Viruses: Contributors to and Mitigators of black band disease in corals

A thesis submitted by

Patrick Buerger

for the degree of Doctor of Philosophy,

at the College of Science and Engineering,

James Cook University.

March, 2017

Australian Institute of Marine Science

AIMS@JCU

James Cook University Australia

Acknowledgements

I wish to express my profound thanks to **Madeleine van Oppen**, **Bette Willis**, **Karen Weynberg** and **Elisha Wood-Charlson** who supported my candidature and gave me the opportunity to conduct my research at the Australian Institute of Marine Science (AIMS) and James Cook University (JCU). It has been a wonderful experience. I am grateful that you took me on as a student and for sharing your expertise on corals, viruses, genomics and coral diseases. Thank you very much.

Thanks a lot also to my supporters at AIMS, who always had an open door and who were always more than keen to help: **Katharina Fabricius**, **Nicole Webster**, **Jason Doyle**, **Andrew Negri**, **Peter Thomas-Hall**, **Florita Flores**, **Grant Milton**, **Emmanuelle Botte**, **Sara Bell**, **Brett Baillie**, **Lisa Peplow**, **Libby Evans-Illidge**, **Sven Uthicke**, **Britta Schaffelke**, **Patrick Laffy**, **Craig Humphrey**, **Carlos Alvarez-Roa**, **Geoff Millar, Julia Strahl**, **Heidi Luter, the SeaSim team**, **Joe Gioffre**, **Neal Cantin**, **Michael Civiello**, **Jose Montalvo**, **Line Bay**, **Liz Howlett, Scott Bainbridge**, **Yui Sato**, **David Bourne**, **Victor Beltran**, and everyone who is not on the list.

I would like to especially thank **Steve Clarke**, **Tim Simmonds**, **Catherine Naum**, **Kate Green**, **Michael Kebben** and **Kevin Blackman** for their generous support, supply with tools and infrastructure as well as (and probably most important) after hours security. Thank you very much for the great times. Working with you was one of the best experiences I had at AIMS.

Kathryn, thank you very much for your patience, understanding, for your presence, and endless support.

Of course, many thanks to my family, friends and everyone who supported me during my candidature and beyond.

Statement of the contribution of others

Stipend support
- James Cook University (JCU)
- AIMS@JCU

Advisors
- Prof. Madeleine van Oppen
  Australian Institute of Marine Science
  University of Melbourne
- Prof. Bette Willis
  James Cook University
- Dr. Elisha Wood-Charlson
  Australian Institute of Marine Science
  University of Hawaii
- Dr. Karen Weynberg
  Australian Institute of Marine Science

Editorial assistance
- Prof. Madeleine van Oppen (AIMS)
- Prof. Bette Willis (JCU)
- Dr. Elisha Wood-Charlson (AIMS)
- Dr. Karen Weynberg (AIMS)

Laboratory assistance
- Dr. Karen Weynberg (AIMS)
- Dr. Elisha Wood-Charlson (AIMS)
- Prof. Madeleine van Oppen (AIMS)
- Lesa Peplow (AIMS)
- Sara Bell (AIMS)

Collaborators
- Carlos Alvarez-Roa (JCU)
- Sebastien Baeklelandt (AIMS)
- Dr. Patrick Laffy (AIMS)
- Dr. Yui Sato (AIMS)

Project costs
- AIMS@JCU (#17625)
- Australian Institute of Marine Science (AIMS)
  Australian Research Council Future Fellowship to MvO (FT100100088)
  SuperScience Fellowship to KW (FS110200034)
- James Cook University (JCU)
- Great Barrier Reef Foundation (GBRF)

Use of infrastructure external to JCU
- Laboratory facilities (AIMS)
- National Sea Simulator (AIMS)
- High performance computing (AIMS)
- High performance cloud computing - Australia's National eResearch Collaboration Tools and Resources project (NeCTAR)

Summary

Coral reefs worldwide have seen a considerable increase in coral disease prevalence and disease-related mortality over the last few decades. Black band disease (BBD) is a common and widespread disease of reef-building corals that causes tissue loss at a rate of up to two cm/day. The polymicrobial disease is characterised by a dense, dark mat comprising numerous bacterial taxa including filamentous cyanobacteria, the most prevalent bacterium in terms of biomass. BBD-associated cyanobacteria contribute to the disease by providing a framework and structure to the other bacteria and by fuelling the disease progression with their photosynthesis products. The cyanobacterium *Roseofilum reptotaenium* of the genus *Oscillatoria* is ubiquitously present in BBD. Despite years of bacterium-centred research on BBD, key questions, such as the direct cause of the initial disease onset and disease mitigation, remain unclear. Viruses are known agents of a number of diseases in the marine environment, but little is known about their roles in coral disease, including BBD.

In this thesis, I examine whether viruses that infect bacteria (bacteriophages) are involved in the virulence of the main BBD-associated cyanobacteria and progression of BBD. This includes: (1) the characterisation of the bacteriophage community that targets BBD-associated cyanobacteria, (2) the establishment of methodologies to cultivate the main BBD cyanobacterium, *R. reptotaenium*, and to isolate lytic bacteriophages that target *R. reptotaenium*, (3) an assessment of the role of lytic bacteriophages as potential mitigators of the disease by lysing *R. reptotaenium* and (4) an examination of lysogenic bacteriophages as potential contributors to the virulence of their hosts by introducing virulence genes into BBD cyanobacteria genomes.

(1) Using amplicon next generation sequencing and bioinformatic analyses, I described the BBD bacteriophage community and showed that bacteriophages are part of the BBD microbial consortium. The bacteriophage community of tissue samples of BBD and healthy corals showed a generally higher variability than the communities prevalent in seawater samples. Nevertheless, several BBD-associated bacteriophage taxa were exclusively abundant in BBD samples and I hypothesised these infect one of the main BBD cyanobacteria; other bacteriophage taxa were more prevalent in healthy coral tissues.

(2) To further clarify the role of BBD-associated bacteriophages, I optimised existing methodologies for the isolation and cultivation of one of the main BBD cyanobacteria, *R. reptotaenium*. The newly established protocols serve as a basis for the isolation of lytic and lysogenic bacteriophages that target *R. reptotaenium*.

(3) Lysis of *R. reptotaenium* cultures was successfully achieved using protocols commonly applied for lytic bacteriophage isolation. Three bacteriophages that potentially target the *R. reptotaenium* were isolated from the culture medium, their genomes sequenced, and used in a phage therapy trial on BBD-affected corals. The isolated bacteriophages are closely related to *Cellulophaga* phages and therefore not known to infect cyanobacteria and not suitable for BBD phage therapy. Currently, phage therapy of BBD might be possible, but requires extensive optimisation in order to successfully mitigate the disease.

(4) To investigate whether lysogenic bacteriophages are present in the *R. reptotaenium* genome and whether they may increase cyanobacterial virulence, I sequenced and assembled the first draft genome of *R. reptotaenium* and in conjunction analysed a previously published genome of the BBD cyanobacterium *Geitlerinema* sp.

BBD_1991. Both cyanobacteria were equipped with adaptive, heritable defence systems that help to prevent bacteriophage infections (clustered regularly interspaced short palindromic repeats – CRISPR-Cas systems). Bacteriophage taxa that previously infected the cyanobacteria were reconstructed by analysing the target sequences (spacers) of the CRISPR-Cas defence systems, and were found to include taxa of the BBD bacteriophage community (*Chapter 2*). In addition, potential prophages were identified in three regions of the *R. reptotaenium* AO1 genome and in five regions of the *Geitlerinema* sp. BBD_1991 genome. These genomic regions contained putative virulence genes relevant of BBD, such as an NAD-dependent epimerase/dehydratase (a gene with homologue function to the third and fourth most expressed gene in BBD), lysozyme/metalloendopeptidases and other genes for lipopolysaccharide modification.

My findings suggest that the BBD disease mat is a hot-spot for phage infections. The presence of CRISPR-Cas defence systems in the cyanobacterial host genomes provides evidence of a constant arms race between BBD-associated cyanobacteria and bacteriophages. Maintaining such a defence system likely reduces the number of successful bacteriophage infections and mortality in the cyanobacteria, supporting the progress of BBD. The disease related genes in potential prophage regions suggest a role of bacteriophages as contributors to the virulence of BBD cyanobacteria. In addition, bacteriophages that successfully lyse cyanobacteria within the mat might be redistributing organic matter to the BBD microbial consortium as part of a viral shunt and also possibly contribute to the progression of the disease. Since the application of antibiotics is not feasible to control BBD due to the complexity of coral-microbial interactions, the use of a lytic bacteriophage in a phage therapy is a promising approach and should be developed further.

This study is one of the first to investigate the multi-faceted role that

bacteriophages have in coral disease and reveals bacteriophages as potential contributors to and mitigators of BBD. Applying the concepts and approaches of this study to other coral diseases is likely to show bacteriophages as hidden drivers, e.g. in the coral disease white syndrome and potentially even coral bleaching. Future studies should focus on characterising the associated virus community of healthy and diseased coral colonies with viral metagenomic approaches to obtain information on baseline communities and investigate individual virus-host interactions with quantitative PCR, classic virology methods and targeted amplicon sequencing. Viruses in coral diseases may hold the key to important answers regarding the virulence and pathogenicity of pathogens, as well as innovative mitigation strategies to improve coral health with phage therapy approaches.

Table of Contents

List of Figures

List of Tables

Chapter 1

General introduction to viruses and their potential roles in coral health and black band disease

1.1 The importance of marine viruses

Marine viruses are the largest, but most poorly explored genetic reservoir on the planet (Suttle 2007). They occur ubiquitously at an average density of 5×10^6 - 15×10^6 viruses per mL of seawater, which represents abundances an order of magnitude higher than those of bacteria in seawater (Bergh et al. 1989). Viruses contribute to biogeochemical cycling via the viral shunt, in which viruses that infect bacteria (bacteriophages) lyse approximately 20 - 25% of marine bacterial communities every day, redistributing organic matter and making it available to other organisms (Wilhelm and Suttle 1999). While marine viruses infect almost all forms of living organisms (Wommack and Colwell 2000), they are mostly known for their pathogenicity and capacity to cause virulent marine diseases (Munn 2006). Examples of marine diseases caused by viruses include white spot syndrome (WSS) in shrimp, which has had deleterious impacts on the shrimp aquaculture industry all over the world (van Hulten et al. 2001; Reddy, Jeyasekaran, and Shakila 2013), and fibropapillomatosis, a herpes tumor-like disease that is caused by tornovirus 1 (STTV1) (Ng et al. 2009). However, marine viruses are more than disease pathogens. In some instances viruses may mitigate rather than cause disease and might also be beneficial to their hosts (van Oppen, Leong, and Gates 2009). Since the 1930's, 'phage therapy' has harnessed the ability of viruses to infect bacteria as a way to treat bacterial diseases in humans, pets, and commercially important species (Adams 1959; Chan, Abedon, and Loc-Carrillo 2013; D'Herelle 1930). Phage therapy has been applied successfully in aquaculture to control bacterial pathogens in species such as shrimp (*Penaeus monodon*), catfish (*Clarias batrachusm*) and ayu fish (*Plecoglossus altivelis*) as a promising alternative to antibiotic treatments (Alisky et al. 1998; Oliveira et al. 2012).

Our knowledge of marine viruses has greatly improved over the last few

decades (Jover et al. 2014; Sánchez-Paz et al. 2014; Suttle 2005), in part because of methodological developments in virus enumeration and identification. Especially, flow cytometry and next generation sequencing now provide powerful new tools to decipher the role of viruses in animal diseases. In this PhD thesis, I investigated viruses as potential contributors to, or mitigators of, the widespread coral disease, black band disease (BBD).

1.2 Coral diseases are an emerging threat

The contributions of diseases to coral mortality has increased considerably in recent years for a number of reasons. Warm seawater temperature anomalies, which have been increasing in frequency (Hoegh-Guldberg 1999; Hughes et al. 2003), have been correlated with higher disease prevalence, possibly due to increased activity of pathogenic bacteria combined with reduced immunocompetence of corals, particularly when bleached (Bruno et al. 2007; Maynard et al. 2015; Ruiz-Moreno et al. 2012). Secondly, the increasing scale of anthropogenic impacts on coral reefs has been linked to higher disease prevalence, as a consequence of high human densities (Aeby et al. 2011), reduced water quality (Sutherland, Porter, and Torres 2004), chemicals of sunscreen creams in seawater (Danovaro et al. 2008) and increased disturbances from tourist activities (Lamb et al. 2014). For example, up to 15 fold higher coral disease prevalence has been reported for brown band disease, white syndrome and black band disease on reefs with tourist platforms in the Great Barrier Reef (Lamb and Willis 2011). Thirdly, cyclones and crown-of-thorns starfish cause injuries on corals that provide entry points for pathogenic microorganisms (e.g. Katz et al. 2014), both disturbances that have increased in frequency and severity over the last decades (De'ath et al. 2012). Despite the increase in prevalence of coral diseases, tools required for rapid disease diagnostics are still lacking and management strategies to prevent and mitigate

coral disease outbreaks are largely inadequate (Pollock et al. 2011). Of prime concern is that causative agents have not been identified for the majority of the coral diseases described (Sheridan et al. 2013). While a few known scleractinian coral pathogens are bacteria (Harvell et al. 2007), the role of viruses in these diseases has barely been examined (e.g. Cervino et al. 2004; Soffer et al. 2014; Lawrence et al. 2015; reviewed by Thurber et al. 2017).

1.3 Black band disease (BBD)

BBD is the earliest described coral disease (Antonius 1973). It infects a wide range of hard coral species (scleractinian corals) and is abundant on coral reefs around the globe (Sutherland et al. 2004). The unique BBD consortium of bacteria (e.g. Frias-Lopez et al. 2002) forms a black mat that is characterised by an anoxic sulfide layer across the coral surface (Glas et al. 2012), which is lethal for the underlying coral tissue and progresses over the colony surface at rates of up to 2 cm per day (Fig. 1.1) (Rützler, Santavy, and Antonius 1983; Taylor 1983). A suite of different bacteria have been identified in close association with the disease, including among others the cyanobacteria species *Roseofilum reptotaenium* (Casamatta et al. 2012) and *Geitlerinema* sp. BBD_1991 (Den Uyl et al. 2016), sulphate-reducing *Desulfovibrio* bacteria, *Cytophaga* spp., Alphaproteobacteria and various other heterotrophic microbes (Cooney et al. 2002; Miller and Richardson 2011; Sato, Willis, and Bourne 2010). In terms of biomass, the filamentous cyanobacterium, *Roseofilum reptotaenium* (Rasoulouniriana) Casamatta, is the most abundant cyanobacterial species in the disease mat (Casamatta et al. 2012; Rasoulouniriana et al. 2009; Richardson and Kuta 2003). This cyanobacterium has been recognised as one of the main pathogens within the BBD microbial consortium because of its ability to penetrate polyp tissues and gastrovascular cavities (Kramarsky-Winter et al. 2014; Richardson et al. 2014),

thereby providing a physical framework and structure for the establishment of the disease (Sato et al. 2016, 2017). In addition, BBD-associated cyanobacteria can produce toxins in some geographic locations, such as microcystins, which may contribute additionally to the progression of the disease (Richardson et al. 2007; Gantar et al. 2009; but see Glas et al. 2010). The cyanobacteria also supplies nutrients to the microbial mat, possibly through fermentation and photosynthesis in the sulphide-rich BBD environment (Sato et al. 2017).

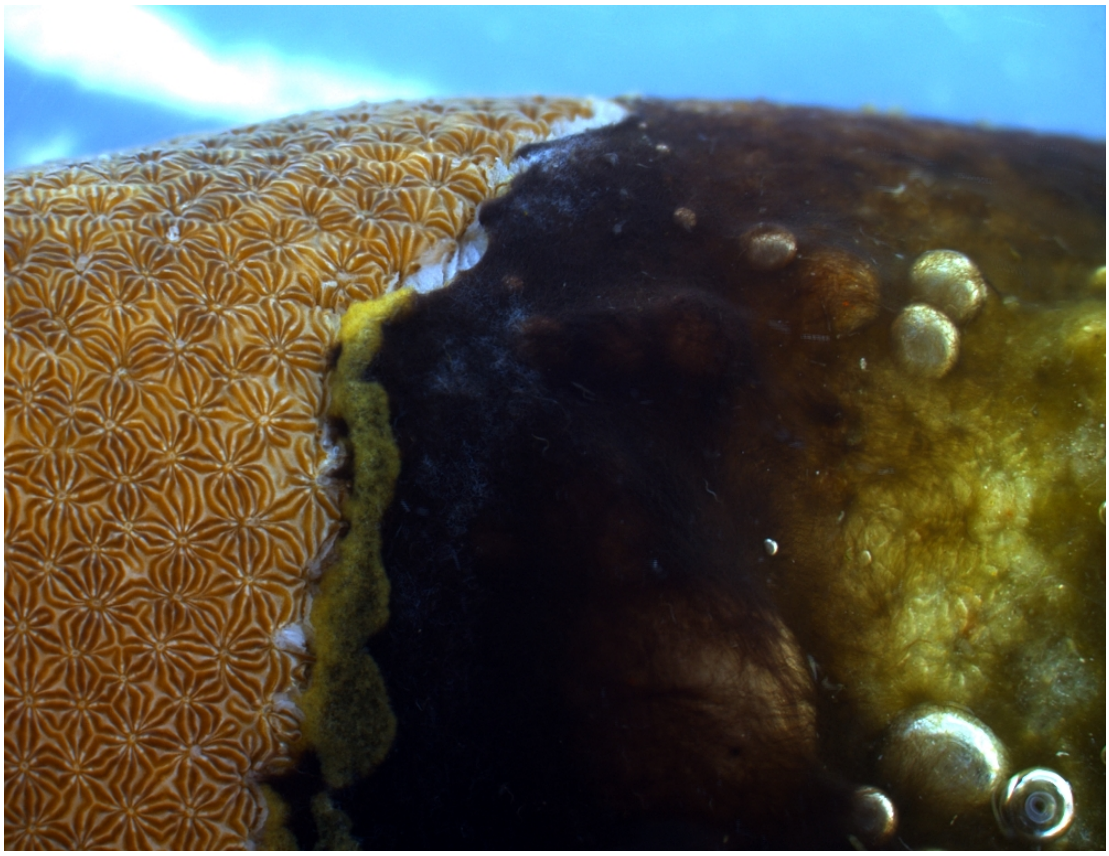


Figure 1.1 Black band disease (BBD) on the coral *Pavona duerdeni*. The BBD microbial mat is dominated by the filamentous cyanobacterium *Roseofilum reptotaenium.* The disease mat kills the underlying coral tissue and progresses at a rate of up to several cm per day.

The recent increase in disease outbreaks has been correlated with a range of stressors listed previously, including increased nutrients and high seawater

temperatures (Aeby et al. 2015; Lamb and Willis 2011; Yang et al. 2014). Despite substantial recent progress in understanding BBD through bacterium-centred research, the role of viruses in BBD has not been considered. Accordingly, answers to key questions, such as the direct cause of the disease onset and strategies to mitigate the disease, remain unclear.

1.4 Viral life history and potential role of viruses in BBD

Viruses may infect bacterial pathogens of BBD and either contribute to the onset or even prevent BBD progression. The infectious stage in the life cycle of a virus is represented by an acellular structure characterised by a layer of proteins surrounding the viral genome (Forterre 2013). Bacteriophages (viruses infecting bacteria) attach to a bacterial host cell, penetrate the bacterial cell wall and inject the bacteriophage DNA into the host cell (labelled A in Fig. 1.2). The lysogenic cycle refers to the integration of this bacteriophage genome into the bacterial host genome; this is called a prophage (labelled B in Fig. 1.2). These prophages, 'temperate bacteriophages', can persist over extended periods of time until a trigger, e.g., an increase in temperature (Wilson et al. 2001) or UV radiation (Lohr, Munn, and Wilson 2007), induces the lytic cycle (Ackermann and DuBow 1987). Following integration into the bacterial genome, the presence of virulence genes in the prophage genome may contribute to the virulence of a pathogen (Brüssow, Canchaya, and Hardt 2004). For example, the pathogenicity of the bacterium *Vibrio cholerae* primarily depends on infection by a lysogenic bacteriophage (CTXphi). The bacteriophage transfers genes that encode for one of the primary virulence factors, in this case the cholera toxin (CT), and converts *V. cholerae* from a non- pathogenic to a pathogenic strain (Faruque and Mekalanos 2003; Waldor and Mekalanos 1996).

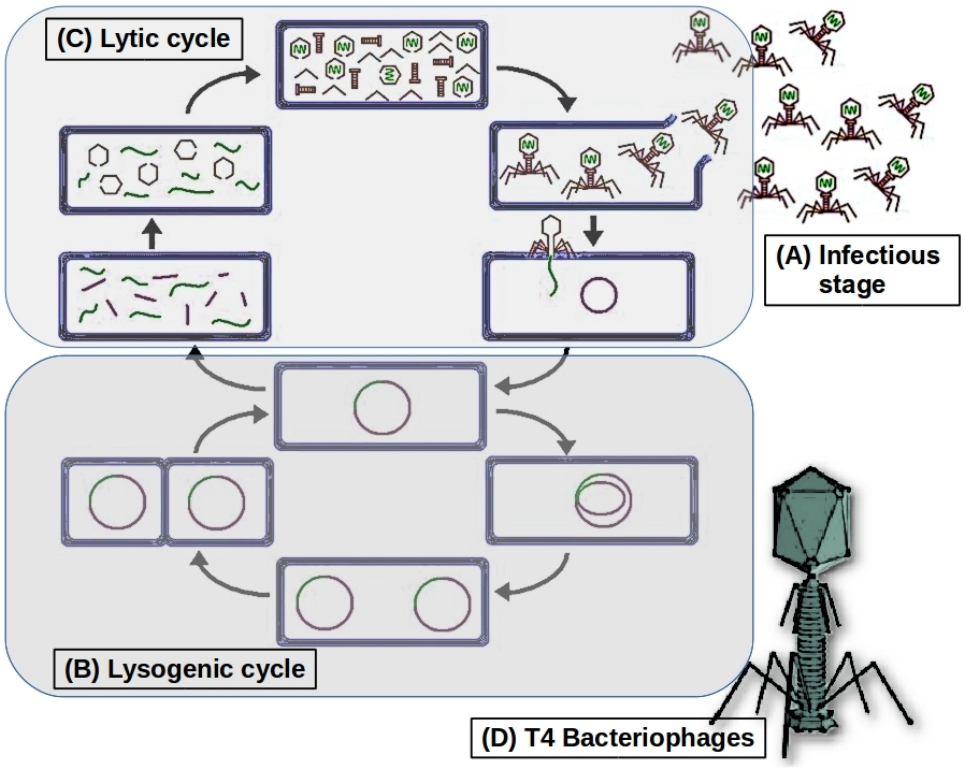


Figure 1.2 Stages in the life cycle of a bacteriophage. (A) Infectious stage: Bacteriophages infect bacteria by attaching onto their surfaces and injecting DNA into the bacterial cell. After DNA injection, the bacteriophage may go into either a lysogenic or a lytic stage. (B) Proviral lysogenic stage: In the lysogenic cycle, the bacteriophage genome integrates into the genome of the host bacterium as a prophage, where it replicates with the host. Environmental factors, such as temperature or UV-light, may trigger a lytic cycle. (C) Vegetative lytic stage: In the lytic cycle, the bacteriophage takes over the cellular machinery of the host to produce new virions and ultimately lyses (disintegrates) the host bacterium to release its progeny phages, which are ready to infect new bacteria. (D) Typical morphology of a T4-bacteriophage (a group of viruses that infect a wide range of bacteria, including cyanobacteria). Image is labelled for reuse with modification, source: commons.wikimedia.org/wiki/File:Phage2.JPG.

The vegetative lytic stage is characterised by the replication of bacteriophages within the host (labelled C in Fig. 1.2), which results in lysis (disintegration) of the host cell and release of newly produced bacteriophages into the environment (Adams 1959; D'Herelle 1930). Lytic bacteriophages have been used for the treatment of coral diseases that are caused by a bacterial pathogen (i.e., phage therapy). For instance, phage therapy has been successfully applied to the control coral pathogens, *Thalassomonas loyana* (Efrony et al. 2007; Efrony, Atad, and Rosenberg 2009), *Vibrio*

coralliilyticus strains YB1 (Atad et al. 2012; Efrony et al. 2007, 2009), and P1 (Cohen et al. 2013). Bacteriophages targeting *T. loyana*, which causes white plague-like disease (WPL) in the coral *Favia favus* in the Red Sea (Barash et al. 2005; Thompson et al. 2006), inhibited WPL disease progression and transmission to other corals, during both a seven-week field experiment (Atad et al. 2012) and a 21-day laboratory experiment (Efrony et al. 2007).

Viruses are typically present in high densities (up to 10^{10} viruses g^{-1}) within cyanobacteria-dominated, photosynthetic microbial mats (Carreira, Piel, et al. 2015; Carreira, Staal, et al. 2015), marine sediments (Carreira et al. 2013; Danovaro and Middelboe 2010; Suttle and Fuhrman 2010) and hypersaline mats (Pacton et al. 2014; de Wit et al. 2015). Recent detection of viruses in metagenomic and metatranscriptomic reads of the BBD microbial community (Arotsker et al. 2016; Sato et al. 2017), suggests that bacteriophages are associated with BBD. Bacteriophages may occur within the BBD mat in one or more of their life cycle stages (Fig. 1.2), i.e., at the infectious stage, the proviral lysogenic stage, or the vegetative lytic stage (McDaniel 2011; Weinbauer 2004). They may be able to infect members of the BBD community via their lysogenic and lytic cycles, and influence virulence and onset of the disease. The bacteriophage community of BBD has not been described in detail and their possible role in BBD etiology has not been considered. As the prevalence of BBD is increasing, bacteriophage research may provide new insights into some of the unanswered questions about BBD etiology, while also providing important knowledge towards the development of management strategies to mitigate this disease.

1.5 Study aims and objectives

To advance current knowledge about causative agents involved in the onset and

development of the virulent coral disease BBD and to develop potential preventative strategies, I examined the diversity and roles of bacteriophages associated with BBD. In particular, I investigated how bacteriophages might contribute to development of the disease, and alternatively, how they might be used to mitigate BBD. Overall, the four objectives and hypotheses listed below (*Chapters 2-5*) describe a sequence of studies designed to clarify the influence of bacteriophages on the main BBD cyanobacteria.

1) Characterise BBD bacteriophage communities that potentially infect the primary component of the BBD consortium, the cyanobacterium *R. reptotaenium* (*Chapter 2*). Hypothesis: Specific bacteriophage types are associated with BBD that infect disease associated cyanobacteria. Cyanophages that are part of the T4-bacteriophages in the family *Myoviridae* are targeted using amplicon sequencing of the major capsid protein gene (*gp23*), an established genetic marker for identification of this group of phages. I compared the diversity of T4-bacteriophages associated with seawater and healthy coral tissue (as an control) and BBD. I analysed samples collected during the transition from winter into summer months to explore temporal patterns that could influence BBD seasonal prevalence patterns and to provide evidence that specific bacteriophage communities are associated with BBD. This study provided a starting point for the detection of candidate bacteriophages that potentially infect BBD cyanobacteria and the first step towards understanding bacteriophage dynamics in BBD.

2) Establish protocols for long-term cultivation of the cyanobacterium *R. reptotaenium*, the primary pathogen in the BBD consortium for use in the isolation of lysogenic and lytic bacteriophages (*Chapter 3*). In this chapter, I present results of experiments designed to optimise culture protocols for the

cyanobacterium *R*. *reptotaenium.* I developed methods to measure cyanobacterial biomass, generate growth curves, and to maximise exponential phases as well as cyanobacterial biomass. The developed cultivation protocols were subsequently used for the isolation of lytic and lysogenic bacteriophages.

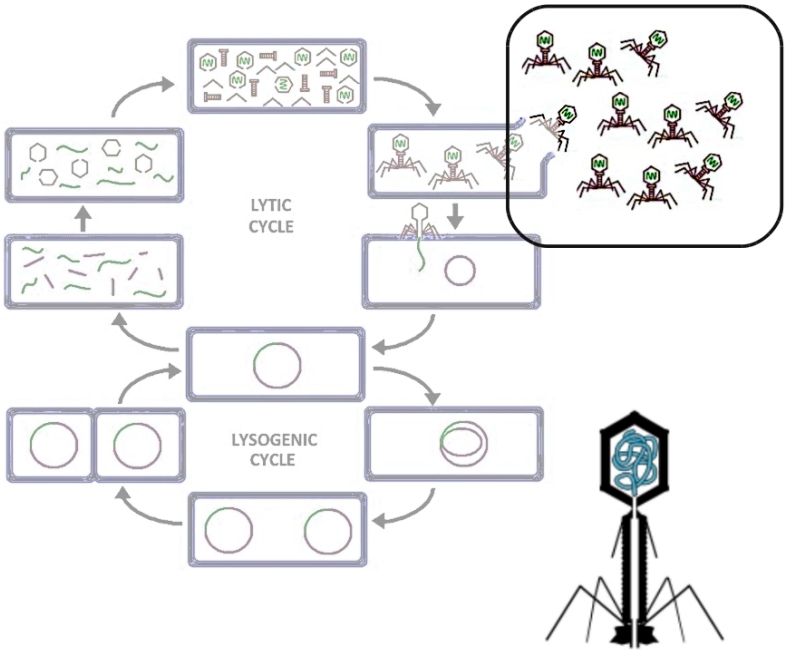
3) Investigate the potential of a lytic bacteriophage that targets *R. reptotaenium* to mitigate BBD (*Chapter 4*). Hypothesis: A lytic bacteriophages that specifically infects *R*. *reptotaenium* will mitigate BBD progression in culture and on an infected coral. In this chapter, I evaluate the potential of a virulent bacteriophage that targets the main BBD cyanobacterium, *R*. *reptotaenium,* to mitigate BBD in a phage therapy. I sequenced the genomes of three bacteriophage isolated from *R*. *reptotaenium* lysate and analysed the genomes to understand host range, the infection process and to assess temporal patterns in bacteriophage abundance. This study assessed if bacteriophages might have potential beneficial roles for disease mitigation strategies in BBD.

4) Examine lysogenic bacteriophages as contributors to the virulence of the main BBD cyanobacterium, *R. reptotaenium* (*Chapter 5*). Hypothesis: Bacteriophages that infect BBD cyanobacteria contribute to their virulence by introducing disease relevant genes. Here, I investigate the influence of lysogenic bacteriophages on the virulence of BBD cyanobacteria *R. reptotaenium* and *Geitlerinem*a sp. BBD_1991. For this chapter, I sequenced and assembled the first draft genome of *R*. *reptotaenium*. I analysed the *R*. *reptotaenium* draft genome as well as a publicly available genome of a second BBD cyanobacterium, *Geitlerinema* sp. BBD_1991, for the presence of bacteriophages (i.e., prophage regions), introduced virulence genes, and bacterial defence mechanisms.

5) Synthesise the results of the previous chapters to elaborate on current BBD pathogenicity models (Chapter 6). In this chapter, I summarise the results of the thesis and provide further details regarding their interpretation and discussion. To fully synthesise the results, I developed an updated BBD pathogenicity model, which includes bacteriophages as either potential contributors to or mitigators of the disease. Additionally, I discuss the multi-faceted roles of viruses in coral health, and provide future research direction to advance the field.

Chapter 2

T4-bacteriophage diversity associated with black band disease in corals



A portion of this chapter is in preparation as:
Buerger P, Weynberg KD, Wood-Charlson EM, Willis BL, van Oppen MJH (in prep) T4-bacteriophage diversity associated with black band disease in corals. Target journal: *Environmental microbiology*.

2.1 Abstract

Marine viruses that infect bacteria (bacteriophages) may mitigate or in some cases promote disease. If lysogenic bacteriophages introduce new genetic material into their bacterial host genome they may increase the virulence of the host bacterium. If lytic bacteriophages kill their host, thereby decreasing the abundance of pathogenic bacteria, they may mitigate disease. Here, I characterised prokaryote and T4-bacteriophage communities associated with black band disease (BBD) using amplicon-based next generation sequencing of *gp23* and 16S rRNA-genes, respectively. To identify candidate phages that might play a role in BBD pathogenesis, communities of T4-like bacteriophages and prokaryotes were compared among samples of BBD lesions, healthy coral tissue, and seawater collected from or near the coral *Montipora hispida* in the central Great Barrier Reef. In accordance with past studies, BBD samples were associated with cyanobacteria (*Oscillatoria*), sulphate-reducing bacteria and Gammaproteobacteria, whereas healthy tissue and seawater were associated with significantly higher abundances of Endozoicimonaceae and Pelagibacteraceae. BBD-associated bacteriophage communities showed 14 bacteriophage OTUs that were uniquely present in BBD samples, including *de novo* OTUs of the taxa Cyanophage PRSM6, *Synechococcus* phage S-SSM4, and *Prochlorococcus* phage P-SSM2. In addition, bacteriophage communities in BBD lesions and healthy coral tissues had higher tetranucleotide variability compared to seawater samples, indicating either more strain variation within tissue samples or that these communities were relatively isolated from those of neighbouring coral colonies and may have developed independently from each other. I hypothesise that bacteriophages unique to BBD lesions contribute to disease progression and virulence by lysing their bacterial hosts and redistributing organic matter, and by introducing new genetic material into their hosts during lysogeny.

2.2 Introduction

Diagnostics and mitigation strategies are urgently needed, but will only be effective if disease etiologies are fully understood (Pollock et al. 2011). However, for the majority of coral diseases described to date, knowledge of pathogens is still limited (Harvell et al. 2007; Sheridan et al. 2013). While most coral disease studies have focused on the role of bacteria in pathogenesis, the influence of viruses on coral health might be substantial (van Oppen, Leong, and Gates 2009; Thurber and Correa 2011), but has received little attention (Cervino et al. 2004; Davy et al. 2006; Lawrence et al. 2015; Pollock et al. 2014; Soffer et al. 2014).

Black band disease (BBD) is prevalent on coral reefs worldwide and infect a wide range of mostly hard-coral species (Page and Willis 2006; Sutherland, Porter, and Torres 2004). A suite of different bacteria have been identified in close association with the disease, such as the cyanobacterium species *Roseofilum reptotaenium*, sulphate-reducing *Desulfovibrio* bacteria, *Cytophaga*, Alphaproteobacteria, and various other heterotrophic microbes (Cooney et al. 2002; Miller and Richardson 2011; Sato, Willis, and Bourne 2010). However, among the members of the black band microbial community, cyanobacteria and *Desulfovibrio* bacteria have been recognised as the main drivers of the disease (Brownell and Richardson 2014; Stanic et al. 2011; Den Uyl et al. 2016). Our understanding of BBD has progressed considerably over the last few decades, especially due to bacterium-centred research, however essential information regarding the causative agent triggering the initial onset of the disease and strategies for managing and mitigating the disease is still missing, and a potential role for viruses in BBD has not been considered. Recent studies have already shown that diverse marine viral assemblages are present within healthy, diseased and stressed coral colonies (Hewson et al. 2012; Marhaver, Edwards, and Rohwer 2008; Soffer et

al. 2014; Thurber et al. 2008). Such community-based approaches are important first steps to characterise coral-associated viruses and identify potential hidden drivers of diseases (Wood-Charlson et al. 2015).

In this study, I explore whether a distinct bacteriophage community is associated with BBD mats infecting corals from the central Great Barrier Reef. I hypothesise that bacteriophages are unique to BBD lesions, healthy coral tissue and seawater. Therefore, I compare bacteriophage communities associated with a temporal sequence of BBD-affected corals (as the disease transitioned from low prevalence in winter to high prevalence in summer) to communities associated with healthy tissue. Since extensive research has been conducted on bacteriophages in seawater, results of the tissue analyses are compared to the respective seawater communities as an outgroup and control. Although a range of different bacterial species are abundant in the BBD mat, I focused on bacteriophages that infect the main BBD cyanobacterial species (cyanophages), as such bacteriophages are likely to influence the disease, or alternatively, are promising candidates for use in disease mitigation. While cyanophages are included in the order Caudovirales and are classified into the families *Myoviridae*, *Siphoviridae*, and *Podoviridae* (reviewed in Shestakov and Karbysheva 2015; Saffermann et al. 1983), the diversity of cyanophages that target BBD associated cyanobacteria is unknown. Bacteriophages that infect marine filamentous cyanobacteria have not been isolated to this date, only bacteriophages infecting freshwater filamentous cyanobacteria, such as Cyanophage N-1 that infects *Nostoc* cyanobacteria (Chénard et al. 2016), and a variety of bacteriophages that infect marine unicellular cyanobacteria, such as *Prochlorococcus* phages P-SSP7, P-SSM2 and P-SSM4 (Sullivan et al. 2005) and *Synechococcus* phage Syn5 (Pope et al. 2007). A large part of the cyanophage diversity is found within

the T4-bacteriophages in the family *Myoviridae* (Comeau and Krisch 2008), which is one of the most diverse bacteriophage groups and shares a conserved gene, the major capsid protein (*gp23*), suitable for genetic identification and diversity estimates (Ackermann and Krisch 1997; Filée et al. 2005). Accordingly, this study explores coral-associated T4-bacteriophage diversity through PCR targeted at the *gp23* gene, followed by amplicon-based next generation sequencing. Although this approach misses the diversity of cyanophages with divergent *gp23* gene sequences, this approach covers BBD-associated cyanophages in one of the most diverse and widely distributed bacteriophage groups, T4-bacteriophages (reviewed in Clokie et al, 2010; Filée et al. 2005; Comeau and Krisch 2008). Additionally, prokaryote communities were analysed to demonstrate that the samples accurately reflect the known respective bacteria species diversity, as identified by analysis of the 16S rRNA gene through amplicon-based next generation sequencing on the same samples.

2.3 Methods

2.3.1 Coral tissue sampling and DNA extractions

A series of tissue samples from healthy and BBD-infected colonies of the coral *Montipora hispida* were collected from reefs surrounding Orpheus Island (18.6376 S, 146.4982 E) at depths of 3 - 5 m during 2007 and 2008 (BBD: n = 6 samples; healthy: n = 4 samples; each sample from a different colony; Appendix 2.1). While BBD samples were collected as the disease transitioned from low prevalence in winter months to high prevalence in summer months, healthy coral tissue samples were collected only at the end of the sampling period. BBD mats with underlying coral tissue and tissue from healthy corals (controls) were removed with separate sterile razor blades (20-50 mg / sample), preserved in 100 % ethanol and stored at -20°C until further processing. Before DNA extraction, ethanol was evaporated from samples for

10 min at 30°C. Whole DNA content of coral tissues and BBD mats (~ 20 mg) was extracted for T4-bacteriophage and prokaryote community analyses with a Mo-Bio Power Plant Pro kit (cat. no. 13400-50), according to the manufacturer's recommendations, with the following modifications. Samples were crushed and lysed by bead beating in Power Plant Pro kit solution PD1 (450 µL), PD2 (50 µL) and RNase A (3 µL, 25 mg/mL) for 60 seconds at max speed (BioSpec 1001, Mini-Beadbeater-96) to disrupt cells and to remove RNA-contamination. Proteinase K (15 µL, 20 mg/mL) was added to each sample, incubated for 1 hour at 56°C, 10 min at 65°C, and bead beaten again for 1 min at max speed to open virus capsid proteins that contain the target DNA. DNA was eluted from columns with 50 µL TE, incubation time 2 x 5 min.

2.3.2 Seawater sampling and DNA extractions

Seawater samples were collected during 2012 - 2015 near healthy (control) and BBD-affected corals to supplement this project (BBD n = 8; Healthy-tissue n = 6; Appendix 2.1). Approximately 20 L of seawater were collected in five 4 L container bags less than 5 cm above the coral surface, while each coral colony (healthy control or diseased) was sampled once. Water samples were stored in carboys on ice water and in the dark, and transported to the Australian Institute of Marine Science (AIMS) before being stored at 4°C. The following day all seawater samples were filtered through 0.22 µm (Sterivex, GV, Cat no. SVGV010RS, Lot. No. 412H6407; Billerica, MA, USA) to capture prokaryote al communities for DNA extraction. The filters were stored at -20°C until further processing. Cells of prokaryotes captured on Sterivex filters were disrupted by replacing seawater with a lysis buffer (1 M Tris-Hcl, 0.5 M EDTA, 25.6 g Sucrose in 100 mL MiliQ) and incubated for 24 h. Disrupted cells (500 µL) were taken for DNA extraction with Mo-Bio Power Plant Pro kit according to the previously described methods. The remaining virus communities in the seawater filtrate, after

removal of prokaryotes with Sterivex filters, were concentrated to approximately 200 mL by tangential flow filtration (TFF, filter cassette: 10 kDa) with subsequent 50 mL back-flush. Viral DNA in concentrated seawater samples was extracted for T4-bacteriophage analyses with a Mo-Bio Power Plant Pro kit, using the methods described above (20 µL of viral concentrate at ~$2x10^6$ - 10^7 viral particles mL^{-1}).

2.3.3 Polymerase chain reaction (PCR)

The T4 *gp23* capsid protein gene was targeted with the primer pair MZIA6 and MZIA1bis (Filée et al. 2005), and the bacteria 16S rRNA gene V3 and V4 regions were amplified with the primers Bakt_341F and Bakt_805R (Herlemann et al. 2011; Klindworth et al. 2013) in a PCR, both with an expected amplicon size of ~500 bp (Appendix 2.2). A touchdown PCR protocol was used to amplify the marker genes with a MyTaq polymerase (Bioline, BIO-25041, for cycle details see Appendix 2.3). The amplified 16S rRNA and T4 *gp23* PCR products were loaded on an agarose gel (TBE, 1.5 %) and amplicons in the range of 250 - 600 bp were excised and purified with a Qiaquick Gel Purification kit (Qiagen). Illumina sequencing adapters were added to the purified DNA in 15 additional amplification cycles (Appendix 2.3). All samples were normalised to a volume of 20 µL at a DNA concentration of 5 ng $µL^{-1}$ and sent for next generation sequencing (Miseq 2x300 v3, Ramaciotti Centre, UNSW). Paired-end sequences were merged with the software PEAR 0.9.5 (average merged reads: 92.2%). Samples with a low read merge of 50% or less were not considered for further analyses because of potential under-representation of associated communities. Merged reads with a phred score below 30 and length below 75 bases were removed. Further, primer sequences and duplicates were removed from data sets with Fastx v0.0.13 to account for PCR amplification bias and to subsequently compare biodiversity detected among samples (bioinformatics pipeline code in Appendix 2.4).

2.3.4 Prokaryote community composition based on 16S rRNA gene sequences

Prokaryote biodiversity among and within samples was analysed using the software package Quantitative Insights Into Microbial Ecology (QIIME version 1.8.0, bioinformatics pipeline code in Appendix 2.4, qiime.org). In brief, sequences with a minimum of 97% sequence similarity were grouped into *de novo* operational taxon units (OTUs), aligned with Python Nearest Alignment Space Termination (PyNAST, Caporaso et al. 2010), and identified taxonomically with the Greengenes 16S rRNA database (version gg_13_8 DeSantis et al. 2006). OTUs with a taxonomic assignment to Chlorophyta were taken out of analyses (data access Appendix 2.5). Species richness (number of unique OTUs) was measured with Chao1 index to compare alpha diversity. Relative OTU diversity (percent composition) was transformed into a Bray-Curtis distance matrix to compare beta diversity and visualised in a multidimensional scaling plot (MDS). OTU assignments were filtered for taxa that were associated with individual sample types (group-specific-OTUs) with the QIIME native script group_significance.py.

2.3.5 T4-bacteriophage community composition based on *gp23* gene sequences

All merged *gp23* reads were uploaded to Metavir for taxonomic identification (viral RefSeq, database from 05.01.2015) (Roux et al. 2011). Absolute values of taxonomy assignments were transformed into relative abundances. The assigned Metavir taxonomy was filtered for taxa that were associated with a respective sample type (see section 2.3.7 below). In addition, bacteriophage *gp23* sequences were analysed in QIIME with the built-in *de novo* OTU picking algorithm to compare respective alpha and beta diversities (UCLUST cut-off at 97% nucleotide sequence similarity, selected as an alternative diversity test to the Metavir taxonomy, with highest diversity separation among sampling groups based on tests across sequence similarities of 45 -

97%; Appendix 2.6). Sequences were aligned with MAFFT and alpha-rarefaction curves were calculated with Chao1 index. Overall OTU abundance was converted into compositional abundances (relative abundance), transformed into a Bray-Curtis matrix and visualised in an MDS plot to visualise overall community differences. Temporal patterns were revealed by plotting the relative OTU abundance over time. Additionally, the dataset was filtered for OTUs that were dominant and uniquely associated with a fraction of the three sample types, with the following criteria: 1) an OTU was considered to be 'uniquely' associated with a sample type if at least 95% of its relative abundance across all samples was associated with one sampling group, and 2) an OTU was considered to be dominant in either BBD or healthy tissue samples if one group contributed at least 50% to the relative abundance of the OTU, and the other group contributed less than 5% of its abundance. To avoid potential sequencing errors, chimeras, and inclusion of under-represented sequences, OTUs below 5% relative abundance across all sampling groups were removed prior to this step.

2.3.6 Nucleotide frequencies

Sequence similarity cut-offs may not accurately reflect taxonomic boundaries, such as boundaries among species or other taxonomic classifications (Erko and Ebers 2006). In addition, sequence similarity cut-offs are not established for virus communities, as species boundaries between viruses are less defined compared to bacteria, genetic substitution can occur frequently and rates can be variable. I therefore used tetranucleotide frequencies of the 16S rRNA and T4 *gp23* data to analyse variability among samples, as an additional measurement that is independent from database matches and clustering algorithms (perl script Appendix 2.7). Tetranucleotide frequencies were transformed into relative percentages to normalise values across samples, then transferred into a Bray-Curtis distance matrix, and visualised in a

principal component analysis. The variance between sample types was analysed by calculating their multivariate dispersions (distance of sample to respective group centroid) and visualised in a box plot (scripts in Appendix 2.7) (Anderson 2006; Anderson, Ellingsen, and McArdle 2006).

2.3.7 Statistics

Relative QIIME OTU assignments for bacteria (16S rRNA), and T4-bacteriophage (*gp23*) Metavir taxonomy were transformed into a Bray Curtis distance matrix and tested for significant differences among the 4 communities sampled (BBD, Healthy-tissue, Seawater (SW)-BBD, SW-Control), with a one-way PERMANOVA (permutations n = 9999, Bonferroni sequentially corrected p-values). Individual QIIME OTU and Metavir taxa assignments were tested for significantly different abundances among sample types (BBD, Healthy-tissue, SW-BBD, SW-Control) (QIIME group_significance.py) with a Kruskal-Wallis test and a false discovery rate (FDR) for p-value correction. Variance of tetranucleotide frequencies among sample types (BBD, Healthy-tissue, SW-BBD, SW-Control) in form of their PCoA multivariate dispersions (distance of sample to respective group centroid) were compared in a permutation test (permutations, n = 9999) with Tukey's HSD post-hoc comparisons (scripts in Appendix 2.7).

2.4 Results

2.4.1 Taxonomy assignments: 16S rRNA gene

Prokaryote communities associated with BBD mats, healthy tissue and seawater formed separate, significantly different clusters on an MDS plot (Fig. 2.1A, PERMANOVA; BBD vs. Healthy-tissue: $p = 0.0012$; BBD vs. SW-BBD: $p = 0.0008$; Healthy-tissue vs. SW-Control: $p = 0.0084$). However, seawater collected above BBD

mats and healthy corals harboured prokaryote communities that were indistinguishable using these methods (SW-BBD vs. SW-Control: $p = 0.8316$). In addition, tissue-associated prokaryote communities showed lower alpha diversity in comparison to seawater prokaryote communities (OTU richness, Fig. 2.2A). Bacterial genera that had significantly higher abundances in BBD mats than in other samples included: the cyanobacterium *Oscillatoria*, the sulphate-reducing bacterium *Desulfovibrio*, *Arcobacter* (Class: Epsilonproteobacteria), and *Thalassomonas* (Class: Gammaproteobacteria). Healthy coral tissue samples included bacterial sequences associated with Endozoicimonaceae (Class: Gammaproteobacteria) and Myxococcales (Class: Deltaproteobacteria). Seawater samples had higher enrichment of Actinobacteria, OCS155, Synechococcaceae and Pelagibacteraceae (Fig. 2.3A, Table 2.1).

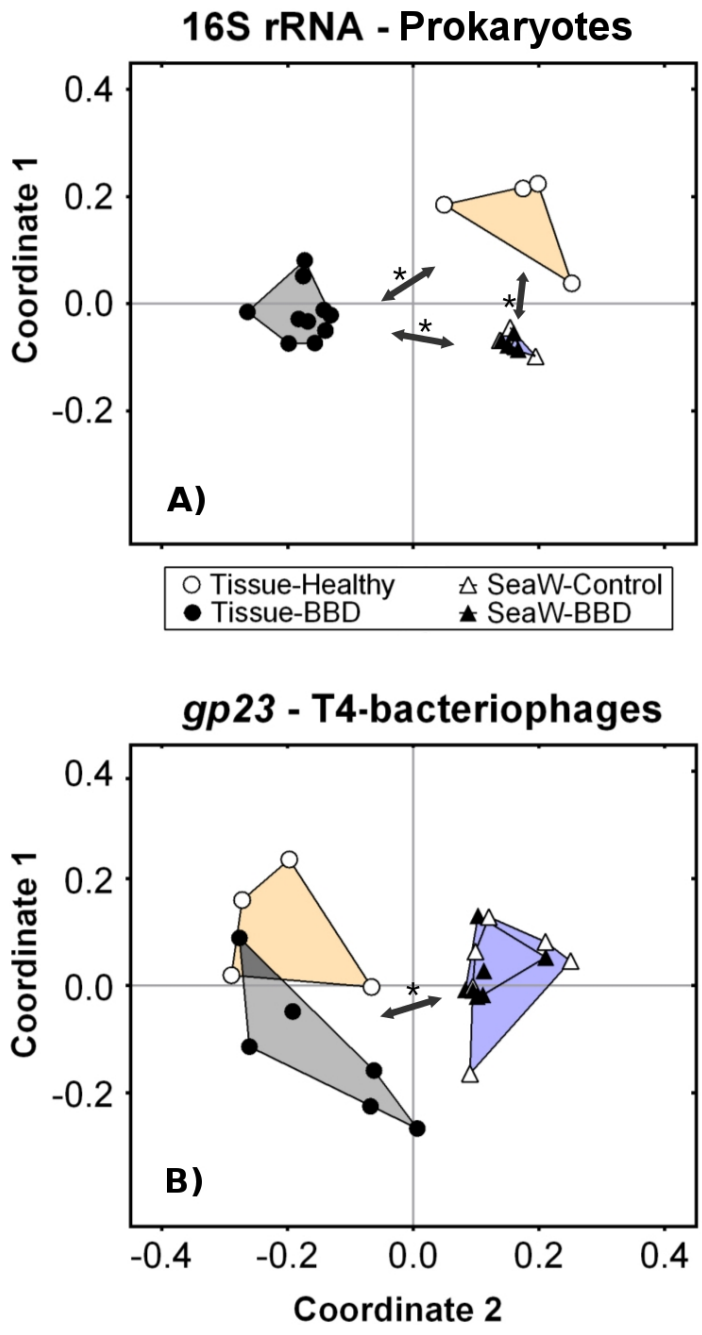


Figure 2.1 Community profiles of prokaryote and bacteriophage OTUs. (A) Prokaryote 16S rRNA communities and (B) T4-bacteriophage *gp23* communities displayed as multidimensional scaling plots. Relative abundances of OTUs (97% sequence similarity) were transformed into a Bray-Curtis distance matrix. Seawater samples (overlapping each other) = SW-Control and SW-BBD; BBD = black band disease mat, Healthy-tissue = coral tissue samples. MDS stress values: 16S rRNA = 0.151; *gp23* = 0.181. * = significant difference, level < 0.05.

