virus PgV-16T, AaV: Aureococcus anophagefferens virus, PacV: Pacmanivirus, FauV: Faustovirus
Supplementary Figure 9: Phylogenetic analysis of DNA or RNA helicases of superfamily II (NCVOG0076; A-C), packaging ATPase (NCVOG0249; D-F), Poxvirus Late Transcription Factor VLTF3-like (NCVOG0262; G-I) from selected NCLDVs. A/D/G: Maximum likelihood phylogenetic tree based on 1000 bootstrap replicates as calculated by RAxML. B/E/B: Maximum likelihood phylogenetic tree based on 1000 bootstrap replicates as calculated by PhyML. C/E/I: Bayesian posterior consensus tree with Bayesian posterior probability on two independent MCMC chains. MVc: Megavirus chilensis, AMoV: Acanthamoeba polyphaga Moumouvirus, ApMV: Acanthamoeba polyphage Mimivirus, CatV: Catovirus, BsV: Bodo saltans virus, HokV: Hokovirus, KloV: Klosneuvirus (KlosnV), IndV: Indivirus, CroV: Cafeteria roenbergensis virus, OLV1: Organic Lake Phycodnavirus 1, OLV2: Organic Lake Phycodnavirus 2, CeV: Chrysochromulina Ericina Virus, PgV: Phaeocystis globosa virus PgV-16T, AaV: Aureococcus anophagefferens virus, VacV: Vaccinia virus.
C. Supplementary information to chapter 4

*Supplementary Figure 10: Pathway analysis of C. destructans according to pathway tools.*
D. Supplementary information to chapter 5

Supplementary Figure 12: Bdellovibrio salishius plaque assay. Left column shows B. salishius infected plates, right column shows the control. Plaques highlighted by arrows.