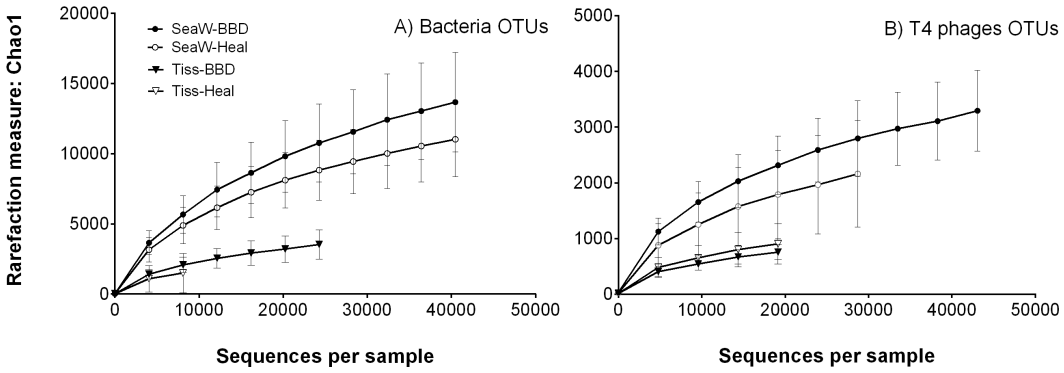


Figure 2.2 Alpha-rarefaction curves for seawater and tissue samples. Seawater bacteria communities had a higher species richness than tissue samples, due to their longer and steeper alpha-rarefaction curve. However, it is not possible to compare T4 with 16S rRNA species richness data in this case, because they target marker genes with a different initial range of biodiversity. It is likely that 16S rRNA primers have a much wider range of targets than the T4-bacteriophage primers.

Table 2.1 Relative abundance and BLAST identification of selected bacteria species. FDR = false discovery rate from statistical comparison, significance threshold: 0.05. Selected taxa are known associates with their respective sample type.

Sample type	Taxon	Relative abundance [%]	FDR	BLAST match (e-value)	Reference accession
BBD	Oscillatoria, *R. reptotaenium*	18.0 ± 7.9	0.033	0.0	KU579387.1
BBD	*Desulfovibrio*	5.6 ± 4.1	0.042	0.0	AF473944.1
BBD	Epsilonproteobacteria *Arcobacter*	9.4 ± 6.5	0.033	0.0	GU472139.1
BBD	Gammaproteobacteria *Thalassomonas*	7.3 ± 7.0	0.033	0.0	HM768482.1
Healthy	Gammaproteobacteria Endozoicimonaceae	10.0 ± 14.7	0.033	0.0	EF466050.1
Seawater	Actinobacteria, OCS155	11.5 ± 4.6	0.040	0.0	KT318695.1
Seawater	Synechococcaceae	7.1 ± 0.9	0.033	0.0	KU867935.1
Seawater	Pelagibacteraceae	7.6 ± 1.8	0.035	0.0	KM223722.1

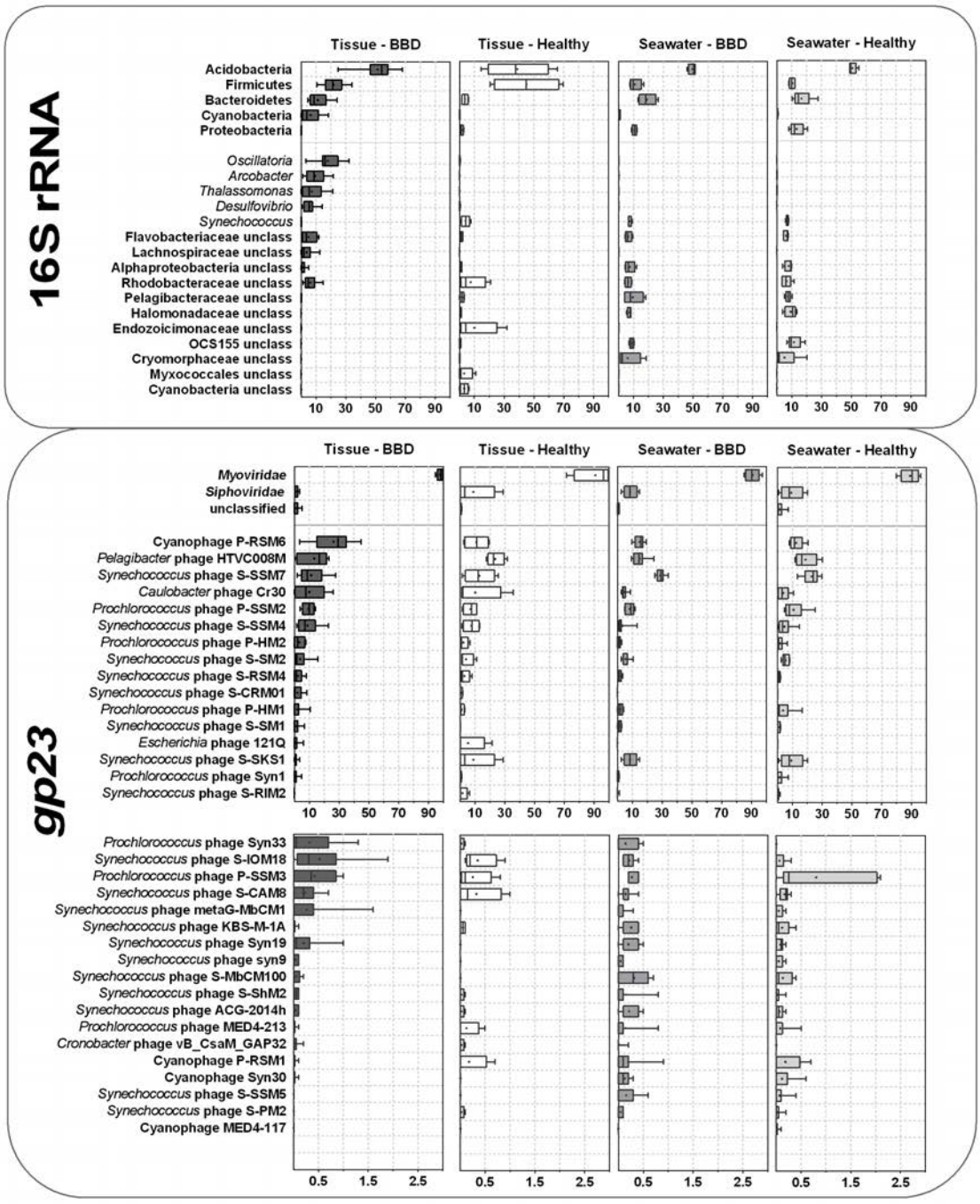


Figure 2.3 Taxonomy of bacteria and T4-bacteriophage communities. Top: Taxonomies assigned to bacteria with QIIME and the Greengenes database shown as phylum and either genus or family (lowest and highest meaningful projections, respectively). Bottom: Taxonomies assigned to bacteriophages with Metavir shown as family and species; bar graphs show higher (<5%) and lower (>5%) abundances.

2.4.2 T4-bacteriophage community composition based on *gp23*

In accordance with the methodology and used primer set used, which targets T4-bacteriophages as part of the *Myoviridae*, the family *Myoviridae* (Order: Caudovirales) was the most abundant bacteriophage group detected in all samples (Fig. 2.3B). BBD samples showed a significantly higher relative abundance of *Myoviridae* and lower abundance of *Siphoviridae* compared to SW samples (PERMANOVA; BBD vs. SW-BBD $p = 0.008$; BBD vs. SW-Control $p = 0.005$; note: only one taxon of *Siphoviridae* detected: *Synechococcus* phage S-SKS1). However, due to generally high variability in the abundances of Metavir taxa among samples, no statistically significant differences were detected between BBD and healthy tissues (BBD vs. Healthy-tissue $p = 0.1846$). The assigned T4-bacteriophage taxonomy, e.g., from Metavir, did not provide enough taxonomic resolution to reveal specific bacteriophage-sample type associations. Based on the current dataset, only OTU clustering at a scale of 97% sequence similarity revealed distinct bacteriophages associated with BBD mat, healthy tissue and seawater samples. OTU sequence similarity clusters at 97% revealed more sample type-specific differences in the composition of T4-bacteriophage communities and detailed taxon associations compared to Metavir assignments. Bacteriophage *gp23* communities formed individual clusters on an MDS plot at 97% sequence similarity (Fig. 2.1B), indicating that overall communities differed significantly between seawater and tissue samples (PERMANOVA; BBD vs. SW-BBD: $p < 0.001$: Healthy-tissue vs. SW-Control: $p = 0.005$).

Tissue-associated bacteriophage communities showed a lower alpha diversity compared to bacteriophages in seawater samples (OTU richness, Fig. 2.2). However, bacteriophage community clusters did not differ significantly between healthy tissue and BBD mats, nor between seawater collected above BBD versus above healthy

corals (PERMANOVA; BBD vs. Healthy-tissue: p = 0.134; SW-BBD vs. SW-Control: p = 0.836). However, the relative contribution of sequences to an OTU revealed 36 unique and dominant taxa in seawater samples, 14 taxa in BBD mats and 9 taxa in healthy coral tissue (Table 2.2). Seawater samples were mostly associated with *Synechococcus* S-SSM7 *de novo* OTUs (n = 9 taxa) and *Pelagibacter* phage HTVC008M *de novo* OTUs (n = 7 taxa), while healthy coral tissue samples had one unique OTUs (each n = 1) from the taxa *Caulobacter* phage Cr30, *Prochlorococcus* phage P-SSM2, Cyanophage PRSM6 and *Synechococcus* phages. In contrast, BBD samples were mostly associated with Cyanophage PRSM6 *de novo* OTUs (n = 4 out of 14). Temporal patterns in the abundance of BBD-associated bacteriophage OTUs were detected, with three OTUs present at high relative abundances in winter but absent in summer, and six OTUs increasing in relative abundance from winter to summer. In addition, 5 BBD-associated OTUs tended towards higher relative abundance in the transition period between winter and summer months, although abundances did not differ significantly among months (Fig. 2.4).

Table 2.2 Relative sequence contribution to OTUs associated with sample types. The listed OTUs are either uniquely abundant (> 95%) or dominant (> 50%) within a sample type. SW-Control and SW-BBD are listed in one column, due to their similarity. The column "# sequences" shows in brackets the numbers of samples that contained the respective OTU.

Sample group	Unique OTUs				Relative OTU sequence contribution		
	#	taxon	OTU	# sequences	Seawater	BBD	Healthy
	1	*Caulobacter* phage Cr30	denovo22551	14872 (10/13)	97.53	1.10	1.36
S	1	Cyanophage PRSM6	denovo27975	3207 (09/13)	100.00	0.00	0.00
E	2		denovo8379	7157 (10/13)	100.00	0.00	0.00
A	3		denovo14889	8576 (06/13)	99.44	0.00	0.56
W	4		denovo30262	1896 (07/13)	98.07	1.93	0.00
A	1	*Pelagibacter* phage HTVC008M	denovo26229	6693 (09/13)	100.00	0.00	0.00
T	2		denovo14891	9329 (11/13)	99.99	0.00	0.01
E	3		denovo14883	3328 (08/13)	99.95	0.00	0.05
R	4		denovo4488	7365 (13/13)	99.76	0.24	0.00
	5		denovo27976	3276 (07/13)	98.72	0.51	0.77
	6		denovo9311	3388 (06/13)	98.39	0.00	1.61
	7		denovo7627	14378 (10/13)	97.29	2.12	0.59
	1	*Prochlorococcus* phage P-SSM2	denovo19412	18759 (09/13)	100.00	0.00	0.00
	2		denovo34103	3714 (07/13)	100.00	0.00	0.00
	3		denovo4490	13301 (10/13)	99.99	0.00	0.01
	1	*Synechococcus* phage S-SKS1	denovo16637	4455 (12/13)	100.00	0.00	0.00
	2		denovo2710	2977 (09/13)	100.00	0.00	0.00
	3		denovo3944	3971 (11/13)	100.00	0.00	0.00
	4		denovo14890	9600 (13/13)	99.98	0.00	0.02
	1	*Synechococcus* phage S-SM2	denovo26102	3630 (06/13)	100.00	0.00	0.00
	2		denovo25008	5550 (11/13)	99.81	0.19	0.00
	3		denovo21071	3426 (11/13)	99.66	0.32	0.02
	4		denovo11421	2143 (04/13)	99.48	0.34	0.18
	1	*Synechococcus* phage S-SSM2	denovo7552	5082 (04/13)	99.96	0.03	0.02
	1	*Synechococcus* phage S-SSM4	denovo15239	6781 (08/13)	100.00	0.00	0.00
	2		denovo26227	2747 (08/13)	100.00	0.00	0.00
	3		denovo2267	3104 (04/13)	99.94	0.06	0.00
	1	*Synechococcus* phage S-SSM7	denovo11432	3956 (11/13)	100.00	0.00	0.00
	2		denovo17714	4011 (07/13)	100.00	0.00	0.00
	3		denovo3468	2718 (11/13)	100.00	0.00	0.00
	4		denovo8374	28704 (13/13)	99.93	0.05	0.02
	5		denovo27971	6559 (09/13)	99.96	0.00	0.04
	6		denovo19236	2683 (12/13)	99.82	0.13	0.05
	7		denovo29656	3259 (11/13)	95.66	3.44	0.91
	8		denovo4492	11958 (12/13)	96.57	1.58	1.84
	9		denovo32871	11688 (12/13)	97.43	0.22	2.35
	1	*Caulobacter* phage Cr30	denovo20855	44338 (2/6)	38.02	60.37	1.61
B	1	Cyanophage PRSM6	denovo25920	4668 (3/6)	0.24	99.72	0.04
B	2		denovo15550	9341 (3/6)	7.58	80.50	11.92
D	3		denovo23699	39580 (5/6)	28.16	70.62	1.22
	4		denovo17715	2585 (5/6)	47.66	52.34	0.00
	1	*Pelagibacter* phage HTVC008M	denovo10787	2328 (5/6)	15.04	74.46	10.50
	1	*Prochlorococcus* phage P-SSM2	denovo17862	2964 (5/6)	0.29	96.74	2.97
	2		denovo496	2223 (2/6)	0.95	96.04	3.01
	1	*Synechococcus* phage S-RSM4	denovo16827	1486 (3/6)	3.39	96.31	0.30
	1	*Synechococcus* phage S-SM2	denovo14888	4418 (2/6)	27.62	72.38	0.00
	1	*Synechococcus* phage S-SSM4	denovo7136	4447 (3/6)	0.01	98.69	1.30
	2		denovo14617	2219 (4/6)	0.00	96.78	3.22
	3		denovo15559	675 (2/6)	8.18	85.52	6.30
	1	*Synechococcus* phage S-SM1	denovo9281	2707 (2/6)	19.67	80.18	0.15
	1	*Caulobacter* phage Cr30	denovo30660	8000 (2/4)	0.04	0.03	99.92
H	1	*Escherichia* phage 121Q	denovo14835	4907 (2/4)	0.01	0.05	99.94
E	1	*Prochlorococcus* phage P-SSM2	denovo22903	6407 (2/4)	12.36	4.55	83.09
A	1	Cyanophage PRSM6	denovo10444	6271 (1/4)	0.03	0.02	99.94
L	1	*Synechococcus* phage S-RSM4	denovo17932	1937 (1/4)	7.31	0.00	92.69
T	1	*Synechococcus* phage S-SKS1	denovo27948	16022 (3/4)	0.33	0.00	99.67
H	1	*Synechococcus* phage S-SM2	denovo1441	2787 (1/4)	29.97	3.40	66.63
Y	1	*Synechococcus* phage S-SSM7	denovo29813	4523 (1/4)	6.40	6.84	86.76
	1	*Synechococcus* phage S-RIM2	denovo28949	5916 (1/4)	37.75	0.00	62.25

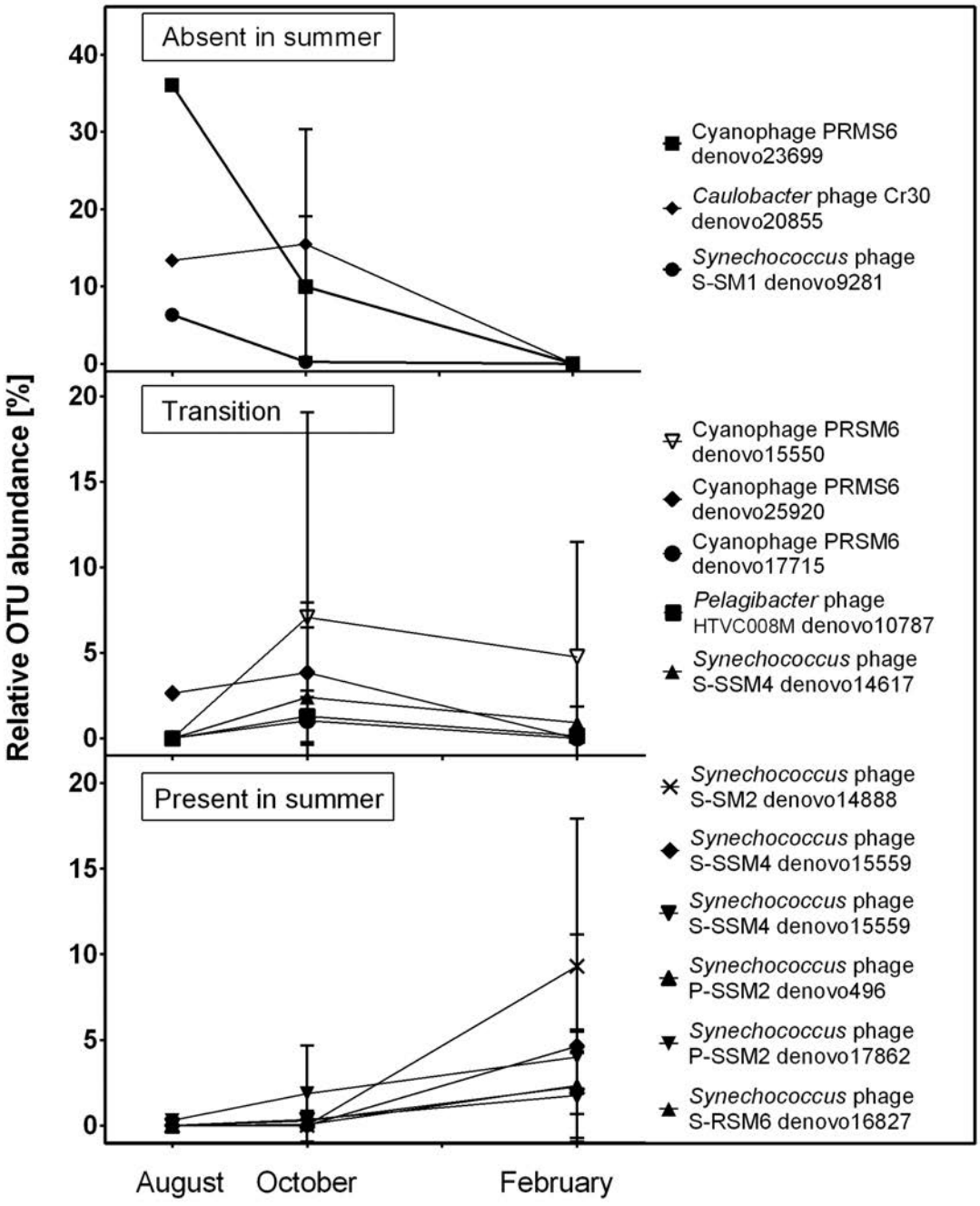


Figure 2.4 Temporal patterns of BBD-specific OTUs. Top: BBD-associated OTUs (Table 2.2) decrease in relative abundance from August (Australian winter) to February (Australian summer). Middle: BBD-associated OTUs that do not show a peak during the transition time from winter to summer. Bottom: BBD relevant OTUs that are present in summer while being absent in winter. Note change in scale (y-axis) for the different graphs to improve the readability.

2.4.3 Nucleotide frequencies

Tetranucleotide frequencies of BBD-associated prokaryotes had the lowest variability of all samples (Fig. 2.5B). While variability of prokaryote communities was significantly lower in BBD samples than in healthy coral tissue samples ($p_{adjusted} < 0.008$, Appendix 2.8), variability of T4-bacteriophage communities was similar between the two sample types ($p_{adj} > 0.999$). However, T4-bacteriophage communities were more variable in BBD than in seawater samples ($p_{adj} < 0.038$). In general, tetranucleotide frequencies had greater variability among tissue samples, which were collected within months of each other, than among seawater samples, which were sampled over a timespan of two years. A larger sample size could have reduced the variability within coral tissue samples. However, the general difference of the variability between seawater samples and tissue is likely to remain also with larger sample size.

2.5 Discussion

Distinct prokaryote communities were associated with each of the three sample types tested, i.e., BBD, healthy coral tissue and seawater. In contrast, T4-bacteriophage communities associated with BBD and healthy coral tissues were similar, but these two communities differed in comparison to seawater communities. In general, coral tissue samples showed higher variability in nucleotide frequencies of associated prokaryote and T4-bacteriophages than seawater samples. A bacteriophage OTU analysis using a 97% sequence similarity cut-off revealed 14 viral OTUs that were uniquely associated with and dominant in BBD samples. Of these 14 BBD-associated bacteriophage OTUs, three were abundant in summer but absent in winter, while seven were more abundant in summer.

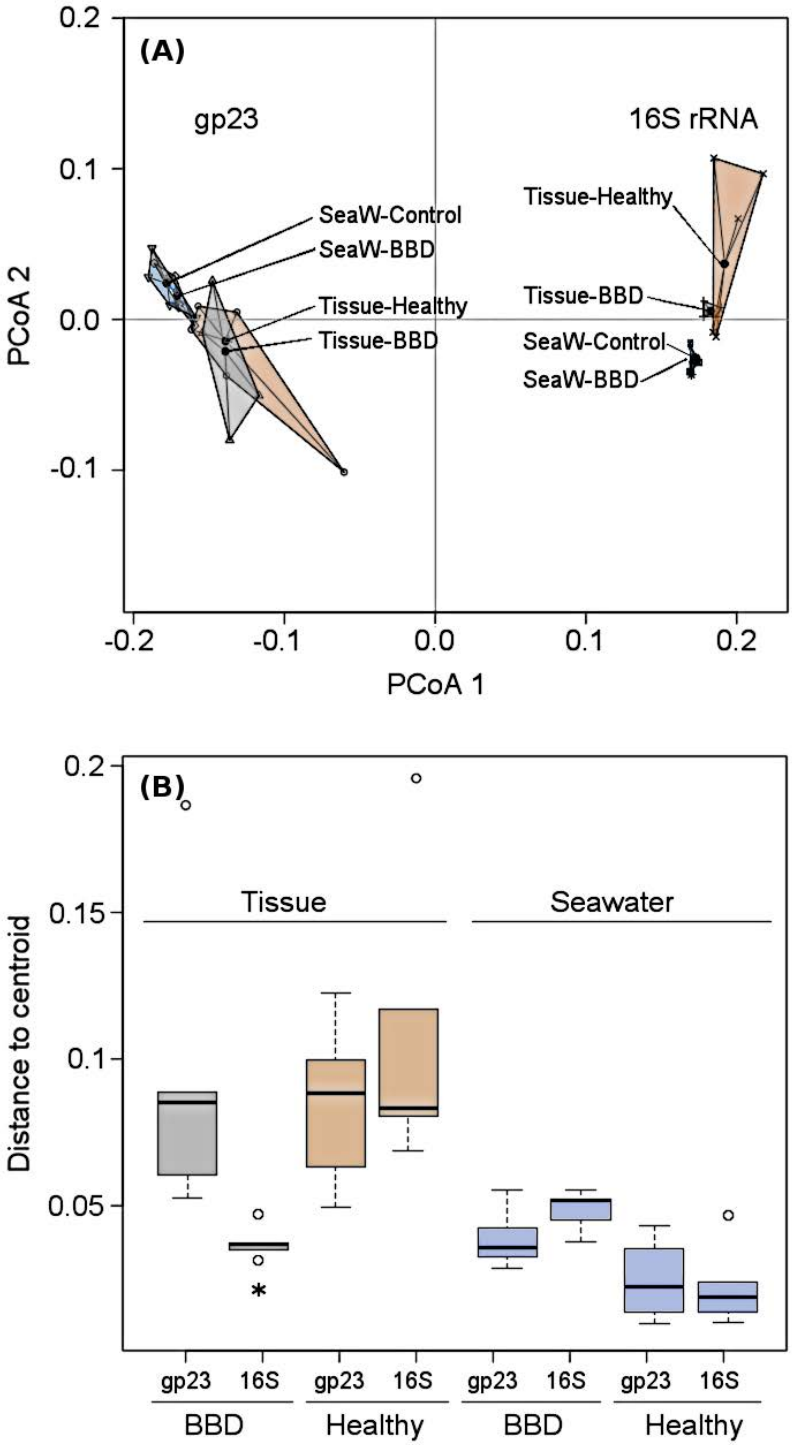


Figure 2.5 Prokaryote (16S rRNA) and T4-bacteriophage (gp23) community differences in tetranucleotide frequencies. A) Principal component analysis (PCoA) of tetranucleotide frequencies. B) Variability of bacteria and T4-bacteriophage tetranucleotide frequencies are shown as boxplots and calculated from the distances of respective samples to their centroid (centre of the respective PCoA of tetranucleotide frequencies. Seawater samples were collected over the time scale of several years, while tissue samples were collected within months.

2.5.1 Prokaryote communities

Patterns in the composition of prokaryote communities in the three sample types analysed were similar to patterns of coral-associated and seawater communities described in previous studies. Consistent with earlier studies, the following groups of bacteria were significantly more prevalent in BBD samples (Fig. 2.3, Table 2.1): the cyanobacterium *R. reptotaenium*, Oscillatoria (Buerger, Alvarez-Roa, et al. 2016; Casamatta et al. 2012), sulphate reducing bacteria *Desulfovibrio* (Deltaproteobacteria) (Cooney et al. 2002), Epsilonproteobacteria (Arotsker et al. 2015) and other Gammaproteobacteria such as *Thalassomonas* (Klaus, Janse, and Fouke 2011). Seawater samples were enriched for bacteria taxa that often occur in plankton and seawater samples, such as Actinobacteria OCS155 (Houghton 2015), Synechococcaceae (Hunter-Cevera et al. 2016) and Pelagibacteraceae (West et al. 2016). Healthy coral tissue samples showed significantly higher occurrence of known coral microbial symbionts such as the family Endozoicimonaceae (Gammaproteobacteria) (Apprill et al. 2012; Bayer et al. 2013; Morrow et al. 2012; Neave et al. 2017) and Myxococcales (Deltaproteobacteria) (Garcia et al. 2013).

2.5.2 BBD bacteriophage community

Patterns in BBD-associated T4-bacteriophage communities differed from those found for BBD-associated bacterial communities. Whereas the tetranucleotide composition of the BBD-associated prokaryote community was remarkably stable over six months (Fig. 2.1C), indicating its consistent importance and contribution to BBD, the T4-bacteriophage community showed a higher level of variability among BBD samples. As expected, BBD samples mainly comprised *Myoviridae* bacteriophages (Order: Caudovirales) and had the lowest relative abundance of *Siphoviridae* (Fig. 2.3B). Within the *Myoviridae*, most of the taxa belonged to Cyanophage PRSM6 and

Synechococcus phage S-SSM4 (Table 2.2), and 14 OTUs were unique to and dominant in BBD samples. Some bacteriophages are known to be locally adapted and to most readily infect bacterial hosts in their respective local environment (Koskella et al. 2011; Vos et al. 2009). Since marine *Synechococcus*-, *Prochlorococcus*- and cyanophages infect marine cyanobacteria (Shestakov and Karbysheva 2015), the unique BBD-associated OTUs might represent host-specific bacteriophages that infect bacteria, such as *R*. *reptotaenium* and *Geitlerinema* sp., within the BBD mat.

Several BBD-associated bacteriophages showed temporal patterns in their abundances, with some having higher relative abundance in winter, while other BBD-specific bacteriophage OTUs increased in relative abundance towards summer (Fig. 2.4). Bacteriophage abundances are known to depend on the abundance of their bacterial host, thus they will increase with environmental conditions that are favourable for the growth and abundance of their host bacteria (Chibani-Chennoufi et al. 2004; Chow et al. 2013). For example, host bacteria in the BBD mat are typically less abundant in winter months due to colder seawater temperatures and less available light, and more abundant in summer months due to warmer seawater temperatures and greater light availability (Sato et al. 2009). Also the characteristics of the BBD microbial mat itself change seasonally, gaining in biomass (Kuehl et al. 2011) and progressing more rapidly in response to higher temperatures and higher light (Boyett, Bourne, and Willis 2007; Sato, Bourne, and Willis 2011). In addition, the relative abundance of bacteriophages may also be directly influenced by environmental factors, such as temperature, UV-light, salinity and nutrients (reviewed in Mojica and Brussaard 2014). Such environmental factors are known to inactivate and degrade bacteriophages and influence latent and lytic cycles within their life cycles. Temporal patterns in the abundance of BBD bacteriophages detected could therefore be related

to the abundance of their bacterial hosts, as well as to environmental factors that disintegrate bacteriophages and influence their lytic and lysogenic cycles.

The presence of bacteriophages able to target cyanobacteria in BBD microbial mats is supported by the results of recent studies. As part of a recent genome sequencing study, a CRISPR-Cas defence system has been detected in the genome of the BBD-associated cyanobacterium *Geitlerinema* sp. BBD_1991 (Buerger, Wood-Charlson, et al. 2016; Den Uyl et al. 2016). CRISPR-Cas defence systems help to prevent viral infections and recognise intruding bacteriophages by storing reference sequences (spacers) in repetitive arrays that match the infecting genetic material (Makarova, Aravind, et al. 2011; Makarova, Haft, et al. 2011). The presence of a CRISPR-Cas system in the cyanobacterium *Geitlerinema* sp. BBD_1991 confirms that bacteriophages are targeting this specific cyanobacterium in BBD (Den Uyl et al. 2016). Furthermore, high-throughput shotgun sequencing of the BBD mat has recovered a higher relative proportion of cyanophage sequences in a metatranscriptomic dataset compared to the pre-disease cyanobacterial patch stage (Sato et al. 2017). Most of the BBD metatransriptomic viral reads were identical to photosynthetic gene sequences D1 from the cyanobacteria dominating the BBD lesion, inferring a phage-host association (Sato et al. 2017). While the most abundant bacteriophage taxon in the corresponding BBD metagenome dataset belonged to *Synechococcus* phage S-SM2 (21%), unfortunately the necessity for a pooled study design meant that samples were not replicated in Sato et al. (2017), and due to high variability in the BBD-associated bacteriophages detected in the current study, any comparison of abundance data should be taken with caution. Nevertheless, the complementary results of recent studies confirm that BBD cyanobacteria are likely to be a target of BBD-associated bacteriophages.

2.5.3 Variability of bacteriophages communities

Bacteriophage communities associated with BBD and healthy coral tissue samples collected over the course of 6 months showed higher variability, in terms of associated taxa (Fig. 2.1A, B) and nucleotide frequencies (Fig. 2.5), than seawater samples collected over two years. Each coral colony may harbour its own individual signature of prokaryote communities (Frias-Lopez et al. 2002), although certain taxa, such as *Endozoicomonas,* can occur commonly in microbial communities associated with corals (Neave et al. 2017). Nevertheless, components of coral-associated microbial communities may develop relatively independently on neighbouring corals due to microhabitat and environmental differences around each coral (Apprill et al. 2012; Morrow et al. 2012). Therefore, the high variability among bacteriophage communities detected in this study may have resulted from environmental differences, but could also have resulted from strain variations associated with locally adapted bacteriophages (Koskella et al. 2011; Vos et al. 2009). In terms of their capacity to infect bacterial hosts, bacteriophages can range from being generalists able to infect a wide range of bacteria, to specialists able to infect only a specific type of bacteria (Flores et al. 2011; Koskella and Meaden 2013). Bacteriophages have been shown to differ in their host ranges in infection experiments, e.g. *Prochlorococcus* phage P-SSP7 only infects a high-light adapted *Prochlorococcus* strain (Sullivan, Waterbury, and Chisholm 2003), in contrast to several other cyanophages (Sullivan et al. 2003) and *Synechococcus* phage Syn9 (Weigele et al. 2007) that have a wider host range and can infect many *Prochlorococcus* and *Synechococcus* strains. Some of the overlap of bacteriophage communities found among BBD, healthy tissue and seawater samples, might therefore be attributed to similar environmental factors or to the presence of generalist bacteriophages. In contrast, some of the variation between communities might be attributable to the presence of specialist bacteriophages unique

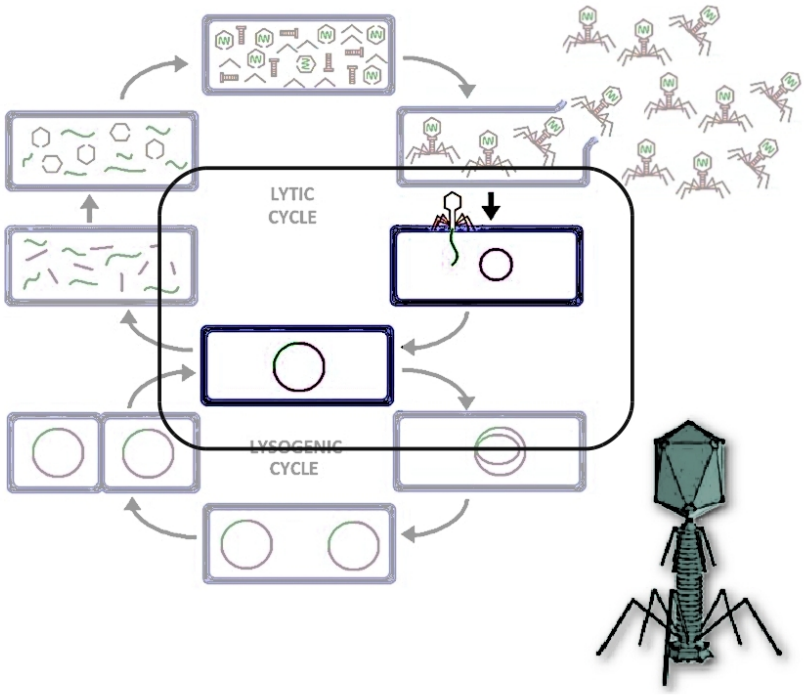
to a particular colony.

2.5.4 Conclusion

Few unique bacteriophage OTUs were only associated with healthy coral tissue and BBD mat samples. The overall variability within sample groups and among sample did mask significant differences between the samples. Nevertheless, due to the specific OTU associations with BBD in the samples tested, I conclude that the BBD-derived T4 cyano-, *Synechococcus* and *Prochlorococcus* phages are likely able to infect cyanobacteria, such as *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991. Future studies should focus on retrieving virome data sets associated with BBD in order to fully analyze bacteriophage communities associated with the disease, as well as on cultivating disease-associated bacteriophages in order to determine their characteristics and host range. This study presents the first evidence of temporal patterns in the abundance of bacteriophage taxa associated with BBD microbial communities and provides amplicon sequences in a public database for future comparative analyses.

Chapter 3

Cultivation of the main BBD cyanobacterium, *Roseofilum reptotaenium*, and of associated bacteriophages



A portion of this chapter is published as:

Buerger P, Alvarez-Roa C, Weynberg KD, Baekelandt S, van Oppen MJH (2016) Genetic, morphological and growth characterisation of a new *Roseofilum* strain (Oscillatoriales, Cyanobacteria) associated with coral black band disease. *PeerJ*. 9;4:e2110.

3.1 Abstract

Black band disease (BBD) is a common disease of reef-building corals with a worldwide distribution that causes tissue loss at a rate of up to 2 cm per day. Critical for a mechanistic understanding of the influence of viruses on the disease etiology is the cultivation of its proposed pathogen, a filamentous cyanobacterium (genus *Roseofilum*), and the isolation of lytic and lysogenic bacteriophages. Here, I optimised existing protocols for the isolation and cultivation of *Roseofilum reptotaenium* using a new strain (AO1) from the central Great Barrier Reef. I demonstrated that the isolation of this strain *via* inoculation onto agar plates was highly effective with a low percentage agar of 0.6% and that growth monitoring was most sensitive with fluorescence measurements of chlorophyll-a (440 / 685 nm). Cell growth curves in liquid and solid media were generated for the first time for this cyanobacterium species and showed best growth rates for the previously untested L1-medium (growth rate k = 0.214 biomass/day; doubling time t_{gen} = 4.67 days). My results suggest that the trace metals contained in L1-medium maximise an increase in biomass over time for this cyanobacterium strain. Since the newly isolated *Roseofilum* strain AO1 is genetically closest to the main BBD-associated cyanobacterium *Pseudoscillatoria coralii*, but in terms of pigmentation and cell size closer to *Roseofilum reptotaenium*, I formally merged the two species into a single taxon by providing an emended species description, *Roseofilum reptotaenium* (Rasoulouniriana) Casamatta emend. In addition, I used the newly established protocols in lytic and lysogenic bacteriophage isolation experiments. *R. reptotaenium* strain AO1 filaments were lysed in bacteriophage enrichments by liquid assays. Lysogenic virus induction for *R. reptotaenium* strain AO1 was not successful with standard methods that involved UV-light and mitomycin C treatment. Following these optimized protocols is recommended for fast isolation and cultivation of *Roseofilum* cyanobacteria, for growth curve generation in strain comparisons, for maximisation of

biomass in genetic studies and for the screening of lytic and lysogenic bacteriophages.

3.2 Introduction

Coral diseases contribute to coral mortality and to the decline of reefs worldwide (Aeby et al. 2015; Bruckner 2015; Frias-Lopez et al. 2003; Page and Willis 2006; Sutherland, Porter, and Torres 2004; Willis, Page, and Dinsdale 2004). Of the over 20 prevalent coral diseases that are known (Sheridan et al. 2013), black band disease (BBD) was the first to be described (Antonius 1973). The BBD microbial community affects a wide range of reef-building coral species and forms a dark mat on the coral surface (Bruckner 2015; Sutherland et al. 2004), which progresses up to several cm per day, and kills the underlying tissue (Rützler, Santavy, and Antonius 1983).

A variety of different bacteria have been identified in the BBD mat, including various species of cyanobacteria, sulphate-reducing *Desulfovibrio* bacteria, *Cytophaga*, *Alphaproteobacteria* and a range of other heterotrophic microbes (Cooney et al. 2002; Miller and Richardson 2011; Sato, Willis, and Bourne 2010). DNA sequence analysis of the 16S rRNA gene indicates that cyanobacteria of the proposed genus *Roseofilum* (Casamatta et al. 2012), closely related to the genus *Oscillatoria*, are found in the disease mat all over the world (Aeby et al. 2015; Arotsker et al. 2015; Casamatta et al. 2012; Cooney et al. 2002; Frias-Lopez et al. 2003; Gantar, Sekar, and Richardson 2009; Glas et al. 2010; Meyer et al. 2015; Miller and Richardson 2011; Rasoulouniriana et al. 2009; Sato et al. 2010; Sussman, Bourne, and Willis 2006). In a non-axenic culture, *R. reptotaenium*, is able to cause a progressing BBD lesion in healthy corals a few days after infection (Casamatta et al. 2012). This species has therefore been suggested to play major roles in the etiology and virulence of the disease (Richardson

et al. 2014). However infections are usually only possible in the presence of sulfate-reducing bacteria as necessary secondary pathogens (Brownell and Richardson 2014). One common approach to study BBD etiology is to experiment with the dominant BBD cyanobacterium in isolation. To this end, a variety of culture conditions have been applied successfully for the isolation and cultivation of the respective BBD cyanobacteria; the most commonly used culturing medium ASNIII (Aeby et al. 2015; Casamatta et al. 2012; Glas et al. 2010; Schwenk 2012; Sussman et al. 2006) and most time-efficient isolation of BBD cyanobacterial filaments *via* phototaxis on agar surfaces towards an unidirectional light source (Aeby et al. 2015; Casamatta et al. 2012; Glas et al. 2010; Stanic et al. 2011; Sussman et al. 2006). However, basic information on optimised cultivation protocols and BBD cyanobacteria species characterisations were not provided in previous studies, such as the agar concentrations used, growth curves under various conditions, replication times and how to maximise biomass. The missing information are essential for the isolation of lytic and lysogenic bacteriophages and are therefore being established for my PhD thesis. Here, I present an optimized protocol for the cultivation of a *Roseofilum* cyanobacterium (strain AO1) associated with BBD that is also applicable to other cyanobacteria present in the disease mat and for lytic and lysogenic bacteriophage isolation.

3.3 Methods

3.3.1 Sample collection

BBD coral colonies (*Pavona duerdeni*) were collected in 3 m seawater depth at Orpheus Island (S 18-34.609 / E 146-29.793) in June 2013 (GBRMPA permit G14/36788.1), transported to the Australian institute of Marine Science and maintained in outdoor aquarium systems at 27°C with shaded, natural sunlight and flow through

seawater supply. The isolation of the BBD associated cyanobacteria started immediately during the days following collection.

3.3.2 Isolation of cyanobacteria

The isolation of the motile, BBD associated cyanobacteria of the clade *Roseofilum* (Casamatta et al. 2012) were target of my study. Since cyanobacteria live in close symbiotic association with other bacteria, the objective was to produce a viable monoclonal (culture from a single filament), but not axenic culture (may contains other close associated or symbiotic bacteria). Approximately 1 cm^2 of raw BBD mat was homogenised in 50 mL of autoclaved seawater that has been filtered through 0.04 µm (Memcor ultrafiltration, Siemens). The mat was homogenised by pipetting the slurry up and down with a sterile 1 mL plastic transfer pipette and then centrifuged at 3,000 *g* for 3 min to select and clean the BBD cyanobacteria. The supernatant, containing the majority of other mat associated bacteria was discharged and the cyanobacterial pellet resuspended in autoclaved seawater. The cyanobacterial pellet was inoculated onto an agar plate to clean cyanobacterial filaments from other contaminating microbes (0.6% bacteriological agar in autoclaved seawater, Oxoid, LP0011 agar no. 1) and incubated (Innova 4230, New Brunswick Scientific) under sideway unidirectional light (50 - 80 µE m^{-2} s^{-1} light intensity) for 6 hours at 30°C (Glas et al. 2010; Richardson and Kuta 2003). Cyanobacteria migrated towards the light on top of the agar while scraping off contaminating bacteria attached to their external cell surface. The cyanobacteria furthest away from the inoculation site were carefully excised (approx. 2 cm^2 of agar) with a sterile scalpel blade in a biosafety cabinet before being transferred to a fresh, solid agar plate. This cleaning step was repeated twice under the previously described incubation conditions. Subsequently, a liquid culture was established by transferring an approx. 2 cm^2 agar piece containing a high density of cyanobacteria into freshly

prepared medium. After genetic identification of the culture (details below), single cyanobacterial filaments were selected for the establishment of a monoculture as follows: a) from liquid medium under an inverted microscope and with a micro-pipette; b) from agar with a stereo-microscope and sterile scalpel. These cultures were grown in liquid L1 medium under a 12 hour light and dark cycle at 30°C with 50 - 80 µE m^{-2} s^{-1} light intensity (Glas et al. 2010; Richardson and Kuta 2003). The cyanobacteria were characterised morphologically with an inverted microscope (Leica, DMI 6000 B) and a regular microscope (Zeiss, Axioskop 2 mot plus, 100x oil immersion objective with micrometer reference), monitored visually *via* a stereo microscope (Wild Heerbrugg, M3Z, Switzerland) and subcultured if required.

3.3.3 Genetic identification of cyanobacteria culture

To verify isolation success of the target species, cleaned cyanobacterial filaments were grown in a liquid culture and their DNA extracted from approx. 50 mg of tissue with a Mo-Bio Power Plant Pro kit (cat. 13400-50) according to the manufacturers recommendations. The V1-V9 region of the 16S rRNA marker gene was amplified with the primers 27f and 1492r (Lane 1991) in a polymerase chain reaction (MyTaq polymerase Fermentas, cycles: 95°C for 3 min, 95°C for 1 min, 55°C for 1 min, 72°C for 1 min with a final extension of 72°C for 5 min). Amplicons were cloned with a TOPO TA Cloning Kit for (pCR 4-TOPO Vector, K4575-02, Thermo Fisher Scientific) and 30 clones were sent for Sanger sequencing with the 27f primer (Macrogen, South Korea). Read quality and base calling control was conducted with the software CodonCode Aligner v4.1.1 (CodonCode Corporation, USA). Cyanobacterium isolation success was verified by determination of the closest match for the sequences with a megablast search against the “nr” database at the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Sequence variations and close association with other

cyanobacteria of the clade *Roseofilum* were visualised in a maximum likelihood tree, generated in MEGA5 (Tamura et al. 2011) with a kimura 2-parameter model and 1000 bootstrap replications.

3.3.4 Absorption spectra of photosynthetic pigments

Approximately 10 mg of exponentially growing cyanobacteria filaments were taken from a liquid culture and centrifuged at 10,000 *g* for 5 min at 4°C to pellet cells (Eppendorf Centrifuge 5430R). The pellet was subsequently resuspended in 1 mL phosphate buffer (0.1 M) and disrupted in freeze and thaw cycles. Cell debris was pelleted at 10,000 *g* for 10 min at 4°C and the supernatant (filtered 0.45 µm) analysed for phycobiliprotein absorbance spectra on a spectrophotometer (Biotek, Synergy H4, 300 - 700 nm with increments of one) (Casamatta et al. 2012).

3.3.5 Solid media preparation and comparison

To determine optimised agar concentrations for most efficient isolation of *R. reptotaenium* AO1, cyanobacterial growth on solid medium was compared among three agar concentrations (final agar concentration: agar [%] 0.6; 1; 1.5). In brief, 250 mL of fresh seawater was mixed with bacteriological agar (Oxoid, LP0011 agar no. 1), autoclaved in a 500 mL Schott bottle, cooled to approx. 40°C and enriched with L1-medium (National Center for Marine Algae and Microbiota, East Boothbay, USA) (Appendix 3.1). Each petri dish was filled with approx. 15 mL of L1 agar and stored inverted at 30°C. *R. reptotaenium* AO1 growth on agar was estimated per cm^2 by averaging filament counts along six radial, equally spaced line transects (originating in the centre of the agar plate, drawn onto the plate towards the wall of the petri dish) and extrapolating the numbers to the overall petri dish area.

3.3.6 Growth measurement optimisation for liquid media

The growth of cyanobacteria in liquid media was assessed *via* time series measurements in 24-well plates (2 mL in each well) using three approaches: 1) optical density (OD) at 750 nm as a pigment independent measurement (Cirés et al. 2011); 2) fluorescence measurements at 440 / 685 nm (Held 2011), and 3) percent coverage of bottom of well.

Optical density and fluorescence readings were averaged from an area scan with 25 measurements per well (Synergy H4, Biotek). *R. reptotaenium* AO1 at different growth stages were pelleted at 5,000 *g* for 5 minutes, dried overnight at 60°C and weighed (AD-4 Autobalance, Perkin-Elmer) to establish the correlation of “biomass - OD 750” and “biomass - fluorescence”. Percent coverage values were calculated from images taken of the well bottom with an inverted microscope at standardised settings (Leica, DMI 6000 B, 5x magnification objective). The pixel count of *R. reptotaenium* AO1 filaments was averaged from five images per well and expressed as percent coverage. Although *R. reptotaenium* AO1 filaments were growing on the well bottom and in suspension, the coverage of the well bottom was taken as a proxy for the overall growth.

3.3.7 Liquid media preparation and comparison

R. reptotaenium AO1 growth was compared among four different media for culture optimisation: ASNIII, L1, F/2, and IMK (detailed recipes in Appendix 3.1). The growth medium ASNIII was prepared in deionised water with complementary vitamin and trace metal solutions (Rippka et al. 1979). Growth medium L1 (1000x concentrate, National

Center for Marine Algae and Microbiota, Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine 04544 USA), F/2 (50x concentrate, Sigma Aldrich Australia PTY Ltd., 14 Anella Ave, Castle Hill NSW 2154) and IMK (premade powder, Wako Chemicals, Richmond, VA, USA) were prepared in autoclaved seawater according to the manufacturers recommendations. All growth media were used in a final 1x dilution and filter-sterilised before use with a cellulose acetate/cellulose nitrate mixed esters membrane (pore size 0.2 µm, Corning, cat. no. 430758) with high protein affinity to remove any possible contaminants. Cultures for media comparison were generated by splitting 20 mL of an exponential growing *R. reptotaenium* AO1 culture (growing homogeneous and in suspension) into four equal parts and pelleting the filaments at 3000 *g* for 3 min (Allegra X-15R, Beckman Coulter). The supernatant was discarded and each pellet was resuspended in 10 mL of freshly prepared growth medium: ASNIII; L1; F/2; IMK, respectively. The fresh *R. reptotaenium* AO1 stocks were distributed randomly into two 24-well plates (1.5 mL per well, wells per medium = 6) and incubated at 30°C in a 12 hour light cycle at 50 - 80 µE m^{-2} s^{-1} (PAR) with shaking at 30 rpm. *R. reptotaenium* AO1 growth curves were assessed by conversion of fluorescence measurements over time (440 / 685 nm) into dry weight and by calculating growth rates $k = \log(10) X_i - X_0 / \log(10) 2 * t$ (where X_i - X_0 is the biomass difference calculated from the end and start of the exponential phase) as well as doubling times for exponential phases (Neidhardt et al. 1990).

3.3.8 Lysogenic virus induction

R. reptotaenium AO1 (L1 medium) was exposed to a UV-light and mitomycin C treatment to induce lysogenic bacteriophages that may be present to enter the lytic cycle. Mitomycin C was added to the cyanobacterial cultures (exposure time 2 hrs) at

final concentrations of 0.1, 0.5 and 1 µg mL^{-1} (3 wells per concentration, 3 plate replicates). After the 2-hour exposure, mitomycin C-treated medium was replaced by fresh L1 growth medium and plates were maintained at control incubation conditions (Paul and Weinbauer 2010). In a separate experiment, cyanobacteria were exposed to UVA and UVB light (intensity 2.5 mW/cm^2) for 1, 2, 4, 10, 15, 20 and 30 min. After UV exposure, cultures were kept at control incubation conditions, 12 hour light and dark cycle at 30°C with 50 - 80 µE m^{-2} s^{-1} light intensity (section 3.3.2). *R. reptotaenium* AO1 growth was monitored for potential lysis with fluorescence readings over time at 440 / 685 nm (Synergy H4, Biotek). Fluorescence readings were averaged from an area scan, with 25 reads per well, as an indicator for the cyanobacterial biomass (section 3.4.3).

3.3.9 Bacteriophage enrichment

A virus homogenate was generated from approximately 5 cm^2 fresh BBD mat, collected in August 2015 at Orpheus Island (S 18-34.609/E 146-29.793) (GBRMPA permit G14/36788.1). The mat was rinsed with sterile, filtered seawater and then homogenised by vortexing and pipette mixing in 10 mL sterile L1 medium. The homogenate was centrifuged at 3,000 *g* for 3 min to remove larger cells debris, the supernatant transferred into a clean tube (step repeated three times) and then filtered through a 0.45 µm PES membrane (Merk Milipore). Virus-like particles (VLPs) recovered with this method (10^4 VLPs mL^{-1}) were not concentrated further because of the limited amount of starting material available, but were used for bacteriophage purification by liquid assays (as an equivalent method to single-plaque isolations on agar). Bacteriophages were inoculated via a dilution series into mid-exponentially growing *R. reptotaenium* AO1 liquid cultures (12 well plates with 2 mL volume, dilutions of VLPs 1:10 - 1:1000) to

enrich and purify bacteriophage strains that infect bacteria present in the culture and to dilute out non-replicating viruses (Middelboe et al. 2010). *R. reptotaenium* AO1 cultures with lysed cells were lighter in colour than controls (i.e., cleared) and showed high VLP numbers of up to 10^7 VLPs mL^{-1} due to bacteriophage replication (measured on a flow cytometer, section 3.3.10). The cleared culture with the lowest starting density of inoculated bacteriophages was enriched and purified a total of three times, as described above. Subsequently, bacteriophages were diluted to extinction (3 times) to a threshold where viral replication is lost (Wilcox and Fuhrman 1994), in order to obtain viral replication derived from a single bacteriophage lineage. Inoculated *R. reptotaenium* AO1 cultures (L1 medium) were incubated at 12 hour light and dark cycle at 30°C with 50 - 80 µE m^{-2} s^{-1} light intensity (3.3.2) and monitored with fluorescence measurements (3.4.3). In addition, 14 virus seawater concentrates were collected for bacteriophage isolations between the years 2012 and 2015 (BBD = 8; Control: 6, Appendix 2.1). A detailed description of seawater collection and concentration with tangential flow filtration is provided in section 2.3.2. To obtain a pure bacteriophage culture of a single lineage, seawater concentrates of 10^6 - 10^7 VLPs mL^{-1} were inoculated into *R. reptotaenium* AO1 cultures in a dilution series (10^2 - 10^8 dilution) and cyanobacteria were monitored with fluorescence measurements as described in section 3.4.3. Media types other than L1 (e.g. ASNIII, F/2 and IMK) were not used in bacteriophage enrichment experiments, because of their inefficiency to promote growth of the cyanobacterial host.

3.3.10 Quantification of virus like particles in cyanobacteria cultures

To quantify viral abundance in bacteriophage enrichments (after every purification step) and lysogenic bacteriophage induction experiments, the abundance of virus-like

particles (VLPs) in the *R. reptotaenium* AO1 culture medium was measured on a flow cytometer (BD FACSVerse) with a 488 nm argon-ion laser following the protocol outlined in Brussaard et al. (2010). In brief, samples were fixed in the dark at a final concentration of 4% glutaraldehyde, incubated for 30 min at 4°C, and stained with 1 x SYBR green I (Invitrogen) at 80°C for 10 min. Samples were counted in a dilution series at a low flow rate with approx. 200 - 400 events per second, using the same VLP standards described by Pollock et al. (2014).

3.3.10 Statistics

Differences in growth curves of *R. reptotaenium* AO1 cyanobacteria on agar and in liquid media were statistically analysed by comparing regression slopes from log phases with a one-way analysis of variance (ANOVA) and a Tukey post-hoc comparison, all assumptions met (Appendix 3.2 - 3.6).

3.4 Results and discussion

3.4.1 Genetic and morphological characterisation of the isolated cyanobacterium

In this study, I optimised the isolation of the main BBD associated *Roseofilum* cyanobacteria *via* phototaxis on agar, and provide cultivation protocols which results in healthy, fast growing and viable filaments (this will not provide an axenic culture, Appendix 3.7). The target cyanobacterium was successfully isolated, since the 16S rRNA gene sequences of my cyanobacterial culture showed a 99% - 100% identity (0 - 4 nucleotide differences within 346 bp) to the publicly available BBD-associated *Roseofilum* reference sequences. Cyanobacteria of the proposed genus *Roseofilum* are known to be the most abundant cyanobacterial species in terms of biomass within the

BBD community (Casamatta et al. 2012; Miller and Richardson 2011; Rasoulouniriana et al. 2009; Sato et al. 2010). However, the exact species identification was not as straight forward compared to the genus classification.

The dominant BBD-associated cyanobacteria were originally classified as *Pseudoscillatoria coralii* Rasoulouniriana et al. (2009). This genus and taxon was argued to be invalid due to incorrect orthography (no Latin description, no type indication) and re-established instead as the new genus and species *Roseofilum reptotaenium* by Casamatta et al. (2012) (International Code of Botanical Nomenclature (ICBN) and International Code of Nomenclature of Prokaryotes (ICNP)). Although *P. coralii* and *R. reptotaenium* share >97% of their 16S rRNA gene sequence, they were not considered the same species. Both taxa were supposed to be maintained as separate species in the genus *Roseofilum* because of differences in trichome dimensions and associated pigments (Table 3.1) (Casamatta et al. 2012). However, a new name for *P. coralii* has not been established, and a detailed taxonomic revision of *P. coralii* within the genus *Roseofilum* has yet to be undertaken (Casamatta et al. 2012).

In terms of genetics, the newly isolated cyanobacterium of the present study showed characteristics of both *P. coralii* and *R. reptotaenium*. The partial 16S rRNA gene sequences of *P. coralii* and the newly isolated cyanobacterium were 100% identical to each other and clustered separately from *R. reptotaenium* (99% identical) within a phylogenetic tree, albeit with low bootstrap support (bootstrap value = 42, Fig. 3.1).

Table 3.1 Comparison of cyanobacteria previously isolated from black band disease, *Roseofilum* clade. Note that *Roseofilum reptotaenium* strain AO1 is genetically most similar to strain BgP10_4s from the Red Sea, but morphologically closer to the Caribbean strains. "Host genus" refers to the diseased coral from which the strain was isolated; na = data not available.

***Roseofilum* strain**	**Trichome/cell width and length [µm]**		**Tip cell shape**	**Dominant pigmentation [nm]**	**Host genus, location**	**Colour of culture**	**Clumping & motility**	**Source**
R. reptotaenium 100-1	2.5 - 4.0	3.0 - 3.9	round, tapered	phycoerythrin 548, 565, 620	*Siderastrea*, Caribbean	dark red/brown	yes	Casamatta et al. 2012
R. reptotaenium 101-1	3.2 - 3.6	3.4 - 4.0	round, tapered	phycoerythrin 548, 565, 620	*Diploria*, Caribbean	dark red/brown	yes	Casamatta et al. 2012
BDA 82.01	4.0 - 4.2	4.0 - 4.5	round, tapered	phycoerythrin 548, 565, 620	*Montastrea*, Bermuda	dark brown/black	yes	Rützler et al. 1983
R. reptotaenium AO1	3.6 - 4.0	3.8 - 4.3	round, tapered	phycoerythrin 548, 568, 620	*Pavona*, GBR	dark brown/black	yes	this study
BBD cyanob. isolate	4.0 - 4.2	na	round, tapered	na	*Montipora,* GBR	na	na	Glas et al. 2010
BgP10_4S (former *P. coralii*)	5.0 - 6.0	na	round, tapered	phycocyanin 336, 436, 666	*Favia,* Red Sea	dark green	yes	Rasoulouniriana et al. 2009

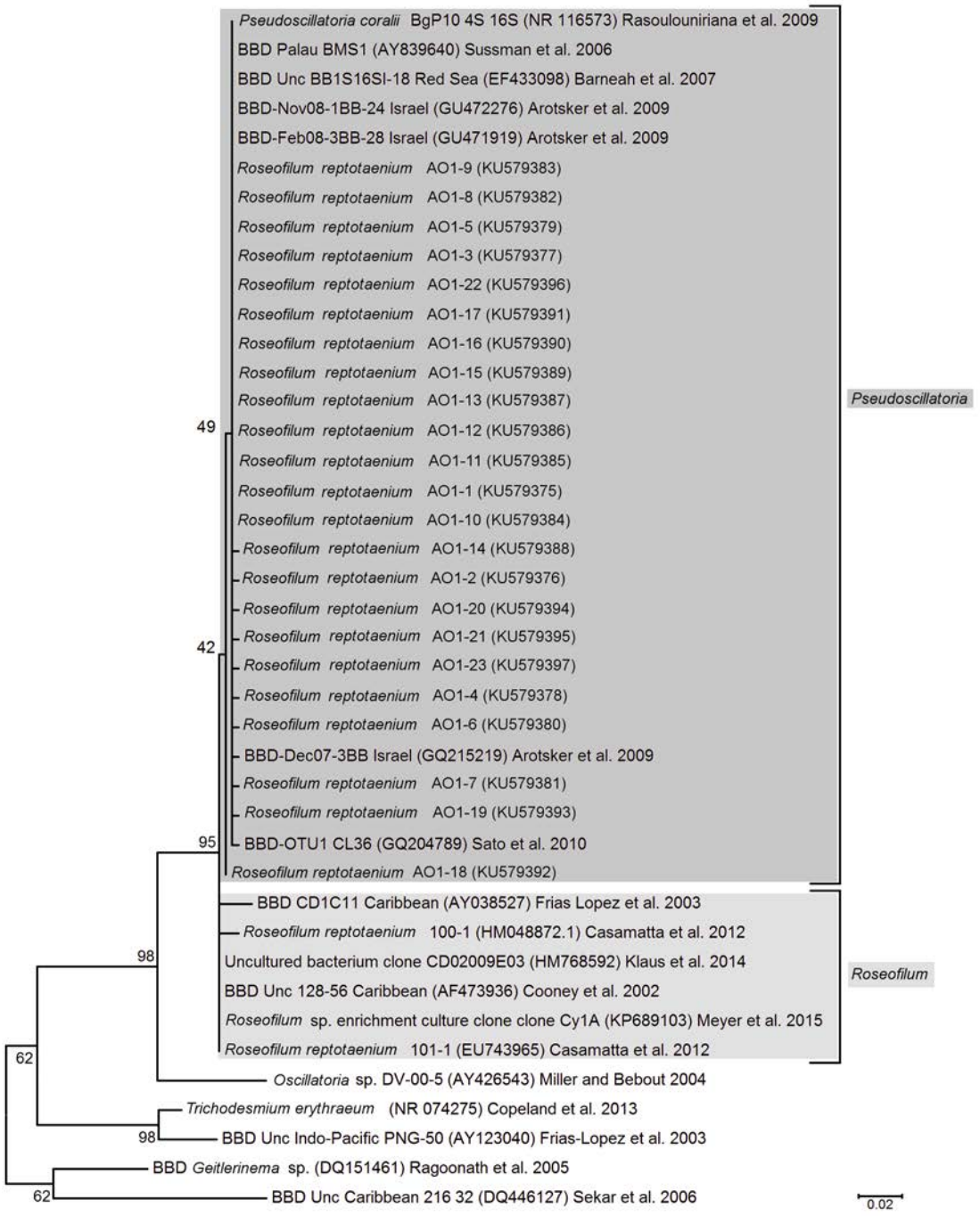


Figure 3.1 Phylogenetic tree of black band disease cyanobacterial partial 16S rRNA gene sequences based on maximum likelihood analysis. Numbers next to branches indicate percentages of replicated trees in which associated sequences grouped together (bootstrap, n = 1000). Reference cyanobacterium sequences were selected based on close blast matches (cultured cyanobacteria in top 20 blast hits) and previous studies of cultured cyanobacteria. All aligned sequences were trimmed to the shortest reference sequence (346 bp). The scale indicates evolutionary distance, calculated using Kimura 2-parameter in MEGA5.

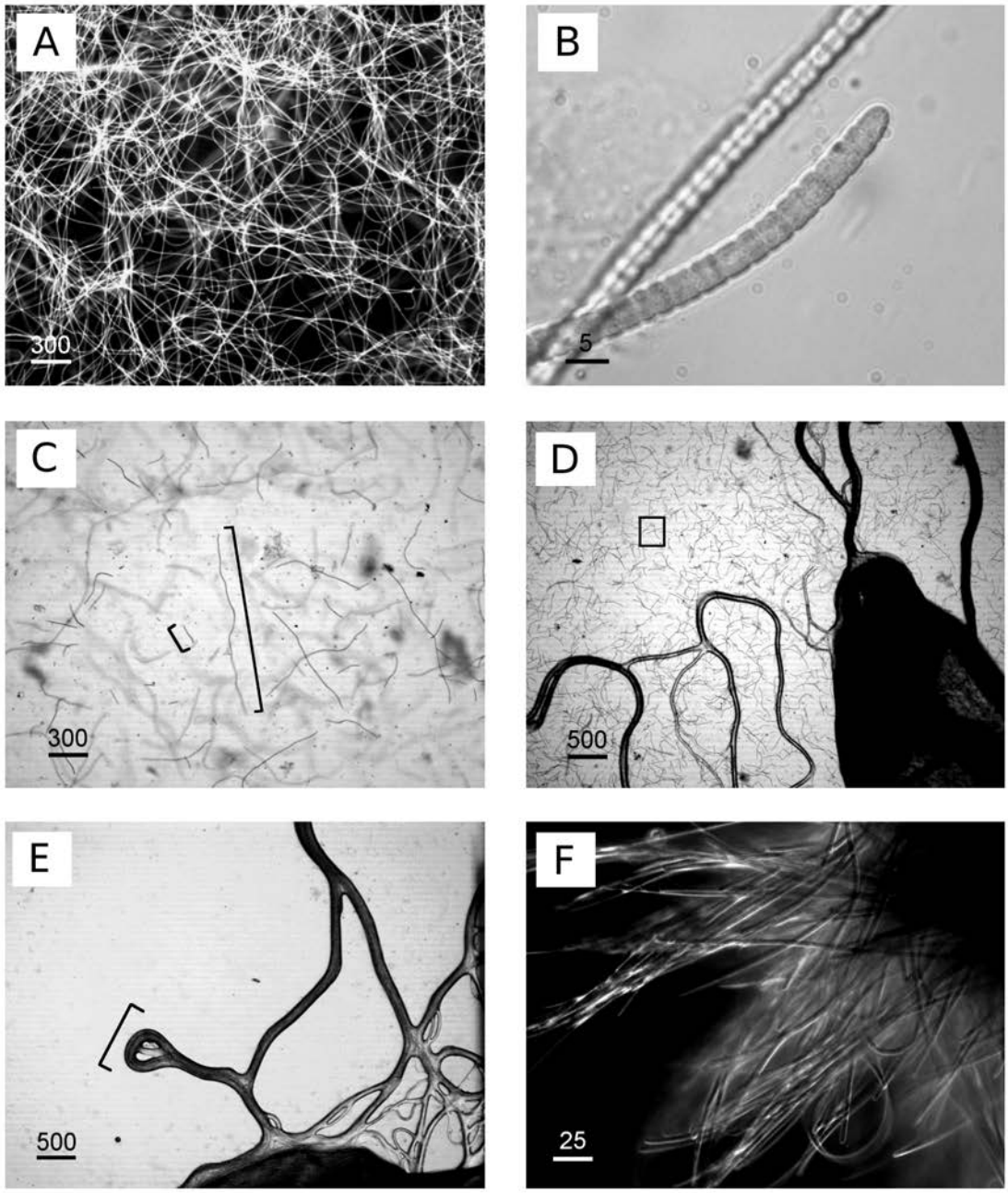


Figure 3.2 Images of cyanobacterium *R. reptotaenium* AO1 filament morphologies on agar and liquid cultures. A) Homogeneous, exponential *R. reptotaenium* AO1 growth in L1 medium, filament length up to 1,200 µm, B) detailed image of isolated cyanobacteria, C) agar 0.6%, *R. reptotaenium* AO1 grew within and not on top of agar, image taken during transition of exponential phase (longer cyanobacteria) to collapse of the culture (shorter cyanobacteria, approx. 100 µm), D) agar 1%, *R. reptotaenium* AO1 grew on top (dark lines) and within agar (square), E) agar concentration 1.5%, *R. reptotaenium* AO1 grew only on top of agar plates in tracks (bracket) in close proximity to the inoculation site without penetrating into the agar, F) clumps / aggregates of *R. reptotaenium* AO1 formed in larger volumes e.g. 250 mL flasks. Scale bars in µm.

In terms of morphology, the unbranched trichomes with rounded and tapered cell tips were up to 4 µm in width (Fig. 3.2A, B), i.e. smaller than *P. coralii*, but slightly larger than *R. reptotaenium* (Table 3.1). Associated pigmentation was most similar to *R. reptotaenium* due to light absorbance peaks at 620, 548 and 565 nm, indicating the presence of phycocyanin and phycoerythrin respectively (predominantly phycoerythrin due to ratio 565:620 (0.123/0.85) = 1.45, Casamatta et al. (2012)) (Fig. 3.3, Table 3.1, Appendix 3.8).

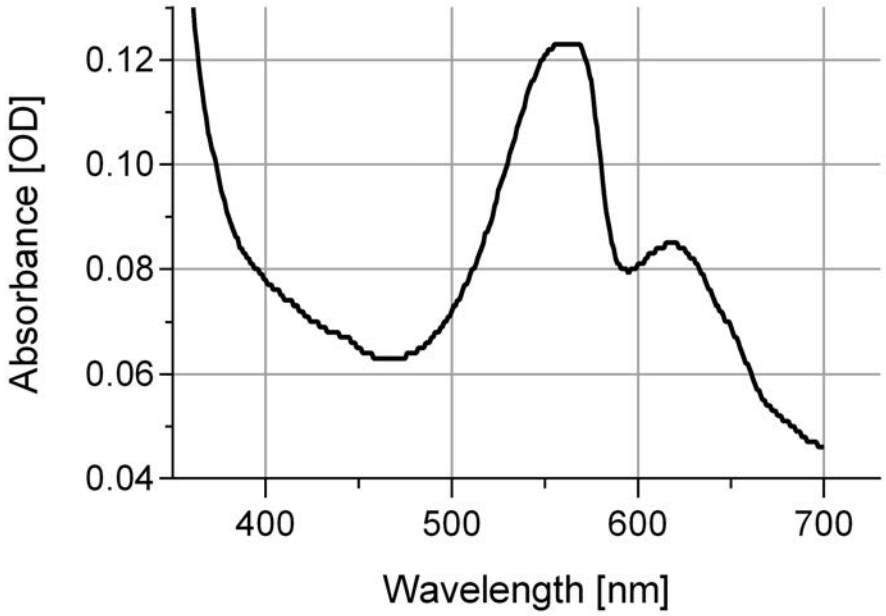


Figure 3.3 Absorbance spectrum for associated phycobiliproteins, *R. reptotaenium* AO1. Note representative peaks for phycoerythrin (548 and 565 nm) and phycocyanin (620 nm). *R. reptotaenium* AO1 was grown in L1 medium.

My results show, that the characteristics used (16S rRNA gene, morphology and pigmentation) to distinguish the BBD-associated *Roseofilum* species do not reliably separate *P. coralii*, *R. reptotaenium* and the cyanobacterium of the present study. The phenotype of strain BgP10_4S is clearly an exception compared to all the other reported *Roseofilum* strains (Table 3.1). However, geographical separation,

environmental or culture-based pressures may have led to the observed distinct phenotypic diversity among BBD associated *Roseofilum* strains (Dvořák et al. 2015). A multiple DNA marker analysis is required to validate taxonomic affinities among the *Roseofilum* strains (Wu, Jospin, and Eisen 2013) and show if strain BgP10_4S is a true, distinct taxon from *R. reptotaenium* (i.e. *P. coralii* with phycocyanin as the main pigment, lacking phycoerythrin pigments, and exhibiting large cell dimensions of 5 - 6 µm).

Nevertheless, due to the overlapping characteristics of the current strains of *P. coralii*, *R. reptotaenium* and the cyanobacterium of the present study, the most parsimonious solution is the integration of the species into a single taxon, as already practised in Richardson (2015). I therefore provide an emended diagnosis for *R. reptotaenium* below and unite the two taxa according to the principle of priority (i.e. the final, valid and earliest taxon description, article 38 and 41 of the International Code of Nomenclature for algae, fungi, and plants ICN), into *Roseofilum reptotaenium* (Rasoulouniriana) ex Casamatta, while *Pseudoscillatoria coralii nom*. *inval*. Rasoulouniriana becomes its synonym. As a consequence, the newly isolated cyanobacterium of the present study has been classified as *Roseofilum reptotaenium* strain AO1 (Australia, Orpheus Island, strain isolate1, Genbank: KU579375 - KU579397, collected at Orpheus Island, Australia, S 18-34.609 / E 146-29.793 in June 2013, 3 m seawater depth, from black band disease on *Pavona* sp. coral) (Oren 2011a, b) and deposited as an epitype (CS-1145) to the Australian National Algae Culture Collection ANACC, Hobart, Tasmania, Australia, www.csiro.au/ANACC. The holotype is deposited at the Brigham Young University Herbarium of Non-Vascular Cryptogams BRY C-53584 (Casamatta et al. 2012).

Emended description of *Roseofilum reptotaenium*

Roseofilum reptotaenium (Rasoulouniriana) ex Casamatta, emend.

Synonym: *Pseudoscillatoria coralii nom. inval.* Rasoulouniriana et al. 2009

Gram-negative, motile cyanobacterium growing epizoic on corals in black, microbial mats that move over the coral surface and kill the underlying tissue (associated with coral black band disease). In culture, filaments appear can appear dark-green to blackish-brown and reach up to 1 mm in length. Unbranched trichomes with thin sheath, no heterocysts, tapered cells tips, cells of of 3.0 - 4.5 µm length. High levels of phenotypic plasticity with variants in terms of cell width and pigmentation (Table 3.1) ranging from 2.5 - 4.0 µm cell width and predominant pigment phycoerythrin in the Caribbean; 3.6 - 4.0 µm cell width on the Great Barrier Reef; to 5.0 - 6.0 µm cell width and pigment phycocyanin in the Red Sea. 16S rRNA gene sequence may show minor nucleotide variations depended on the sample location (Fig. 3.1). Optimal growth conditions from 25°C - 30°C, pH 7 and 8, salinity 5 to 5.5% (w/v). For further details access full formal description of the genus *Roseofilum* and the species *R. reptotaenium* in Casamatta et al. (2012) Phycologia 51:489-499, and Rasoulouniriana et al. (2009) Dis Aquat Organ 87:91-96.

3.4.2 Solid media comparison

The separation of cyanobacteria from the microbial mat was the most time-efficient with motility *via* phototaxis on an agar surface towards an unidirectional light source (Vaara, Vaara, and Niemela 1979). Previous studies that isolated BBD cyanobacteria from the microbial mat did not indicate used agar concentrations (Aeby et al. 2015; Casamatta et al. 2012; Glas et al. 2010; Stanic et al. 2011; Sussman et al. 2006), but reported e.g.

motility over the agar surface with up to 5 cm per day (Sussman et al. 2006), or almost no motility at all (Glas et al. 2010). The use of a 0.6% soft agar in my study resulted in faster motility (up to 8 cm in 6 hr) compared to higher percentage agars (1.5% with almost no gliding at all) (ANOVA Appendix 3.4, 0.6% vs. 1.0%, p = 0.0325; 0.6% vs. 1.5%, p = 0.0186). *R. reptotaenium* AO1 on 0.6% agar spread within the entire agar plate with three times as many filaments after 7 days compared to the next successful treatment of 1% (Figs. 3.2C - E and 3.4). There was no spread or cell replication of *R. reptotaenium* AO1 on 1.5% agar plates. A higher percentage and stickier agar (1.5%) might be more effective to scrape off contaminants from motile cyanobacteria than a lower percentage agar (0.6%) (Rippka, Waterbury, and Stanier 1981). However, due to the reduced growth and reduced motility of cyanobacteria on the tested higher percentage agars, I recommend using a lower percentage agar with at least three repetitive steps to clean up *R. reptotaenium* AO1 filaments.

Two additional cyanobacteria have been isolated with a low percentage agar of 0.6% and deposited into the sequence reference database Genbank the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Based on top 5 blast hits (97% - 99% identity), cyanobacteria species 1 (KU720412) was close related to *Leptolyngbya* sp. (KJ206339.1), *Oscillatoria limnetica* (AF410934.1) and *Phormidium* sp. (JF837333.1), while the second cyanobacteria species (KU720413) was closest to *Limnothrix* sp. (DQ889938.1). These two species were only able to move through a 0.6% agar and could be potentially missed during the isolation process if a higher percentage agar was used.

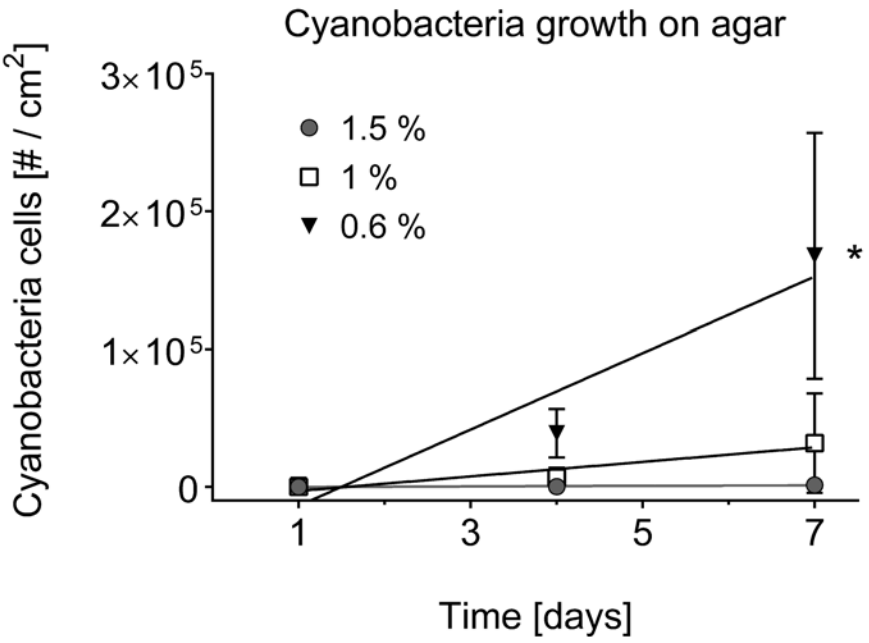


Figure 3.4 *R. reptotaenium* AO1 growth on various agar concentrations. Cell counts are displayed over time in days. A square centimetre of exponentially growing *R. reptotaenium* AO1 cyanobacteria in agar was inoculated on each of the plates (n = 3 agar plates for every agar concentration) and bacterial cell counts (#) monitored over time. The use of 0.6% agar resulted in a significantly higher *R. reptotaenium* AO1 growth rate (indicated by asterisk, 0.6% vs. 1.0%, p = 0.0332, at significance level of 0.05) compared to the other agar concentrations of 1% and 1.5%.

3.4.3 Growth measurement optimisation

To date, growth of BBD associated cyanobacteria has only been qualitatively assessed by visual inspection of cultures (Glas et al. 2010; Sussman et al. 2006), or by biomass measurements following four weeks after inoculation (Gantar et al. 2009). The formation of clumps and aggregates (Richardson et al. 2014), with no homogeneous cell distribution makes it difficult to estimate cell numbers and replication in culture. In the present study, the undisturbed *R. reptotaenium* AO1 culture in liquid medium showed homogeneous growth with no clumping behaviour in smaller volumes of up to 5 mL per well (Fig. 3.2A), and could be monitored for cell replication *via* various methods: optical density [OD 750], fluorescence of chlorophyll a [440 / 685 nm] and well coverage measurements [%] (Fig. 3.5). All measurements indicated exponential growth from day

1 to day 8. Coverage [%] showed a long stationary phase (approx. duration 16 days) with no decline in *R. reptotaenium* AO1 abundance, whereas fluorescence and optical density measurements indicated a decline of the culture (after day 8). Coverage [%] was not considered further, due to unreliable measurements and the inefficiency of this labour-intensive approach.

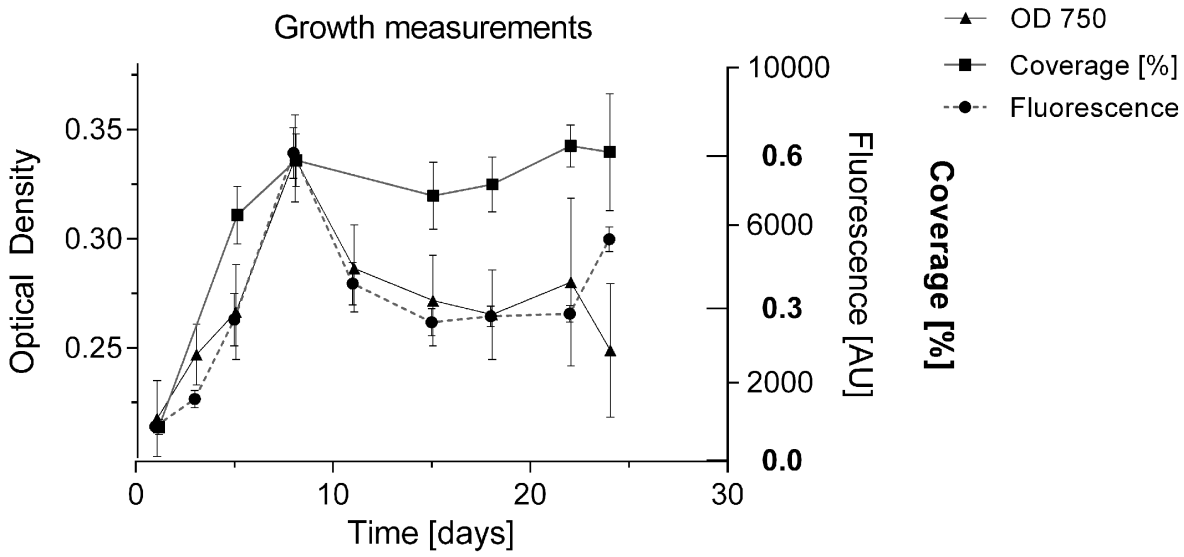


Figure 3.5 Comparison of methods for measuring growth, *R. reptotaenium* AO1. Cyanobacterial cell number was monitored for 24 days with different methods (% surface coverage of well bottom; fluorescence of chlorophyll-a 440 / 685 nm; optical density 750 nm). Y-axes of the respective measurements were adjusted to compare methods by setting start and maximum values within the same range.

Fluorescence was used for subsequent measurements as it provided a more sensitive detection capability with the possibility to distinguish between alive and dead *R. reptotaenium* AO1 filaments, measuring a second growth peak after 25 days and offered more consistency across measurements (Fig. 3.6). The linear relationship between biomass (dry weight in g/L) and fluorescence (Appendix 3.2, 3.3) (equation: $y = 0.00011x + 0.04863$; $r^2 = 0.794$) allowed the conversion of measured fluorescence values into biomass for the calculation of growth rates (k) and doubling times (t_{gen}) for

subsequent experiments and comparison with other strains. Clumping behaviour of *R. reptotaenium* AO1 filaments was observed only after the cultures had been disturbed (e.g., shaking during transportation or growth measurements). However, clumped *R. reptotaenium* AO1 filaments were capable of homogenising overnight once returned to the constant incubator environment. In addition, previously reported ring formations and continuous clumping behaviour (Richardson et al. 2014) were observed, but only in larger volumes of > 200 mL (Fig. 3.2F).

3.4.4 Liquid media comparison

Various media types have been compared for the cultivation of cyanobacteria of the *Roseofilum* clade (Glas et al. 2010; Sussman et al. 2006) with most of the reports of successful growth using ASNIII medium (Casamatta et al. 2012; Glas et al. 2010; Richardson and Kuta 2003; Sussman et al. 2006). In the present study, *R. reptotaenium* AO1 filaments in ASNIII had a growth rate of k = 0.078 biomass/day and time to double dry weight t_{gen} = 12.85 days, with a survival time of 50 days (Fig. 3.6). Although ASNIII enabled *R. reptotaenium* AO1 to survive longer compared to the other tested media types, I observed reduced motility and short filaments in a similar range as observed for the unsuccessful media types F/2 and IMK (approximately 100 - 300 µm). Since cyanobacterium species are known for filament length variation and shorter filaments during nutrient depletion (Gibson and Smith 1982; Kruskopf and Du Plessis 2006; Smith and Gilbert 1995), I was particularly interested in a media type that enhanced filament length and increased biomass, in order to provide the healthiest and dense bacteria culture for the bacteriophage isolation. Improved growth results were achieved with the media type L1, which, to the best of my knowledge, has not previously been tested to grow BBD cyanobacteria. *R. reptotaenium* AO1 in L1 medium appeared as longer

filaments of up ~1200 µm in length (Fig. 3.2C), survived ~25 days without additional nutrient supply (Fig. 3.6) and started to grow exponentially within the first day to 8 days after inoculation (growth rate k = 0.214 biomass/day; time to double dry weight t_{gen} = 4.67 days; ANOVA Appendix 3.4, 3.5; slope comparison L1 vs. ASNIII: p = 0.0001).

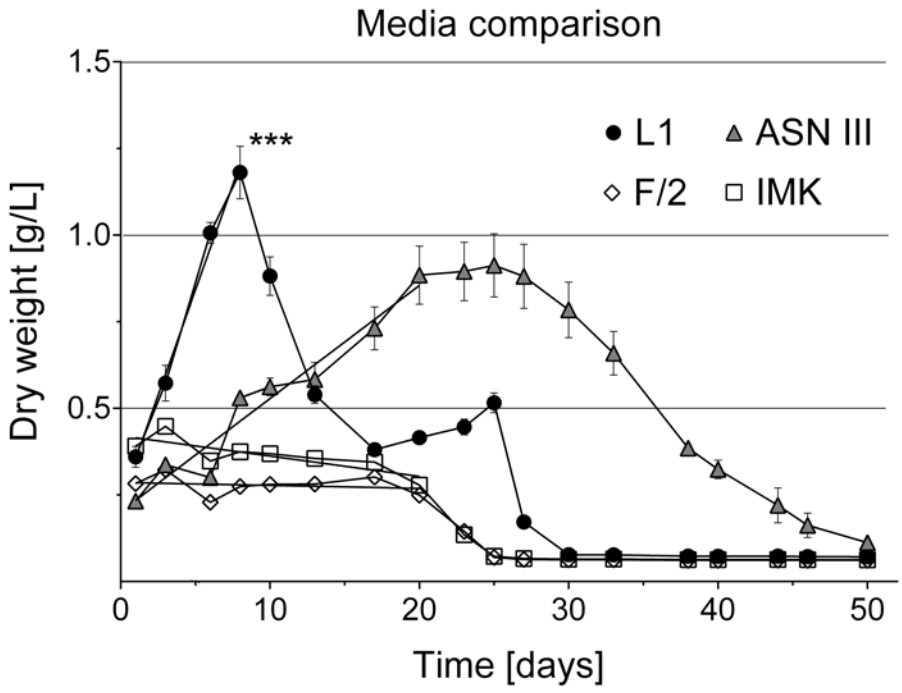


Figure 3.6 *R. reptotaenium* AO1 growth curves in different liquid culture media. Cyanobacterial growth was tested in four different media types (L1, ASNIII, F/2 and IMK). *R. reptotaenium* AO1 cultures in L1 medium grew exponentially, reached significantly higher cell densities (indicated by asterisk, p = 0.0001 at significance level of 0.05) and collapsed earlier without additional nutrient supply than in the other media types (n = 6 for each media type, volume 1.5 mL).

R. reptotaenium AO1 filaments in L1 medium reached twice the amount of biomass after 20 days compared to the ASNIII cultures, if L1 nutrients were re-supplied by inoculation of N, P, trace metals and vitamins with a final x1 concentration every 3 days (biomass at day 20 for L1 = 1.78 ± 0.2 g, compared to ASNIII = 0.88 ± 0.21 g). Interestingly, the only difference between the chemical components of L1-medium (best growth) and F/2-medium (almost no observed growth) was the presence of selenous

acid (H_2SeO_3), nickel (II) sulfate hexahydrate ($NiSO_4 \cdot 6H_2O$), sodium orthovanadate (Na_3VO_4) and potassium chromate (K_2CrO_4) in the former. Due to the differences in growth of cyanobacteria in L1 and F/2 media, it is likely that the presence of one or more of these trace metals is essential for maximising the growth potential of *R. reptotaenium* AO1. Since these trace metals are not present in ASNIII medium (Appendix 3.1), it is possible that L1 enhances growth as well for close related Caribbean and Red Sea strains of *R. reptotaenium*, that have so far only been cultivated in ASNIII (Casamatta et al. 2012; Glas et al. 2010; Richardson and Kuta 2003; Sussman et al. 2006).

3.4.5 Lytic and lysogenic virus induction

Standard techniques to isolate bacteriophages, such as plaque assays, were not applicable here, due to the characteristics of the main BBD filamentous cyanobacteria, i.e. clumping behaviour and high motility. Therefore, bacteriophage isolation had to be conducted by liquid serial dilution to extinction assays (Middelboe et al. 2010) with the developed methods for cyanobacteria cultivation (*Chapter 3*). Inoculation with the 14 seawater concentrates did not result in an increase of VLPs and did not induce lysis of cyanobacteria in culture, as indicated by flow cytometry measurements and absence of clearance (data not shown). Lysis of *R. reptotaenium* AO1 cultures was only possible with a 0.45 µm pre-filtered virus concentrate, which was prepared directly from the BBD mat. Approximately two days after virus inoculation of the BBD-mat concentrate into the cultures, *R. reptotaenium* AO1 lost their characteristic clumping behaviour. Loss of biofilm formation and swarming behaviour can be characteristics of bacteria infected by bacteriophages, such as *Pseudomonas aeruginosa* (Zegans et al. 2009). Lysis of *R. reptotaenium* AO1 cultures occurred 5 - 8 days after inoculation with the virus concentrate during bacteriophage enrichment and purification. Lysed parts of *R.*

reptotaenium AO1 filaments were no longer fluorescent and appeared as empty hulls, scattered in individual segments across the bottom of the well (Fig. 3.7). The lysate with the lowest VLP starting density that showed clearance of the host bacteria (1:1000 dilution, with 10^3 VLPs mL^{-1}) was enriched in two additional purification steps to dilute out non-replicating bacteriophages and to increase the abundance of potentially 'rare' viruses. In the following two purification steps, lysis occurred up to a starting density of 10^3 VLPs mL^{-1} (Fig. 3.8A). Three populations of VLPs were measured at the end of the initial three enrichment and purification steps by flow cytometry (total of 10^7 VLPs mL^{-1}, initial VLP 10^4 dilution, Fig. 3.8B), indicating that potentially three bacteriophage populations were purified in the cultures. Lysates that were filtered to 0.22 µm prior to inoculation did not show subsequent lysis, while inoculation with 0.45 µm filtered lysates did result in lysis. In addition, while lysis in the dilution to extinction purification occurred up to a VLP dilution of 10^5, it was not possible to split bacteriophage populations to obtain single individual populations that derived from a single bacteriophage lineage in the dilutions to extinction.

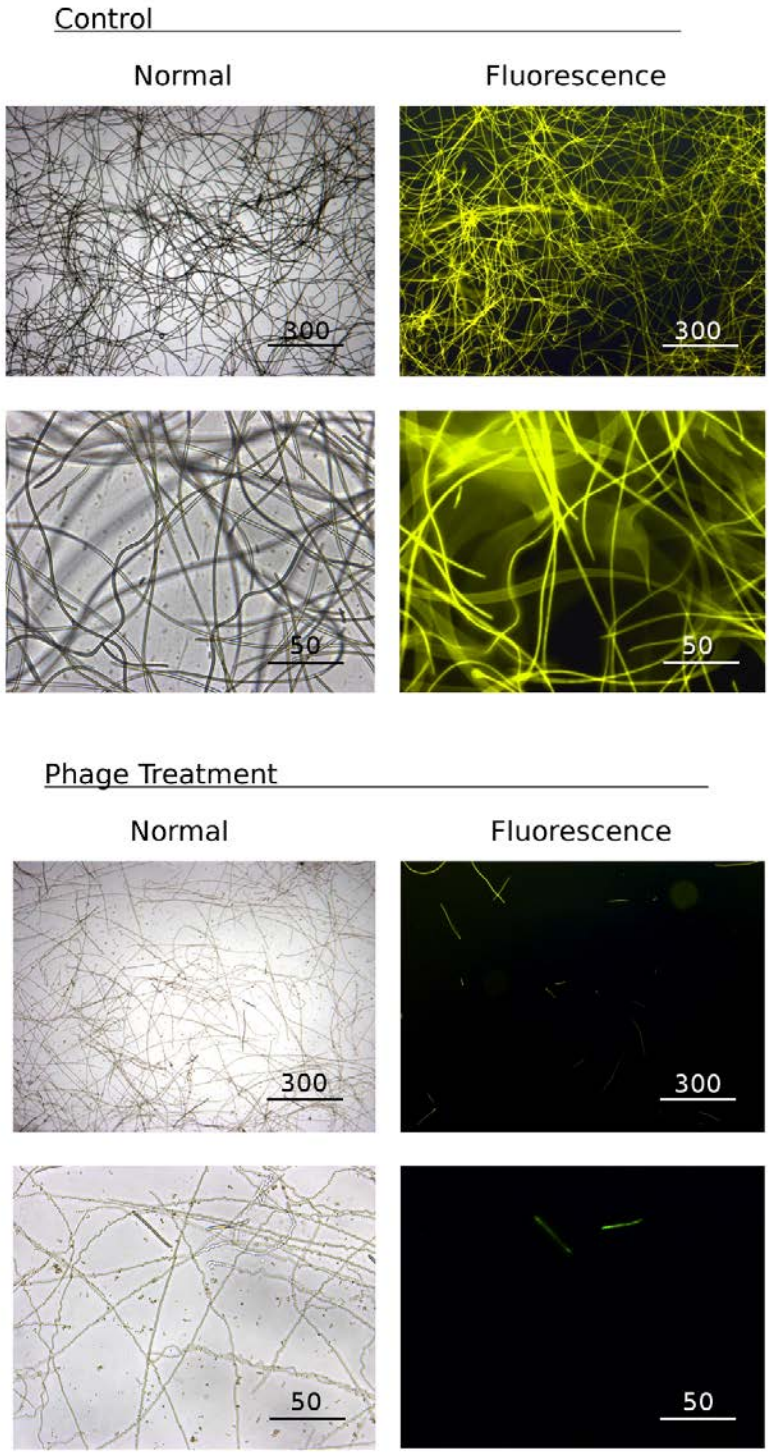


Figure 3.7 Potential lysis of cyanobacteria *R. reptotaenium* AO1 cultures. Cyanobacteria cultures (control and bacteriophage treatment) are shown with normal light and fluorescence light. While cyanobacteria in control cultures show fluorescent filaments, cyanobacteria after lysis are scattered into individual segments and show no fluorescence. Scale bars in µm.

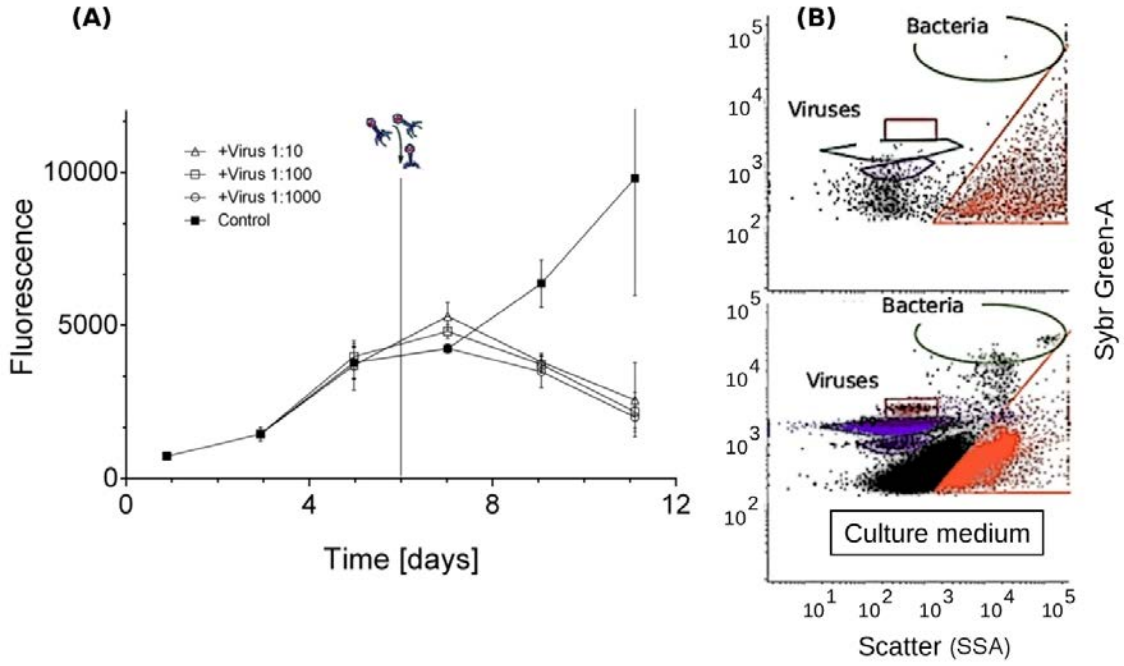


Figure 3.8 Results of lytic virus enrichment and purification experiments. (A) Bacteriophages were enriched by liquid assays in *R. reptotaenium* AO1 cultures. Graph shows the third enrichment and purification step with fluorescence over time for the respective treatments (control, initial bacteriophage dilution 1:10, 1:100, 1:1000). Vertical line indicates the time point of bacteriophage inoculation. (B) Cytograms of particle size distributions from flow cytometry measurements after treatment, top plot: control; bottom plot: culture medium of third enrichment step with 1:1000 dilution after 11 days. In accordance with standard references (Pollock et al. 2014), gate P4 = bacteria populations, and gates P5 - P7 = virus like particles (VLPs).

It was not possible to induce a lysogenic virus from our *R. reptotaenium* AO1 cultures. Although cyanobacterial cell densities in cultures declined over time in response to the mitomycin C and UV treatments, I did not detect any evidence of bacteriophage induction in the flow cytometry counts (Fig. 3.9). I interpret this to mean that the mitomycin C and UV treatments caused the cyanobacteria to disintegrate and to reduce their fluorescence compared to controls, instead of cell lysis through bacteriophage induction.

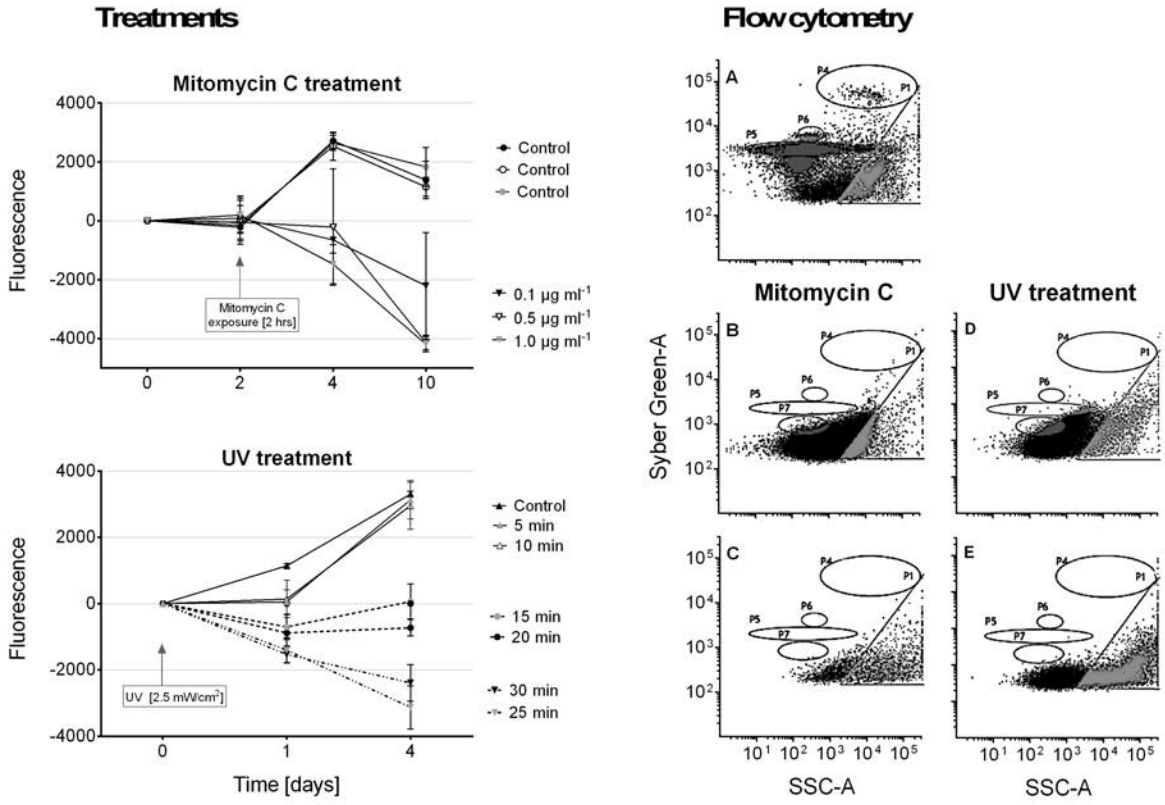


Figure 3.9 Lysogenic virus induction using mitomycin C and UV-light treatments. Panel on the left shows *R. reptotaenium* AO1 growth, measured as fluorescence, in response to the respective treatments (mitomycin C and UV light). Mitomycin C exposure was 2 hrs at the respective concentrations (e.g., 0.1 µg mL^{-1}). UV treatment intensity (UVA and UVB) was 2.5 mW/cm^2 at the respective exposure times (e.g., 5 min). Samples were normalized by calculating differences from starting values. Panels on the right show examples of particle size distributions from flow cytometry measurements after treatment. In accordance with standard references (Pollock et al. 2014), gate P4 = bacteria populations, and gates P5 - P7 = virus like particles (VLPs). A) reference plot of bacteria and virus populations, lytic enrichment, B) 2 hour mitomycin C exposure at 1 µg mL^{-1}, C) mitomycin C control measurement, D) 30 min UV exposure, E) UV treatment control measurement.

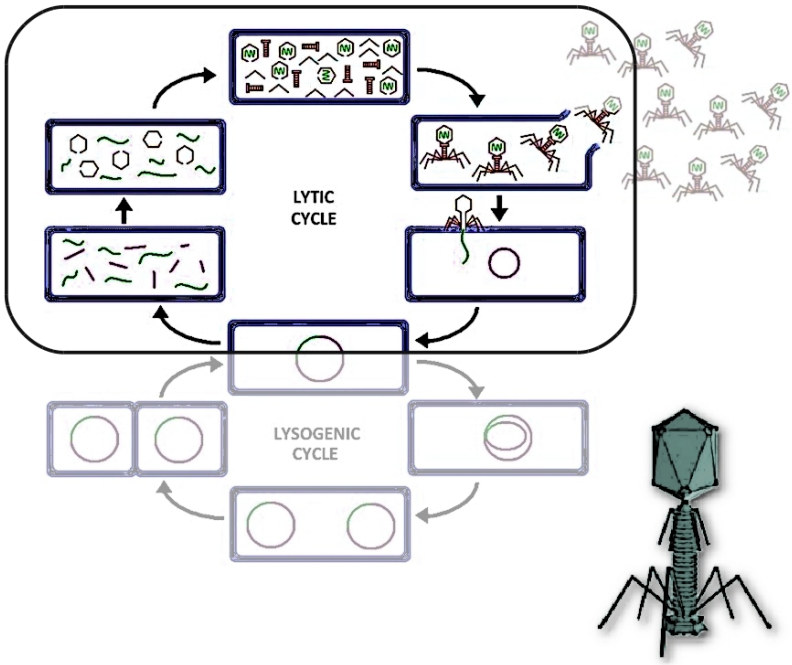
3.4.6 Conclusion

Here, I presented an optimised cultivation protocol for the main BBD cyanobacterium *R. reptotaenium* (strain AO1) and formally united the taxa *P. coralii* and *R. reptotaenium* in an emended species description into *Roseofilum reptotaenium* (Rasoulouniriana) Casamatta emend. Healthy, fast growing and viable *R. reptotaenium* AO1 cultures were established on a low percentage 0.6% L1 agar by transferring a dense cyanobacteria

agar pellet onto a new plate every 7 - 10 days and in L1 liquid medium (250 mL with subculturing every month and fresh addition of nutrients to prolong exponential phases and increase biomass). The *R. reptotaenium* AO1 species isolation with a low percentage agar (0.6%) resulted in faster and easier gliding filaments and enabled recovery of two additional cyanobacteria species from BBD samples. The homogeneous growth of *R. reptotaenium* AO1 filaments in smaller volumes of < 5 mL, if undisturbed, allowed the generation of growth curves for the first time for BBD associated cyanobacteria. My media comparison showed, that the commonly used growth medium ASNIII did not result in optimal growth conditions while L1 maximised biomass for the tested *Roseofilum* species. Maximising biomass of the cultured cyanobacteria is essential for any downstream genomics and bacteriophage infection experiments. Lytic bacteriophage enrichment and purification as well as lysogenic induction experiments are feasible due to these new cultivation protocols. While the purification of lytic bacteriophages resulted in three VLP populations, indicated by flow cytometry, the lysogenic induction did not result in measurable bacteriophage replication. The protocols developed in this chapter provide the methodological basis for the following two chapters (*4* and *5*) on phage therapy and genome sequencing of bacteriophages and the cyanobacterial host in BBD.

Chapter 4

A feasibility assessment of phage therapy to treat black band disease in corals



4.1 Abstract

Mitigation of coral disease is currently limited since antibiotics and immunisation treatments are not suitable for environmental applications and alternative methods are not well established. Phage therapy, which uses bacteriophages to lyse pathogenic bacteria, provides an alternative treatment with the potential to cure and prevent the spread of diseases caused by bacterial pathogens. Although black band disease (BBD) is a polymicrobial disease, it may be possible to decrease its progression with a lytic bacteriophage that lyses the main BBD-cyanobacterium, *R. reptotaenium*. In this study, I used bacteriophages, isolated from BBD-affected corals and purified by infection of a *R. reptotaenium* culture (*Chapter 3*), to attempt to mitigate BBD. Although results from *Chapter 3* demonstrated that a non-axenic culture of *R*. *reptotaenium* AO1 was successfully lysed, phage therapy of BBD on corals did not show reduced disease progression. Genome sequence analysis of the bacteriophages in the phage enrichment revealed GC-content, gene homologies and genome sizes related to *Cellulophaga* phages (*Podoviridae*) and *Persicivirga* phages (*Siphoviridae*). Since the cultures were not axenic and *Cellulophaga* phages and *Persicivirga* phages are known to infect Bacteroidetes, it is likely that *R*. *reptotaenium* was not the host for these bacteriophages. Instead, observed lysis of *R*. *reptotaenium* was likely due to a potential toxin or lysin released after inoculation. In addition, bacteriophage genes identified in this analysis showed sequence homology to factors involved in lysogeny and virulence. Successful phage therapy is most effective when lytic viruses are used, suggesting that the bacteriophages isolated in this study are not suitable for BBD phage therapy.

4.2 Introduction

Phage therapy, the treatment of a bacterial disease with a lytic bacteriophage, has been used to treat a range of bacterial diseases. Over the past decades, phage therapy has been successfully applied to mitigate human diseases in clinical trials (reviewed in Abedon et al. 2011), diseases in aquaculture (reviewed in Oliveira et al. 2012), and shown to prevent the progression of several coral diseases, such as white syndrome and white plaque-like disease (Atad et al. 2012; Cohen et al. 2013; Efrony et al. 2007; Efrony et al. 2009). In phage therapy, the bacteriophage infects the target host, replicates itself, and lyses the host bacterium, subsequently spreading progeny phages to any remaining host cells (Adams 1959; D'Herelle 1930). During this process, the abundance of replicating bacteriophages is inversely proportionally to the decreasing abundance of susceptible host bacterium (Jensen et al. 2006; Teplitski and Ritchie 2009; Weld, Butts, and Heinemann 2004), eventually removing the majority of the disease causing bacteria.

Over the last few decades, phage therapy has also successfully treated coral diseases, at least at relatively small experimental scales, but the potential for long-term protection against certain pathogens was evident. In some studies, bacteriophages remained active and associated with corals for several weeks after the decline in bacterial host abundance from the coral tissue (Atad et al. 2012; Efrony et al. 2007, 2009). For example, infection by *Vibrio coralliilyticus* was prevented up to a month after seawater surrounding the diseased corals had been inoculated with bacteriophages Efrony et al. (2007). In addition, after 21 days, bacteriophages targeting *Thalassomonas loyana* were detected on the coral at densities similar to initial phage therapy inoculation, even though the seawater was completely exchanged three times during the experiment (Efrony et al. 2007). Further, a field-based phage therapy

experiment in the Red Sea showed that bacteriophages remained on the coral for seven weeks (Atad et al. 2012). While phage therapy experiments were successfully conducted for coral diseases caused by single bacterial pathogen, phage therapy as a treatment for complex bacterial assemblages, such as biofilms, is less well established, but has nevertheless shown to be possible (Gabisoniya et al. 2016; Loc-Carrillo and Abedon 2011; Motlagh, Bhattacharjee, and Goel 2016). Phage therapy of BBD has been considered before, but was unsuccessful due to difficulties in cultivating BBD associated cyanobacteria (Schwenk 2012).

An important part of developing phage therapy treatments is the characterisation of prospective bacteriophages, through approaches such as genome analysis, assessment of abundance and distribution patterns in the environment, and and examination of life history strategies, e.g., lytic replication and bacterial lysis (Loc-Carrillo and Abedon 2011; Chan et al. 2013; Brüssow 2012). The genome sequence of a lytic bacteriophage provides information on the infection mechanisms and taxonomy. Additional, genome sequence analysis is recommended and will reveal possible host virulence factors carried by the bacteriophage, such as lysogeny genes, which could potentially increase the virulence of an infected pathogenic bacterium (Amarillas et al. 2016; Bardina et al. 2016). In this chapter, I present annotated genome sequences of three lytic bacteriophages isolated from a culture of the dominant pathogen of BBD, *R. reptotaenium* (*Chapter 3*) and test whether inoculation of BBD corals with this enriched viral concentrate was an effective treatment for the disease.

4.3 Methodology

4.3.1 Phage therapy

BBD-affected corals (*Pavona duerdeni*) (BBD n = 10; healthy = 5) were collected from Orpheus Island (S 18-34.609; E 146-29.793) in 3 - 5 m water depth and kept in aquaria at water temperatures of 29 - 30°C and light conditions of approximately 400 µE m^{-2} s^{-1}. BBD-affected coral colonies were fragmented (7.7 ± 1.4 cm diameter) and fragments were randomly allocated to the following treatments, each treatment with four colony replicates: BBD+phage, BBD-control, Healthy+phage, Healthy-control. Due to fragment number restrictions, BBD+phage treatment tank was replicated three times simultaneously and, other treatments tanks were replicated once. Smaller vessels (600 mL volume) with 500 mL 0.04 µm filtered seawater (FSW) were used for bacteriophage inoculation and incubation, in order to obtain high densities of bacteriophages with the available lysates to treat BBD-affected corals. BBD-affected and healthy corals were inoculated with a cocktail of putative three isolated bacteriophage populations determined by flow cytometry (10^7 VLPs mL^{-1}, *Chapter 3*) at a phage titer previously shown to be effective at causing lysis of *R. reptotaenium* cultures cyanobacteria (10^4 VLPs mL^{-1}, *Chapter 3*). Therefore, 1 mL lysate was added to vessels containing BBD-affected corals (final density: 10^5 - 10^4 VLPs mL^{-1}). When compared to previous phage therapy experiments (Atad et al. 2012; Efrony et al. 2007), the final bacteriophage density used in this experiment was ~10x greater and I used a cocktail of putative three bacteriophages instead of a single bacteriophage. Bacteriophage cocktails can be more effective than using a single isolated phage to overcome host resistance (Chan et al. 2013). The lysate was inoculated directly onto the BBD mat and coral colony with a disposable, sterile 1 mL pipette. Fragments of control treatments received 1 mL of FSW (0.04 µm) instead of bacteriophage lysate. All coral fragments were kept in 100 mL FSW (0.04 µm) for 24 hrs to allow the bacteriophages to adsorb

to their host bacteria, and then coral fragments were transferred into aquaria containing 20 L FSW (0.04 µm) (Fig. 4.1). The BBD phage therapy trial was conducted with a bacteriophage adsorption time of 24 hrs because a) cyanobacteria in culture lose their clumping behaviour after 24 hr of inoculation with bacteriophage (a putative indicator of successful infection, (*Chapter* 3; Zegans et al. 2009) and b) seawater quality in the smaller vessels decreased and had to be replaced in order to prevent damage to bacteriophages and coral. Criteria for successful phage therapy were (1) reduced BBD progression, and (2) increase in VLPs measured by flow cytometry. In culture conditions, lysis of exponentially growing BBD cyanobacteria was effective at 30 ºC and light conditions of 50 - 80 µE m^{-2} s^{-1} (*Chapter 3*). However, it was not feasible to replicate these incubator conditions for BBD-affected corals during phage therapy as the relatively high seawater temperatures of 30°C might cause additional stress to the corals and BBD progression would be maximised, with progression rates of up to several millimetres per day (Sato, Bourne, and Willis 2011). Since lysis effects in cyanobacteria cultures are visual 5 - 8 days after bacteriophage inoculation (*Chapter 3*), I kept BBD-affected corals for 14 days under light conditions of (200 µE m^{-2} s^{-1}) and seawater temperature of (28 - 29 ºC) in order to see clear treatment effects. The location of the coral within the tanks was randomly rearranged every other day to account for possible unequal light and temperature distributions. The aquaria were closed, non-circulating system in a temperature controlled room to ensure maximum control over water parameters (10 L of seawater was replaced at day 7) and aerated by means of air stones to keep water circulating and saturated with oxygen. BBD progression was monitored by photographing the coral colonies every other day, Canon PowerShot G16 (settings: macro, no flash) and measuring the progression according to the images in cm. At the end of the experiment, all seawater and equipment was soaked in fresh water with 1% bleach for 24 hrs to disinfect and kill

potential pathogenic microbes.

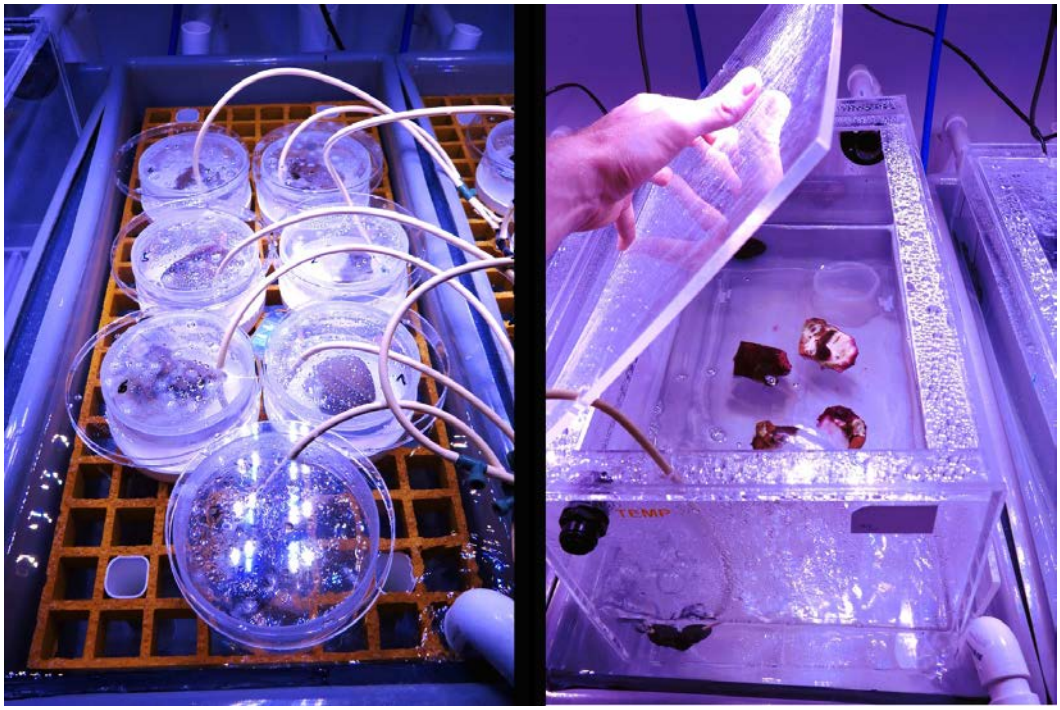


Figure 4.1 Aquarium setup for BBD phage therapy. Coral fragments were kept in 500 mL for 24 hrs for bacteriophage inoculation (left image), then moved to 20 L tanks and observed for signs of successful phage therapy (right image). Aeration ensured oxygen saturation and kept water moving within respective tanks.

4.3.2 DNA extraction for metagenome sequencing

Lysates (6 mL each) of *R. reptotaenium* AO1 cultures with putative three VLP populations were filtered through 0.45 µm and centrifuged at 3,000 *g* for 3 min to remove larger cell debris. Supernatants were subsequently concentrated through a centrifugal spin column with a 30 kDa membrane (Amicon filters, Milipore) and 1 mL of the retentate was recovered. The concentrated lysate was treated with DNase TURBO (6 µL, 2 U/µL) and RNase (4 µL, 5 µg/mL) in 110 µL 10x DNase buffer (Ambion) and incubated at 37°C for 30 min to remove free DNA and RNA prior to viral nucleic acid extraction. DNA was purified from 500 µL lysate with a column-based extraction kit

(Macherey-Nagel, DNA extraction), according to the manufacturers recommendation, with an initial proteinase K (20 mg μL^{-1}) step to digest viral capsids. Total DNA content was amplified in a Random Priming-mediated Sequence- Independent Single-Primer Amplification (RP-SISPA) approach modified from (Weynberg et al. 2014). In brief, 5 µL of DNA extraction was mixed with 1 µL dNTPs (2.5 mM), 1.5 µL PCR buffer (10x, New England Biolabs Buffer II), 1.5 µL FR26RV-N primer (10 µM, GCCGGAGCTCTGCAGATATCNNNNNN) and 5 µL DNase-free distilled water. The dsDNA was separated and labelled with the RP-SISPA primer by incubating the mix for 3 min at 94°C and for 3 min on ice. 1 µL of Klenow fragment (3 - 5 exo-, 5U/µL) was added to the reaction and incubated for 60 min at 37°C to amplify the first DNA stand. The second DNA strand was labelled by adding 1 µL dNTPs (2.5 mM) and 1.5 µL FR26RV-N primer, incubation at 94°C, for 3 min on ice and 60 min at 37°C. The reaction was terminated at 75°C for 20 min. The labelled DNA strands were amplified in five replicates with each 25 µL reactions by mixing 1 µL of DNA template with 15.25 µL PCR-grade water, 2.5 µL reaction buffer (10x, LA buffer, Clontech), 4 µL dNTPs (2.5 mM), 2 µL FR20RV primer (10 µM, GCCGGAGCTCTGCAGATATC) and 0.25 µL TaKaRa LA HS Taq (5 U/µL, Scientifix). Polymerase chain reaction had the following steps: denaturation 10 min at 95°C, 30 cycles at 95°C for 30 sec, 60°C for 1 min, 72°C for 1.5 min, and a final extension at 72°C for 13 min. A reconditioning PCR step was conducted by mixing 55.25 µL PCR water with 10 µL reaction buffer (10x, LA buffer, Clontech), 16 µL dNTPs (2.5 mM), 8 µL FR20RV primer (10 µL stock concentration) and 0.75 µL TaKaRa LA HS Taq (5 U/µL, Scientifix). The PCR conditions were as previously described with only 5 amplification cycles. The five RP-SISPA reactions were cleaned with a Qiagen MiniElute PCR purification kit, pooled, and sequenced with a Nextera XT library preparation and 2x300 paired-end V3 chemistry on a MiSeq Illumina platform at the Ramaciotti Centre, Sydney.

4.3.3 Genome analysis

The paired-end sequence reads obtained in the Illumina MiSeq platform were merged and subsequently assembled with the default parameters of the software CLC genomics workbench (version 8.5.1, CLC Bio). Bioinformatic statistics to assess the final assembly were generated with gnx-tools (version 0.1+20120305, github.com/mh11/gnx-tools), and included for example N50 value and total assembly length. Circular genome projections were generated with pDRAW32 (www.acaclone.com). Open reading frame (ORF) identification (minimum 75 nt) and gene annotation was conducted using Artemis (Rutherford et al. 2000). Genes were annotated with a local BLASTx 2.2.30+ (Altschul et al. 1990) for homologies with the viral RefSeq database at NCBI (version: 11.01.2017, www.ncbi.nlm.nih.gov) and a recently developed virus marker database within the bioinformatics pipeline HoloVir (Laffy et al. 2016). Additionally, gene annotations were completed with an automated annotation pipeline RAST (Rapid Annotations using Subsystems Technology, Overbeek et al. 2014) and manual BLASTp to find homologies in the 'nr' database at NCBI. The best *e*-values across the identification methods were considered for gene annotations. Genes coding for tRNAs were predicted with the software tRNAscan-SE 2.0 (Lowe and Eddy 1997), and subsequently analysed with a BLASTn to identify homologues in the viral RefSeq database. Best and most frequent taxon database matches of ORF and tRNAs were used to identify taxonomic affiliations of the respective contigs. According to the taxonomic affiliations, genome comparisons among respective reference genomes and assembled contigs at amino acid level was conducted with progressiveMAUVE (Darling, Mau, and Perna 2010) to determine overall genetic similarities.

Annotated genes were also scanned for specific marker genes, such as auxiliary metabolic genes (AMGs) that are common among cyanophages (list of genes in Shestakov and Karbysheva 2015; Crummett et al. 2016). Contigs with similar taxonomic affiliations and GC-content were identified and binned as putative bacteriophage genomes. BLASTn was used to compare the assembled virus contigs to several local data sets: a) T4 bacteriophage community members (*Chapter 2*), b) potential host bacteria in the cyanobacterium cultures (*Chapter 3* and *5*), c) CRISPR-Cas spacers of *R. reptotaenium* AO1 and another cyanobacterium associated with BBD *Geitlerinema* sp. BBD_1991 (*Chapter 5;* Den Uyl et al. 2016), and d) viral reads from a recent metagenomic study (Sato et al. 2017).

4.4 Results

4.4.1 Phage therapy

Phage therapy treatments on BBD-affected corals showed no difference between the progression of BBD in treatments with and without bacteriophages (Fig. 4.2). Seawater quality in 500 mL vessels decreased during incubation time for all BBD-affected corals, controls and treatments (this was not quantified but was noticeable by sulfide odour). Controls that were inoculated with the bacteriophage cocktail showed no signs of BBD.

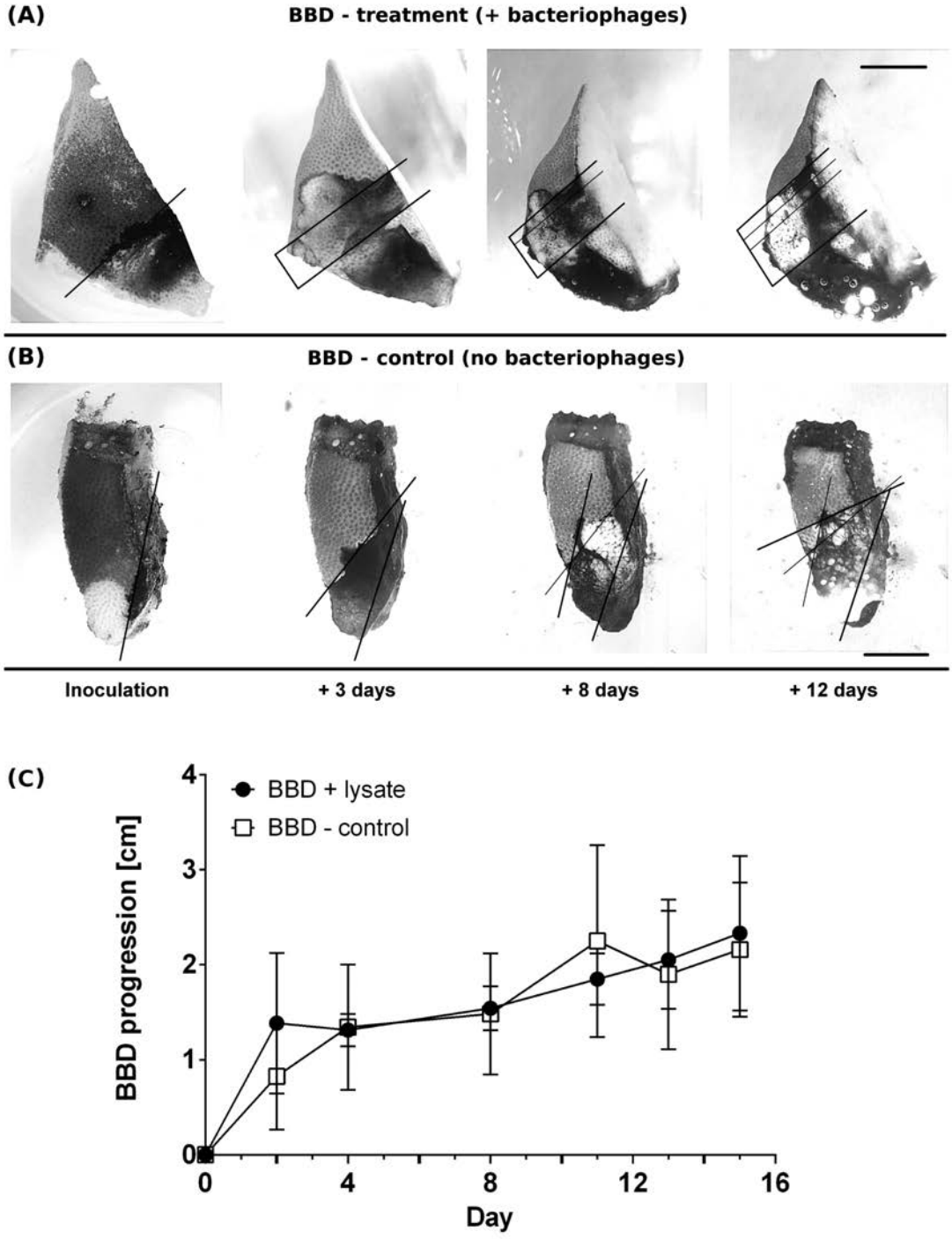


Figure 4.2 BBD progression during phage therapy trial. Coral with BBD in (A) bacteriophage treatment, (B) control with no bacteriophage inoculation. Among all fragments and treatments, no difference in BBD progression has been detected (scale bars = 2 cm). Lines on the BBD-affected corals mark the lesion front at the respective day. The start line and the most recent line are highlighted in bold. (C) BBD progression in cm is shown over time.

4.4.2 Bacteriophage genome descriptions

Assembly of the sequence reads (Appendix 4.1) resulted in three relatively large contigs with high coverage (26,089x - 4,824x) and 879 smaller contigs with relatively low coverage (average 13.01x ± 5.97, min 1.86x, max 41.37x) (Table 4.1). Based on the putative three dominant viral populations from flow cytometer analyses (*Chapter 3*), the assembly of 3 main viral contigs was predicted.

Contig_1 had a GC-content of 34.2%, with 196 predicted ORFs, and the longest reading frame (ORF) consisting of 2075 amino acids (Table 4.1). About a third of the ORFs (n = 64 out of 196) could be annotated with homologies to sequences of known bacteriophage genes (Fig. 4.3A, Appendix 4.2). In addition, a total of 20 tRNAs were detected (Appendix 4.3). The most frequent assignment of the translated ORF annotations were to *Cellulophaga* phages, *Podoviridae* (n = 39 out of 196, 19.9%, Table 4.2), while other bacteriophage genera showed comparably few hits (e.g. genera of phages infecting cyanobacteria = 4, Table 4.2). The overall alignment of contig_1 with the most closely related available reference genomes (i.e. *Cellulophaga* phage phi4:1, 27 homologous genes out of 39) showed generally more dissimilarities than overlapping genomic regions (Appendix 4.4, 4.5).

Table 4.1 Genomic descriptions of contigs and references. Three relatively large contigs were assembled with high coverage, while contigs 4-882 showed considerably shorter with lower coverage. Open reading frames (ORF) were predicted only for the largest three contigs. Maximum ORF sizes are presented in number of amino acids [aa] and compared to the existing data of respective reference genomes. Taxonomic affiliations were assigned according to the best and most frequent BLASTx hits (Table 4.2). Genome characteristics of the closest available reference species are shown in the second section of the table 'Reference genomes'. The %-value of orthologue genes for the viral contig 1-3 and their respective reference genome is shown in the column 'Closest reference to'.

Contig #	**Coverage**	**Length [bp]**	**GC content [%]**	**ORF #**	**ORF max [aa]**	**tRNAs**	**Affiliated taxonomy**
Contig_1	26,089	148,900	34.2	196	2,075	20	*Podoviridae*
Contig_2	63,033	75,795	41.9	92	1,528	16	*Podoviridae*
Contig_3	4,824	41,880	38.5	57	747	0	*Siphoviridae*
Contig_4-882	43.83 ± 7.71	13.01 ± 5.98	3,89 ± 3,56	NA	NA	NA	NA
Reference genomes							**Closest reference to**
Cellulophaga phage phi4:1	NA	145,700	33.0	198	2,290	24	contig_1 (19.9%)
Cellulophaga phage phi38:1	NA	72,534	38.1	101	1,692	16	contig_2 (26.1%)
Persicivirga phage P12024S	NA	35,700	35.7	59	755	0	contig_3 (12.3%)

Contig_2 had a total length of 75,795 bp, a GC-content of 41.9%, 92 predicted ORFs and the longest ORF at 1528 amino acids (Table 4.1, Fig. 4.3B, Appendix 4.2) and 16 tRNAs (Appendix 4.3). While taxonomic affiliations showed best BLASTx hits to *Cellulophaga* phages, *Podoviridae* (35 out of 92 ORFs, 38.1%, Table 4.2), the overall genome alignment with closest reference genomes (e.g., *Cellulophaga* phage phi38:1, 13 homologous genes) showed more dissimilarities than overlapping regions (Fig. 4.4).

Contig_3 was the smallest of the three isolated virus genomes with 57 ORFs and a total length of 41,880 bp. Sequence homology was detected for 35 of the ORFs (31.6%) (Fig. 4.3C, Appendix 4.2, 4.3), with the closest taxonomic affiliation to *Persicivirga* phages, *Siphoviridae* (n = 7 out of 57, 12.3%, Table 4.2). The GC-content of contig_3 was 38.5% with no tRNAs detected and only marginal similarities in genome alignment to closest reference sequences (Fig. 4.4).

Table 4.2 Taxonomic identification of bacteriophage genomes. Best annotation matches were counted for taxon associations (full annotation Appendix 4.2). A contig was taxonomically identified according to the most frequent and best BLASTx hits (highlighted in bold).

Contig #	Taxon reference	Sum of annotation matches
Contig_1	***Cellulophaga* phage**	**39**
	Cellulophaga phage phi4:1	27
	Cellulophaga phage phi17:2	9
	Other *Cellulophaga* phages	3
	cyanophage *Synechococcus* phage *Prochlorococcus* phage	4
	Cyanophage S-TIM5	2
	Other phages infecting cyanobacteria	2
Contig_2	***Cellulophaga* phage**	**25**
	Cellulophaga phage phi38:1	13
	Cellulophaga phage phi46:3	3
	Other *Cellulophaga* phages	9
	cyanophage *Synechococcus* phage *Prochlorococcus* phage	2
	Synechococcus phage S-RSM4	1
	Synechococcus phage S-SSM7	1
	Staphylococcus phage	2
	Staphylococcus phage SA11	1
	Staphylococcus phage vB_SauM_Romulus	1
Contig_3	***Persicivirga* phage**	**7**
	Persicivirga phage P12024S	5
	Persicivirga phage P12024L	2
	Cellulophaga phage	5
	Cellulophaga phage phi10:1	1
	Cellulophaga phage phi14:2	1
	Cellulophaga phage phi19:1	1
	Cellulophaga phage phi19:3	1
	Cellulophaga phage phi38:1	1
	cyanophage *Synechococcus* phage *Prochlorococcus* phage	0

Several smaller contigs also showed gene similarities to *Cellulophaga* phages. For example, on contig_26, a gene was identified coding for chaperonin GroEL (*e*-value: 1*e*-084, YP_008241386.1, *Cellulophaga* phage phi38:1), and a gene similarity to an exonuclease of *Cellulophaga* phage phi38:1 (*e*-value: 9*e*-033, YP_008241385.1). However, since these contigs are rather short (e.g. contig_26: 6,431 bp) with relatively

low coverage (e.g. contig_26: 16.7x) and these exact genes are also found on the larger contigs 1-3, I did not include them as additional parts of the viral genomes (contig 1-3).

Auxiliary metabolic genes (AMGs) commonly associated with bacteriophages infecting cyanobacteria were not found on contigs 1-3. While a few AMGs were found on seven of the smaller, low coverage contigs (such as ferredoxin *petF* and phosphate starvation-inducible protein *Phoh*, Appendix 4.6), no homologs were detected for viral photosynthesis genes (psbA, psbD). The longest contigs that contained cyanophage marker genes were relatively short with 11,800 bp and 11,579 bp compared to the main contigs and compared to the complete genome sizes of cyanophages (usually >100,000 bp, min ~40,000 bp, www.ncbi.nlm.nih.gov). Since these contigs also showed partially not very good *e*-values matches to the cyanophage marker genes (>e-024, in 5 out of 7 markers), and had different GC-contents and coverage to each other, they were unlikely to correspond to an unassembled cyanophage in the data set.

In order to assess the completeness of the assemblies, it is possible that the three predominant viral genomes are not fully assembled into one complete contig but are instead spread across several smaller contigs. In support of this notion, a few smaller contigs showed gene similarities to *Cellulophaga* phages. For example, on contig_26, a gene was identified coding for chaperonin GroEL (*e*-value: 1*e*-084, YP_008241386.1, *Cellulophaga* phage phi38:1), and a gene similar to an exonuclease of *Cellulophaga* phage phi38:1 (*e*-value: 9*e*-033, YP_008241385.1). However, since these contigs are rather short (e.g. contig_26: 6431 bp) with low coverage (e.g. contig_26: 16.7x) and these genes are also found on the larger contigs 1-3, I did not include as part of the viral genomes (contig 1-3).

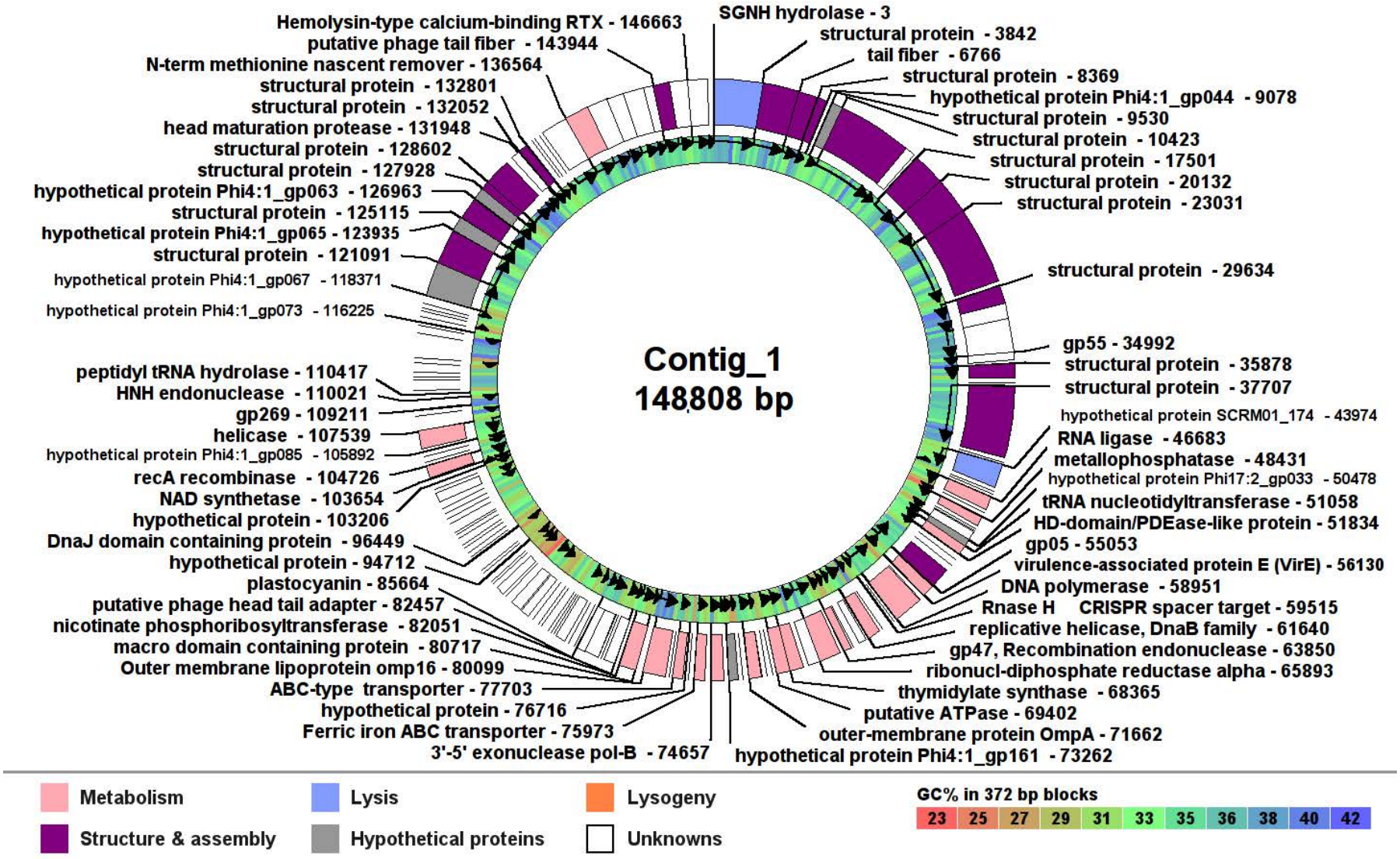


Figure 4.3A Circular genome representations of contigs 1-3. GC-content is shown for the individual genome regions in colour. Open reading frames (ORF) are indicated with their respective reading direction and colour according to their function. Annotation ticks mark the start of the reading frame. Detailed annotations in Appendix 4.2. A) Contig_1. B) Contig_2. C) Contig_3.

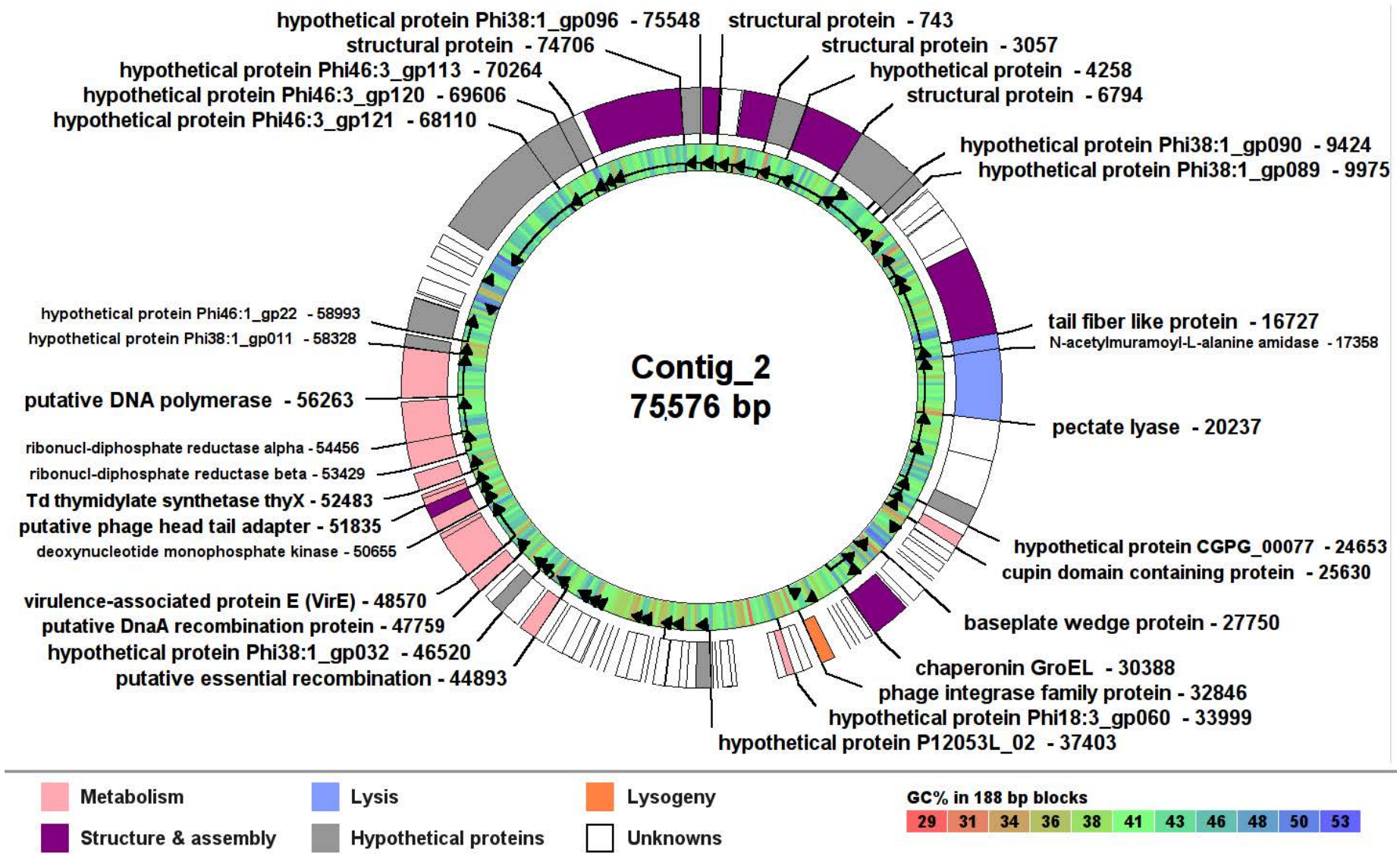


Figure 4.3B Circular genome representations of contigs 1-3. GC-content is shown for the individual genome regions in colour. Open reading frames (ORF) are indicated with their respective reading direction and colour according to their function. Annotation ticks mark the start of the reading frame. Detailed annotations in Appendix 4.2. A) Contig_1. B) Contig_2. C) Contig_3.

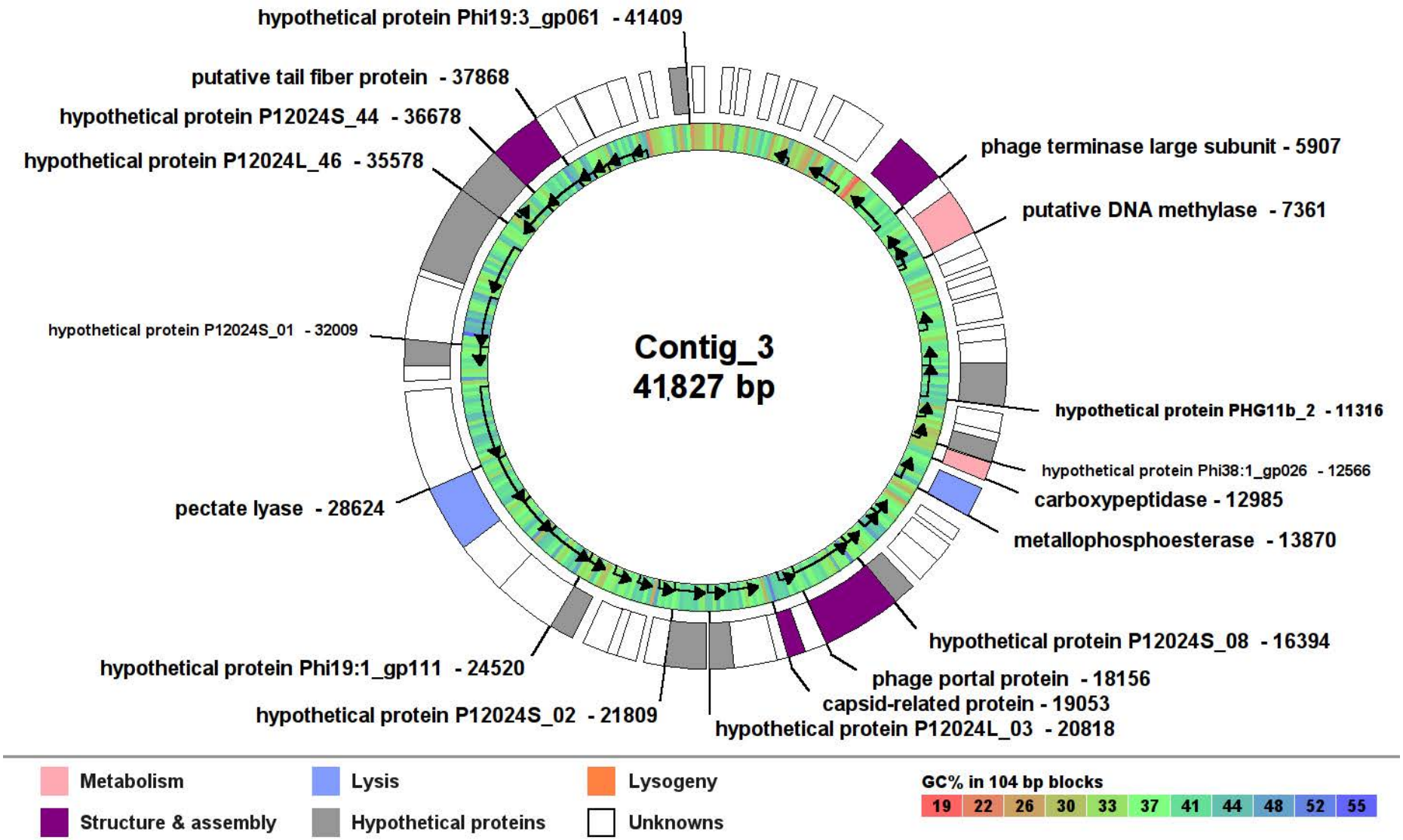


Figure 4.3C Circular genome representations of contigs 1-3. GC-content is shown for the individual genome regions in colour. Open reading frames (ORF) are indicated with their respective reading direction and colour according to their function. Annotation ticks mark the start of the reading frame. Detailed annotations in Appendix 4.2. A) Contig_1. B) Contig_2. C) Contig_3.

The data sets were scanned for evidence of horizontal gene transfer from the bacterial hosts into the virus as well as possible host contamination. Several genes in smaller, low coverage contigs showed high nucleotide similarities to genes in the main bacteria present in the culture from which the phage nucleic acids were extracted: *R. reptotaenium* AO1, an Alphaproteobacterium and *Cytophagaceae* sp., and similarities to genes of the genome of *Geitlerinema* sp. BBD_1991 (*Chapter 5*). Genes with high similarity to *Cytophagaceae* and *Alphaproteobacterium* included translation elongation factors, ATP synthase alpha chains, Na(+)-translocating NADH-quinone reductase subunits and exonuclease ABC subunits (*e*-value ranging from 0.0 - 1*e*-13, Appendix 4.7). Other genes of assembled short contigs had matches in cyanobacteria genomes, such as contig_783 (length 1265 bp, coverage 4.71), with a particularly high GC-content of 57.0%, was similar to a sulfurtransferase of *Geitlerinema* sp. BBD_1991 (Chapter 5; Den Uyl et al. 2016), a tRNA of *R. reptotaenium* AO1 and a conserved protein domain of DNA-binding transcriptional regulator AraC (*e*-value 9.37*e*-07, PRK10572, BLASTx). In addition, several genes matched an Alphaproteobacteria rRNA gene subunit, another bacteria that was present in the cyanobacterium culture (Appendix 4.5). Since only genes of short contigs matched to bacterial genes, I considered them bacterial contamination that had not been fully removed before DNA extraction and sequencing. Detection of genes that had been horizontally transferred between the bacterial host and a virus would have co-occurred on the larger viral contigs 1-3.

Using tBLASTx analyses, several genes of contigs 1-3 showed amino acid similarities to viral reads in a recently published BBD-metagenomic data set (Sato et al. 2017), such as genes coding for a HNH endonuclease, ribonucleoside-diphosphate reductase alpha subunit as well as genes, thymidylate synthetase thyX and DNA

methylase (full tBLASTx gene similarity list in Appendix 4.7). For example, the gene encoding for HNH endonuclease on contig_1 was similar to several viral reads (best blast hit: *e*-value 8e-07). The assigned functions and taxonomic origins of the respective metagenome reads (identified with a BLASTp) confirmed the annotations of the viral contigs in this study. However, taxonomic affiliations were not always consistent and included *Vibrio* phage pVp-1, cyanophages and *Synechococcus* phages (Appendix 4.7). There were no perfect matches between the data sets (best *e*-value 8e-07) and the gene matches to the metavirome were not to *Cellulophaga* phages, indicating that the three main viral contigs were not detected in the BBD-metagenome. Unsurprising, no similarities via BLASTn and tBLASTx were detected between the viral contigs and the T4 bacteriophage data set (*Chapter 2*).

4.5 Discussion

BBD phage therapy was not feasible with the current bacteriophages isolated via the cultured cyanobacterium, *R*. *reptotaenium*. Genome sequence analyses of the isolated bacteriophages showed three phages that are related to *Cellulophaga* and *Persicivirga* phages, respectively (Holmfeldt et al. 2013; Kang, Jang, and Cho 2012). Since the culture was monoclonal from a single cyanobacterial filament, but not axenic (it contained other bacteria, such as *Alphaproteobacterium* and *Cytophagaceae* sp., *Chapter 3*, *5*), bacterial hosts may have been present in the culture. The phages isolated in this study likely do not infect the cyanobacterium *R*. *reptotaenium*, and are therefore unsuitable for a BBD phage therapy.

4.5.1 Bacteriophage genome features

The results presented in this chapter on genome homology, GC-content, and genome size allow for a rough taxon assignment for contigs 1-3 (Table 4.1, 4.2), with contig_1

and contig_2 most likely belonging to *Cellulophaga* phages (*Podoviridae*) (Holmfeldt et al. 2013) and contig_3 to *Persicivirga* phages (*Siphoviridae*) (Kang, Jang & Cho 2012). A minimum of 40% orthologous proteins is recommended prior to classification into an existing bacteriophage genus (Lavigne et al. 2008). The percent similarity of orthologous proteins on contig_1 and contig_2 to *Cellulophag*a phage 4:1 and 38:1 was 19.9% and 26.1%, respectively, while contig_3 showed 12% protein similarity to *Persicivirga* phages (Table 4.1, 4.2). Thus, these phages likely represent novel bacteriophage genera. Nevertheless, for a proper taxonomic classification of the bacteriophages, further information is required, e.g. regarding their morphology (capsid structure and symmetry), exact size of the closed genome, burst size and host range (Büchen-Osmond 2003), which was beyond the scope of this study.

Contig_1 showed genetic similarities to *Cellulophaga* phage phi4:1, proposed genus Cba41likevirus (Holmfeldt et al. 2013), based on genome length, number of ORFs, longest ORF, number of tRNAs, GC-content, and gene annotations. Contig_2 was assigned to *Cellulophaga* phage phi38:1, proposed genus *Cba40likevirus* (Holmfeldt et al. 2007, 2013), using the same criteria. *Cellulophaga* phages (*Podoviridae*) can carry exceptionally large genomes of up to 150 kb (contig_1: 148,900 bp, contig_2: 75,795 bp), while 93% of podoviruses have genomes smaller than 70 kb (Holmfeldt et al. 2013). In addition, *Cellulophaga* phages encode for numerous tRNAs (Holmfeldt et al. 2013), which holds true for contigs 1 and 2 in this study (Table 4.1, Appendix 4.3). Some bacteriophages are known to encode their own tRNAs, such as *Cellulophaga* phages, to expand their codon usage in order to match the host's GC-content and to enable infection of a wider range of bacterial hosts (Enav, Béjà, and Mandel-Gutfreund 2012), as well as to promote bacteriophage replication within the host (Delesalle et al. 2016). *Cellulophaga* phages typically encode for one of

the two chaperonin genes, GroEL and Cpn10 (Holmfeldt et al. 2014), which are probably involved in folding viral capsid proteins (Hildenbrand and Bernal 2012). While no chaperonin gene was detected on contig_1, GroEL was detected on contig_2. Thymidylate synthetase and ribonucleotide-diphosphate reductase (RNR), genes involved in *de novo* synthesis of nucleotides, were identified on contig_1 and contig_2. These genes are uncommon for podoviruses (Dwivedi et al. 2013; Holmfeldt et al. 2013), but have been found in *Cellulophaga* phages previously (Holmfeldt et al. 2013). The gene VirE on contig_1 and contig_2 was annotated as a virulence factor, but may also be involved in DNA replication and host interaction (Holmfeldt et al. 2013). A phage integrase gene was detected on contig_2. This gene plays a key role in the integration of bacteriophage genomes into the host genome (Bellanger et al. 2014; Landy and Ross 1977), and suggests the potential for lysogeny of contig_2. No other genes that may be related to a lysogenic life history were detected and the functionality of the integrase gene was not assessed.

Contig_3 showed 12.3% orthologous protein similarity to *Persicivirga* phages, which are proposed to belong to *Siphoviridae* (Kang et al. 2012). Core genes for structural assembly were similar to *Persicivirga* phages, such as phage terminase large subunit, a phage portal protein, capsid-related protein and a putative tail fibre protein (Appendix 4.2). However, other genes commonly encoded for by *Persicivirga* phages genomes were missing, such as VRR-NUC domain protein (nuclease superfamily) and a YqaJ-like viral recombinase, both likely required for host interactions. Since they were not detected on any other smaller contigs in the data set, it is possible that the bacteriophage may not be fully functional or uses other genes for host interactions. A pectate lyase on contig_3, which has not been detected as part of a *Persicivirga* phage genome before (Kang et al. 2012), has been found on a

Cellulophaga phage phi13:2 and may be involved in host recognition and cell wall penetration (Holmfeldt et al. 2014).

4.5.2 Potential hosts of virus contig 1-3 in cyanobacteria cultures

The isolated bacteriophages (affiliated to *Cellulophaga* phages and *Persicivirga* phages) are not known to infect cyanobacteria, but infect bacteria of the phylum Bacteroidetes, such as *Cellulophaga baltica* (Holmfeldt et al. 2013) and *Persicivirga* sp. respectively (Kang et al. 2012). This brings into question what agent caused lysis of the cyanobacterium *R. reptotaenium* AO1 in culture. Lysis of *R. reptotaenium* AO1 was successful after a 0.45 µm pre-filtration step, but was not observed following a 0.22 µm filtration step (*Chapter 3*), a commonly used filtration size that allows many viruses to pass through. However, I can exclude the possibility that the filtration step of 0.22 µm removed some of the viruses from the lysate, since I did not filter sterilise the lysate that was sequenced and subsequently did not assemble a genome of a characteristically large virus. It is more likely that filtering lysates through 0.22 µm removed most bacterial contamination and prevented the transfer of large cell debris, while filtering through 0.45 µm left most of the bacterial contamination in the filtrate. Accordingly, the culture was not axenic and contained an *Alphaproteobacterium, Cytophagaceae* sp. (*Chapter 3*, *5,*) as well as potentially other bacteria that have been introduced with the inoculation of the 0.45 µm pre-filtered lysate from BBD. Therefore, bacteria other than the cyanobacterium *R. reptotaenium* AO1 could have been infected during the bacteriophage enrichment. The cyanobacteria may have been lysed in a secondary reaction, e.g. due to the enrichment and release of toxins and lysins from the lysis of other infected bacteria (Fenton et al. 2010; Yoong et al. 2004). A potential gene located on virus contig_2 that could have caused such a reaction is coding for an amidase (Appendix 4.2), which is known to be a lysate with broad range effect on

peptidoglycan of gram-negative bacteria (Drulis-Kawa et al. 2012; Pastagia et al. 2013) and also carried by cyanophages (e.g. Sullivan et al. 2005).

Cellulophaga phages have previously been detected in association with BBD before, e.g. in a recent metagenome and metatranscriptome that compared the taxonomic compositions and gene expression profiles of BBD with a pre-disease cyanobacterial patch (CP) (Sato et al. 2017). Among the taxa that were identified in the metagenomic reads were *Cellulophaga* phage phi10:1 and phiST. My BLAST analysis that compared the isolated viruses of the present study with the metagenomic and metatranscriptomic reads did not show sequence similarities between the respectively associated *Cellulophaga* phages (Appendix 4.7), indicating the novelty of the isolated viruses presented in this thesis. To highlight the potential involvement of *Cellulophaga* phages in BBD, the relative abundance of *Cellulophaga* phages was higher in fully developed BBD compared to the pre-disease CP, while their known host bacteria of the phylum Bacteroidetes were relatively less abundant in BBD compared to CP (Sato et al. 2017). However, in this case, relative abundances should be taken with caution, since metagenomic and metatranscriptomic samples were pooled and not replicated, and overlapping regions between the two studies were potentially not present, since metagenomic reads were relatively short (100 - 192 bp).

Currently, phage therapy of BBD requires further optimisation in order to successfully mitigate the disease, e.g. obtaining highly specific, lytic viruses for the major BBD-cyanobacteria in a culture setting to inoculate a bacteriophage cocktail. Bacteriophage infections experiments should be conducted at different temperatures, to combine optimal but environmentally relevant conditions with shorter bacteriophage incubation times. Additionally, healthy corals that have been pre-conditioned with lytic

BBD-cyanophages may increase resistance to BBD and should be tested in infection experiments. An ideal candidate bacteriophage for a phage therapy is highly efficient in killing their host (virulent), infects only one particular bacterium (host specific) and does not carry genes for lysogeny or virulence (Keen and Adhya 2014; Nobrega et al. 2015). In contrast, the bacteriophages in this study do not appear to infect cyanobacteria, but are generalists infecting a wider range of bacteria of the phylum (Bacteroidetes) that possess potential lysogeny genes as well as potential virulence. Since it is unknown to what extent and how the bacterium *Cellulophaga* sp. contribute to BBD and influence coral health, *Cellulophaga* phages should not be used in a phage therapy to treat BBD. The isolated bacteriophages are therefore undesirable candidates for BBD phage therapy.

While a bacteriophage infecting *R*. *reptotaenium* was not isolated with the current methodology, future studies should obtain axenic monoclonal cultures of *R. reptotaenium.* Axenic monoclonal cultures that provide only a single host during bacteriophage isolation, will increase the likelihood of isolating a cyanophage that infects *R*. *reptotaenium* and will prevent cyanobacteria lysis due to unknown reasons such as secondary toxin release of other bacterium infections.

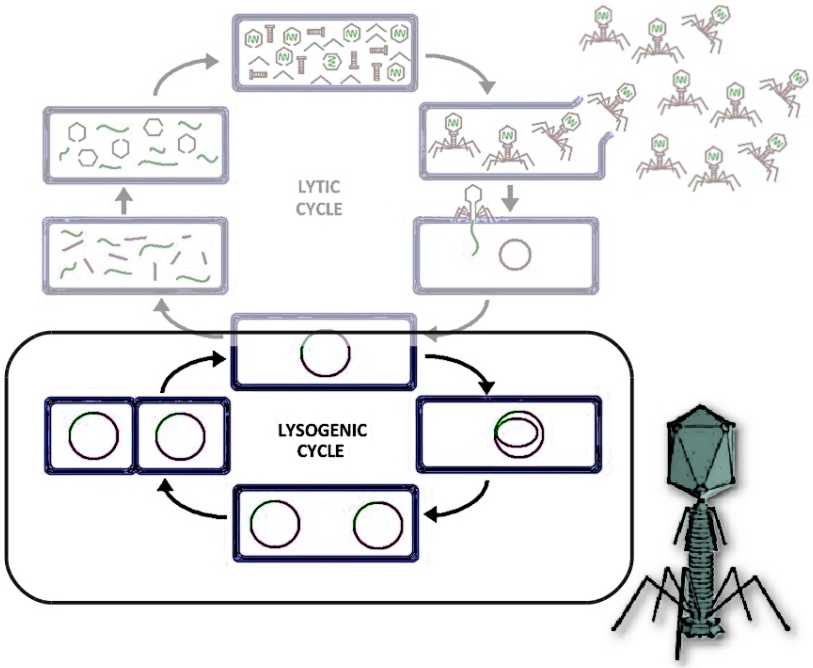
4.5.4 Conclusion

Phage therapy on BBD should be conducted with bacteriophages, such as cyanophages, that are highly virulent and infect *R*. *reptotaenium*. A treatment effect in current phage therapy experiment was not observed, because the applied bacteriophages did not infect *R*. *reptotaenium*. The cyanobacterium was lysed in culture most likely due to additional toxins and lysates that were released during lysis of other bacteria. In order to establish a successful phage therapy protocol, different bacteriophages have to be isolated and preferably used in a cocktail of several

different cyanophages to overcome potential host resistances among BBD-cyanobacteria.

Chapter 5

Lysogenic bacteriophages as potential contributors to black band disease virulence



A portion of this chapter is published as:
Buerger P, Wood-Charlson EM, Weynberg KD, Willis BL and van Oppen MJH (2016) CRISPR-Cas defence system and potential prophages in cyanobacteria associated with the coral black band disease *Frontiers Microbiology* 7:2077 10.3389/fmicb.2016.02077

5.1 Abstract

Understanding how pathogens maintain their virulence is critical to developing tools to mitigate disease in animal populations. Since lysogenic bacteriophages can be contributors to the virulence of bacteria, I sequenced and assembled the first draft genome of *Roseofilum reptotaenium* AO1, the dominant cyanobacterium underlying pathogenicity of the virulent coral black band disease (BBD), and analysed parts of the BBD-associated *Geitlerinema* sp. BBD_1991 genome *in silico.* Both cyanobacteria are equipped with an adaptive, heritable CRISPR-Cas defence system type I-D and have potential virulence genes located within several prophage regions. The defence system helps to prevent infection by viruses and mobile genetic elements via identification of short fingerprints of the intruding DNA, which are stored as templates in the bacterial genome, in so-called "clustered regularly interspaced short palindromic repeats" (CRISPRs). Analysis of CRISPR target sequences (protospacers) revealed an unusually high number of self-targeting spacers in *R. reptotaenium* AO1 and extraordinary long CRIPSR arrays of up to 260 spacers in *Geitlerinema* sp. BBD_1991. The self-targeting spacers are unlikely to be a form of autoimmunity; instead these target an incomplete lysogenic bacteriophage. Since lysogenic virus induction experiments with mitomycin C and UV light did not reveal an actively replicating virus population in *R. reptotaenium* AO1 cultures, I propose that phage functionality is compromised or excision could be blocked by the CRISPR-Cas system. Potential prophages were identified in three regions of *R. reptotaenium* AO1 and five regions of *Geitlerinema* sp. BBD_1991, containing putative BBD relevant virulence genes, such as an NAD-dependent epimerase/dehydratase (a homologue in terms of functionality to the third and fourth most expressed gene in BBD), lysozyme/metalloendopeptidases and other lipopolysaccharide modification genes. To date, viruses have not been considered to be a component of the BBD consortium or a contributor to the virulence

of *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991. I suggest that the presence of virulence genes in potential prophage regions, and the CRISPR-Cas defence systems are evidence of an arms race between the respective cyanobacteria and their bacteriophage predators. The presence of such a defence system likely reduces the number of successful bacteriophage infections and mortality in the cyanobacteria, facilitating the progress of BBD.

5.2 Introduction

Diseases have become a major contributor to coral mortality over the last few decades (Bourne et al. 2009; Rosenberg et al. 2007; Willis, Page, and Dinsdale 2004). Black band disease (BBD) is one of the most widely reported ones (Aeby et al. 2015; Barneah et al. 2007; Green and Bruckner 2000; Johan et al. 2015; Page and Willis 2006). The disease consists of a microbial consortium of cyanobacterial species, sulphate -reducing and -oxidizing bacteria, *Alphaproteobacteria, Cytophaga*, as well as other heterotrophic bacteria (Cooney et al. 2002; Miller and Richardson 2011; Sato, Bourne, and Willis 2011). Important questions, such as the direct cause of BBD onset and progression drivers, remain unclear (Sato et al. 2016). Although the role of viruses, in particular bacteriophages, can provide new insights into the drivers and causation of a coral disease (Weynberg et al. 2015), BBD etiology has most commonly been studied by investigating the functions of associated bacteria, while the role of bacteriophages has not been considered.

Lysogenic bacteriophages change a bacterium's behaviour and virulence by integrating new genetic material into their genome (Brüssow, Canchaya, and Hardt 2004). For example, the bacteriophage (CTXphi) transfers primary virulence factors into *Vibrio cholerae* that contribute to the bacterium's pathogenicity, such as the cholera

toxin (CT). This infection converts *V. cholerae* from a non-pathogenic to a pathogenic strain (Faruque and Mekalanos 2003; Waldor and Mekalanos 1996). Lysogenic conversion may also transform *Vibrio coralliilyticus* into a coral pathogen triggering the coral disease white syndrome, since some of its virulence factors are found on pathogenicity islands that contain toxin genes homologous to those of the *V. cholerae* prophage (Weynberg et al. 2015).

Conversely, bacteria can prevent bacteriophage infection by maintaining defence mechanisms, such as 'clustered regularly interspaced short palindromic repeats' (CRISPR) associated systems. Short sequences (spacers) that match the foreign target DNA and RNA sequences (protospacers) are stored in repetitive regions, known as CRISPR arrays (reviewed in Makarova et al. 2011a,b). An operon of associated genes codes for proteins (Cas) that detect and cleave the foreign DNA, guided by the transcribed spacers (crRNAs). CRISPR-Cas systems are known to be prevalent in ~10% of currently sequenced bacterial genomes (Burstein et al. 2016), but seem to be more widespread among cyanobacterial genomes (detected in 68.3%, 86 out of 126 cyanobacterial genomes, Cai et al. 2013).

In this study, I test for the existence of interactions between bacteriophages and the main BBD cyanobacteria, *R. reptotaenium* and *Geitlerinema* sp. BBD_1991. I evaluate the likelihood that the cyanobacteria are target of bacteriophage predation, which could influence their virulence and the development of BBD pathogenicity. By sequencing the genome of *R. reptotaenium* AO1 (*Chapter 3*, Buerger et al. 2016) and retrieving the available genome for *Geitlerinema* sp. BBD_1991, I analyse the two data sets *in silico* for prophage integrations, bacteriophage-host interactions, and host defence mechanisms against virus infections.

5.3 Methodology

5.3.1 DNA-extraction and sequencing

The main BBD-associated cyanobacterium, *R. reptotaenium* (Rasoulouniriana) Casamatta, was isolated and cultured as described in *Chapter 3*. DNA was extracted from 50 mg (dry weight) samples of the cyanobacterial biomass with a Mo-Bio Power Plant Pro DNA extraction kit (cat. 13400-50), according to the manufacturer's recommendations, with the following small modifications. Samples were: 1) bead-beaten in Power Plant Pro kit solution PD1 (450 µL), PD2 (50 µL), and RNase A (3 µL, 25 mg/mL) for 1 min at max speed with (Fastprep-25 5G, MP Biomedicals); 2) incubated for 1 hour at 56°C and 10 min at 65°C with proteinase k (15 µL, 20 mg/mL); and 3) eluted with 2 x 50 µL TE (10mM Tris/Hcl, pH 8.5, 0.1mM EDTA). Approximately 2.5 µg of purified DNA (Zymo genomic DNA clean and concentrator) was sent for next generation sequencing with a Truseq library preparation on a MiSeq 2 x 300 V3 (Ramaciotti Centre, University of New South Wales).

5.3.2 Genome assembly and annotation

Paired-end sequences were merged with PEAR 0.9.5, using default parameters. Low quality reads (phred score < 33 within 95% of the sequence) identified by Fastx v0.0.13 and below 100 bp were removed. Reads were assembled into contigs with SPAdes 3.5.0, k-mer range = 55, 99, 127 (Bankevich et al. 2012). Contig bins were created based on marker genes, nucleotide composition, and contig abundance (minimum contig length 1000 bp) with MaxBin-1.4.2 (Wu et al. 2014), and taxonomically identified to genus level using Kraken v0.10.5-beta (Wood and Salzberg 2014). A circular genome view was created with the software CGView (Stothard and Wishart 2005). Assemblies were annotated using RAST (Rapid Annotations using Subsystems Technology, Overbeek et al. 2014), submitted to NCBI GenBank (accession numbers in

results section and Appendix 5.1) and filtered for genes that were relevant for bacteriophages (weblinks for bioinformatics tools, Appendix 5.1). To compare the *R. reptotaenium* AO1 genome analysis with currently available BBD genomic data, an additional metagenome was retrieved from *Geitlerinema* sp. BBD_1991 (Den Uyl et al. 2016). Although *Geitlerinema* sp. may not play a key role in the pathogenicity of BBD in comparison to *R. reptotaenium,* the cyanobacterium is contributing to the BBD environment by oxidising sulfide, mixed-acid fermentation and detoxification of reactive oxygen species and resistance to antibiotics (Den Uyl et al, 2016). The bioinformatics comparison of prophages and CRISPR-Cas systems between the two BBD cyanobacteria provides insights into the presence and activity of bacteriophages in BBD.

5.3.3 CRISPR-Cas systems

Clustered regularly interspaced short palindromic repeats (CRISPR), associated Cas genes, and direct repeats were identified within the respective cyanobacterium genome bins of *R. reptotaenium* AO1 (*Chapter 3*, Buerger et al. 2016) and *Geitlerinema* sp. BBD_1991 (Den Uyl et al. 2016) with CRISPRfinder, CRISPRdb and CRISPRcompar (Grissa, Vergnaud, and Pourcel 2007). Only confirmed CRISPR arrays were considered. To assess protospacers and potential self-targeting spacers, CRISPR arrays were compared to: a) publicly available databases: viral RefSeq, plasmid RefSeq and genbank-phage accessed through CRISPRtarget online tool, as well as b) local databases of the assembled genomic bins, by retrieving best possible BLASTn matches with default parameters for short sequences (-gapopen 10, -gapextend 2, -dust no, -reward 1, -penalty -1, -word_size 7, -qcov_hsp_perc 100, Biswas et al. 2013). For both analyses, only protospacers with the CRISPRtarget default minimum matching score of 20 were considered as a possible categorical match (bacteriophage,

plasmid, unknown). A less stringent matching score (i.e., 18 and 19) was only considered if protospacer matches were relevant to the respective environment (e.g. cyanophage). Cas gene assignments, self-targeted genes, and open reading frames adjacent to CRISPR arrays were verified against the NCBI nr database to known protein sequences (tBLASTx) using Artemis (version 16, Rutherford et al. 2000).

5.3.4 Prophage analyses

The assembled cyanobacterium genome contigs of *R. reptotaenium* AO1 (*Chapter 3*, Buerger et al. 2016) and *Geitlerinema* sp. BBD_1991 (Den Uyl et al. 2016) were analysed for prophage gene signatures with PHAST (Zhou et al. 2011), PHASTER (Arndt et al. 2016) and VIRsorter (Roux et al. 2015), using default parameters for submission as metagenomic contigs. Although incomplete prophages are reported in this paper, only complete prophage signatures were considered as potentially functional prophages. Annotations of potential prophage regions were checked for unrecognised phage genes, toxicity genes and genes of virulence with Blast2GO against the swissprot database (Conesa et al. 2005) and BLASTp against the NCBI nr database.

5.4 Results

5.4.1 Genome assembly and gene annotation

The sequences retrieved from the *Roseofilum* culture were assembled using the software SPAdes 3.5.0 and resulted in three genomic bins that were submitted for automated annotation with RAST. The three bins were taxonomically identified as cyanobacterium *R. reptotaenium* AO1 (RAST 564709.3; NCBI GenBank project

accession number MLAW00000000), *Alphaproteobacterium* (RAST 28211.29), and *Cytophagaceae* sp. (RAST 89373.4). The *R. reptotaenium* AO1 genome bin matched the expected BBD cyanobacterial target and was chosen for further analyses. The *R. reptotaenium* AO1 genome bin consisted of 134 contigs, with a total length of 5,826,181 bp (Table 5.1).The full annotation contained 5,491 features, with 39 possible missing genes (data access, RAST and NCBI GenBank, Appendix 5.1).

Table 5.1 Assembly details of the draft genome of *Roseofilum reptotaenium* AO1. Bioinformatic statistics such as N50 values and sequence length were generated with the software gnx-tools (version 0.1+20120305).

Parameter	Assembly details
Sequencing library	Truseq library
Sequencing platform	MiSeq 2x300 V3
Assembly software	SPAdes 3.5.0
K-mer length	55 - 127
Total number of sequences	134
Total length [bp]	5,826,181
Shortest sequence [bp]	1032
Longest sequence [bp]	308,535
Total number of Ns	305
N50	94,947 (20 sequences)
GC content [%]	44.81
Coverage	300 - 600 x

5.4.2 CRISPR-Cas systems

A CRISPR-Cas immune system type I-D, also known as CASCADE (CRISPR associated complex for antiviral defence), was identified within the *R. reptotaenium* AO1 genome (Fig. 5.1). CRISPR type I systems are known to target DNA only, not RNA, and require a specific protospacer-adjacent motif (PAM) on the target sequence next to the protospacer to be functional (reviewed in Jiang & Marraffini 2015). The associated genes, adjacent to CRISPR array #2, were homologues of the known Cas genes, such as Cas3 helicase, Cas10d, Cas7, Cas5, Cas6, Cas2, and Cas1 (respective functions in Appendix 5.2, modified from Makarova et al. 2011a,b).

Compared to other cyanobacterium genomes, the *R. reptotaenium* AO1 CRISPR-Cas system is of average complexity, with 7 array loci and a total of 100 spacers (median of type I systems contain approx. 92 spacers and 3 CRISPR loci; Cai et al. 2013).

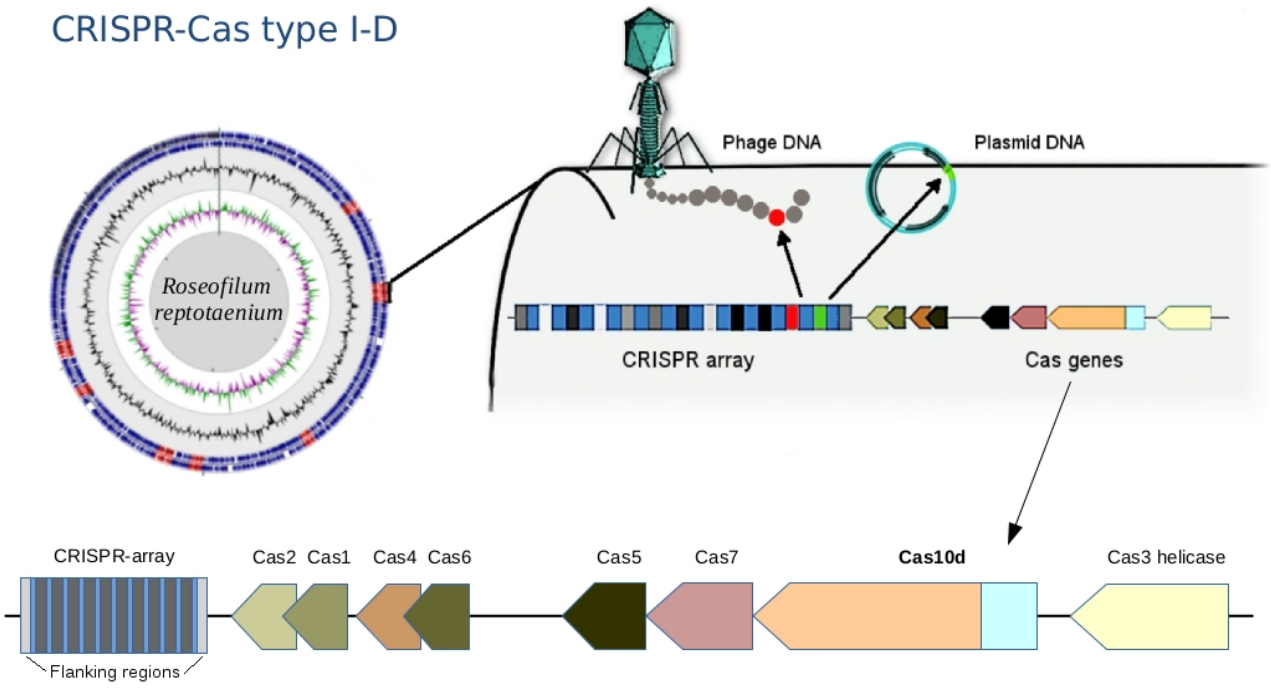


Figure 5.1 Draft genome assembly and illustration of the CRISPR-Cas system in *Roseofilum reptotaenium* AO1. The adaptive, heritable CRISPR-Cas system defends the cyanobacterium against bacteriophage infections, plasmids and mobile genetic elements. The gene Cas10d is representative of the CRISPR-type I-D, commonly found in other cyanobacteria. CRISPR arrays are marked in red on the genome contigs, displayed in a circular view. The Cas gene arrangement is indicated at the bottom of the figure.

Horizontal gene transfer has been inferred as one of the main methods for distributing parts of the CRISPR-Cas systems among bacteria, based on high sequence similarity of direct repeats (Godde and Bickerton 2006). In *R. reptotaenium* AO1, all direct repeats of the CRISPR arrays were 37 bp long (Table 5.2) and showed high similarity to direct repeats of other cyanobacterial species, such as the marine *Rivularia* sp. PCC 7116 strain (Appendix 5.3). The respective CRISPR-flanking regions

were found to be specific for *R. reptotaenium* AO1, with no BLAST homologies to other known sequences.

In total, seven CRISPR arrays containing a total of 100 unique spacers were identified in *R. reptotaenium* AO1. Identical matches (protospacers) to eight spacer sequences were found on one cyanobacterial genome contig (contig 93, coverage 558, length 31,342 bp). The protospacers on contig 93 match spacers in conserved middle parts of the CRISPR arrays (average spacer position: 8.75 out of 15.88, Appendix 5.4). One of the target regions on contig 93 codes for a DnaB domain-containing helicase (related to a bacteriophage multi-domain, phage_DnaB, BLASTp *e*-value <6.28E-20), while all others code for hypothetical proteins (Appendix 5.4). Analysis of the assembly's de Bruijn graphs showed that contig 93 is not circular shaped, but can be connected to other contigs, and is therefore not part of a plasmid or linear extra-chromosomal element (Appendix 5.5). Contig 93 was not integrated into other contigs, likely due to multiple connection possibilities, strain variations, or repetitive regions. Although a phage integrase was detected on one of the adjacent contigs (contig 184, pfam00589, *e*-value <8.47E-03, Appendix 5.6), no other genes related to bacteriophages or mobile genetic elements were uncovered.

Results from protospacer matches to public databases (40% of spacers had potential protospacer matches) showed that protospacers were mostly unrelated to cyanobacteria or the marine environment (Appendix 5.7), indicating that spacer sequences were novel and not represented in publicly available databases. Protospacer origins were non-redundant and matched to a diversity of plasmids and viruses, with multiple hits to *Synechococcus* phages, a *Cyanothece* sp. plasmid, and a *Sinorhizobium fredii* plasmid (number of hits were 3, 2 and 2, respectively, Appendix

5.7). Some of the cyanobacterial reference CRISPR-spacers and putative prophage regions were originated from bacteriophage taxa that were also detected in the T4-bacteriophage dataset (*Chapter 2*), such as *Synechococcus* phage S-ShM2, Syn19, S-SM2 and S-SKS1. In addition, a CRISPR-Cas spacer (NoG2_49) from *Geitlerinema* sp. BBD_1991 was genetically similar to an RNase encoding gene, which was located on one of the assembled bacteriophage contigs (score: 18, contig_1, *Chapter 4*, Appendix 5.8).

Table 5.2 CRISPR-Cas spacers of *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991. The total length of the CRISPR array is given from the start to the end of the respective direct repeats (DR). Only complete CRISPR arrays were considered. CDS numbers refer to overall nucleotides with contigs names in sequence.

CRISPR array # *R. reptotaenium* AO1	Position 1st CRISPR, contig number	length [bp]	DR length [bp]	Spacers #	Spacer average length [bp]
1	171508..171547, contig_15	1518	37	20	37.1
2 + Cas I-D	446355..446395, contig_18	1071	37	14	36.9
3	2179117..2179156, contig_37	933	37	12	37.8
4	2351308..2351339, contig_38	553	37	7	36.9
5	3024497..3024532, contig_46	844	37	11	36.5
6	3582097..3582130, contig_55	2022	37	27	36.6
7	4115583..4115616, contig_65	693	37	9	36.0
Total		average	average	sum	Total average
7		1091	37	100	36.8
CRISPR array # *Geitlerinema* sp. BBD_1991	Position 1st CRISPR, contig number	length [bp]	DR length [bp]	Spacers #	Spacer average length [bp]
1 + Cas I-D	384474..384507, BBD_1000996	19166	37	260	37.6
2	1172325..1172357, BBD_1000999	7340	37	100	36.0
3 + Cas III-U	1589539..1589575, BBD_1001002	7461	37	101	36.5
4 + Cas III-U	1597098..1597137, BBD_1001002	4589	37	62	36.4
5 + Cas I-MYXAN	2022334..2022369,	2114	36	29	35.7

	BBD_1001004				
6 + Cas III-B	2315188..2315226, BBD_1001007	5320	36	71	38.4
7 + Cas genes	2341519..2341555, BBD_1001007	1725	35	23	38.5
8 + Cas genes	2577530..2577580, BBD_1001010	246	25	3	48.7
9	3423177..3423219, BBD_1001016	243	30	3	41.0
10 + Cas genes	4126767..4126803, BBD_1001024	985	35	13	38.1
11 + Cas genes	4127858..4127894, BBD_1001024	1568	35	21	38.0
Total		average	average	sum	Total average
11		4614	34.5	686	38.5

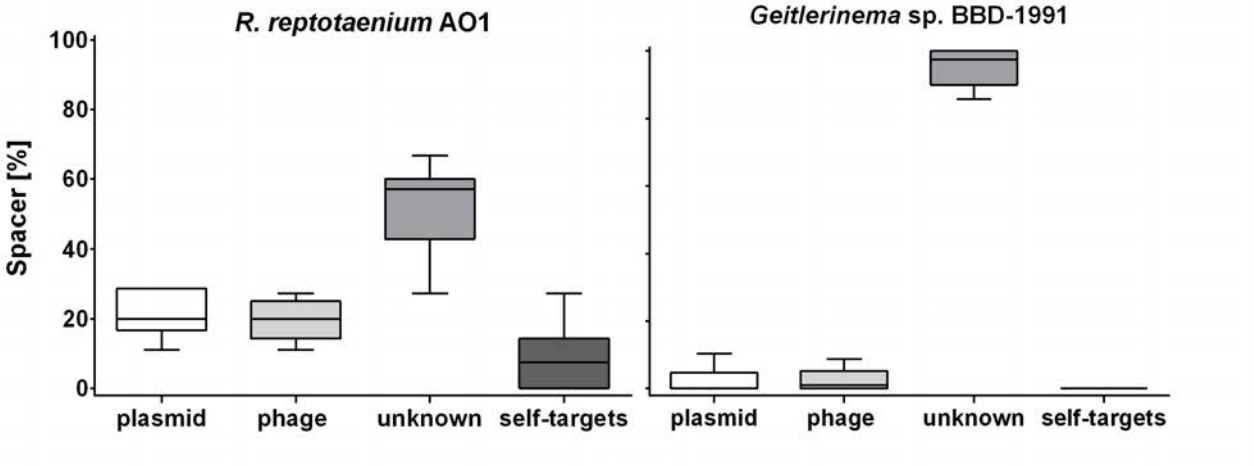


Figure 5.2 Potential CRISPR-Cas spacer targets within BBD. Most of the spacer targets (protospacers) could not be identified by matches to CRISPR sequences in publicly available databases. An unusually high number of protospacers were located within a short region of the *R. reptotaenium* AO1 genome. Spacer targets in % were calculated for the respective categories within each of the seven CRISPR arrays and visualised as replicates in a boxplot (detailed protospacer Appendix 5.7). Bar = mean, whiskers = min to max values, box = 25th and 75th percentiles.

Eleven CRISPR arrays were detected in the genome of *Geitlerinema* sp. BBD_1991 (Den Uyl et al. 2016), which had an extraordinarily high number of spacers (n = 260), approximately six times as many spacer sequences compared to *R.*

reptotaenium AO1 (Table 5.2). While most of the target sequences (protospacers) were unknowns (95% ± 6), no identical spacer sequences were identified between the two genomes and also no self-targeting spacers were found in *Geitlerinema* sp. BBD_1991 (Fig. 5.2, Appendix 5.7). Cas genes were adjacent to almost all CRISPR arrays of Geitlerinema sp. BBD_1991, whereas *R. reptotaenium* AO1 had only one CRISPR array with adjacent Cas genes (Table 5.2, Appendix 5.2). Several genes representative of different CRISPR-Cas types were detected in *Geitlerinema* sp. BBD_1991, such as types I-D, II-U, III-B and I-MYXAN. The direct repeat sequences of *Geitlerinema* sp. BBD_1991 were genetically similar to *R. reptotaenium* AO1, and closely related to other cyanobacteria species, such as *Crinalium epipsammum* and *Synechococcus* sp. (Appendix 5.3).

5.4.3 Prophage analyses

Three potential prophages were detected in *R. reptotaenium* AO1 (R1-R3, Table 5.3, Appendix 5.9, 5.10). These were classified as incomplete and questionable prophages, because some of the required genes to fully form an assembled virus, such as genes coding for a tail and capsid, were missing from the contigs. It is noteworthy that the tools PHAST, PHASTER, and VIRsorter did not identify prophage signatures in the same regions of the genome, which may suggest that the ability to capture prophage signatures is still limited. The potential prophage regions of the *R. reptotaenium* AO1 genome carried genes that could be involved in processes other than virion assembly, such as the genes DNA adenine methyltransferase, transketolase, GDP-D-mannose 4,6-dehydratase, D,D-heptose 7-phosphate kinase, phosphoheptose isomerase, and a ADP-L-glycero-D-mannoheptose-6-epimerase (Appendix 5.10). Some of these associated genes are involved in the non-oxidative pentose phosphate pathway (GDP-D-mannose 4,6-dehydratase), which provides energy during replication of virus when

photosynthesis is not present (Shestakov and Karbysheva 2015) as well as in receptor modification of lipopolysaccharides to prevent infection of the same bacteriophage (Kropinski et al. 2007). However, potential cyanobacterial virulence factors that were located on prophage regions included lysozyme/metalloendopeptidases (e.g. region R1, *e*-value 7.00E-017, Appendix 5.10), genes potentially involved in lipopolysaccharide production (phosphoheptose isomerase, ADP-L-glycero-D-mannoheptose-6-epimerase and an NAD-dependent epimerase in region R2, *e*-value 3.37E-029, Appendix 5.10). Genes related to bacteriophages functionality were detected by RAST outside potential prophage regions and scattered across the *R. reptotaenium* AO1 contigs, such as a phage endolysin gene, phage tail protein, a T4-like virus tail tube protein gp19, a phage associated DNA primase, a putative prophage protein, a phage shock protein and a phage integrase (Appendix 5.9).

Table 5.3 Details of prophage detection. Different tools were used for prophage detection in the draft genome of *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991. Gene details in Appendix 5.10-5.11. Classifications (PHAST, PHASTER): Intact, questionable or incomplete prophage = scoring from lowest 0 - highest 150, Table 1 in Zhou et al. 2011. VIRsorter category 3 = sequence similar to virus genome structure, but without sequence similarity to known viruses. CDS indicates the number of coding sequences. Overall GC-content *R. reptotaenium* AO1: 44.81%; *Geitlerinema* sp. BBD_1991: 50.38%. Annotation details in Appendix 5.10-5.11.

#	Contig node #	Software	Classification score		Length [bp]	CDS	GC [%]
***R. reptotaenium* AO1**							
R1	contig_72	PHAST	incomplete	20	7,100	8	45.1
		PHASTER	-	-	-	-	-
		VIRsorter	-	-	-	-	-
R2	contig_41	PHAST	-	-	-	-	-
		PHASTER	incomplete	10	103,700	15	42.2
		VIRsorter	-	-	-	-	-
R3	contig_93	PHAST	-	-	-	-	-
		PHASTER	-	-	-	-	-
		VIRsorter	category 3		31,342	9	44.4
***Geitlerinema* sp. BBD_1991**							
G1	BBD_1000996	PHAST	incomplete	20	19,800	7	53.29
		PHASTER	incomplete	20	19,800	7	53.21
		VIRsorter	-	-	-	-	-

G2	BBD_1001009	PHAST	incomplete	40	9,000	10	50.07
		PHASTER	-	-	-	-	-
		VIRsorter	-	-	-	-	-
G3	BBD_1001028	PHAST	questionable	90	16.600	16	51.12
		PHASTER	questionable	80	16.600	16	51.01
		VIRsorter	-	-	-	-	
(G4)	BBD_1001072-4	PHAST	questionable	80	20,900	29	49.14
		PHASTER	-	-	-	-	-
		VIRsorter	-	-	-	-	-
G5	BBD_1001065	PHAST	-	-	-	-	-
		PHASTER	-	-	-	-	-
		VIRsorter	category 3		19,845	11	56.32

By comparison, five potential prophages were detected in *Geitlerinema* sp. BBD_1991 (Table 5.3, Appendix 5.11) (Den Uyl et al. 2016). Prophage_G1 region, recognised by PHAST and PHASTER (not VIRsorter), was the only region flanked by integration sites AttL and AttR, but was missing phage-related assembly genes. Phage-related genes were detected by PHAST and PHASTER mainly in the regions G3, resulting in a high prediction score (scored 90, questionable prophage) including genes such as phage baseplate, tail tube and tail sheath. In addition, the prophage_G4 region stretched over three consecutive contigs (BBD_1001072 - BBD_1001074) that were not necessarily connected to each other in the genomic assembly and therefore likely represent a false positive result. Nevertheless, a gap-less assembly may rearrange the respective contig connections into a fully intact prophage. Genes directly located within the *Geitlerinema* sp. BBD_1991 potential prophage regions (Appendix 5.11), which are likely to be involved in processes other than phage assembly, include photosytem II components, such as PsbE, PsbF and PsbJ (prophage_G1), genes that contribute to Fe(II) transport systems, such as iron permease FTR1, to a non-oxidative pentose phosphate pathway, such as the GDP-D-mannose 4,6-dehydratase and other genes, such as a phosphoribosylglycinamide formyltransferase, orotate phosphoribosyltransferase, guanine deaminase and a biotin-(acetyl-CoA-carboxylase)

ligase BirA. Potential virulence genes located in *Geitlerinema* sp. BBD_1991 prophage regions include a lysozyme/metalloendopeptidase (region G3 and G4, see annotations Table S1 in Den Uyl et al. 2016), a peptidase S8/S53 (subtilase family protease, PHAST in region G4) and lipopolysaccharide modification genes, such as GDP-L-fucose synthase and glycosyl transferase (PHAST in region G4).

5.5 Discussion

My analyses reveal that the genome of the cyanobacterium *R. reptotaenium* AO1, the dominant member of the microbial consortium causing black band disease in corals, and that of *Geitlerinema* sp. BBD_1991 are equipped with CRISPR-Cas adaptive defence systems and several prophage regions that contain potential virulence-related genes. CRISPR-Cas systems thrive in environments with a high incidence of phage predation (Jiang et al. 2013), and can be multi-functional, ranging from preventing bacteriophage infections, interfering with the uptake of plasmids and mobile genetic elements (Barrangou et al. 2007), to the control of gene expression (Hatoum-Aslan and Marraffini 2014) and support of DNA repair mechanisms (Babu et al. 2011).

5.5.1 CRISPR-Cas self-targets

A section of the *R. reptotaenium* AO1 genome is targeted by an unusually high number of spacer sequences (8 out of 100 spacers, Appendix 5.7). Although spacer sequences of the CRISPR-Cas systems are usually thought to target foreign genetic material, approximately 1 in 250 spacer sequences can be self-targeting and match to particular regions of the host's genome (Stern et al. 2010). Typically, a large proportion of these self-targeting spacers are located within the most recent obtained first or second position of a CRISPR array and may be acquired accidentally, resulting in

autoimmunity through digestion of the host's genetic code or inactivation of the CRISPR array (Stern et al. 2010). However, the self-targeting CRISPR-Cas spacers of *R. reptotaenium* AO1 do not result in autoimmunity and are not inactivated for the following reasons. The presence of the self-targeting spacers is unlikely to be accidental, because they originate from relatively conserved middle parts of the CRISPR arrays (average spacer position: 8.75 out of 15.88, Appendix 5.4). They target loci that are not randomly distributed across the genome, but are located in a narrow region of 31,342 bp (contig 93). In addition, the CRISPR arrays are still active, because up to three self-targeting spacers are present on single CRISPR arrays, indicating that they were acquired as separate events. These results suggest a secondary regulatory role of this CRISPR-Cas system in which cleavage of the host genome is prevented, possibly by the lack of PAM recognition sequences on the target regions or through another unknown process.

5.5.2 Contig 93, a potential prophage

In *R. reptotaenium* AO1, contig 93 was detected to be an incomplete prophage by VIRsorter (Table 5.2, Appendix 5.10). Self-targeting spacers are known to target environmental lysogenic bacteriophages or prophages of other bacteria, but rarely their own prophage signatures (Briner et al. 2015; Hargreaves et al. 2014; Touchon and Rocha 2010). The potential prophage is classified as incomplete, because only a single gene (DnaB domain-containing helicase) was detected as having a phage origin (Phage_cluster_71 PFAM-AAA_25, coding for a DNA repair protein), while the contig lacks essential genes for virus replication and assembly, such as capsid, head, or tail genes. It is possible that essential parts of the potential bacteriophage could be spread over multiple assembled contigs. Indeed, according to the assembly de Bruijn graphs (Appendix 5.5), contig 93 was separated from other contigs due to multiple possibilities

for continued assembly and probably unresolved repetitive regions. However, the only bacteriophage-related gene on adjacent contigs is a phage integrase (contig 184, pfam00589, *e*-value <4E-46, Appendix 5.6), which is required for site-specific DNA excision and integration (Fogg et al. 2014).

The area of *R. reptotaenium* AO1 contig 93 could resemble a bacteriophage that is still hidden in the genome bin. This hypothesis was tested in an experiment to induce a potential hidden prophage with mitomycin C and UV treatments (*Chapter 3*). Although cyanobacterial biomass declined in response to the treatments, the lack of any measurable virus replication events indicated that the observed cell degradation was probably due to the respective treatments, rather than viral lysis. Alternatively, it has been shown that a CRISPR-Cas system of *Escherichia coli* can prevent prophage induction without killing the host bacterium when both processes, virus replication and CRISPR-Cas defence, are activated at similar time points (Edgar and Qimron 2010). Although unconfirmed, the authors suggest possible regulation by promoters of cas proteins during a stress response, such as the sigma factor σ^{32} of *E. coli*, which could silence CRISPR activity until required in order to prevent prophage induction (Edgar and Qimron 2010). Consequently, even if contig 93 was part of a prophage, as indicated by the presence of multiple spacers that target the incomplete prophage region, it might not have been possible to induce it.

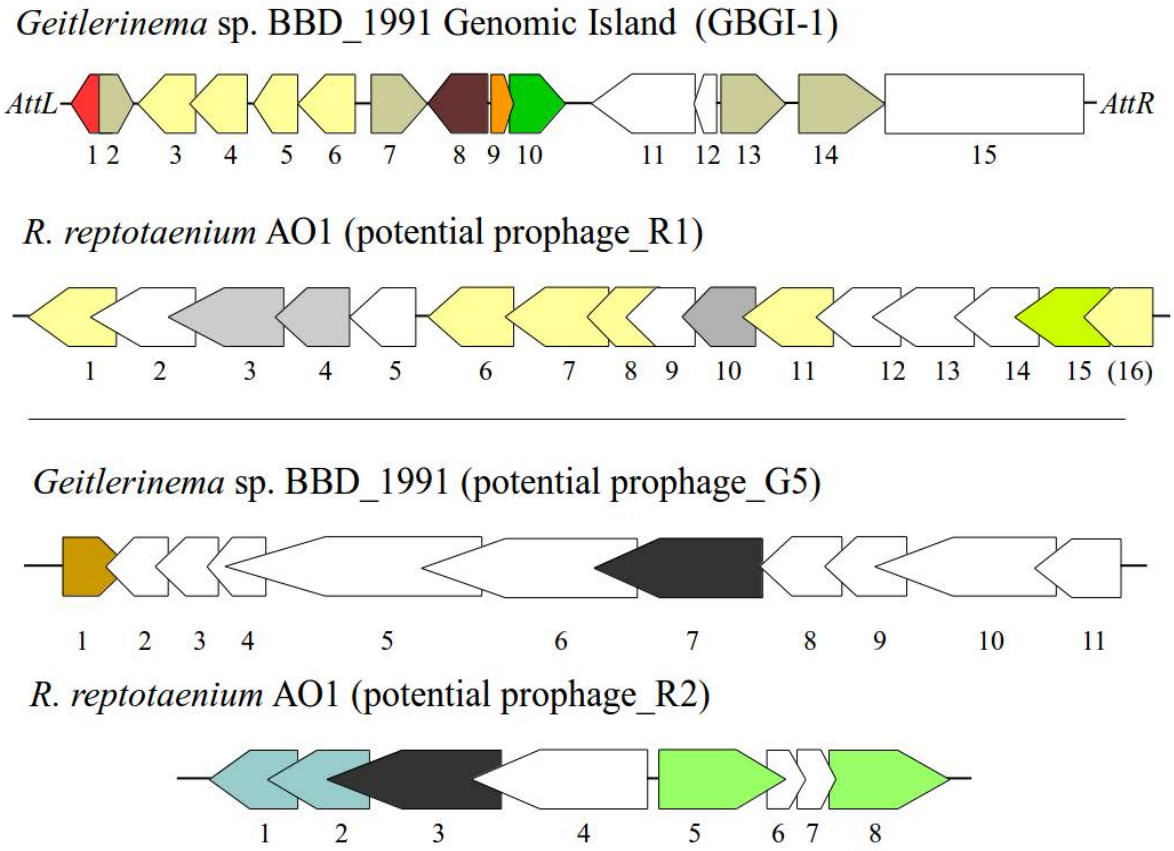
5.5.3 Functional role of potential prophage regions

Contig 93 contains predicted coding regions for 39 hypothetical proteins with unknown functions, seven of which match directly to CRISPR-Cas spacers (Appendix 5.4, 5.7). Other known genes on contig 93 code for a Rec-D like helicase, DnaB helicase, a putative proteinase, and a DNA-damage inducible protein. CRISPR-Cas systems can

be multi-functional and involved in processes other than defence against foreign genetic material, such as expression regulation of pathogenicity genes leading to increased virulence (Hatoum-Aslan and Marraffini 2014; Louwen et al. 2014), biofilm formation (Zegans et al. 2009), and DNA repair (Babu et al. 2011). The CRISPR-Cas1 protein (YgbT) of *E. coli* is known to interact with Rec repair proteins, which can increase resistance to DNA damage (Babu et al. 2011). In *R*. *reptotaenium* AO1, the proteins encoded by contig 93 may be interacting with the CRISPR-Cas system for increased DNA repair. Such increased DNA repair would be beneficial for *R. reptotaenium* during UV-induced DNA damage, for example in high light environments, a condition that would lead to more rapid BBD progression (Sato, Willis, and Bourne 2010). However, further research is required to tease apart alternative hypotheses about the function of the CRISPR-Cas system in *R. reptotaenium* AO1.

Several genes that represent potential virulence factors were located within other prophage regions of the *R*. *reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991 genomes. Although some of the gene annotation differed between RAST and the prophage detection tools (PHAST and PHASTER), their functional roles were consistently associated with virulence factors of other known bacteria, such as the involvement in lipopolysaccharide production and coding for lysozyme/metalloendopeptidases. The third and fourth most highly expressed cyanobacterial gene within BBD that encodes for an NAD-dependent epimerase/dehydratase (Arotsker et al. 2016) was homologous in terms of its functionality to a gene located in *R*. *reptotaenium* AO1 prophage region_R2 (PHAST annotation, Appendix 5.10, *e*-value <3.37E-029). NAD-dependent epimerase/dehydratases have been linked to increased virulence in the bacterium *Pectobacterium carotovorum* causing soft rot disease of vegetables (Islam 2016). *P.*

carotovorum with an intact NAD-dependent epimerase/dehydratase (wcaG gene) showed increased secretion of virulence associated exoenzymes and caused 21.5% - 26.7% macerated tissue, while mutants with a disrupted gene caused only 5.8% - 6.5% tissue damage on vegetables (Islam 2016). A reason for the increased virulence could be the homology of both genes, the NAD-dependent epimerase/dehydratase of *P. carotovorum* (Islam 2016) and *R. reptotaenium* AO1, to the wcaG gene coding for a GDP-fucose synthetase, which are virulence associated exoenzymes likely to be involved in the production of lipopolysaccharides and colanic acid (Fry et al. 2000). The same respective gene in *R. reptotaenium* AO1 prophage_R2 region has been annotated by RAST as a rhamnose containing glycans subsystem coding for an UDP-glucose 4-epimerase (EC 5.1.3.2). This particular annotation is also a known virulence factor, which is located within a gene cluster that produces endotoxic lipopolysaccharides in *Campylobacter* spp. (Fry et al. 2000). A deactivation of the UDP-glucose 4-epimerase coding gene resulted in the expression of incomplete lipopolysaccharides and a virulence reduction (Fry et al. 2000). In addition, both cyanobacterial genomes had other virulence-associated genes in potential prophage regions involved in lipopolysaccharide production (Fig. 5.3).



Genes potentially involved in:

Lipopolysaccharide production

GBGI-1_3	GDP-fucose synthetase	[*Geitlerinema* sp. PCC 7105], BLASTp 0.0, WP_017662405.1
GBGI-1_4	GDP-mannose 4,6-dehydratase	[*Leptolyngbya valderiana*], 0.0, WP_063716818
GBGI-1_5	Galactosyl-1-phosphate transferase	[*L. valderiana*], 8e-146, WP_063716819.1
GBGI-1_6	Glycosyl transferase family 1	[*L. valderiana*], 0.0, WP_063716824.1
R1_1	NAD-dependent epimerase/dehydratase, PP_02287, PHAST 3.37e-029, phage (gi815854730)	
R1_6	GDP-D-mannose 4,6-dehydratase	PP_02292, PHAST 5.64e-040, phage (gi815854729)
R1_7	D,D-heptose 7-phosphate kinase	PP_02293, PHAST 5.36e-044, phage (gi725949173)
R1_8	Phosphoheptose isomerase	PP_02294, PHAST 1.69e-015, phage (gi725949171)
R1_11	ADP-L-glycero-D-mannoheptose-6-epimerase, PP_02297, PHAST 5.82e-011, (gi815854739)	
(R1_16)	O-antigen export system, permease protein, RAST2:fig|564709.3.peg.2774	

Metalloendopeptidases

G4_20	Peptidase M23/M37 family [*Phormidium* sp.], BLASTp 2e-124, Den Uyl., 2016 BBD_100107313
G5_7	Metalloendopeptidase-like membrane protein [*O. acuminata*], BLASTp 2e-19 WP_015152078.1
R2_3	Lysozyme/metalloendopeptidase, PHAST 7.00e-017, PHAGE_Microc_Ma_LMM01_NC_008562

Figure 5.3 Genetic structure of prophage regions and their potential virulence factors. Genes with similar function are colour coded, e.g. yellow = genes potentially involved in lipopolysaccharide production, dark grey = metalloendopeptidases, white = hypothetical proteins, other colours = see annotations Appendix 5.10-5.11. Genomic island of *Geitlerinema* sp. BBD_1991 is shown on top with the attachment sites AttL and AttR. BLASTp and PHAST results are given with *e*-values and protein accession numbers.

For *R*. *reptotaenium* AO1 these were: GDP-D-mannose 4,6-dehydratase (Webb et al. 2004), phosphoheptose isomerase, ADP-L-glycero-D-mannoheptose-6-epimerase and an O-antigen export system (Brooke and Valvano 1996); and for *Geitlerinema* sp. BBD_1991: GDP-L-fucose synthase (Mäki and Renkonen 2004), GDP-D-mannose 4,6-dehydratase (Webb et al. 2004), glycosyl transferase (Davies et al. 2013). In *Geitlerinema* sp. BBD_1991 some of these genes are located within prophage_G1 region, that contained an integrase (G1_1), recombinase (G1_2) and was flanked by phage attachment sites AttL/AttR (Fig. 5.3, Appendix 5.11), a known arrangement of genomic and pathogenicity islands (Bellanger et al. 2014). Several studies have suggested that cyanobacterial lipopolysaccharides are less endotoxic than classic lipopolysaccharides (reviewed in Gemma et al. 2016).

However, the high expression of NAD-dependent epimerase/dehydratase (Arotsker et al. 2016) leads to the assumption that lipopolysaccharides could play an important role in the virulence of BBD associated cyanobacteria. Other virulence associated genes coding for lysozyme/metalloendopeptidases (superfamily peptidase M23, zinc metallopeptidases, BLASTp *e*-value <2.23E-36), were present in both genomes (prophage_R1 in Appendix 5.10; G4 BBD_100107313 and G5 BBD_10010657 in Appendix 5.11). Although unreported for cyanobacteria, metalloendopeptidases are known virulence factors in a wide range of bacterial pathogens causing tissue damage (Miyoshi and Shinoda 1997).

Potential virulence genes that had overlapping functionality were found in prophage regions of both genomes, *R*. *reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991. To date, viruses have not been considered as a contributor to the virulence of BBD associated bacteria. My results strongly suggest that genetic material related to

lysogenic bacteriophages contributes to the virulence of *R*. *reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991.

5.5.4 Conclusions

Here I show that the cyanobacterium *R. reptotaenium* AO1, the dominant microbe within the BBD consortium, and *Geitlerinema* sp. BBD_1991, a less abundant member of the consortium, acquire resistance against bacteriophages by maintaining adaptive, heritable CRISPR-Cas defence systems. It is not surprising to find CRISPR-Cas systems and potential prophage regions in BBD associated cyanobacteria. Microbial mats, such as BBD, can be hot-spots for bacteriophage-host interactions due to the high abundance of bacteria and associated viruses (Carreira, Piel, et al. 2015; Carreira, Staal, et al. 2015; Heidelberg et al. 2009). The most abundant microbes within microbial mats are likely to be targets for bacteriophage infections (Thingstad 2000), such as the cyanobacterium within the BBD consortium. Such constant phage predation pressure can result in an arms race and initiate evolution of host defence mechanisms, such as CRISPR-Cas systems (Held et al. 2013; Levin 2010; Stern and Sorek 2011). CRISPR-Cas systems are not an absolute barrier and can be overcome frequently by rapidly evolving bacteriophages (Andersson and Banfield 2008), also reflected by the multiple potential prophage regions within the *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991 genomes. Some of the genes located in potential prophage regions are coding for known virulence factors and indicate that bacteriophages and foreign genetic material are likely contributors to the virulence of *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991. Maintaining CRISPR-Cas systems is costly, but crucial for protection against a potentially high number of lytic and lysogenic bacteriophages present in the BBD consortium. While a lytic infection could decrease the abundance and biomass of cyanobacteria in the mat, a lysogenic

conversion could introduce new genetic material that might change the phenotypic characteristics of the infected cyanobacteria. Under both scenarios, an infection might alter the functional role of the cyanobacterium within the BBD mat, i.e., increase- or decrease its virulence. The detected CRISPR-Cas systems and potential prophage regions are evidence of a close interaction between bacteriophages and their host and highlight viruses as functional members of the BBD microbial consortium and as possible contributors to the virulence of the BBD-associated cyanobacteria *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991.

Chapter 6

General discussion: Potential roles of bacteriophages in black band disease and coral health

6.1 General discussion

Research presented in this thesis investigates the roles that viruses play in the virulent coral disease, black band disease (BBD). The results significantly enhance our understanding of how some viral processes may contribute to the virulence of BBD while others may mitigate its impact. Providing evidence of the contributions that viruses make to disease development can be difficult, particularly as methodologies for virus cultivation and community assessments are more complex than approaches used in bacterial analyses. Therefore, while summarising the research described in *Chapters 2-5*, I also discuss how the results address my original research objectives or if they required a modified approach. As a culmination to this body of research, I synthesise insights gained throughout this thesis to develop a model for BBD pathogenicity that includes pathways by which bacteriophages influence the disease development. In addition, I describe the multi-faceted roles of viruses in coral health more generally, and provide future research directions that will progress the field.

6.1.1 Discussion of applied methodology and limitations of the study

Assessing the diversity of environmental virus communities on both spatial and temporal scales is challenging. While the abundance and diversity of bacteria communities can be estimated with universal marker genes, such as the 16S rRNA gene (Erko and Ebers 2006), viruses lack such broad ecological marker genes (reviewed by Sullivan 2015). Thus, characterising the diversity of virus communities requires either whole metavirome approaches (Thurber et al. 2009; Weynberg et al. 2014) or marker gene analyses that concentrate on only a fraction of the community, such as bacteriophages that encode for the major capsid protein *gp23* (*Chapter 2*; Filée et al. 2005). In addition, assessment and detection of viruses are impeded by the sparse population of public database with (marine) viral sequences, the complexity of

virus taxonomic frameworks, and poorly established methodologies for bioinformatic analyses in comparison to bacteria. Therefore, characterisation of the T4-bacteriophage community based on the *gp23* marker gene (*Chapter 2*) provides insights into spatial and temporal patterns in the diversity of BBD-associated bacteriophages, but does not provide a full characterisation of all BBD-associated viruses. Given that the *gp23* marker gene analysis in this study excluded other virus and cyanophages taxa that may contribute to BBD, future studies should analyse whole metaviromes for a more holistic approach. Ideally, such studies would follow the development of BBD pathogenesis for a full year and include replicates within each season.

To further understand correlations between BBD development and BBD-associated bacteriophage communities, I also investigated the main cyanobacterial pathogen and its bacteriophages through culture-based studies (*Chapters 3-5*). Culture-based methodologies are still the "gold standard" for investigations of bacteriophage-host interactions in virology (Leland and Ginocchio 2007) and include the use of plaque assays, soft-agar overlay techniques to isolate bacteriophages (Kropinski et al. 2009), as well as flow cytometry to enumerate virus particles and host bacteria (Brussaard 2004). However, isolation and cultivation of viruses can be difficult if protocols for axenic cultivation of their host bacteria are not established, as was the case for the host cyanobacterium *R. reptotaenium* in this study. Because of special characteristics of the main filamentous cyanobacterium associated with BBD (i.e., long filamentous body shape, clumping behaviour, propensity to attach to surfaces, active motility and photo taxis), standard cultivation techniques were not appropriate. Methodologies for cultivating this cyanobacterium had to be adjusted and optimised, and unfortunately did not yield an axenic cyanobacterium culture even after extensive

optimisation. Nevertheless, the protocols developed represent a significant advance for studies of the BBD cyanobacterium because they 1) simplified isolation of the cyanobacterium, 2) maximised growth rates and health of filaments compared to media used previously, and 3) enabled, for the first time, the generation of BBD cyanobacterial growth curves. In addition, the cyanobacterial cultures were submitted to an open access algae culture facility (*R. reptotaenium* AO1, CS-1145, Australian National Algae Culture Collection ANACC, Hobart, Tasmania, Australia, www.csiro.au/ANACC), to enable other laboratories to conduct comparative studies, investigate physiological and molecular strain variations, as well as develop lytic and lysogenic bacteriophage infection experiments. Besides optimising cultivation protocols that enable more consistent classification of BBD cyanobacteria, my research provides an emended species description of *R. reptotaenium*, in which the BBD associated cyanobacterium *Pseudoscillatoria coralii* has been formally merged into the taxon *R. reptotaenium* (*Chapter 2*). An important objective for future studies should be the development of axenic cultures of *R. reptotaenium*. Extensive filtering and use of antibiotics may yield an axenic cyanobacterium culture, which should increase the possibility of isolating a cyanophage and would make downstream phage therapy more feasible. However, it is possible that the viability of *R. reptotaenium* is reliant on the presence of heterotrophic bacteria in the cyanobacterial sheath (Hube, Heyduck-Söller, and Fischer 2009; Praveen Kumar et al. 2009), hence axenic culture of this species may be not possible to maintain, demonstrated in Richardson et al. (2014).

BBD is a polymicrobial disease requiring the concurrent presence of several agents that collectively cause the disease (Brownell and Richardson 2014), thus phage therapy may not be as effective for BBD as for diseases caused by a single infectious agent. The isolation of a cyanophage that infects *R. reptotaenium* and the subsequent

development of phage therapy for BBD was more challenging than initially suspected, probably due to the presence of CRISPR-Cas defence systems detected in the two primary BBD cyanobacteria, *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991 (*Chapter 5*). CRISPR-Cas defence systems are designed to prevent bacteriophage infections, thus they are likely to have impeded isolation of bacteriophages. Ideally, for a successful phage therapy, several bacteriophages should be applied simultaneously in a cocktail to overcome potential host resistance and to decrease the abundance of multiple bacteria. Additional target species for BBD phage therapy could be *Desulfovibrio* and heterotropic bacteria that also contribute to BBD pathogenesis and potentially do not have CRISPR-Cas defence systems. With this in mind, the sample collection time for the isolation of a lytic bacteriophage might be critical for the bacteriophage isolation, due to the high variability observed in the BBD bacteriophage community composition, and its temporal abundance patterns (*Chapter 2*). In summary, BBD phage therapy may be possible, but it will require rigorous optimisation of all of these factors.

Only two studies have considered the influence of lysogenic conversions on the virulence of bacterial coral pathogens, *V. coralliilyticus* associated with white syndromes (Weynberg et al. 2015) and *R. reptotaenium* associated with BBD (*Chapter 5*). Genome sequencing of *R. reptotaenium* provided evidence of the presence of several prophages encoding for potential virulence factors, and demonstrated that bacteriophages have infected these BBD pathogens in the past (*Chapter 5*). Sequencing the genomes directly from bacterial cultures usually ensures minimal bacterial sequence contamination during the assembly and results in more complete and reliable contigs, compared to metagenomic datasets from environmental samples. Therefore, the assembled first draft genome of the main BBD cyanobacterium, *R.*

reptotaenium AO1 (*Chapter 5*), provides a scaffold for metagenomic analyses, a reference genome for gene expression and meta-transcriptomic studies, and significantly advances data analyses for “omics” approaches looking to better understand BBD.

6.1.2 A new model of BBD pathogenicity

Prior to my PhD research, the virus community associated with BBD had received little attention and was not considered to be part of the disease process in the latest models of BBD pathogenicity (Arotsker et al. 2016; Sato et al. 2016). Based on results from my dissertation, I have expanded recent BBD pathogenicity models by including bacteriophages and their potential roles as both contributors to pathogen virulence and as mitigators of disease impact (Fig. 6.1). In brief, according to the latest BBD pathogenicity models, cyanobacteria (primarily *R. reptotaenium*) form the main biomass component of the BBD mat and provide a framework for other bacteria, such as heterotrophic and sulfate-reducing bacteria. A gradient in sulfide and anoxic conditions, as a result of anaerobic sulfate-reducers and microbial respiration, respectively, peaks at the bottom of the mat causing necrosis of the underlying coral tissue. Disease progression is promoted by organic input into the mat from necrotic coral tissue and growth of heterotrophic bacteria, and is positively correlated to higher seawater temperatures and light intensities (e.g. summer compared to winter conditions). Coral tissue degradation fuels the growth of heterotrophic bacteria and the progression of the disease, while desulfuration of organic matter may further contribute to the accumulation of sulfide (Arotsker et al. 2016; Sato et al. 2016).

A few cyanophage and *Synechococcus* phage OTUs are uniquely associated with the BBD mat (*Chapter 2*), likely because they represent locally adapted and highly

specific bacteriophage strains, which infect only certain BBD cyanobacteria. Temporal patterns in their abundances are influenced by the abundance of their hosts and by seasonal patterns in environmental factors (Short and Suttle 2003), particularly UV-light, which degrades the DNA in bacteriophage capsids (Jacquet and Bratbak 2003) (Fig. 6.1, point 1). The high sulfide conditions of BBD itself may also partially interfere with the presence of viruses and simply degrade susceptible bacteriophages. However, the abundant CRISPR-Cas defence mechanisms of *R*. *reptotaenium* and *Geitlerinema* sp. show that bacteriophages are highly active within BBD and that the sulfide gradient may act as a barrier in deeper layers of the mat but not as an complete exclusion mechanism.

Despite the unique association of a few OTUs with the BBD mat, the bacteriophage community was relatively large (*Chapter 2*). Nevertheless, the presence of bacteriophages shapes and accelerates the evolution of pathogenic bacteria in biofilms by enforcing a strong selection pressure (Davies et al. 2016) and result in an arms race between the BBD bacteriophage community and the BBD bacteria (Stern and Sorek 2011) (Fig. 6.1, point 2). BBD-associated cyanobacteria like *R*. *reptotaenium* and *Geitlerinema* sp. BBD_1991 are exposed to high viral predation pressure, as indicated by the presence of highly complex CRISPR-Cas systems in their genomes (Fig. 6.1, point 3). The extraordinary length of CRISPR-arrays in *Geitlerinema* sp. BBD_1991 (*Chapter 5*), which encompasses up to 260 spacer sequences in one CRISPR-array – to the best of my knowledge this is the longest sequence so far reported - suggests that BBD is a hot-spot for virus infections. The presence of the CRISPR-Cas systems ultimately reduces successful bacteriophage infections (e.g. Barrangou et al. 2007; Rath et al. 2015), and as a consequence, maintaining a CRISPR-Cas system (likely the cyanobacteria's main protective defence mechanism

against bacteriophage infection) enables the cyanobacteria to continue their core functions within the mat. Further, the possibility that the CRISPR-Cas system is a pre-requisite for cyanobacterial infections leading to BBD, should be explored in future research, e.g. by knocking out of essential CRISPR-Cas genes, and subsequently monitoring successful bacteriophage infections and BBD-progression.

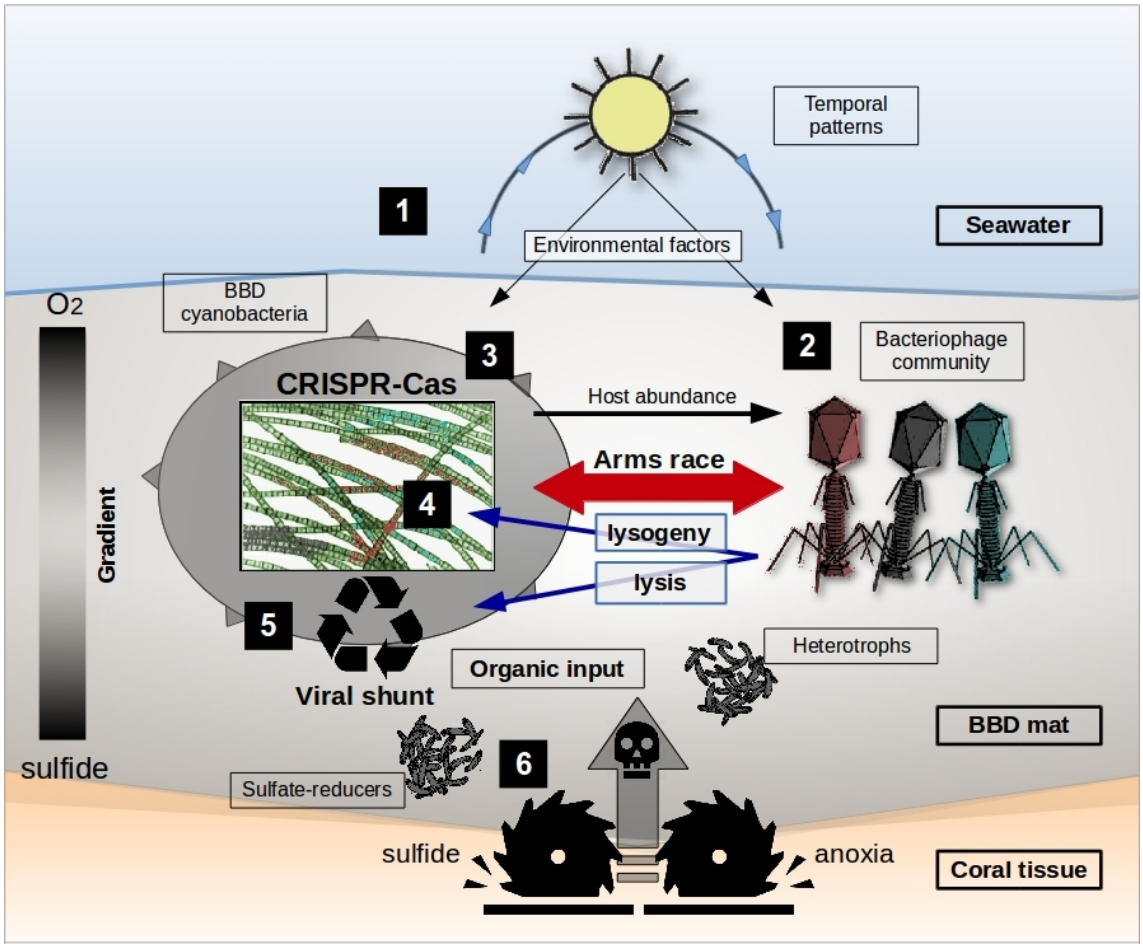


Figure 6.1 Model illustrating how bacteriophages influence BBD pathogenicity. (1) Seasonal patterns influence BBD prevalence and progression, i.e., higher temperature and light availability promote growth of heterotrophs and progression of the mat. Environmental factors (e.g. UV-light and salinity) as well as bacterial host abundance influence the BBD-associated bacteriophage community with respect to diversity and abundance. (2) The BBD-associated community of bacteriophages is complex, with few OTUs specific to the BBD mat, and high variability among BBD-affected coral colonies. Nevertheless, the rapid evolution of the bacteriophage community and selection pressure on the BBD bacteria community, results in a bacteriophage-bacterium arms race. (3) Host cyanobacteria have complex CRISPR-Cas systems that defend against bacteriophage infections, indicating that BBD is a hot-spot for virus infections. Since cyanobacterial CRISPR-Cas systems reduce the prevalence of viral infections, they allow the cyanobacteria to continue their role in BBD virulence. Despite the presence of

CRISPR-Cas defense systems, bacteriophage infections may still occur. (4) Lysogeny may contribute to the virulence of cyanobacteria, by introducing new genetic material into the cyanobacterial genome. (5) Lytic infections may reduce the abundance of cyanobacteria in the BBD mat and contribute to disease mitigation, although recycling organic matter in a viral shunt through lysis of cyanobacteria would also contribute to build-up of the biomass of remaining cyanobacteria. (6) Anoxic and sulfide gradients in BBD disintegrate coral tissue, with resulting organic input fueling further development of the mat.

Nevertheless, bacteriophages frequently overcome host defenses (CRISPR-Cas systems) by rapid evolution of their own genetic code (Andersson and Banfield 2008; Levin et al. 2013), and expression of genes that inhibit CRISPR-Cas activities (Bondy-Denomy et al. 2013). Such a successful infection can introduce new genetic material into the bacterial host genome, such as toxin genes, lysogenic bacteriophages may increase the virulence of infected cyanobacteria (*Chapter 5*). Lysogeny can be more prevalent when conditions are not optimal for the host bacterium and bacterial productivity is low (Brum et al. 2015; Payet and Suttle 2013), which, for BBD cyanobacteria, would equate to winter months when seawater temperatures and light levels are lower (Sato et al. 2009). This means that lysogeny might be more prevalent in BBD bacteria during winter months. Conversely, bacteriophages may switch to lytic replication during summer, when conditions favor growth of their host bacteria.

Although lysis generally reduces the abundance of infected cyanobacteria, lysis of a limited number of cyanobacteria within the mat may not necessarily interfere with the disease. Some bacteriophages are known to use a form of quorum sensing during bacterial infection to limit the number of lytic infections and to prevent host populations from collapse (Erez et al. 2017). Organic matter from the lysed cyanobacteria would be recycled via the viral shunt within the mat and become available for other BBD bacteria (Fig. 6.1 point 4). Within plankton communities, viral lysis recycles an estimated 6 - 26% of primary produced carbon, including lysis of 2 -

10% of phytoplankton and 20 - 30% of bacterioplankton, whereas heterotrophs and other bacteria typically recycle only 9 - 10% of primary production (Wilhelm and Suttle 1999). Therefore, viral lysis in BBD could represent a short-cut for providing organic input derived from cyanobacterial photosynthetic products and coral tissue lysis, and additionally fuel progression of the disease. Such nutrient recycling within cyanobacterial mats can also be an important factor in the establishment and onset of the BBD microbial community, especially in nutrient poor environments (Varin et al. 2010). This hypothesis could be tested to a certain extent with infection experiments using culture conditions in which the presence of bacteriophages are minimal and can be manipulated.

Conversely, successful lysis of most of the cyanobacteria would impede their functionality and reduce the impact of the disease. This process would disrupt the framework provided by the cyanobacteria, which is required for disease development (Sato et al. 2016). It would also reduce the capacity of the cyanobacteria to supply energy and nutrients to the microbial mat through fermentation and photosynthesis in the sulfide-rich BBD environment (Sato et al. 2017; Den Uyl et al. 2016).

My research provides new insights into the potential mechanisms that lead to the onset of BBD by considering the influence of bacteriophages on BBD pathogenicity. It is conceivable that bacteriophages can trigger the onset of BBD due to recycling of nutrients and potential gene transfer. If bacteriophages are a hidden driver and necessary component of successfully progressing BBD, the addition of bacteriophages as part of a phage therapy would provide additional fuel for disease development instead of ultimately stopping its progression. My research provides the first steps towards a better understanding of the role of bacteriophages in BBD.

6.1.3 Virus research in coral disease and future directions

Coral bacterial communities play an important role in coral health and disease as probiotic agents and bacterial pathogens (Bourne et al. 2009; Krediet et al. 2013). The probiotic theory suggests that the coral's immune function may be aided by symbiotic microorganisms, especially bacteria that serve as probiotics (benefit coral health), e.g., by enhancing nutrient availability (Reshef et al. 2006) and by the production of antimicrobial compounds that are effective against external bacteria only (demonstrated for soft corals (Kelman et al. 2006) and for sponges (Thakur, Anil, and Müller 2004)). However, little is known about the contribution of eukaryotic viruses and bacteriophages to the coral holobiont.

The mechanisms by which bacteriophages interfere with or contribute to disease pathogenesis are primarily indirect (the virus on its own does not influence the coral animal or *Symbiodinium*, but interacts with the bacterial community, which then secondarily influences coral health) (Fig. 6.2). Lysis of a probiotic bacterium in the holobiont might open up a niche for a pathogenic bacterium, which then may cause disease in the coral (Fig. 6.2A). Lytic bacteriophages can reduce the abundance of pathogenic bacteria by lysing their cells in a lytic cycle (Fig. 6.2B), which probably also occurs naturally in the environment and within the coral holobiont. Bacteriophages that target specific pathogenic bacteria may form part of the natural microbiome of corals and may confer some disease resistance (Atad et al. 2012; Marhaver, Edwards, and Rohwer 2008; Reyes et al. 2012; Stern et al. 2012) by preventing bacteria from excessive proliferation (Bourne et al. 2009), similar to the probiotic theory described for symbiotic bacteria (Fig. 6.2C). Bacteriophages may also influence coral diseases indirectly by increasing the virulence of an external bacterial pathogen (Fig. 6.2D). After successfully establishing lysogeny and transferring new genetic material into the host

bacteria, the bacterial pathogen may cause disease in the coral holobiont.

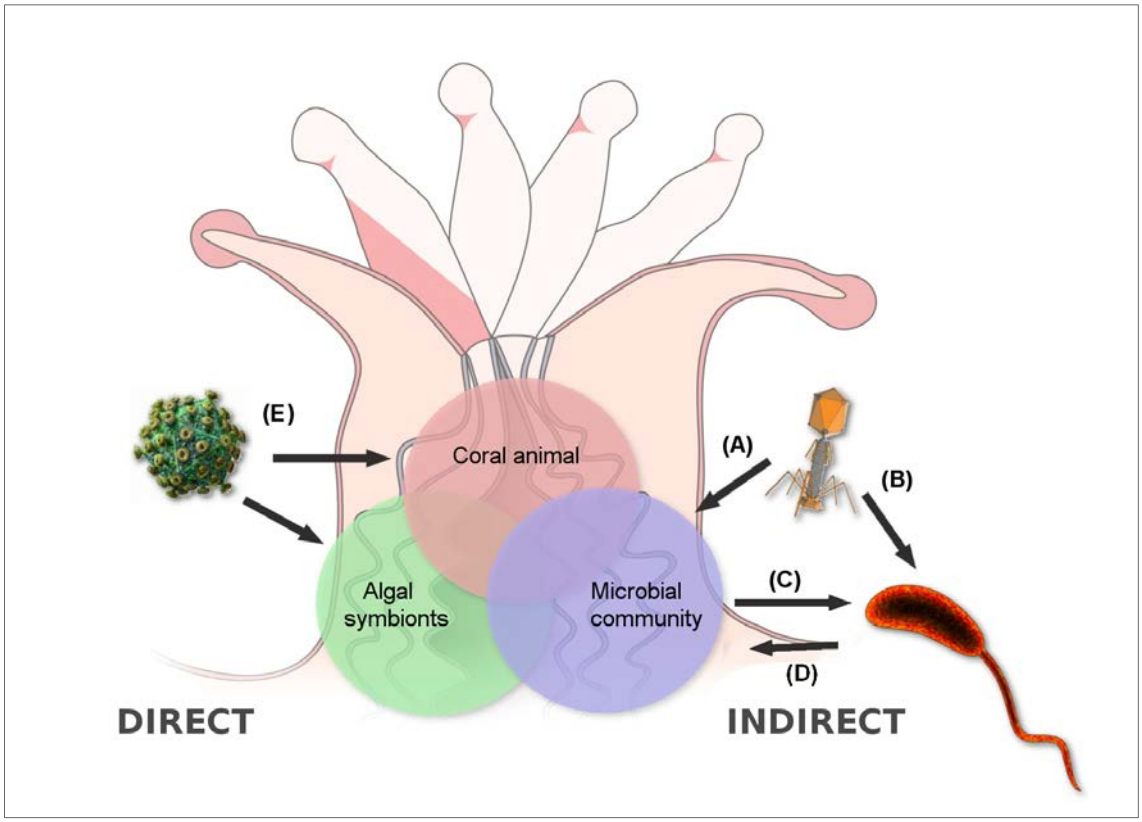


Figure 6.2 Viruses in coral health. Viruses can influence coral health through both indirect (A-D) and direct (E) processes that either contribute to or interfere with disease pathogenesis. Indirect processes include viruses that interact with the microbial community, which then have a secondary influence on the coral animal or algal endosymbionts (*Symbiodinium* spp.) (i.e., bacteriophages). Direct processes include viruses that interact with the coral animal and *Symbiodinium* (i.e. eukaryotic viruses). A) A bacteriophage may lyse a probiotic bacterium, which could open up a niche for a coral pathogen. B) Bacteriophages may infect pathogenic bacteria, reducing the impact of a disease. In this case the source of the virus is external from the coral holobiont, e.g., applied in phage therapy. C) The bacteriophages may be internal as a part of the coral holobiont and its microbiome. Lysis of a potential pathogen may protect the coral animal indirectly from a bacterial disease. D) A bacteriophage might also increase the virulence of an infected bacterium though gene transfer and cause disease in the coral. E) Eukaryotic viruses might target either the coral animal or *Symbiodinium* and can cause a disease, e.g., as suggested in the case of virus-induced coral bleaching and yellow blotch disease (here, the virus on its own would cause the disease, therefore a direct interaction). Source of coral polyp image, available for reuse with modification: commons.wikimedia.org/wiki/File3ACoral_ polyp_fi.svg.

Occasionally bacterial pathogens have been known to lose their ability to induce disease, and it is feasible (but untested) that bacteriophage infections are

responsible. For instance, the bacterium *Vibrio shiloi* was known as the causative agent for coral bleaching of *Oculina patagonica* in the Mediterranean (Kushmaro et al. 1996, 2001). Since 2003, however, an infection with *V. shiloi* no longer causes coral bleaching (Rosenberg et al. 2007). *V. shiloi* can still be isolated from bleached corals, but seems to be present as a result of opportunistic colonisation rather than being a primary cause of bleaching (Ainsworth et al. 2008). Corals possess an innate immune system, but not an adaptive immune system, and therefore are not able to develop resistance through antibody production (Bosch 2013; Cooper 2010; Palmer and Traylor-Knowles 2012). Thus, interactions among microbial community members might be crucial for coral health. Viruses represent the most abundant member of the microbial community in seawater and also likely in the coral holobiont. They are potential contributors to and mitigate of coral disease and therefore likely to play a leading role as drivers of coral health.

6.2 Conclusion

Although over 20 coral diseases have been described, the etiologies of most of these diseases remain unknown (Harvell et al. 2007; Sheridan et al. 2013). While viruses are known contributors to many marine diseases (Munn 2006), their roles in coral diseases pathogenesis are relatively unexplored. Conversely, a number of viral processes might also be beneficial to coral health (van Oppen, Leong, and Gates 2009). The current scarcity of virus-related studies of corals is due to innumerable difficulties associated with environmental virus research. These include poorly populated public sequence databases and inconsistencies in methodologies and genomic sequence data analyses (Wood-Charlson et al. 2015). Future research efforts into understanding the roles of viruses in coral diseases should consider a multi-faceted approach, such as ones presented in this thesis. Approaches that include best practices and optimised

protocols are required, such as those established for 1) viral metagenomics to characterise and describe virus communities (Weynberg et al. 2014), 2) flow cytometry for virus enumeration (e.g. Brussaard 2004; Pollock et al. 2010), 3) laboratory techniques to isolate bacteriophages by liquid assays and plaque assays (*Chapter 3*; Middelboe, Chan & Bertelsen 2010), 4) bioinformatic pipelines that are designed for virus sequence data (Laffy et al. 2016; Lorenzi et al. 2011; Roux et al. 2011, 2014), and 5) approaches for detection of viruses and pathogenic bacteria that occur at low abundances in the environment (Dang and Sullivan 2014). This PhD project provides novel data regarding the influence of bacteriophages on one particular coral disease (BBD) and highlights the influence of bacteriophages on coral health. Determining the role of viruses in coral health and disease will begin to close vast knowledge gaps, reveal more information about the ecological importance of marine viruses in the coral holobiont, and provide new ways to manage coral diseases on the reef.

7. Bibliography

Abedon, S. T., Kuhl, S. J., Blasdel, B. G. and Kutter, E. M. (2011) 'Phage treatment of human infections'. Bacteriophage, 1(2):66–85.

Ackermann, H. W. and DuBow, M. S. (1987) 'Viruses of prokaryotes: General properties of bacteriophages'. CRC Press, University of Michigan.

Ackermann, H. W. and Krisch, H. M. (1997) 'A catalogue of T4-type bacteriophages'. Archives of Virology, 142:2329–2345.

Adams, M. H. (1959) 'Bacteriophages'. Interscience Publishers Inc., New York, http://archive.org/details/bacteriophages00adam.

Aeby, G. S., Williams, G. J., Franklin, E. C., Haapkyla, J., Harvell, C. D., Neale, S., Page, C. A., Raymundo, L., Vargas-Ángel, B., Willis, B. L., Work, T. M. and Davy, S. K. (2011) 'Growth anomalies on the coral genera *Acropora* and *Porites* are strongly associated with host density and human population size across the Indo-Pacific'. PLoS ONE, 6(2):e16887.

Aeby, G. S., Work, T. M., Runyon, C. M., Shore-Maggio, A., Ushijima, B., Videau, P., Beurmann, S. and Callahan, S. M. (2015) 'First Record of black band disease in the Hawaiian Archipelago: response, outbreak status, virulence, and a method of treatment'. PLoS ONE, 10(3):e0120853.

Ainsworth, T. D., Fine, M., Roff, R. and Hoegh-Guldberg, O. (2008) 'Bacteria are not the primary cause of bleaching in the mediterranean coral *Oculina patagonica*'. The ISME Journal, 2(1):67–73.

Alisky, J., Iczkowski, K., Rapoport, A. and Troitsky, N. (1998) 'Bacteriophages show promise as antimicrobial agents'. The Journal of Infection, 36:5–15.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool'. Journal of Marine Biology, 215:403–10.

Amarillas, L., Chaidez, C., González-Robles, A., Lugo-Melchor, Y. and León-Félix, J. (2016) 'Characterization of novel bacteriophage phiC119 capable of lysing multidrug-resistant Shiga toxin-producing *Escherichia coli* O157:H7'. PeerJ, 4:e2423.

Anderson, M. J. (2006) 'Distance-based tests for homogeneity of multivariate dispersions'. Biometrics, 62(1):245–253.

Anderson, M. J., Ellingsen, K. E. and McArdle, B. H. (2006) 'Multivariate dispersion as a measure of beta diversity'. Ecology Letters, 9(6):683–693.

Andersson, A. F. and Banfield, J. F. (2008) 'Virus population dynamics and acquired virus resistance in natural microbial communities'. Science, 320(5879):1047–50.

Antonius, A. (1973) 'New observations on coral destruction in reefs'. 10th Tenth Meeting of the Association of Island Marine Laboratories of the Caribbean, 10:3.

Apprill, A., Marlow, H. Q., Martindale, M. Q. and Rappé, M. S. (2012) 'Specificity of associations

between bacteria and the coral *Pocillopora meandrina* during early development'. Applied and Environmental Microbiology, 78(20):7467–75.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y. and Wishart, D. S. (2016) 'PHASTER: a better, faster version of the PHAST phage search tool'. Nucleic Acids Research, 44:1–6.

Arotsker, L., Kramarsky-Winter, E., Ben-Dov, E. and Kushmaro, A. (2016) 'Microbial transcriptome profiling of black band disease in a *Faviid* coral during a seasonal disease peak'. Diseases of Aquatic Organisms, 118(1):77–89.

Arotsker, L., Kramarsky-Winter, E., Ben-Dov, E., Siboni, N. and Kushmaro, A. (2015) 'Changes in the bacterial community associated with black band disease in a Red Sea coral, *Favia* sp., in relation to disease phases'. Diseases of Aquatic Organisms, 116:47–58.

Atad, I., Zvuloni, A., Loya, Y. and Rosenberg, E. (2012) 'Phage therapy of the white plague-like disease of *Favia Favus* in the Red Sea'. Coral Reefs, 31(3):665–70.

Babu, M., Beloglazova, N., Flick, R., Graham, C., Skarina, T., Nocek, B., Gagarinova, A., Pogoutse, O., Brown, G., Binkowski, A., Phanse, S., Joachimiak, A., Koonin, E. V., Savchenko, A., Emili, A., Greenblatt, J., Edwards, A. M. and Yakunin, A. F. (2011) 'A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair'. Molecular Microbiology, 79(2):484–502.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. a. and Pevzner, P. (2012) 'SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing'. Journal of Computational Biology, 19(5):455–477.

Barash, Y., Sulam, R., Loya, Y. and Rosenberg, E. (2005) 'Bacterial strain BA-3 and a filterable factor cause a white plague-like disease in corals from the Eilat coral reef'. Aquatic Microbial Ecology, 40(2):183–189.

Bardina, C., Colom, J., Spricigo, D. A., Otero, J., Sánchez-Osuna, M., Cortés, P. and Llagostera, M. (2016) 'Genomics of three new bacteriophages useful in the biocontrol of *Salmonella*'. Frontiers in Microbiology, 7(4):1–25.

Barneah, O., Ben-Dov, E., Kramarsky-Winter, E. and Kushmaro, A. (2007) 'Characterization of black band disease in Red Sea stony corals'. Environmental Microbiology, 9(8):1995–2006.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. and Horvath, P. (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes'. Science, 315(5819):1709–12.

Bayer, T., Neave, M. J., Alsheikh-hussain, A., Aranda, M., Yum, L. K., Mincer, T. and Hughen, K. (2013) 'The microbiome of the Red Sea coral *Stylophora pistillata* is dominated by tissue-associated *Endozoicomonas* bacteria'. Applied and Environmental Microbiology, 79(15):4759–4762.

Bellanger, X., Payot, S., Leblond-Bourget, N. and Guédon, G. (2014) 'Conjugative and mobilizable genomic islands in bacteria: Evolution and diversity'. FEMS Microbiology

Reviews, 38(4):720–760.

Bergh, Ø., BØrsheim, K., Bratbak, G. and Heldal, M. (1989) 'High abundance of viruses found in aquatic environments'. Nature, 340:467–468.

Biswas, A., Gagnon, J. N., Brouns, S. J. J., Fineran, P. C. and Brown, C. M. (2013) 'CRISPRTarget: bioinformatic prediction and analysis of crRNA targets'. RNA Biology, 10(5):817–27.

Bondy-Denomy, J., Pawluk, A., Maxwell, K.L. and Davidson, A.R. (2012) 'Bacteriophage Genes That Inactivate the CRISPR/Cas Bacterial Immune System'. Nature 493(7432):429–32.

Bosch, T. C. G. (2013) 'Cnidarian-microbe interactions and the origin of innate immunity in metazoans'. Annual Review of Microbiology, 67:499–518.

Bourne, D. G., Garren, M., Work, T. M., Rosenberg, E., Smith, G. W. and Harvell, C. D. (2009) 'Microbial disease and the coral holobiont'. Trends in Microbiology, 17(12):554–62.

Boyett, H. V., Bourne, D. G. and Willis, B. L. (2007) 'Elevated temperature and light enhance progression and spread of black band disease on staghorn corals of the Great Barrier Reef'. Marine Biology, 151(5):1711–1720.

Briner, A. E., Lugli, G. A., Milani, C., Duranti, S., Turroni, F., Gueimonde, M., Margolles, A., Van Sinderen, D., Ventura, M. and Barrangou, R. (2015) 'Occurrence and diversity of CRISPR-Cas systems in the genus *Bifidobacterium*'. PLoS ONE, 10(7):e0133661.

Brooke, J. S. and Valvano, M. A. (1996) 'Molecular cloning of the *Haemophilus influenzae* gmhA (lpcA) gene encoding a phosphoheptose isomerase required for lipooligosaccharide biosynthesis'. Journal of Bacteriology, 178(11):3339–3341.

Brownell, A. C. and Richardson, L. L. (2014) 'Sulfate reducing bacteria as secondary and necessary pathogens in black band disease of corals'. Revista de Biología Tropical, 62(9):249–257.

Bruckner, A. (2015) 'History of coral disease research. In: Woodley, C. M., Downs, C. A., Bruckner, A., Porter, J. W., Galloway, S. B. (eds) Diseases of coral. Hoboken: John Wiley & Sons. 52–84.

Brum, J. R., Hurwitz B. L., Schofield O., Ducklow H. W, and Sullivan M. B. (2015) 'Seasonal time bombs: dominant temperate viruses affect southern ocean microbial dynamics'. The ISME Journal, 10(2):1–13.

Bruno, J. F., Selig, E. R., Casey, K. S., Page, C. A., Willis, B. L., Harvell, C. D., Sweatman, H. and Melendy, A. M. (2007) 'Thermal stress and coral cover as drivers of coral disease outbreaks'. PLoS Biology, 5(6):1220–1227.

Brussaard, C. P. D. (2004) 'Optimization of procedures for counting viruses by flow cytometry'. Applied and Environmental Microbiology, 70(3):1506–13.

Brussaard, C. P. D., Payet, J. P., Winter, C. and Weinbauer, M. G. (2010) 'Quantification of aquatic viruses by flow cytometry'. Manual of Aquatic Viral Ecology, 102.

Brüssow, H. (2012) 'What is needed for phage therapy to become a reality in Western medicine?'. Virology, 434(2):138–142.

Brüssow, H., Canchaya, C. and Hardt, W. (2004) 'Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion'. Microbiology and Molecular Biology Reviews, 68(3):560–602.

Büchen-Osmond, C. (2003) 'The universal virus database ICTVdB'. Computing in Science and Engineering, 5(3):16–25.

Buerger, P., Alvarez-Roa, C., Weynberg, K. D., Baekelandt, S. and van Oppen, M. J. H. (2016) 'Genetic, morphological and growth characterisation of a new *Roseofilum* strain (Oscillatoriales, Cyanobacteria) associated with coral black band disease'. PeerJ, 4, e2110.

Buerger, P., Wood-Charlson, E. M., Weynberg, K. D., Willis, B. L. and van Oppen, M. J. H. (2016) 'CRISPR-Cas defense system and potential prophages in cyanobacteria associated with the coral black band disease'. Frontiers in Microbiology, 7,2077.

Burstein, D., Sun, C. L., Brown, C. T., Sharon, I., Anantharaman, K., Probst, A. J., Thomas, B. C. and Banfield, J. F. (2016) 'Major bacterial lineages are essentially devoid of CRISPR-Cas viral defence systems'. Nature Communications, 7(10613):10613.

Cai, F., Axen, S. D. and Kerfeld, C. A. (2013) 'Evidence for the widespread distribution of CRISPR-Cas system in the phylum cyanobacteria'. RNA biology, 10(5):687–93.

Caporaso, J. G., Bittinger, K., Ushman, F. D., Desantis, T. Z., Andersen, G. L. and Knight, R. (2010) 'PyNAST: A flexible tool for aligning sequences to a template alignment'. Bioinformatics, 26(2):266–267.

Carreira, C., Larsen, M., Glud, R. N., Brussaard, C. P. D. and Middelboe, M. (2013) 'Heterogeneous distribution of prokaryotes and viruses at the microscale in a tidal sediment'. Aquatic Microbial Ecology, 69(3):183–192.

Carreira, C., Piel, T., Staal, M., Stuut, J.-B. W., Middelboe, M. and Brussaard, C. P. D. (2015) 'Microscale spatial distributions of microbes and viruses in intertidal photosynthetic microbial mats'. SpringerPlus, 4(1):239.

Carreira, C., Staal, M., Middelboe, M. and Brussaard, C. P. D. (2015) 'Counting viruses and bacteria in photosynthetic microbial mats'. Applied and Environmental Microbiology, 81(6):2149–2155.

Casamatta, D., Stanic, D., Gantar, M. and Richardson, L. L. (2012) 'Characterization of *Roseofilum reptotaenium* (Oscillatoriales, cyanobacteria) gen. et sp. nov. isolated from Caribbean black band disease'. Phycologia, 51:489–499.

Cervino, J. M., Hayes, R., Goreau, T. J. and Smith, G. W. (2004) 'Zooxanthellae regulation in yellow blotch/band and other coral diseases contrasted with temperature related bleaching: In situ destruction vs expulsion'. Symbiosis, 37(1–3):63–85.

Chan, B. K., Abedon, S. T. and Loc-Carrillo, C. (2013) 'Phage cocktails and the future of phage therapy'. Future Microbiology, 8(6):769–83.

Chenard, C., Wirth, J. F., Suttle, C. A. (2016) 'Viruses infecting a freshwater filamentous cyanobacterium (*Nostoc* sp.) encode a functional CRISPR array and a proteobacterial DNA polymerase B'. Mbio, 7(5):1–11.

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M. and Bru, H. (2004) 'Phage-host interaction: an ecological perspective'. Journal of Bacteriology, 186(12):3677–3686.

Chow, C.-E. T., Kim, D. Y., Sachdeva, R. and Caron, D. A. (2013) 'Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists'. The ISME Journal, 8:816–829.

Cirés, S., Wörmer, L., Timón, J., Wiedner, C. and Quesada, A. (2011) 'Cylindrospermopsin production and release by the potentially invasive cyanobacterium *Aphanizomenon ovalisporum* under temperature and light gradients'. Harmful Algae, 10(6):668–675.

Cohen, Y., Joseph Pollock, F., Rosenberg, E. and Bourne, D. G. (2013) 'Phage therapy treatment of the coral pathogen *Vibrio coralliilyticus*'. MicrobiologyOpen, 2(1):64–74.

Comeau, A. M. and Krisch, H. M. (2008) 'The capsid of the T4 phage superfamily: The evolution, diversity, and structure of some of the most prevalent proteins in the biosphere'. Molecular Biology and Evolution, 25(7):1321–32.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. and Robles, M. (2005) 'Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research'. Bioinformatics, 21(18):3674–3676.

Cooney, R. P., Pantos, O., Le Tissier, M. D. A., Barer, M. R., O'Donnell, A. G. and Bythell, J. C. (2002) 'Characterization of the bacterial consortium associated with black band disease in coral using molecular microbiological techniques'. Environmental Microbiology, 4(7):401–13.

Cooper, E. L. (2010) 'Evolution of immune systems from self/not self to danger to artificial immune systems (AIS)'. Physical Life Review, 7(1):55–78.

Crummett, L. T., Puxty, R. J., Weihe, C., Marston, M. F. and Martiny, J. B. H. (2016) 'The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses'. Virology, 499:219–229.

D'Herelle, F. (1930) 'Elimination du bacteriophage dans les symbioses bacterie-bacteriophage'. *Comptes Rendus* des Séances de la Société de Biologie et de ses filiales Société de Biologie, 104, 1254.

Dang, V. T. and Sullivan M. B. (2014) 'Emerging methods to study bacteriophage infection at the single-cell level'. Frontiers in Microbiology, 5(724):1–8.

Danovaro, R. and Middelboe, M. (2010) 'Separation of free virus particles from sediments in aquatic systems'. Manual of Aquatic Viral Ecology, 74–81.

Danovaro, R., Bongiorni, L., Corinaldesi, C., Giovannelli, D., Damiani, E., Astolfi, P., Greci, L. and Pusceddu, A. (2008) 'Sunscreens cause coral bleaching by promoting viral infections'. Environmental Health Perspectives, 116(4):441–7.

Darling, A. E., Mau, B. and Perna, N. T. (2010) 'ProgressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement'. PLoS ONE, 5(6).

Davies, E. V., James, C.E., Williams, D., O'Brien, S., Fothergill, J.L., Haldenby, S., Paterson, S., Winstanley, C. and Brockhurst, M.A. (2016) 'Temperate phages both mediate and drive adaptive evolution in pathogen biofilms'. Proceedings of the National Academy of Sciences

of the United States of America, 113(29):8266–71.

Davies, M. R., Broadbent, S. E., Harris, S. R., Thomson, N. R. and Marjan, W. Van der (2013) 'Horizontally acquired glycosyltransferase operons drive *Salmonellae* lipopolysaccharide diversity'. PLoS ONE, 9(6):e1003568.

Davy, S. K., Burchett, S. G., Dale, A. L., Davies, P., Davy, J. E., Muncke, C. and Wilson, W. H. (2006) 'Viruses: agents of coral disease?'. Diseases of Aquatic Organisms, 69:101–110.

de Wit, R., Gautret, P., Bettarel, Y., Roques, C., Marlière, C., Ramonda, M., Nguyen Thanh, T., Tran Quang, H. and Bouvier, T. (2015) 'Viruses occur incorporated in biogenic high-Mg calcite from hypersaline microbial mats'. PLoS ONE, 10(6):e0130552.

De'ath, G., Fabricius, K. E., Sweatman, H. and Puotinen, M. (2012) 'The 27-year decline of coral cover on the Great Barrier Reef and its causes'. Proceedings of the National Academy of Sciences of the United States of America, 109(44):17995–9.

Delesalle, V. A., Tanke, N. T., Vill, A. C. and Krukonis, G. P. (2016) 'Testing hypotheses for the presence of tRNA genes in mycobacteriophage genomes'. Bacteriophage, 6(3):e1219441.

Den Uyl, P. A., Richardson, L. L., Jain, S. and Dick, G. J. (2016) 'Unraveling the physiological roles of the cyanobacterium *Geitlerinema* sp. BBD and other black band disease community members through genomic analysis of a mixed culture'. PLoS ONE, 11(6):e0157953.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G. L. (2006) 'Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB'. Applied and Environmental Microbiology, 72(7):5069–5072.

Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A. and Lavigne, R. (2012) 'Learning from bacteriophages - advantages and limitations of phage and phage-encoded protein applications'. Current Protein & Peptide Science, 13(8):699–722.

Dvorak, P., Poulickova, A., Hasler, P., Belli, M., Casamatta, D. A. and Papini, A. (2015) 'Species concepts and speciation factors in cyanobacteria, with connection to the problems of diversity and classification'. Biodiversity and Conservation, 24(4):739–757.

Dwivedi, B., Xue, B., Lundin, D., Edwards, R. and Breitbart, M. (2013) 'A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes'. BMC Evolutionary Biology, 13, 33.

Edgar, R. and Qimron, U. (2010) 'The *Escherichia* coli CRISPR system protects from λ lysogenization, lysogens, and prophage induction'. Journal of Bacteriology, 192(23):6291–6294.

Efrony, R., Atad, I. and Rosenberg, E. (2009) 'Phage therapy of coral white plague disease: Properties of phage BA3'. Current Microbiology, 58(2):139–145.

Efrony, R., Loya, Y., Bacharach, E. and Rosenberg, E. (2007) 'Phage therapy of coral disease'. Coral Reefs, 26:7–13.

Enav, H., Béjà, O. and Mandel-Gutfreund, Y. (2012) 'Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts'. The ISME

Journal, 6(3):619–28.

Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., Melamed, S., Leavitt, A., Savidor, A., Albeck, S. and Amitai, G. (2017) 'Communication between viruses guides lysis-lysogeny decisions'. Nature, 541(7638):488–93.

Erko, S. and Ebers, J. (2006) 'Taxonomic parameters revisited: tarnished gold standards'. Microbiology Today, 33:152–155.

Faruque, S. M. and Mekalanos, J. J. (2003) 'Pathogenicity islands and phages in *Vibrio cholerae* evolution'. Trends in Microbiology, 11(11):505–510.

Fenton, M., Ross, P., Mcauliffe, O., O'Mahony, J. and Coffey, A. (2010) 'Recombinant bacteriophage lysins as antibacterials'. Bioengineered Bugs, 1(1):9–16.

Filée, J., Tétart F., Suttle C. A, and Krisch H. M. (2005) 'Marine T4-Type bacteriophages, a ubiquitous component of the dark matter of the biosphere'. Proceedings of the National Academy of Sciences of the United States of America, 102(35):12471–76.

Flores, C., Meyer, J., Valverde, S., Farr, L. and Weitz, J. S. (2011) 'Statistical structure of host–phage interactions'. Proceedings of the National Academy of Sciences of the United States of America, 108(28):288–297.

Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M. and Smith, M. C. M. (2014) 'New applications for phage integrases'. Journal of Molecular Biology, 426(15):2703–2716.

Forterre, P. (2013) 'The virocell concept and environmental microbiology'. The ISME Journal, 7(2):233–6.

Frias-Lopez, J., Bonheyo, G. T., Jin, Q. S. and Fouke, B. W. (2003) 'Cyanobacteria associated with coral black band disease in Caribbean and Indo-Pacific reefs'. Applied and Environmental Microbiology, 69(4):2409–2413.

Frias-Lopez, J., Zerkle, A. L., Bonheyo, G. T. and Fouke, B. W. (2002) 'Partitioning of bacterial communities between seawater and healthy, black band diseased, and dead coral surfaces'. Applied and Environmental Microbiology, 68(5):2214–2228.

Fry, B. N., Feng, S., Chen, Y. Y., Newell, D. G., Coloe, P. J. and Korolik, V. (2000) 'The galE gene of *Campylobacter jejuni* is involved in lipopolysaccharide synthesis and virulence'. Infection and Immunity, 68(5):2594–2601.

Gabisoniya, T. G., Loladze, M. Z., Nadiradze, M. M., Chakhunashvili, N. K., Alibegashvili, M. G., Tamarashvili, N. G. and Pushkina, V. A. (2016) 'Effects of bacteriophages on biofilm formation by strains of *Pseudomonas aeruginosa*'. Applied Biochemistry and Microbiology, 52(3):293–297.

Gantar, M., Sekar, R. and Richardson, L. L. (2009) 'Cyanotoxins from black band disease of corals and from other coral reef environments'. Microbial Ecology, 58:856–864.

Garcia, G. D., Gregoracci, G. B., de O Santos, E., Meirelles, P. M., Silva, G. G. Z., Edwards, R., Sawabe, T., Gotoh, K., Nakamura, S., Iida, T., de Moura, R. L. and Thompson, F. L. (2013) 'Metagenomic analysis of healthy and white plague-affected *Mussismilia braziliensis* corals'. Microbial Ecology, 65(4):1076–86.

Gemma, S., Molteni, M. and Rossetti, C. (2016) 'Lipopolysaccharides in cyanobacteria: a brief overview'. Advances in Microbiology, 6:391–397.

Gibson, C. and Smith, R. (1982) 'Freshwater plankton. In: Carr, N. G., Whitton, B. A. (eds) The biology of cyanobacteria. Oxford: Blackwell Scientific Publication'. 463–489.

Glas, M. S., Motti, C., Negri, A. P., Sato, Y., Froscio, S., Humpage, A. R., Krock, B., Cembella, A. and Bourne, D. G. (2010) 'Cyanotoxins are not implicated in the etiology of coral black band disease outbreaks on Pelorus Island, Great Barrier Reef'. FEMS Microbiology Ecology, 73:43–54.

Glas, M. S., Sato, Y., Ulstrup, K. E. and Bourne, D. G. (2012) 'Biogeochemical conditions determine virulence of black band disease in corals'. The ISME Journal, 6(8):1526–34.

Godde, J. S. and Bickerton, A. (2006) 'The repetitive DNA elements called CRISPRs and their associated genes: Evidence of horizontal transfer among prokaryotes'. Journal of Molecular Evolution, 62(6):718–729.

Green, E. P. and Bruckner, A. W. (2000) 'The significance of coral disease epizootiology for coral reef conservation'. Biological Conservation, 96(3):347–361.

Grissa, I., Vergnaud, G. and Pourcel, C. (2007) 'The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats'. BMC Bioinformatics, 8(172):172.

Hargreaves, K. R., Flores, C. O., Lawley, T. D. and Clokie, M. R. J. (2014) 'Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen'. mBio, 5(5):1–10.

Harvell, D., Jordán-Dahlgren, E., Merkel, S., Rosenberg, E., Raymundo, L., Smith, G., Weil, E. and Willis, B. (2007) 'Coral disease, environmental drivers, and the balance between coral and microbial associates'. Oceanography, 20(1):172–195.

Hatoum-Aslan, A. and Marraffini, L. A. (2014) 'Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens'. Current Opinion in Microbiology, 17(1):82–90.

Heidelberg, J. F., Nelson, W. C., Schoenfeld, T. and Bhaya, D. (2009) 'Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes'. PLoS ONE, 4(1):e4169.

Held, N. L., Childs, L. M., Davison, M., Weitz, J. S., Whitaker, R. J. and Bhaya, D. (2013) 'CRISPR-Cas systems to probe ecological diversity and host-viral interactions'. In: Barrangou, R. and van der Oost, J. (eds) CRISPR-Cas systems: RNA-mediated adaptive Immunity in bacteria and archaea. Springer Berlin / Heidelberg:221–250.

Held, P. (2011) 'Monitoring of Algal Growth Using their Intrinsic Properties'. Biofuel Research Application, 1–5.

Herlemann, D. P., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J. and Andersson, A. F. (2011) 'Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea'. The ISME Journal, 5(10):1571–9.

Hewson, I., Brown, J. M., Burge, C. A., Couch, C. S., LaBarre, B. A., Mouchka, M. E., Naito, M.

and Harvell, C. D. (2012) 'Description of viral assemblages associated with the *Gorgonia ventalina* holobiont'. Coral Reefs, 31(2):487–491.

Hildenbrand, Z. L. and Bernal, R. A. (2012) 'Chaperonin-mediated folding of viral proteins'. Advances in Experimental Medicine and Biology, 726:307–324.

Hoegh-Guldberg, O. (1999) 'Climate Change, coral bleaching and the future of the world' s coral reefs'. Marine and Freshwater Research, 50:839–866.

Holmfeldt, K., Howard-Varona, C., Solonenko, N. and Sullivan, M. B. (2014) 'Contrasting genomic patterns and infection strategies of two co-existing Bacteroidetes podovirus genera'. Environmental Microbiology, 16(8):2501–2513.

Holmfeldt, K., Middelboe, M., Nybroe, O. and Riemann, L. (2007) 'Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts'. Applied and Environmental Microbiology, 73(21):6730–6739.

Holmfeldt, K., Solonenko, N., Shah, M., Corrier, K., Riemann, L., Verberkmoes, N. C. and Sullivan, M. B. (2013) 'Twelve previously unknown phage genera are ubiquitous in global oceans'. Proceedings of the National Academy of Sciences of the United States of America, 110(31):12798–803.

Houghton, K. A. (2015) 'Responses in bacterioplankton production and community structure after exposure to oil and dispersant in the northeastern Gulf of Mexico'. University of West Florida.

Hube, A. E., Heyduck-Söller, B. and Fischer, U. (2009) 'Phylogenetic classification of heterotrophic bacteria associated with filamentous marine cyanobacteria in culture'. Systematic and Applied Microbiology, 32(4):256–65.

Hughes, T. P., Baird, A. H., Bellwood, D. R., Card, M., Connolly, S. R., Folke, C., Grosberg, R., Hoegh-Guldberg, O., Jackson, J. B. C., Kleypas, J., Lough, J. M., Marshall, P., Nyström, M., Palumbi, S. R., Pandolfi, J. M., Rosen, B. and Roughgarden, J. (2003) 'Climate change, human impacts, and the resilience of coral reefs'. Science, 301(5635):929–933.

Hunter-Cevera, K. R., Post, A. F., Peacock, E. E. and Sosik, H. M. (2016) 'Diversity of *Synechococcus* at the Martha's Vineyard Coastal Observatory: insights from culture isolations, clone libraries, and flow cytometry'. Microbial Ecology, 71(2):276–289.

Islam, R. (2016) 'NAD-dependent epimerase/dehydratase affects cell surface properties, virulence and extracellular enzyme production in the soft rot phytopathogen, *Pectobacterium carotovorum'*. Tennessee State University.

Jacquet, S. and Bratbak, G. (2003) 'Effects of ultraviolet radiation on marine virus-phytoplankton Interactions'. FEMS Microbiology Ecology, 44(3):279–89.

Jensen, M. A., Faruque, S. M., Mekalanos, J. J. and Levin, B. R. (2006) 'Modeling the role of bacteriophage in the control of cholera outbreaks'. Proceedings of the National Academy of Sciences of the United States of America, 103(12):4652–7.

Jiang, W. and Marraffini, L. A. (2015) 'CRISPR-Cas: New tools for genetic manipulations from bacterial immunity systems'. Annual Review of Microbiology, 69(1):150724172101001.

Jiang, W., Maniv, I., Arain, F., Wang, Y., Levin, B. R. and Marraffini, L. A. (2013) 'Dealing with the evolutionary downside of CRISPR immunity: bacteria and beneficial plasmids'. PLoS Genetics, 9(9):e1003844.

Johan, O., Bengen, D. G., Zamani, N. P. and Sweet, M. J. (2015) 'The distribution and abundance of black band disease and white syndrome in Kepulauan Seribu, Indonesia'. HAYATI Journal of Biosciences, 22(3):105–112.

Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W. and Weitz, J. S. (2014) 'The elemental composition of virus particles: implications for marine biogeochemical cycles'. Nature Reviews Microbiology, 12(7):519–28.

Kang, I., Jang, H. and Cho, J.-C. (2012) 'Complete genome sequences of two *Persicivirga* bacteriophages, P12024S and P12024L'. Journal of Virology, 86(16):8907–8908.

Katz, S. M., Pollock, F. J., Bourne, D. G. and Willis, B. L. (2014) 'Crown-of-thorns starfish predation and physical injuries promote brown band disease on corals'. Coral Reefs, 33(3):705–716.

Keen, E. C. and Adhya, S. L. (2014) 'Phage therapy: current research and applications'. Clinical Infectious Diseases, 61(1):141–142.

Kelman, D., Kashman Y., Rosenberg E., Kushmaro A, and Loya Y. (2006) 'Antimicrobial activity of Red Sea corals'. Marine Biology, 149(2):357–63.

Klaus, J. S., Janse, I. and Fouke, B. W. (2011) 'Coral black band disease microbial communities and genotypic variability of the dominant cyanobacteria (CD1C11)'. Bulletin of Marine Science, 87(4):795–821.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F. O. (2013) 'Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies'. Nucleic Acids Research, 41(1):1–11.

Koskella, B. and Meaden, S. (2013) 'Understanding bacteriophage specificity in natural microbial communities'. Viruses, 5(3):806–823.

Koskella, B., Thompson, J. N., Preston, G. M. and Buckling, A. (2011) 'Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria'. The American Naturalist, 177(4):440–51.

Kramarsky-Winter, E., Arotsker, L., Rasoulouniriana, D., Siboni, N., Loya, Y. and Kushmaro, A. (2014) 'The possible role of cyanobacterial filaments in coral black band disease Pathology'. Microbial Ecology, 67:177–185.

Krediet, C. J., Ritchie K. B., Paul V. J, and Teplitski M. (2013) 'Coral-associated micro-organisms and their roles in promoting coral health and thwarting diseases'. Proceedings of the Royal Society B: Biological Sciences, 280:20122328.

Kropinski, A. M., Sulakvelidze, A., Konczy, P. and Poppe, C. (2007) '*Salmonella* phages and prophages - genomics and practical aspects'. Methods in Molecular Biology, 394:133–175.

Kropinski, A. M., Mazzocco, A., Waddell, T. E., Lingohr, E, and Johnson, R. P. (2009) 'Enumeration of bacteriophages by double agar overlay plaque assay'. Bacteriophages: methods and protocols, Volume 1: isolation, characterization, and interactions, 501:69–76.

Kruskopf, M. and Du Plessis, S. (2006) 'Growth and filament length of the bloom forming *Oscillatoria simplicissima* (Oscillatoriales, Cyanophyta) in varying N and P concentrations'. Hydrobiologia, 556(1):357–362.

Kuehl, K., Jones, R., Gibbs, D. and Richardson, L. L. (2011) 'The roles of temperature and light in black band disease (BBD) progression on corals of the genus *Diploria* in Bermuda'. Journal of Invertebrate Pathology, 106(3):366–370.

Kushmaro, A., Banin, E., Loya, Y., Stackebrandt, E, and Rosenberg, E. (2001) '*Vibrio shiloi* sp. nov., the causative agent of bleaching of the coral *Oculina patagonica*'. International Journal of Systematic and Evolutionary Microbiology 51(4):1383–88.

Kushmaro, A., Loya, Y., Fine, M, and Rosenberg, E. (1996) 'Bacterial infection and coral bleaching'. Nature, 380(6573):396.

Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Weynberg, K. D., Botté, E. S., van Oppen, M. J. H., Webster, N. S. and Rattei, T. (2016) 'HoloVir: A workflow for investigating the diversity and function of viruses in invertebrate holobionts'. Frontiers in Microbiology, 7(822).

Lamb, J. B. and Willis, B. L. (2011) 'Using coral disease prevalence to assess the effects of concentrating tourism activities on offshore reefs in a tropical marine park'. Conservation Biology, 25(5):1044–52.

Lamb, J. B., True, J. D., Piromvaragorn, S. and Willis, B. L. (2014) 'Scuba diving damage and intensity of tourist activities increases coral disease prevalence'. Biological Conservation, 178:88–96.

Landy, A. and Ross, W. (1977) 'Viral integration and excision: structure of the Lambda att sites'. Science, 197(4309):1147–1160.

Lane, D. J. (1991) '16S/23S rRNA sequencing. In: Stackebrandt, E., Goodfellow, M. (eds) Nucleic acid techniques in bacterial systematics'. John Wiley & Sons. 125-175.

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H. W. and Kropinski, A. M. (2008) 'Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools'. Research in Microbiology, 159(5):406–414.

Lawrence, S. A., Davy, J. E., Wilson, W. H., Hoegh-Guldberg, O. and Davy, S. K. (2015) '*Porites* white patch syndrome: associated viruses and disease physiology'. Coral Reefs, 34(1):249–257.

Leland, D. S. and Ginocchio, C. C. (2007) 'Role of cell culture for virus detection in the age of technology'. Clinical Microbiology Reviews, 20(1):49–78.

Levin, B. R. (2010) 'Nasty viruses, costly plasmids, population dynamics, and the conditions for establishing and maintaining CRISPR-mediated adaptive immunity in bacteria'. PLoS Genetics, 6(10):1–12.

Levin, B. R., Moineau, S., Bushman, M. and Barrangou, R. (2013) 'The Population and Evolutionary Dynamics of Phage and Bacteria with CRISPR-Mediated Immunity'. PLoS Genetics, 9(3).

Loc-Carrillo, C. and Abedon, S. (2011) 'Pros and cons of phage therapy'. Bacteriophage, 1(2):111–114.

Lohr, J., Munn, C. B. and Wilson, W. H. (2007) 'Characterization of a latent virus-like infection of symbiotic zooxanthellae'. Applied and Environmental Microbiology, 73(9):2976–81.

Lorenzi, H. A., Hoover, J., Inman, J., Safford, T., Murphy, S., Kagan, L. and Williamson, S. J. (2011) 'The Viral MetaGenome Annotation Pipeline (VMGAP): an automated tool for the functional annotation of viral metagenomic shotgun sequencing data'. Standards in Genomic Sciences, 4(3):418–29.

Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P. and van der Oost, J. (2014) 'The role of CRISPR-Cas systems in virulence of pathogenic bacteria'. Microbiology and Molecular Biology Reviews, 78(1):74–88.

Lowe, T. M. and Eddy, S. R. (1997) 'tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence'. Nucleic Acids Research, 25:955–964.

Makarova, K. S., Aravind, L., Wolf, Y. I. and Koonin, E. V (2011) 'Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems'. Biology Direct, 6(1):38.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J. M., Wolf, Y. I., Yakunin, A. F., van der Oost, J. and Koonin, E. V (2011) 'Evolution and classification of the CRISPR–Cas systems'. Nature Reviews Microbiology, 9(6):467–477.

Mäki, M. and Renkonen, R. (2004) 'Biosynthesis of 6-deoxyhexose glycans in bacteria'. Glycobiology, 14(3):1–15.

Marhaver, K. L., Edwards, R. A. and Rohwer, F. (2008) 'Viral communities associated with healthy and bleaching corals'. Environmental Microbiology, 10(9):2277–86.

Maynard, J., van Hooidonk, R., Eakin, C. M., Puotinen, M., Garren, M., Williams, G., Heron, S. F., Lamb, J., Weil, E., Willis, B. L. and Harvell, C. D. (2015) 'Projections of climate conditions that increase coral disease susceptibility and pathogen abundance and virulence'. Nature Climate Change, 5, 688.

McDaniel, L. D. (2011) 'Viruses of Cyanobacteria'. In: Hurst, C. J. (ed.) Studies in Viral Ecology: Microbial and Botanical Host Systems. John Wiley & Sons: Chichester, UK:169–187.

Meyer, J. L., Gunasekera, S. P., Scott, R. M., Paul, V. J. and Teplitski, M. (2015) 'Microbiome shifts and the inhibition of quorum sensing by black band disease cyanobacteria'. The ISME Journal, 10(5):1–13.

Middelboe, M., Chan, A. M. and Bertelsen, S. K. (2010) 'Isolation and life cycle characterization of lytic viruses infecting heterotrophic bacteria and cyanobacteria'. Manual of Aquatic Viral Ecology, 118–133.

Miller, A. W. and Richardson, L. L. (2011) 'A meta-analysis of 16S rRNA gene clone libraries from the polymicrobial black band disease of corals'. FEMS Microbiology Ecology, 75:231–241.

Miyoshi, S. and Shinoda, S. (1997) 'Bacterial metalloprotease as the toxic factor in infection'. Journal of Toxicology: Toxin Reviews, 16(4):177–194.

Mojica, K. D. A. and Brussaard, C. P. D. (2014) 'Factors affecting virus dynamics and microbial

host-virus interactions in marine environments'. FEMS Microbiology Ecology, 89(3):495–515.

Morrow, K. M., Moss, A. G., Chadwick, N. E. and Liles, M. R. (2012) 'Bacterial associates of two Caribbean coral species reveal species-specific distribution and geographic variability'. Applied and Environmental Microbiology, 78(18):6438–6449.

Motlagh, A. M., Bhattacharjee, A. S. and Goel, R. (2016) 'Biofilm control with natural and genetically-modified phages'. World Journal of Microbiology and Biotechnology. Springer Netherlands, 32(4):1–10.

Munn, C. B. (2006) 'Viruses as pathogens of marine organisms—from bacteria to whales'. Journal of the Marine Biological Association of the UK, 86(3):453.

Neave, M. J., Rachmawati, R., Xun, L., Michell, C. T., Bourne, D. G., Apprill, A. and Voolstra, C. R. (2017) 'Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales'. The ISME Journal, 11(1):1–15.

Neidhardt, F., Ingraham, J. L., Schaechter, S., Neidhart, F. C., Ingraham, J. L. and Schaechter, M. (1990) 'Physiology of the bacterial cell: a molecular approach'. Sinauer Associates.

Ng, T. F. F., Manire, C., Borrowman, K., Langer, T., Ehrhart, L. and Breitbart, M. (2009) 'Discovery of a novel single-stranded DNA virus from a sea turtle fibropapilloma by using viral metagenomics'. Journal of Virology, 83(6):2500–9.

Nobrega, F. L., Costa, A. R., Kluskens, L. D. and Azeredo, J. (2015) 'Revisiting phage therapy: New applications for old resources'. Trends in Microbiology, 23(4):185–191.

Oliveira, J., Castilho, F., Cunha, A. and Pereira, M. J. (2012) 'Bacteriophage therapy as a bacterial control strategy in aquaculture'. Aquaculture International, 20(5):879–910.

Oren, A. (2011a) 'Cyanobacterial systematics and nomenclature as featured in the International Bulletin of Bacteriological Nomenclature and Taxonomy / International Journal of Systematic Bacteriology / International Journal of Systematic and Evolutionary Microbiology'. International Journal of Systematic and Evolutionary Microbiology, 61:10–15.

Oren, A. (2011b) 'Naming Cyanophyta/Cyanobacteria – a bacteriologist's view'. Fottea, 11(1):9–16.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. and Stevens, R. (2014) 'The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)'. Nucleic Acids Research, 42(D1):D206–D214.

Pacton, M., Wacey, D., Corinaldesi, C., Tangherlini, M., Kilburn, M. R., Gorin, G. E., Danovaro, R. and Vasconcelos, C. (2014) 'Viruses as new agents of organomineralization in the geological record'. Nature Communications, 5, 4298.

Page, C. A. and Willis, B. L. (2006) 'Distribution, host range and large-scale spatial variability in black band disease prevalence on the Great Barrier Reef, Australia'. Diseases of aquatic organisms, 69:41–51.

Palmer, C. V. and Traylor-Knowles, N. (2012) 'Towards an integrated network of coral immune

mechanisms'. Proceedings of the Royal Society B: Biological Sciences, 279(1745):4106–14.

Pastagia, M., Schuch, R., Fischetti, V. A. and Huang, D. B. (2013) 'Lysins: The arrival of pathogen-directed anti-infectives'. Journal of Medical Microbiology, 62(PART10):1506–1516.

Paul, J. H. and Weinbauer, M. (2010) 'Detection of lysogeny in marine environments'. Manual of Aquatic Viral Ecology. ASLO:30–33.

Payet, J. P. and Suttle, C. A. (2013) 'To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status'. Limnology and Oceanography, 58(2):465–74.

Pollock, F. J., M. Wood-Charlson, E., van Oppen, M. J. H., Bourne, D. G., Willis, B. L. and Weynberg, K. D. (2014) 'Abundance and morphology of virus-like particles associated with the coral *Acropora hyacinthus* differ between healthy and white syndrome-infected states'. Marine Ecology Progress Series, 510:39–43.

Pollock, F. J., Morris, P. J., Willis, B. L. and Bourne, D. G. (2011) 'The urgent need for robust coral disease diagnostics'. PLoS Pathogens, 7(10):e1002183.

Pollock, F. J., Morris, P. J., Willis, B. L, and Bourne, D. G. (2010) 'Detection and quantification of the coral pathogen *Vibrio Coralliilyticus* by real-time PCR with TaqMan fluorescent probes'. Applied and Environmental Microbiology, 76(15):5282–86.

Pope, W. H., Weigele, P. R., Chang, J., Pedulla, M. L., Ford, M. E., Houtz, J. M., Jiang, W., Chiu, W., Hatfull, G. F., Hendrix, R. W., King, J. (2007) 'Genome Sequence, Structural Proteins, and Capsid Organization of the Cyanophage Syn5: A 'Horned' Bacteriophage of Marine *Synechococcus*'. J Mol Biol, 368(4):966–81.

Praveen Kumar, R., Vijayan, D., Antony, L. M., Kumar, C. M, and Thajuddin, N. (2009) 'Phylogenetic diversity of cultivable bacteria associated with filamentous non-hetrocystous marine cyanobacteria'. Journal of Algal Biomass Utilization, 1(1):86–101.

Rasoulouniriana, D., Siboni, N., Ben-Dov, E., Kramarsky-Winter, E., Loya, Y. and Kushmaro, A. (2009) '*Pseudoscillatoria coralii* gen. nov., sp nov., a cyanobacterium associated with coral black band disease (BBD)'. Diseases of Aquatic Organisms, 87(1–2):91–96.

Rath, D., Amlinger, L., Rath, A. and Lundgren, M. (2015) 'The CRISPR-Cas immune system: biology, mechanisms and applications'. Biochimie, 117:119–28.

Reddy, A. D., Jeyasekaran, G. and Shakila, R. J. (2013) 'Morphogenesis, pathogenesis, detection and transmission risks of white spot syndrome virus in shrimps'. Fisheries and Aquaculture Journal, 2013:1–13.

Reshef, L., Koren, O., Loya, Y., Zilber-Rosenberg, I. and Rosenberg, E. (2006) 'The coral probiotic hypothesis'. Environmental Microbiology, 8(12):2068–73.

Reyes, A., Semenkovich, N. P., Whiteson, K., Rohwer, F. and Gordon, J. I. (2012) 'Going viral: Next-generation sequencing applied to phage populations in the human gut'. Nature Reviews Microbiology, 10(9):607–17.

Richardson, L. L. and Kuta, K. G. (2003) 'Ecological physiology of the black band disease cyanobacterium *Phormidium corallyticum*'. FEMS Microbiology Ecology, 43:287–298.

Richardson, L. L., Miller, A. W. and Blackwelder, P. (2015) 'Cyanobacterial-associated colored-band diseases of the Atlantic/Caribbean. In: Woodley, C. M., Downs, C. A., Bruckner, A. W., Porter, J. W., Galloway, S. B. (eds) Diseases of coral. John Wiley & Sons. 345–353.

Richardson, L. L., Sekar, R., Myers, J. L., Gantar, M., Voss, J. D., Kaczmarsky, L., Remily, E. R., Boyer, G. L. and Zimba, P. V. (2007) 'The presence of the cyanobacterial toxin microcystin in black band disease of corals'. FEMS Microbiology Letters, 272(2):182–187.

Richardson, L. L., Stanić, D., May, A., Brownell, A., Gantar, M. and Campagna, S. (2014) 'Ecology and physiology of the pathogenic cyanobacterium *Roseofilum reptotaenium*'. Life, 4(4):968–987.

Rippka, R., Deruelles, J., Waterbury, J., Herdman, M. and Stanier, R. (1979) 'Generic assignments, strain histories and properties of pure cultures of cyanobacteria'. Journal of Genetic Microbiology, 111:1–61.

Rippka, R., Waterbury, J. and Stanier, R. Y. (1981) 'Isolation and purification of cyanobacteria: some general principles. In: The Prokaryotes: A Handbook on Habitats, Isolation and Identification of Bacteria'. Springer Berling Heidelberg, 212-220.

Rosenberg, E., Koren, O., Reshef, L., Efrony, R. and Zilber-Rosenberg, I. (2007) 'The role of microorganisms in coral health, disease and evolution'. Nature Reviews Microbiology, 5(5):355–62.

Roux, S., Enault, F., Hurwitz, B. L. and Sullivan, M. B. (2015) 'VirSorter: mining viral signal from microbial genomic data'. PeerJ, 3:e985.

Roux, S., Faubladier, M., Mahul, A., Paulhe, N., Bernard, A., Debroas, D. and Enault, F. (2011) 'Metavir: A web server dedicated to virome analysis'. Bioinformatics, 27(21):3074–3075.

Roux, S., Tournayre, J., Mahul, A., Debroas, D, and Enault, F. (2014) 'Metavir 2: New tools for viral metagenome comparison and assembled virome analysis'. BMC bioinformatics, 15:76.

Ruiz-Moreno, D., Willis, B. L., Page, A. C., Weil, E., Cróquer, A., Vargas-Angel, B., Jordan-Garza, A. G., Jordán-Dahlgren, E., Raymundo, L. and Harvell, C. D. (2012) 'Global coral disease prevalence associated with sea temperature anomalies and local factors'. Diseases of Aquatic Organisms, 100(3):249–261.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M. A. and Barrell, B. (2000) 'Artemis: sequence visualization and annotation'. Bioinformatics, 16(10):944–945.

Rützler, K., Santavy, D. and Antonius, A. (1983) 'The black band diseases of Atlantic reef corals. III. Distribution, ecology, and development'. Marine Ecology, 4(4):301-319.

Sánchez-Paz, A., Muhlia-Almazan, A., Saborowski, R., García-Carreño, F., Sablok, G. and Mendoza-Cano, F. (2014) 'Marine Viruses: the beneficial side of a threat'. Applied Biochemistry and Biotechnology, 174(7):2368–2379.

Safferman, R. S., Cannon R. E., and Desjardins, P. R. (1983) 'Classification and Nomenclature of Viruses of Cyanobacteria'. Intervirology, 19:61–66.

Sato, Y., Bourne, D. G. and Willis, B. L. (2011) 'Effects of temperature and light on the progression of black band disease on the reef coral, *Montipora hispida*'. Coral Reefs,

30(3):753–761.

Sato, Y., Bourne, D. G., Willis, B. L. (2009) 'Dynamics of seasonal outbreaks of black band disease in an assemblage of *Montipora* species at Pelorus Island (Great Barrier Reef, Australia)'. Proceedings of the Royal Society B: Biological Sciences, 276(1668):2795–2803.

Sato, Y., Civiello, M., Bell, S. C., Willis, B. L. and Bourne, D. G. (2016) 'Integrated approach to understanding the onset and pathogenesis of black band disease in corals'. Environmental Microbiology, 18(3):752–765.

Sato, Y., Ling, E. Y. S., Turaev, D., Laffy, P., Weynberg, K. D., Rattei, T., Willis, B. L. and Bourne, D. G. (2017) 'Unraveling the microbial processes of black band disease in corals through integrated genomics'. Scientific reports, 7(40455):1–14.

Sato, Y., Willis, B. L. and Bourne, D. G. (2010) 'Successional changes in bacterial communities during the development of black band disease on the reef coral, *Montipora hispida*'. The ISME Journal, 4(2):203–14.

Schwenk, Y. (2012) 'Bacteriophage therapy of black band coral disease'. Master Thesis, Australian Institute of Marine Science.

Sheridan, C., Kramarsky-Winter, E., Sweet, M., Kushmaro, A. and Leal, M. C. (2013) 'Diseases in coral aquaculture: Causes, implications and preventions'. Aquaculture, 396–399:124–135.

Shestakov, S. V and Karbysheva, E. A. (2015) 'The role of viruses in the evolution of cyanobacteria'. Medical Hypotheses, 135(2):115–127.

Short, S. M. and Suttle C. A. (2003) 'Temporal dynamics of natural communities of marine algal viruses and eukaryotes'. Aquatic Microbial Ecology, 32(2):107–19.

Smith, A. D. and Gilbert, J. J. (1995) 'Spatial and temporal variability in filament length of a toxic cyanobacterium (*Anabaena affinis*)'. Freshwater Biology, 33(1):1–11.

Soffer, N., Brandt, M. E., Correa, A. M. S., Smith, T. B. and Thurber, R. V. (2014) 'Potential role of viruses in white plague coral disease'. The ISME Journal, 8(2):271–283.

Stanic, D., Oehrle, S., Gantar, M. and Richardson, L. L. (2011) 'Microcystin production and ecological physiology of Caribbean black band disease cyanobacteria'. Environmental Microbiology, 13:900–910.

Stern, A. and Sorek, R. (2011) 'The phage-host arms race: Shaping the evolution of microbes'. BioEssays, 33(1):43–51.

Stern, A., Keren, L., Wurtzel, O., Amitai, G. and Sorek, R. (2010) 'Self-targeting by CRISPR: Gene regulation or autoimmunity?'. Trends in Genetics, 26(8):335–340.

Stern, A., Eran, M., Itay, Ti., Or, S. and Sorek R. (2012) 'CRISPR targeting reveals a reservoir of common phages associated with the human gut microbiome'. Genome Research, 22:1985–94.

Stothard, P. and Wishart, D. S. (2005) 'Circular genome visualization and exploration using CGView'. Bioinformatics, 21(4):537–539.

Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. and Chisholm, S. W. (2005) 'Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations'. PLoS Biology, 3(5):e144.

Sullivan, M. B., Waterbury, J. B. and Chisholm, S. W. (2003) 'Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*'. Nature, 424(6952):1047–1051.

Sullivan, M. B. (2015) 'Viromes, not gene markers, for studying double-stranded DNA virus communities'. Journal of Virology, 89(5):2459–61.

Sussman, M., Bourne, D. G. and Willis, B. L. (2006) 'A single cyanobacterial ribotype is associated with both red and black bands on diseased corals from Palau'. Diseases of aquatic organisms, 69:111–8.

Sutherland, K. P., Porter, J. W. and Torres, C. (2004) 'Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals'. Marine Ecology Progress Series, 266:273–302.

Suttle, C. A. (2007) 'Marine viruses - major players in the global ecosystem'. Nature reviews. Microbiology, 5(10):801–12.

Suttle, C. A. (2005) 'Viruses in the sea'. Nature, 437(7057):356–61.

Suttle, C. A. and Fuhrman, J. A. (2010) 'Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy'. Manual of Aquatic Viral Ecology, 145–153.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) 'MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods'. Molecular Biology and Evolution, 28:2731–2739.

Taylor, D. L. (1983) 'The black band disease of Atlantic reef corals'. Marine Ecology, 4(4):301–319.

Teplitski, M. and Ritchie, K. (2009) 'How feasible is the biological control of coral diseases?'. Trends in Ecology & Evolution, 24(7):378–85.

Thakur, N. L., Arga, C. A. and Müller, W. E. G. (2004) 'Culturable epibacteria of the marine sponge *Ircinia fusca*: temporal variations and their possible role in the epibacterial defense of the host'. Aquatic Microbial Ecology, 37(3):295–304.

Thingstad, T. F. (2000) 'Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems'. Limnology and Oceangraphy, 45(6):1320–1328.

Thompson, F. L., Barash, Y., Sawabe, T., Sharon, G., Swings, J. and Rosenberg, E. (2006) '*Thalassomonas loyana* sp. nov., a causative agent of the white plague-like disease of corals on the Eilat coral reef'. International Journal of Systematic and Evolutionary Microbiology, 56:365–8.

Thurber, R. L. V. and Correa, A. M. S. (2011) 'Viruses of reef-building scleractinian corals'. Journal of Experimental Marine Biology and Ecology, 408(1–2):102–113.

Thurber, R. V, Barott, K. L., Hall, D., Liu, H., Rodriguez-Mueller, B., Desnues, C., Edwards, R. a, Haynes, M., Angly, F. E., Wegley, L. and Rohwer, F. L. (2008) 'Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral *Porites*

compressa'. Proceedings of the National Academy of Sciences of the United States of America, 105(47):18413–8.

Thurber, R. V., Payet, J. P., Thurber, A. R. and Correa, A. M. S. (2017) 'Virus–host interactions and their roles in coral reef health and disease'. Nature Reviews Microbiology, 1:1-12.

Thurber, R. V., Haynes, M., Breitbart, M., Wegley, L. and Rohwer, F. (2009) 'Laboratory procedures to generate viral metagenomes'. Nature Protocols, 4(4):470–83.

Touchon, M. and Rocha, E. P. C. (2010) 'The small, slow and specialized CRISPR and anti-CRISPR of *Escherichia* and *Salmonella*'. PLoS ONE, 5(6).

Vaara, T., Vaara, M. and Niemela, S. (1979) 'Two improved methods for obtaining axenic cultures of cyanobacteria'. Applied and Environmental Microbiology, 38(5):1011–1014.

van Hulten, M. C., Witteveldt, J., Peters, S., Kloosterboer, N., Tarchini, R., Fiers, M., Sandbrink, H., Lankhorst, R. K. and Vlak, J. M. (2001) 'The white spot syndrome virus DNA genome sequence'. Virology, 286:7–22.

van Oppen, M. J. H., Leong, J. A. and Gates, R. D. (2009) 'Coral-virus interactions: A double-edged sword?'. Symbiosis, 47(1):1–8.

Varin, T., Lovejoy, C, Jungblut, A. D., Warwick, F. V. and Corbeil, J. (2010) 'Metagenomic profiling of arctic microbial mat communities as nutrient scavenging and recycling systems'. Limnology and Oceanography, 55(5):1901–11.

Vos, M., Birkett, P. J., Birch, E., Griffiths, R. I. and Buckling, A. (2009) 'Local adaptation of bacteriophages to their bacterial hosts in soil'. Science, 325(5942):833.

Waldor, M. and Mekalanos, J. (1996) 'Lysogenic conversion by a filamentous phage encoding cholera toxin'. Science, 272(5270):1910–1914.

Webb, N. A., Mulichak, A. M., Lam, J. S., Rocchetta, H. L. and Garavito, R. M. (2004) 'Crystal structure of a tetrameric GDP- D -mannose 4, 6-dehydratase from a bacterial GDP- D -rhamnose biosynthetic pathway'. Protein Science, 13:529–539.

Weigele, P. R., Pope, W. H., Pedulla, M. L., Houtz, J. M., Smith, A. L., Conway, J. F., King, J., Hatfull, G. F., Lawrence, J. G. and Hendrix, R. W. (2007) 'Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*'. Environmental Microbiology, 9(7):1675–1695.

Weinbauer, M. (2004) 'Ecology of prokaryotic viruses'. FEMS microbiology reviews, 28:127–181.

Weld, R. J., Butts, C. and Heinemann, J. A. (2004) 'Models of phage growth and their applicability to phage therapy'. Journal of Theoretical Biology, 227(1):1–11.

West, N. J., Lepère, C., Manes, C. L. de O., Catala, P., Scanlan, D. J. and Lebaron, P. (2016) 'Distinct spatial patterns of SAR11, SAR86, and Actinobacteria diversity along a transect in the ultra-oligotrophic South Pacific Ocean'. Frontiers in Microbiology, 7(3):1–16.

Weynberg, K. D., Voolstra, C. R., Neave, M. J., Buerger, P. and van Oppen, M. J. H. (2015) 'From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen'. Scientific Reports, 5, 17889.

Weynberg, K. D., Wood-Charlson, E. M., Suttle, C. A. and van Oppen, M. J. H. (2014) 'Generating viral metagenomes from the coral holobiont'. Frontiers in Microbiology, 5(5):206.

Wilcox, R. M. and Fuhrman, J. A. (1994) 'Bacterial viruses in coastal seawater: lytic rather than lysogenic production'. Marine Ecology-Progress Series, 114:35-35.

Wilhelm, S. W. and Suttle, C. A. (1999) 'Viruses and nutrient cycles in the sea'. BioScience, 49(10):781–88.

Willis, B. L., Page, C. A. and Dinsdale, E. A. (2004) 'Coral disease on the Great Barrier Reef'. In: Rosenberg, E. L. Y. (ed.) Coral Health and Disease. Springer: 69–104.

Wilson, W. H., Francis, I., Ryan, K. and Davy, S. K. (2001) 'Temperature induction of viruses in symbiotic dinoflagellates'. Aquatic Microbial Ecology, 25:99–102.

Wommack, K. E. and Colwell, R. R. (2000) 'Virioplankton: Viruses in Aquatic Ecosystems'. Microbiology and Molecular Biology Reviews, 64(1):69.

Wood-Charlson, E. M., Weynberg, K. D., Suttle, C. A., Roux, S. and van Oppen, M. J. H. (2015) 'Metagenomic characterization of viral communities in corals: mining biological signal from methodological noise'. Environmental microbiology, 17(10):3440–3449.

Wood, D. E. and Salzberg, S. L. (2014) 'Kraken: ultrafast metagenomic sequence classification using exact alignments'. Genome biology, 15(3):R46.

Wu, D., Jospin, G. and Eisen, J. A. (2013) 'Systematic identification of gene families for use phylogeny-driven, "'markers'" for phylogenetic and their, ecological studies of bacteria and archaea and major subgroups'. PLoS ONE, 8(10):e77033.

Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. and Singer, S. W. (2014) 'MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm'. Microbiome, 2(1):26.

Yang, H. Q., Shen, J. W., Fu, F. X., Wang, Y. and Zhao, N. (2014) 'Black band disease as a possible factor of the coral decline at the northern reef-flat of Yongxing Island, South China Sea'. Science China Earth Sciences, 57(4):569–578.

Yoong, P., Schuch, R., Nelson, D. and Fischetti, V. A. (2004) 'Identification of a broadly active phage lytic enzyme with lethal activity against antibiotic-resistant *Enterococcus faecalis* and *Enterococcus faecium*'. Journal of Bacteriology, 186(14):4808–4812.

Zegans, M. E., Wagner, J. C., Cady, K. C., Murphy, D. M., Hammond, J. H. and O'Toole, G. A. (2009) 'Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of *Pseudomonas aeruginosa*'. Journal of Bacteriology, 91(1):210–219.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. and Wishart, D. S. (2011) 'PHAST: A fast phage search tool'. Nucleic Acids Research, 39(2):1–6.

Appendix - Chapter 2

T4 bacteriophage communities

Appendix 2.1 List of coral tissue samples. Seawater samples were collected in 3 to 5 m depth just above healthy and diseased corals (*Pavona* sp.). Tissue samples were collected from healthy and diseased coral species *Montipora hispida*, in a transition time from an Australian Winter to an Australian summer (August to February). 20 L of seawater were collected per sampling trip. Multiple tissue samples were collected, however only samples with successful amplification are shown (*gp23*, 16S rRNA, nd = not successful amplification).

Sampling time	Sample #	Sample type	Amplification success	
2007 08 - Winter	1	BBD	*gp23*	16S
2007 10 - Winter	2	BBD	*gp23*	16S
2007 10 - Winter	3	BBD	*gp23*	16S
2007 10 - Winter	4	BBD	*gp23*	16S
2007 10 - Winter	5	BBD	nd	16S
2007 11 - Summer	6	BBD	nd	16S
2007 11 - Summer	7	BBD	nd	16S
2007 12 - Summer	8	BBD	nd	16S
2008 02 - Summer	9	BBD	*gp23*	16S
2008 02 - Summer	10	BBD	*gp23*	16S
2008 02	1	Healthy-tissue	*gp23*	16S
2008 02	2	Healthy-tissue	*gp23*	16S
2008 02	3	Healthy-tissue	*gp23*	16S
2008 02	4	Healthy-tissue	*gp23*	16S
2012 07	1	Seawater-BBD	*gp23*	nd
2012 08	2	Seawater-BBD	*gp23*	nd
2013 05	3	Seawater-BBD	*gp23*	nd
2014 05	4	Seawater-BBD	*gp23*	16S
2014 06	5	Seawater-BBD	*gp23*	16S
2014 10	6	Seawater-BBD	*gp23*	16S
2015 04	7	Seawater-BBD	*gp23*	nd
2015 02	8	Seawater-BBD	*gp23*	16S
2012 02	1	Seawater-Control	*gp23*	nd
2014 05	2	Seawater-Control	*gp23*	16S
2014 06	3	Seawater-Control	*gp23*	16S
2014 10	4	Seawater-Control	*gp23*	16S
2015 02	5	Seawater-Control	*gp23*	16S
2015 04	6	Seawater-Control	*gp23*	16S

Appendix 2.2 PCR primers. Illumina adapters are printed bold, template specific printed with underline.

Illumina adapter + T4 bacteriophage *gp23* primer	
MZIA6	
5' **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCG**CGGTTGATTTCCAGCATGATTTC	3'
MZIA1bis	
5' **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GATATTTGIGGIGTTCAGCCIATGA	3'
Illumina adapter + bacterial 16S rRNA primer	
Bakt_341F	
5' **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**CCTACGGGNGGCWGCAG	3'
Bakt_805R	
5' **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**GACTACHVGGGTATCTAATCC	3'

Appendix 2.3 PCR protocol. T4 virus community (gene gp23) was amplified in a 1st PCR with regular, locus specific primers (no Illumina adapters) in a touchdown protocol. PCR products were purified with Qiaquick gel purification. Illumina adapters were added in a 2nd PCR with low amplification cycles. Bacteria communities (16S rRNA gene) were amplified with a locus specific PCR and a subsequent PCR to attached Illumina MiSeq adapters.

PCR step	Cycle	Temperature [°C]	Time	No of cycles
1st PCR	Denaturation	95	10 min	
Touchdown	Annealing	60 -> 45	30 sec	15
	Extension	72	45 sec	
gp23	Denaturation	95	1 min	
	Annealing	45	30 sec	20
	Extension	72	45 sec	
	Final extension	72	7 min	
Gel purification with Qiaquick				
2nd PCR	Denaturation	95	1 min	
	Annealing	50	30 sec	15
	Extension	72	45 sec	
	Final extension	72	7 min	

PCR step	Cycle	Temperature [°C]	Time	No of cycles
1st PCR	Denaturation	95	5 min	
	Denaturation	95	30 min	
16S rRNA	Annealing	55	30 sec	30
	Extension	72	30 sec	
	Final extension	72	7 min	
Gel purification with Qiaquick				
2nd PCR	Denaturation	95	1 min	
	Annealing	55	30 sec	15
	Extension	72	45 sec	
	Final extension	72	7 min	

Appendix 2.4 16S rRNA bacteria and T4 bacteriophage composition. Raw data QIIME analysis. K = kingdom; c = class; o = order; f = family; g = genus. Values in [%]. To simplify data, this table does not include OTUs that appeared only once among samples as well as OTUs that had an overall abundance of <0.1%.

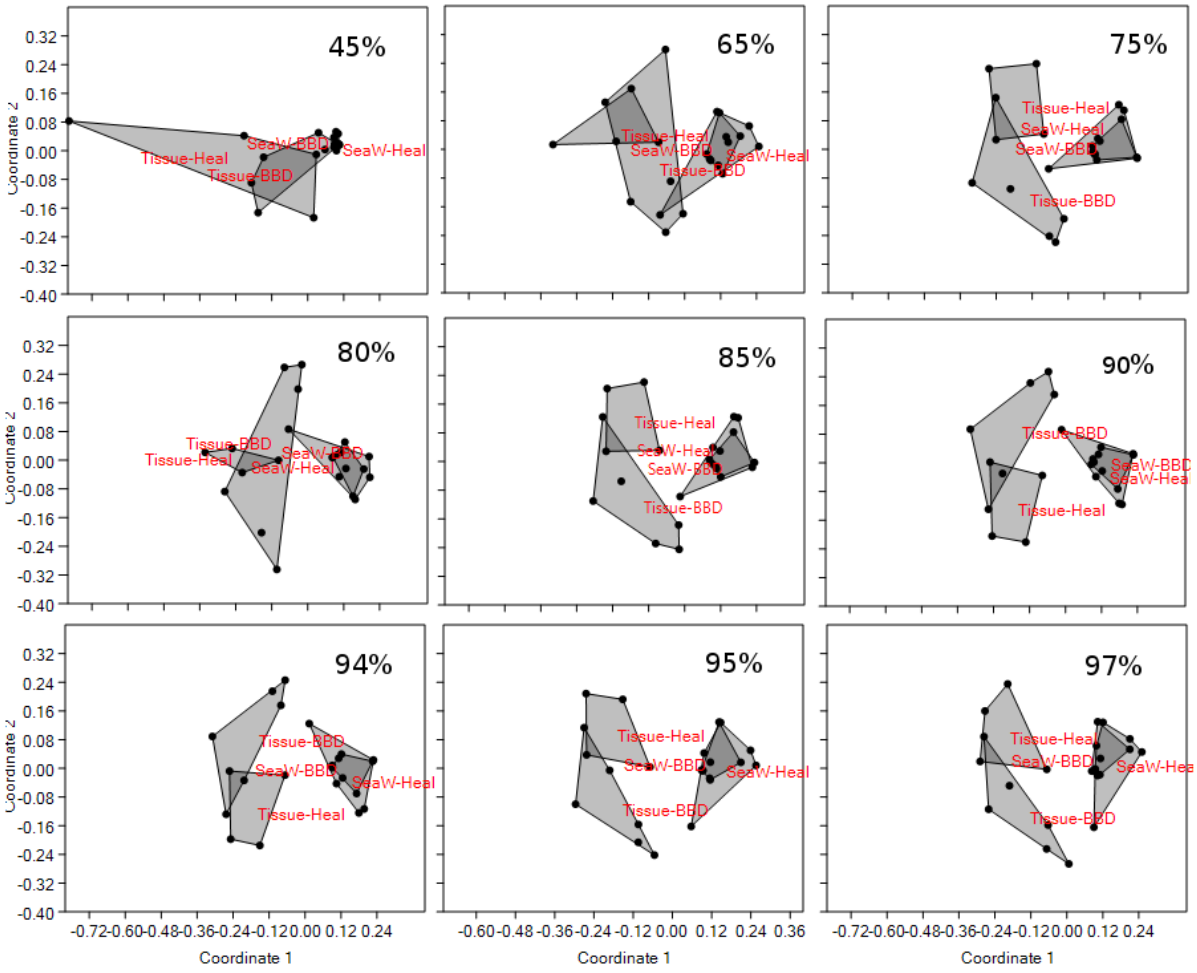
```
# START
############################################################
#####################  QC PIPELINE  ########################
############################################################

####__Unzip files_____________________________
# gunzip input > output
gunzip 1_S1_L001_R1_001.fastq.gz > 1_R1.fastq

####_QC on raw files_____________________________
fastqc -f fastq FILENAME
# examine data

####__Merge reads_____________________________
# Install PEAR
# http://sco.h-its.org/exelixis/web/software/pear/doc.html
sudo apt-get install build-essential autoconf automake libtool
git clone https://github.com/xflouris/PEAR.git
cd PEAR
./autogen.sh
./configure
make
sudo make install
#### installed PEAR ####
# Merge reads
# pear -f LEFTfile -r RIGHTfile -o outputfile.fastq
time pear -f 1_S1_L001_R1_001.fastq -r 1_S1_L001_R2_001.fastq -o 1_Merged.fastq
time pear -f 2_S2_L001_R1_001.fastq -r 2_S2_L001_R2_001.fastq -o 2_Merged.fastq
#### note: merging reads after QC not possible: sequences do not overlap. ####

####_QUALITY BASED TRIMMING_____________________________
fastq_quality_filter -v -q 30 -p 95 -i FILE -o FILE

####_Cut primers_____________________________
# check with "head FILENAME" that primer sequences are on each side of the sequence.
# each primer is 25 bp long, so cut on each side 25 base pairs to cut out the primer sequences
# quality encoding: -Q = First base to keep .-f = Minimum length to kep. -m = clip bp at end of
read. -t = short summary
# clip first 25 bp from all sequences, forward and reverse primer cut together
cat FILE.fastq | fastx_trimmer -v -Q 33 -f 25 -m 70 | fastx_trimmer -v -Q 33 -t 25 -m 70 >
FILEnoprimer.fastq

####_Minimum length of 100 bp, collapse files_____________________________
# java -jar trimmomatic-0.32.jar SE 3_collapsed.fasta 3_collapsed.fasta MINLEN:XY
java -jar path.to/Trimmomatic-0.32/trimmomatic-0.32.jar SE FILE.fastq FILE.min100.fastq
MINLEN:100 && fastx_collapser -v -i FILE.min100.fastq -o FILE.min100.fasta

#####
# summary example for sample:
# Before Pear:
# 955,641 for each file of the paired ends
#
# After Pear:
```

```
# 841,979 merged reads (450,619 + 391,360 from the merged file split into two, QC did not
work on full file)
#
# Into QC trim pipeline
# Input: 841,979 sequences
# OutputL: 382,389 sequences
#
# fastx_collapser:
# Input: 382389 sequences (representing 382389 reads)
# Output: 94185 sequences (representing 382389 reads)
#
# Total leftover at the end: roughly 10 %
#####


############################################################
#####################  QC PIPELINE  ########################
############################################################
# END

# START
############################################################
#####################  QIIME  PIPELINE  ####################
############################################################
# Further explanations, see QIIME tutorials:
# http://qiime.org/
# http://qiime.org/tutorials/tutorial.html


###_Make mapping file_____________________________
#according to the following scheme in a text file:
#SampleID       BarcodeSequence         LinkerPrimerSequence   InputFileName   Description

SeaXY.1                                 SeaW-T4-BBD-MMYY_collapsed.fasta    SeaXY.1

###_Add QIIME labels_____________________________
add_qiime_labels.py -v -i ~/path.to/FILE.collapsed.fasta -m ~/path.to/mapping.txt -c
InputFileName -o ~/path.to/

###_De novo picking_____________________________
pick_otus.py -i ~/path.to/combined_seqs.fna -o ~/path.to/pick_otus_97_percent_rev/ -m uclust -t
-s 0.97 -z
#
make_otu_table.py -i ~/path.to/pick_otus_97_percent_rev/combined_seqs_otus.txt -o
~/path.to/pick_otus_97_percent_rev/otu_table.biom
#
group_significance.py -i ~/path.to/pick_otus_97_percent_rev/otu_table.biom -m
~/path.to/mapping.txt -c Treatment -s kruskal_wallis -o
~/path.to/pick_otus_97_percent_rev/kw_ocs.txt --biom_samples_are_superset
--print_non_overlap

###_Taxonomy (only 16S rRNA)_____________________
pick_rep_set.py -i otus/combined_seqs_otus.txt -f combined_seqs.fna -o rep_set1.fna
#
assign_taxonomy.py -i ~/path.to/pick_otus_97_percent_rev/rep_set97.fna -r
~/path.to/Greengenes/gg_13_5_otus/rep_set/97_otus.fasta -t
~/path.to/Greengenes/gg_13_5_otus/taxonomy/97_otu_taxonomy.txt -o
```

```
~/path.to/pick_otus_97_percent_rev/

###_Sequence alignment_____________________________
align_seqs.py -i ~/path.to/pick_otus_97_percent_rev/rep_set97.fna -m mafft -o
~/path.to/pick_otus_97_percent_rev/mafft_alignment/
#
filter_alignment.py -i
~/path.to/pick_otus_97_percent_rev/pynast_alignment/rep_set97_aligned.fasta -o
~/path.to/pick_otus_97_percent_rev/pynast_alignment/filtered_alignment/

###_Alpha_rarefaction_____________________________
make_phylogeny.py -i
~/path.to/pick_otus_97_percent_rev/pynast_alignment/filtered_alignment/rep_set97_aligned_pfi
ltered.fasta -o ~/path.to/pick_otus_97_percent_rev/rep_phylo.tre
#
alpha_rarefaction.py -i ~/path.to/pick_otus_97_percent_rev/otu_table.biom -o
~/path.to/pick_otus_97_percent_rev/arare_max99/ -f -m ~/path.to/mapping.txt -p
~/path.to/alpha_params.txt -t ~/path.to/pick_otus_97_percent_rev/rep_phylo.tre


################################################################
#################### QIIME PIPELINE ############################
################################################################
# END
```

Appendix 2.5 Operational taxon unit tables of 16S rRNA and T4 bacteriophage communities. Full tables are not included in the thesis due to their length, please access the the data through the website below.

OTU tables
16S rRNA and T4 bacteriophage OTU tables **https://figshare.com/s/3959e20f4f9137a07c93**



Appendix 2.6 T4 bacteriophage sequence similarity tests. Sequence similarity threshold was chosen based on most distinct community patterns among sample groups. Overall OTU abundance was converted into relative abundances, transformed into a Bray-Curtis matrix and visualised in an MDS plot.

Appendix 2.7 Tetranucleotide frequencies. Script *TetraNUCL-script.sh* combines all sequences from respective file of T4 data set into one file – just nucleotides in the file, no sequence identifiers. The perl script run the nucleotide frequencies on the data and saves it as the file: *tetranucs_out.txt.* All frequencies have been combined into a spreadsheet matrix and transformed into percentage values.

Terminal: sh TetraNUCL-script.sh

File: TetraNUCL-script.sh

```
# delete sequence names by combining all nucleotides in one file, run perl script
#!/bin/bash
for C in Sample-names
do
   grep -v '^>' ""$C".fasta" > ""$C"-REDUCED.fasta"
         mkdir ""$C"-REDUCED"
         cp nuclfreq.pl ""$C"-REDUCED"
         cp ""$C"-REDUCED.fasta" ""$C"-REDUCED"
         cd ""$C"-REDUCED"
         perl nuclfreq.pl ""$C"-REDUCED.fasta"
         cd ..
done
```

PERL SCRIPT:

Terminal: see TetraNUCL-script.sh script

File: nuclfreq.pl

```
# Source: http://alrlab.research.pdx.edu/aquificales/bioinformatics_scripts.html
#!/usr/bin/perl
################################################################################
########################
### Get Tetranucleotide Frequencies
### Usage: get_tetranucleotide_frequencies.pl <fasta file>
### This program takes a fasta file as it's first (and only) parameter.
###
### It returns a tab delimited file (tetranucs_out.txt)
###
### Jennifer Meneghin
### July 31, 2012
################################################################################
########################
#-------------------------------------------------------------------------------------------------------------------------
#Deal with passed parameters
#-------------------------------------------------------------------------------------------------------------------------
if ($#ARGV == -1) {
   usage();
   exit;
}
$fasta_file = $ARGV[0];
$out_file = "tetranucs_out.txt";
unless ( open(IN, "$fasta_file") ) {
   print "Got a bad fasta file: $fasta_file\n\n";
   exit;
}
unless ( open(OUT, ">$out_file") ) {
   print "Couldn't create $out_file\n";
   exit;
}
print "Parameters:\nfasta file = $fasta_file\noutput file = $out_file\n\n";
```

```
#-----------------------------------------------------------------------------------------------------------------------
#The main event
#-----------------------------------------------------------------------------------------------------------------------
$seq = "";
while (<IN>) {
    chomp;
    if (/^>/) {
            #finish up previous line.
            if (length($seq) > 0) {
                &process_it;
            }
            #start new line.
            $id = $_;
            $id =~ s/^>(.+?)\s.+$/$1/g;
            print "ID = $id\n";
    }
    else {
            $seq = $seq . uc($_);
    }
}

#finish up last line.
&process_it;

print "Sorting...";

%fourmers;
%records;
for $i (sort keys %tetranucs) {
    @parts = split(/\t/, $i);
    $record = $parts[0];
    $fourmer = $parts[1];
    if ($fourmers{$fourmer}) {
            $fourmers{$fourmer} = $fourmers{$fourmer} + 1;
    }
    else {
            $fourmers{$fourmer} = 1;
    }
    if ($records{$record}) {
            $records{$record} = $records{$record} + 1;
    }
    else {
            $records{$record} = 1;
    }
}

print "Printing...";

print OUT "Tetranucleotide";
for $j (sort keys %records) {
    print OUT "\t$j";
}
print OUT "\n";
for $i (sort keys %fourmers) {
    print OUT "$i";
    for $j (sort keys %records) {
            $key = $j . "\t" . $i;
```

```
            if ($tetranucs{$key}) {
                print OUT "\t$tetranucs{$key}";
            }
            else {
                print OUT "\t0";
            }
    }
    print OUT "\n";
}

close(IN);
close(OUT);

sub usage {
    print "Get Tetranucleotide Frequencies\n";
    print "Usage: get_tetranucleotide_frequencies.pl <fasta file>\n";
    print "This program takes a fasta file as it's first (and only) parameter.\n\n";
    print "It returns a tab delimited file (tetranucs_out.txt) of tetranucleotides. (columns = records,
rows = tetranucleotide counts.)\n\n";
    print "Jennifer Meneghin\n";
    print "July 31, 2012\n\n";
}

sub process_it {
    @letters = split(//, $seq);
    for $i (0..$#letters-3) {
            $tetra = $letters[$i] . $letters[$i+1] . $letters[$i+2] . $letters[$i+3];
            $key = $id . "\t" . $tetra;
            if ($tetranucs{$key}) {
                $tetranucs{$key} = $tetranucs{$key} + 1;
            }
            else {
                $tetranucs{$key} = 1;
            }
    }
    $seq = "";
    $id = "";
}

#
R-SCRIPT for OLIGO NUCLEOTIDE FREQUENCY ANALYSIS
IN R:
# Source: http://cc.oulu.fi/~jarioksa/softhelp/vegan/html/betadisper.html
library(vegan)

# load data, needs to be in columns,,
nuclfreqs <- read.csv("~/nuclfreqs.csv", header = T, sep = ",", strip.white = T)

str(nuclfreqs)

# Bray-Curtis distances between samples
dis <- vegdist(nuclfreqs)

# Data (each sample) needs to be in rows, respective frequencies in columns, define replicates
with following command:
groups <- factor(c(rep(1,5), rep(2,5), rep(3,5), rep(4,5), rep(5,7), rep(6,5), rep(7,4), rep(8,5)),
labels = c("Tissue-BBD-gp23","Tissue-Healthy-gp23","Tissue-BBD-16S", "Tissue-Healthy-16S",
```

```
"SeaW-BBD-gp23", "SeaW-Control-gp23", "SeaW-BBD-16S", "SeaW-Control-16S"))
"

## Calculate multivariate dispersions
mod <- betadisper(dis, groups)
mod

## Perform test
anova(mod)

## Permutation test for F - RESULT
permutest(mod, pairwise = TRUE)

## Tukey's Honest Significant Differences
(mod.HSD <- TukeyHSD(mod))
plot(mod.HSD)

## Plot the groups and distances to centroids on the
## first two PCoA axes
plot(mod)

## Draw a box plot of the distances to centroid for each group
boxplot(mod)


RESULTS:
> permutest(mod, pairwise = TRUE)

Permutation test for homogeneity of multivariate dispersions
Permutation: free
Number of permutations: 999

Response: Distances
           Df   Sum Sq    Mean Sq      F  N.Perm    Pr(>F)
Groups      7 0.041473 0.0059247 6.9585   999     0.001 ***
Residuals  33 0.028097 0.0008514
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Pairwise comparisons:
(Observed p-value below diagonal, permuted p-value above diagonal)
```

	BBD-T4-Tss	Heal-T4-Tss	BBD-16S-Tss	Heal-16S-Tss	BBD-T4-SW	Control-T4-SW	BBD-16S-SW	Control-16S-SW
BBD-T4-Tss		0.77	0.029	0.699	0.011	0.077	0.025	0.01
Heal-T4-Tss	0.74382		0.001	0.437	0.004	0.025	0.002	0.003
BBD-16S-Tss	0.0413413	0.0074043		0.013	0.852	0.019	0.083	0.056
Heal-16S-Tss	0.6670737	0.3923101	0.0132514		0.005	0.02	0.007	0.003
BBD-T4-SW	0.0177941	0.0024908	0.8330091	0.0040889		0.06	0.069	0.027
Control-T4-SW	0.0870427	0.0265895	0.0291038	0.0285004	0.0703261		0.006	0.001
BBD-16S-SW	0.0363573	0.0070904	0.1036929	0.014373	0.0747808	0.0131679		0.862
Control-16S-SW	0.0176501	0.0026809	0.0644268	0.00595	0.0399346	0.0070085	0.8545311	

```
>
> ## Tukey's Honest Significant Differences
> (mod.HSD <- TukeyHSD(mod))
  Tukey multiple comparisons of means
```

95% family-wise confidence level

Fit: aov(formula = distances ~ group, data = df)

$group

Appendix 2.8 Nucleotide frequencies, adjusted p-values. Adjusted p-values are taken from the Tukey's honest significant differences R-script, see lines above.

Comparison	difference	lower	upper	**p-adjusted**
Heal-T4-Tss-BBD-T4-Tss	-0.009291181	-0.06895097	0.0503686069	0.9995652
BBD-16S-Tss-BBD-T4-Tss	-0.058404927	-0.11806471	0.0012548610	0.0585377
Heal-16S-Tss-BBD-T4-Tss	0.014794989	-0.04486480	0.0744547774	0.9918471
BBD-T4-SW-BBD-T4-Tss	**-0.057353794**	**-0.11258802**	**-0.0021195630**	**0.0372489**
Control-T4-SW-BBD-T4-Tss	-0.047039051	-0.10669884	0.0126207370	0.2110588
BBD-16S-SW-BBD-T4-Tss	**-0.072015906**	**-0.13529467**	**-0.0087371448**	**0.0167203**
Control-16S-SW-BBD-T4-Tss	**-0.073929851**	**-0.13358964**	**-0.0142700629**	**0.0071039**
BBD-16S-Tss-Heal-T4-Tss	-0.049113746	-0.10877353	0.0105460420	0.1707466
Heal-16S-Tss-Heal-T4-Tss	-.024086170	-0.03557362	0.0837459584	0.8904223
BBD-T4-SW-Heal-T4-Tss	-0.048062613	-0.10329684	0.0071716181	0.1261276
Control-T4-SW-Heal-T4-Tss	-0.037747870	-0.09740766	0.0219119180	0.4688177
BBD-16S-SW-Heal-T4-Tss	-0.062724725	-0.12600349	0.0005540362	0.0534089
Control-16S-SW-Heal-T4-Tss	**-0.064638670**	**-0.12429846**	**-0.0049788819**	**0.0260877**
Heal-16S-Tss-BBD-16S-Tss	**0.073199916**	**0.01354013**	**0.1328597044**	**0.0078946**
BBD-T4-SW-BBD-16S-Tss	0.001051133	-0.05418310	0.0562853640	1.0000000
Control-T4-SW-BBD-16S-Tss	0.011365876	-0.04829391	0.0710256640	0.9983983
BBD-16S-SW-BBD-16S-Tss	-0.013610979	-0.07688974	0.0496677822	0.9965671
Control-16S-SW-BBD-16S-Tss	-0.015524924	-0.07518471	0.0441348641	0.9891661
BBD-T4-SW-Heal-16S-Tss	**-0.072148783**	**-0.12738301**	**-0.0169145523**	**0.0039518**
Control-T4-SW-Heal-16S-Tss	**-0.061834040**	**-0.12149383**	**-0.0021742524**	**0.0378120**
BBD-16S-SW-Heal-16S-Tss	**-0.086810895**	**-0.15008966**	**-0.0235321342**	**0.0021997**
Control-16S-SW-Heal-16S-Tss	**-0.088724840**	**-0.14838463**	**-0.0290650523**	**0.0007697**
Control-T4-SW-BBD-T4-SW	0.010314743	-0.04491949	0.0655489737	0.9985897
BBD-16S-SW-BBD-T4-SW	-0.014662112	-0.07378682	0.0444625998	0.9918477
Control-16S-SW-BBD-T4-SW	-0.016576057	-0.07181029	0.0386581738	0.9756048
BBD-16S-SW-Control-T4-SW	-0.024976855	-0.08825562	0.0383019062	0.9012094
Control-16S-SW-Control-T4-SW	-0.026890800	-0.08655059	0.0327689881	0.8236970
Control-16S-SW-BBD-16S-SW	-0.001913945	-0.06519271	0.0613648159	1.0000000

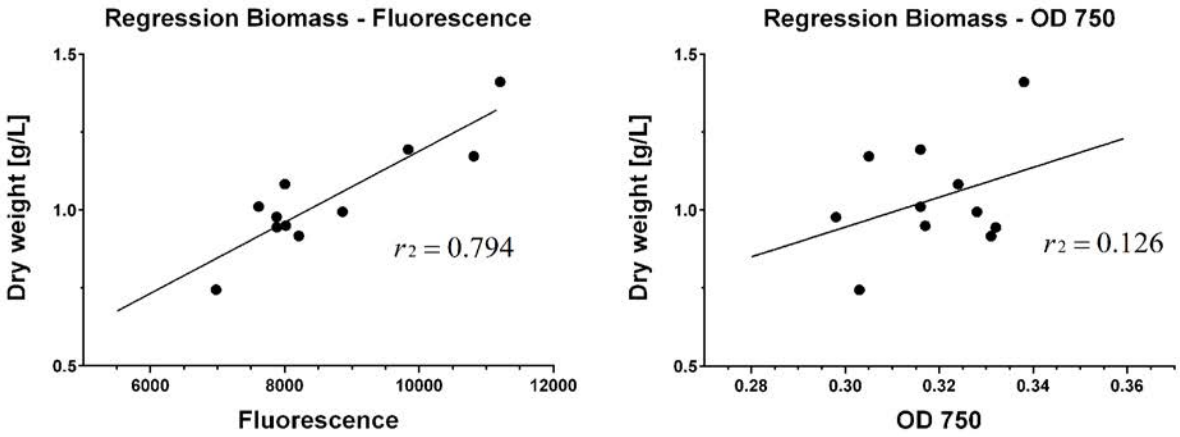
Appendix - Chapter 3

Methods development

Appendix 3.1 Media recipes for L1, F/2, ASNIII and IMK.

	L1	F/2	ASN III	IMK
MACRO-NUTRIENTS [g/L]				
$NaNO_3$	75	75	0.75	200
$NaH_2PO_4 \cdot H_2O$	5	5	-	
$Na_2SiO_3 \cdot 9H_2O$	30	30	-	
$NaHPO_4$	-	-	-	1.4
K_2HPO_4	-	-	-	5
NH_4Cl	-	-	-	2.68
NaCl	-	-	24.95	-
$MgCl_2 \cdot 7H_2O$	-	-	2	-
Kcl	-	-	0.5	-
$K_2HPO_4 \cdot 3H_2O$	-	-	0.02	-
$MgSO_4 \cdot 7H_2O$	-	-	3.5	-
$CaCl_2 \cdot 2H_2O$	-	-	0.5	-
disodium EDTA	-	-	0.5	-
Citric-acid	-	-	0.003	-
$Fe(II)NH_2$+citrate	-	-	0.003	-
Na_2CO_3	-	-	0.02	-
TRACEMETALS [g/L for 1000x stock]				
H_3BO_3	-	-	2.86	-
Fe-EDTA	-	-	-	5.2
Mn-EDTA	-	-	-	0.332
$FeCl_3 \cdot 6H_2O$	3.15	3.15	-	-
Na_2-EDTA	-	-	-	37.2
Na_2-EDTA$\cdot 2H_2O$	4.36	4.36	-	-
$CuSO_4 \cdot 5H_2O$	0.25	1	0.079	0.0025
$Na_2MoO_4 \cdot 2H_2O$	3	1	0.39	0.0073
$ZnSO_4 \cdot 7H_2O$	1	1	0.22	0.023
$CoCl_2 \cdot 6H_2O$	1	1	0.04	-
$MnCl_2 \cdot 4H_2O$	1	1	1.81	-
H_2SeO_3	**1**	**-**	**-**	**0.0017**
$NiSO_4 \cdot 6H_2O$	**1**	**-**	**-**	**-**
Na_3VO_4	**1**	**-**	**-**	**-**
K_2CrO_4	**1**	**-**	**-**	**-**
VITAMINS [g/L for 1000x stock]				
Vitamin B_{12}	1	1	-	0.0015
Biotin	0.1	0.1	0.1	0.0015
Thiamine HCl	0.2	0.2	0.2	0.2
Cyanocobalamin	-	-	1	-
$MnCl_2 \cdot 4H_2O$	-	-	-	0.18

Appendix 3.2 Correlation between fluorescence, OD and biomass (dry weight in g/L).



Appendix 3.3 Regression line details of agar comparison.

Agar [%]	Slope	Different from zero [p-value]
0.6	27,732 ± 7,406	0.0072
1	5,285 ± 2,763	0.0974
1.5	212.8 ± 135	0.1589

Appendix 3.4 ANOVA details of regression comparison of agar concentrations. Relevant post-hoc multiple comparisons are written in the results section of the main article.

	SS	DF	MS	F (Dfn, DFd)	P value
Treatment (between agar conc.)	$8.579 * 10^{-8}$	2	$4.290 * 10^{8}$	F (2, 3) = 20.59	P = 0.0177
Residual (within agar conc.)	$6.250 * 10^{-7}$	3	$2.083 * 10^{7}$		
Total	$9.204 * 10^{-8}$	5			

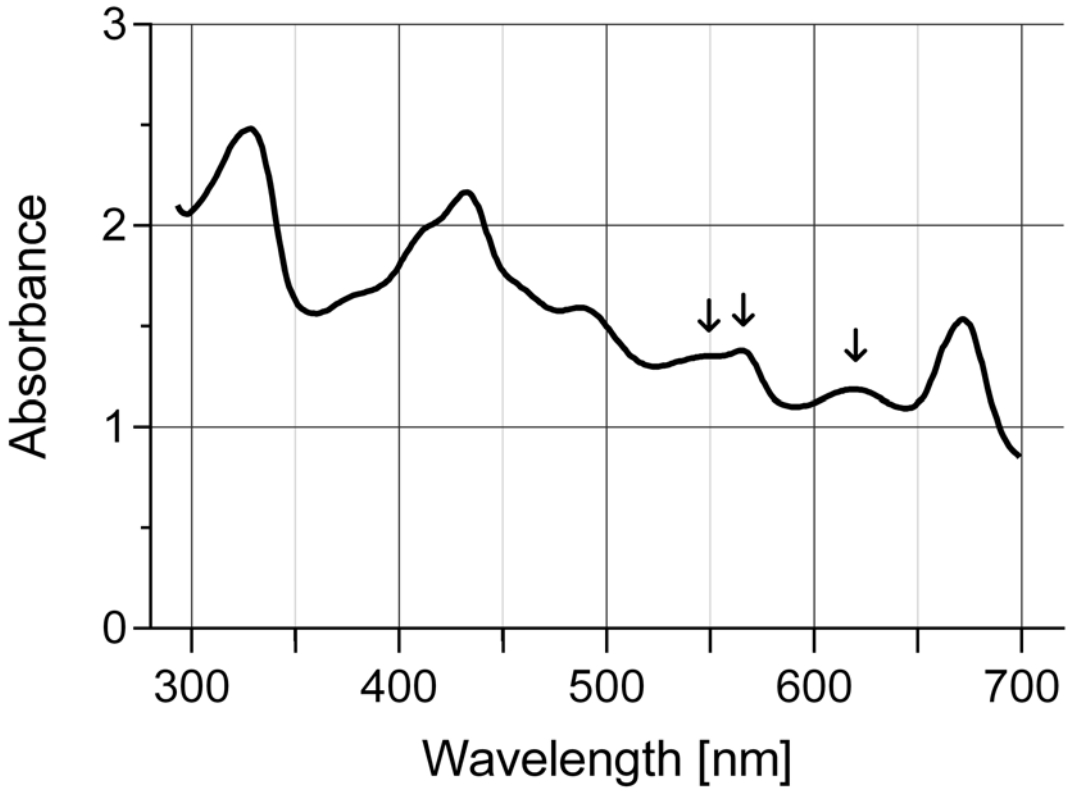
Appendix 3.5 Regression line details of liquid medium comparison.

Medium	Slope	Different from zero [p-value]
ASNIII	0.0326 ± 0.0033	< 0.001
IMK	-0.0059 ± 0.0017	0.0138
L1	0.1215 ± 0.0075	0.0037
F/2	-0.0009 ± 0.0017	0.6418

Appendix 3.6 ANOVA details of regression comparison of liquid medium comparisons. Relevant post-hoc multiple comparisons are written in the results section of the main article.

	SS	DF	MS	F (Dfn, DFd)	P value
Treatment (between media)	0.021	3	0.0696	F (3, 4) = 381.7	P < 0.0001
Residual (within medium)	$7.3 * 10^{-5}$	4	0.00015		
Total	0.021	7			

Appendix 3.1 My cyanobacterium culture was not axenic, since I identified a second taxon in the 16S rRNA sequences of the cyanobacterium cultures (KU720414) which matched an *Alphaproteobacterium* MBIC3865 (AB015896.1, 99% identity, e-value = 0). Cyanobacteria are known to form close associations with epibiotic bacteria, including members of the Alphaproteobacteria (Hube et al. 2009; Praveen Kumar et al. 2009). It was therefore not surprising to find this bacterium in our culture. Achieving and maintaining an axenic state of a *Roseofilum* culture may not be possible, because it may rely on the association and presence of symbiotic bacteria (Richardson et al. 2014).



Appendix 3.8 Cell mass absorption spectrum of cyanobacterial culture in liquid medium. Representative peaks for phycoerythrin (548 and 565 nm) and phycocyanin (620 nm) are marked with arrows in the graph. Measurements and growth in L1 medium.

Appendix - Chapter 4

Phage therapy and bacteriophage genomes

Appendix 4.1 Assembly details of the draft genomes of bacteriophages infecting bacteria in the culture of *Roseofilum reptotaenium* AO1. Statistics of bioinformatic data analyses such as N50 values, sequence length and GC content were generated with gnx-tools (software version 0.1+20120305).

Parameter	Assembly details
Sequencing library	Nextera XT library
Sequencing platform	MiSeq 2x300 V3
Assembly software	CLC genomics workbench version 8.5.1
K-mer length	automatic
Total number of sequences	882
Total length [bp]	3,686,000
Shortest sequence [bp]	837
Longest sequence [bp]	148,780
Total number of Ns	1,152
N50	6,118

Appendix 4.2 Annotation details of largest contigs 1 – 3. ORF = Open reading frames. Q. cover = query coverage. Accession number refers to the NCBI database entry.

Contig #	ORF	ORF Size	ORF start	ORF end	+-	q. cover	*E*-value	Function	Annotation	Query title	Accession #
Conitg1	1	3863	3	3866		1320	2E-022	Lysis	SGNH hydrolase	*Cellulophaga* phage phi13:1	AGO49035.1
	2	2978	3842	6820		779	3E-061	Structure and assembly	structural protein	*Cellulophaga* phage phi17:2	YP_008241543.1
	3	1529	6766	8295		1040	5E-070	Structure and assembly	tail fiber	*Cellulophaga* phage phi4:1	YP_008240634.1
	4	716	8369	9085		674	8E-081	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240633.1
	5	479	9078	9557		464	7E-032	Hypothetical protein	hypothetical protein Phi4:1_gp044	*Cellulophaga* phage phi4:1	YP_008240632.1
	6	896	9530	10426		866	2E-099	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240631.1
	7	6020	10423	16443		5534	0	Structure and assembly	structural protein	*Cellulophaga* phage phi17:2	YP_008241538.1
	10	2630	17501	20131		1613	7E-159	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240628.1
	11	2948	20132	23080		2930	0	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240627.1
	12	6224	23031	29255		764	3E-038	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240625.1
	14	1424	29634	31058		1418	2E-069	Structure and assembly	structural protein	Cyanophage Syn5	YP_001285462.1
			34992	35648				Metabolism	gp55	*Synechococcus* Syn5	NC_009531
	19	1184	35878	37062		698	3E-044	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240625.1
	22	5690	37707	43397		5645	0	Structure and assembly	structural protein	*Cellulophaga* phage phi17:2	YP_008241531.1
	24	1856	43974	45830		1751	2E-091	Hypothetical protein	hypothetical protein SCRM01_174	*Synechococcus* phage S-CRM01	YP_004508607.1
	28	1043	46683	47726		782	2E-019	Metabolism	RNA ligase	*Clavibacter* phage CN1A	YP_009004237.1
	31	758	48431	49189		662	6E-027	Lysis	metallophosphatase	*Cellulophaga* phage phi4:1	YP_008240606.1
	36	620	50478	51098		617	9E-058	Hypothetical protein	hypothetical protein Phi17:2_gp033	*Cellulophaga* phage phi17:2	YP_008241528.1
	37	773	51058	51831		722	4E-060	Metabolism	tRNA nucleotidyltransferase poly(A) polymerase	*Cellulophaga* phage phi17:2	YP_008241527.1
			51834	52352			2E-005		putative HD-domain/PDEase-like protein	*Caulobacter* phage rogue	NC_019408
	41	1616	53437	55053	c	1472	1E-051	Structure and assembly	gp05	Phage phiJL001 ID=279383	YP_223929.1
	42	950	55180	56130	c	917	7E-058	Metabolism	virulence-associated protein E (VirE)	*Cellulophaga* phage phi17:1	YP_008241342.1
	44	2333	56618	58951	c	1148	1E-026	Metabolism	DNA polymerase	*Burkholderia* phage AH2	YP_006561157.1
			59066	59515	c		3E-015	Metabolism	putative ribonuclease H like protein – CRISPR spacer target	*Pseudomonas* phage PaMx25	ALH23775.1
	47	1052	60588	61640	c	953	2E-058	Metabolism	replicative helicase, DnaB family	*Cellulophaga* phage phi4:1	YP_008240785.1
			63182	63850			2E-005	Metabolism	gp47, Recombination endonuclease	*Burkholder* phage phi1026b	NC_005284
	53	1754	64139	65893	c	1652	2E-099	Metabolism	ribonucleoside-diphosphate reductase alpha subunit	Cyanophage S-TIM5	YP_007006137.1
	56	1073	67292	68365	c	1064	6E-118	Metabolism	thymidylate synthase	*Cellulophaga* phage phi14:2	YP_008242256.1
			68359	69402			9E-008		putative ATPase	*Salmonella* phage 7_11	NC_015938
	61	1001	70661	71662	c	974	1E-060		outer-membrane protein OmpA domain	*Cellulophaga* phage phi17:2	YP_008241666.1
	64	728	72534	73262	c	692	6E-069	Hypothetical protein	hypothetical protein Phi4:1_gp161	*Cellulophaga* phage phi4:1	YP_008240749.1
	66	980	73677	74657	c	929	2E-018	Metabolism	3'-5' exonuclease pol-B	*Cellulophaga* phage phi17:2	YP_008241659.1
	69	833	75961	75128	c	175	2E-021		Ferric iron ABC transporter, ATP-binding protein	NA	Fig|6666666.235665.peg.75
	70	296	76420	76716	c	272	1E-012	Hypothetical protein	hypothetical protein	*Staphylococcus* phage SA1 ID=694060	ACZ55601.1
	72	770	77703	76933	c	252	5E-039		ABC-type nitrate/sulfonate/ bicarbonate transporter	NA	Fig|6666666.235665.peg.79
	74	1811	80099	78288	c	101	6E-019		Outer membrane lipoprotein omp16 precursor	NA	Fig|6666666.235665.peg.81
	76	536	80181	80717	c	467	2E-032	Hypothetical protein	macro domain containing protein	*Cellulophaga* phage phi4:1	YP_008240590.1

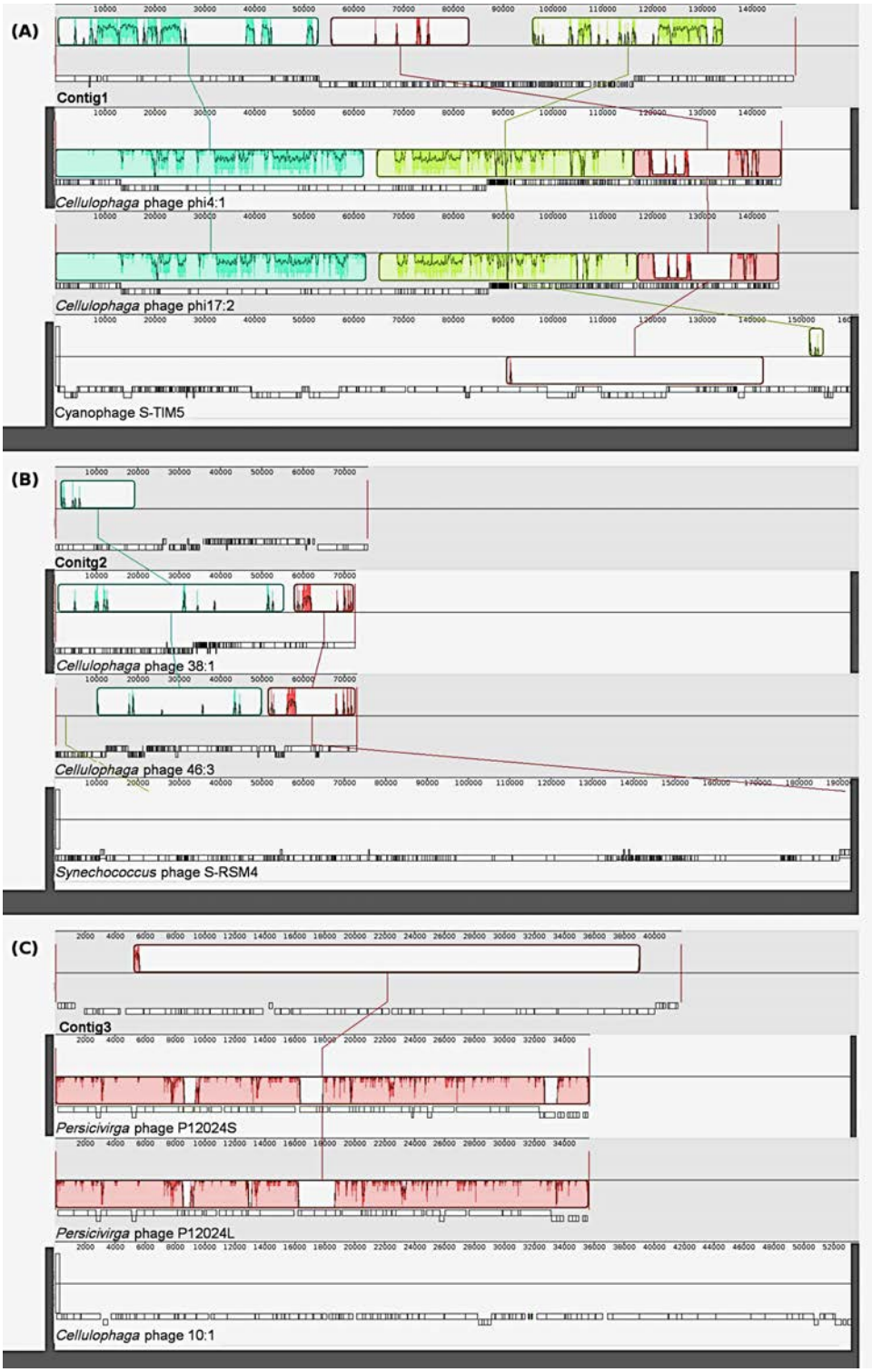
	77	1388	80663	82051	c	1364	3E-059	Metabolism	putative nicotinate phosphoribosyltransferase	*Caulobacter* phage CcrRogue	YP_006989188.1
			82038	82457			2E-009	Structure and assembly	putative phage head tail adapter	*Enterobacter* phage EFDG1	NC_029009
	84	272	85664	85936	c	200	2E-007		plastocyanin	*Salmonella* phage 100268_sal2	gi100187, NC_031902
	101	494	94218	94712	c	416	2E-008	Hypothetical protein	hypothetical protein	*Pseudomonas* phage KPP10 ID=582345	YP_004306843.1
	104	500	95949	96449	c	482	2E-041		DnaJ domain containing protein	*Cellulophaga* phage phi4:1	YP_008240708.1
	119	416	102790	103206	c	251	0.0003	Hypothetical protein	hypothetical protein	*Pseudomonas* phage LUZ24 ID=484895	YP_001671911.1
	120	467	103187	103654	c	440	9E-052		NAD synthetase	*Cellulophaga* phage phi4:1	YP_008240677.1
	121	992	103734	104726	c	788	2E-020	Metabolism	recA recombinase	*Microcystis* phage Ma-LMM01	YP_851022.1
	124	464	105428	105892	c	425	4E-035	Hypothetical protein	hypothetical protein Phi4:1_gp085	*Cellulophaga* phage phi4:1	YP_008240673.1
	126	1454	106085	107539	c	1430	0	Metabolism	helicase	*Cellulophaga* phage phi17:2	YP_008241576.1
			109014	109211			2E-009		gp269	*Bacillus* sp. Phage	gi593777725, NC_023719
	129	536	109485	110021	c	485	3E-017	Metabolism	HNH endonuclease	Cyanophage S-TIM5	YP_007006139.1
			110417	110707			3E-010		peptidyl tRNA hydrolase domain protein	*Mycobacterium* phage ArcherS7	NC_021348
	140	437	116225	116662		353	2E-030	Hypothetical protein	hypothetical protein Phi4:1_gp073	*Cellulophaga* phage phi4:1	YP_008240661.1
	145	2693	118371	121064		2573	9E-146	Hypothetical protein	hypothetical protein Phi4:1_gp067	*Cellulophaga* phage phi4:1	YP_008240655.1
	146	2813	121091	123904		2774	0	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240654.1
	147	1157	123935	125092		884	4E-068	Hypothetical protein	hypothetical protein Phi4:1_gp065	*Cellulophaga* phage phi4:1	YP_008240653.1
	148	1721	125115	126836		1703	0	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240652.1
	149	968	126963	127931		953	2E-042	Hypothetical protein	hypothetical protein Phi4:1_gp063	*Cellulophaga* phage phi4:1	YP_008240651.1
	150	662	127928	128590		647	6E-077	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240650.1
	151	2138	128602	130740		2120	0	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240649.1
			131948	132085			4E-006	Structure and assembly	head maturation protease	*Halovi*. Phage HSTV_1	NC_021471
	155	755	132052	132807		647	4E-058	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240645.1
	156	548	132801	133349		509	3E-031	Structure and assembly	structural protein	*Cellulophaga* phage phi4:1	YP_008240644.1
	161	1964	136564	138528		248	5.6	Metabolism	Removes N-terminal methionine from nascent prot.	*Chlamydia muridarum* MopnTet14 ID=658600	WP_010229916
	166	1328	143944	145272		632	1E-048	Structure and assembly	putative phage tail fiber	*Cellulophaga* phage phi18:1	YP_008240966.1
	169	1640	146663	148303		387	9E-016		Hemolysin-type calcium-binding: RTX domain	NA	Fig|6666666.235665.peg.190
Total	169										
Conitg2	1	695	48	743	c	686	9E-032	Structure and assembly	structural protein	*Cellulophaga* phage phi38:1	YP_008241476.1
	3	1352	1705	3057	c	1271	3E-105	Structure and assembly	structural protein	*Cellulophaga* phage phi38:1	YP_008241474.1
	4	1193	3065	4258	c	0	8E-050	Hypothetical protein	hypothetical protein	uncultured Mediterranean phage	ANS05174.1
	5	2462	4332	6794	c	2399	7E-136	Structure and assembly	structural protein	*Cellulophaga* phage phi38:1	YP_008241472.1
	6	2648	6776	9424	c	2558	0	Hypothetical protein	hypothetical protein Phi38:1_gp090	*Cellulophaga* phage phi38:1	YP_008241471.1
	7	563	9412	9975	c	527	8E-012	Hypothetical protein	hypothetical protein Phi38:1_gp089	*Cellulophaga* phage phi38:1	YP_008241470.1
	13	3689	13038	16727	c	455	7E-030	Structure and assembly	tail fiber like protein	*Synechococcus* phage S-RSM4	YP_003097275.1
	14	653	16705	17358	c	614	8E-077	Lysis	N-acetylmuramoyl-L-alanine amidase	*Cellulophaga* phage phi38:1	YP_008241464.1
	15	2906	17331	20237	c	1160	6E-017	Lysis	pectate lyase	*Cellulophaga* phage phi10:1	YP_008242013.1
	18	788	23865	24653	c	269	1E-007	Hypothetical protein	hypothetical protein CGPG_00077	*Cellulophaga* phage phiST	YP_007673458.1
	20	530	25100	25630	c	458	1E-036		cupin domain containing protein	*Cellulophaga* phage phi4:1	YP_008240647.1
			27532	27750		0	7E-005	Structure and assembly	baseplate wedge protein	*Shigella* phage SHFML_26	gi100097, NC_031011
	29	1619	28769	30388	c	1490	1E-069	Structure and assembly	chaperonin GroEL	*Cellulophaga* phage phi38:1	YP_008241386.1
	34	575	32271	32846	c	177	1E-014	Lysogeny	phage integrase family protein	*Cellulophaga* phage phi46:1	YP_008241860.1
	35	341	33999	34340		347	7E-029	Hypothetical protein	hypothetical protein Phi18:3_gp060	*Cellulophaga* phage phi18:3	YP_008241253.1
	44	620	37403	38023		470	3E-016	Hypothetical protein	hypothetical protein P12053L_02	*Celeribacter* phage P12053L	YP_006560887.1

	61	686	44893	45579		380	1E-039	Metabolism	putative essential recombination function protein	*Clostridium* phage phiMMP02	YP_006990534.1
	64	584	46520	47104		413	1E-041	Hypothetical protein	hypothetical protein Phi38:1_gp032	*Cellulophaga* phage phi38:1	YP_008241413.1
		614	47759	48373		0	3E-010	Metabolism	putative DnaA recombination protein	*Aeromonas* phage vB_AsaM_56	gi422937508, NC_019527
	68	1952	48570	50522		788	4E-026	Metabolism	virulence-associated protein E (VirE)	*Cellulophaga* phage phi38:1	YP_008241397.1
	70	650	50655	51305		302	5E-014	Metabolism	putative deoxynucleotide monophosphate kinase	*Streptomyces* phage R4	YP_006990178.1
	72	449	51835	52284		359	4E-030	Hypothetical protein	hypothetical protein Phi14:2_gp071, putative phage head tail adapter	*Cellulophaga* phage phi14:2	YP_008242284.1
	73	695	52483	53178		581	3E-030	Metabolism	Td thymidylate synthetase thyX	*Synechococcus* phage S-SSM7	YP_004324368.1
	75	1046	53429	54475		968	2E-113	Metabolism	ribonucleotide-diphosphate reductase beta subunit	*Staphylococcus* phage vB_SauM_Romulus	YP_007677561.1
	76	1667	54456	56123		884	0	Hypothetical protein	putative ribonucleoside-diphosphate reductase subunit alpha	*Sinorh*. phage phiM9	gi966199320, NC_028676
	77	2048	56263	58311		1844	1E-056	Metabolism	putative DNA polymerase	*Edwardsiella* phage MSW-3	YP_007348961.1
	78	521	58328	58849		398	6E-019	Hypothetical protein	hypothetical protein Phi38:1_gp011	*Cellulophaga* phage phi38:1	YP_008241392.1
	79	1340	58993	60333		896	1E-055	Hypothetical protein	hypothetical protein Phi46:1_gp22	*Cellulophaga* phage phi46:1	YP_008241829.1
	87	4583	63527	68110	c	1982	3E-066	Hypothetical protein	hypothetical protein Phi46:3_gp121	*Cellulophaga* phage phi46:3	YP_008241164.1
	88	1514	68092	69606	c	656	1E-023	Hypothetical protein	hypothetical protein Phi46:3_gp120	*Cellulophaga* phage phi46:3	YP_008241163.1
	89	674	69590	70264	c	389	5E-020	Hypothetical protein	hypothetical protein Phi46:3_gp113	*Cellulophaga* phage phi46:3	YP_008241156.1
	91	3986	70720	74706	c	3980	0	Structure and assembly	structural protein	*Cellulophaga* phage phi38:1	YP_008241478.1
	92	836	74712	75548	c	830	9E-042	Hypothetical protein	hypothetical protein Phi38:1_gp096	*Cellulophaga* phage phi38:1	YP_008241477.1
Total	92										
Contig 3	8	1211	4696	5907	c	1166	8E-081	Structure and assembly	phage terminase large subunit, Phage packaging machinery	*Riemerella* phage RAP44	YP_007003633.1
	10	1067	6294	7361	c	974	7E-032	Metabolism	putative DNA methylase	*Streptococcus* phage MM1	NP_150145.1
	18	974	10342	11316	c	731	3E-027	Hypothetical protein	hypothetical protein PHG11b_2	*Flavobacterium* phage 11b	YP_112478.1
	21	452	12114	12566	c	410	2E-013	Hypothetical protein	hypothetical protein Phi38:1_gp026	*Cellulophaga* phage phi38:1	YP_008241407.1
			12599	12985	c	0	2E-007		carboxypeptidase	*Pseudomonas* phage H105/1	gi327198556, NC_015293
	23	674	13196	13870	c	635	4E-072	Lysis	metallophosphoesterase	*Cellulophaga* phage phi14:2	YP_008242299.1
	27	539	15855	16394	c	497	9E-016	Hypothetical protein	hypothetical protein P12024S_08	*Persicivirga* phage P12024S	YP_006560348.1
	28	1793	16363	18156	c	1658	4E-078	Structure and assembly	phage portal protein	*Persicivirga* phage P12024S	YP_006560346.1
			18655	19053	c	0	4E-016	Structure and assembly	capsid-related protein	*Sinorh*. phage phiN3	Gi971758981, NC_028945
	33	539	20279	20818	c	530	1E-023	Hypothetical protein	hypothetical protein P12024L_03	*Persicivirga* phage P12024L	YP_006560402.1
	34	926	20883	21809	c	899	5E-030	Hypothetical protein	hypothetical protein P12024S_02	*Persicivirga* phage P12024S	YP_006560342.1
	39	560	23960	24520	c	536	1E-016	Hypothetical protein	hypothetical protein Phi19:1_gp111	*Cellulophaga* phage phi19:1	YP_008241804.1
	42	1508	27116	28624	c	1406	8E-061	Lysis	pectate lyase	*Cellulophaga* phage phi10:1	YP_008242013.1
	45	596	31413	32009	c	497	3E-033	Hypothetical protein	hypothetical protein P12024S_01	*Persicivirga* phage P12024S	YP_006560341.1
	48	1982	33596	35578	c	1961	9E-053	Hypothetical protein	hypothetical protein P12024L_46	*Persicivirga* phage P12024L	YP_006560445.1
	49	1103	35575	36678	c	503	7E-006	Hypothetical protein	hypothetical protein P12024S_44	*Persicivirga* phage P12024S	YP_006560384.1

	50	1205	36663	37868	c	716	1E-013	Structure and assembly	putative tail fiber protein	*Rhodobacter* phage RcapMu	YP_004934701.1
	56	398	41011	41409	c	380	3E-012	Hypothetical protein	hypothetical protein Phi19:3_gp061	*Cellulophaga* phage phi19:3	YP_008240846.1
Total	57										

Appendix 4.3 tRNA matches to 'nr' database. Taxonomic affiliations of tRNAs were obtained with a BLASTn against the 'nr' database. No tRNAs were detected for contig 3. Only best BLAST matches are reported (< 1E-004).

Contig #	tRNA desription (postion), length [bp], score	BLASTn match	Score	Query cov. [%]	E-value	Ident. [%]	Accession nr.
Contig1	trna1-GluTTC (115065-114991) Glu (TTC) 75 bp Sc: 56.2	*Roseofilum reptotaenium* AO1	76.8	80	3E-011	90	RAST2:fig|564709.3.rna.44
	trna2-GlnTTG (114954-114880) Gln (TTG) 75 bp Sc: 47.4	*Acinetobacter* phage vB_phiAbaA1	51.8	52	4E-005	90	KJ628499.1
	trna3-HisGTG (114733-114645) His (GTG) 89 bp Sc: 33.0	no hits					
	trna4-LeuTAG (114640-114559) Leu (TAG) 82 bp Sc: 53.9	*Listeria* phage LP-124	75.2	100	3E-012	81	KJ094031.2
	trna5-LeuTAA (114552-114469) Leu (TAA) 84 bp Sc: 39.0	*Enterococcus* phage EFDG1	50	38	1E-004	94	KP339049.1
	trna6-ArgACG (114463-114389) Arg (ACG) 75 bp Sc: 44.6	*Cellulophaga* phage phi4:1	55.4	73	3E-006	82	KC821632.1
	trna7-LysTTT (114007-113936) Lys (TTT) 72 bp Sc: 34.4	*Bacteriophage* T5 strain st0	51.8	87	4E-005	78	AY692264.1
	trna8-IleGAT (113721-113647) Ile (GAT) 75 bp Sc: 62.6	*Pseudomonas* phage PAK_P1	48.2	85	5E-004	78	KC862297.1
	trna9-GlyTCC (113640-113567) Gly (TCC) 74 bp Sc: 45.9	no hits					
	trna10-AspGTC (112842-112770) Asp (GTC) 73 bp Sc: 43.6	*Bacillus anthracis* str. Ames	386	95	6E-009	80	AE016879.1
	trna11-ThrTGT (112296-112221) Thr (TGT) 76 bp Sc: 55.2	*Cellulophaga* phage phi4:1	105	92	5E-010	81	KC821632.1
	trna12-ValTAC (112212-112139) Val (TAC) 74 bp Sc: 66.4	*Acinetobacter* phage AM24	62.6	79	2E-008	85	KY000079.1
	trna13-ArgTCT (111259-111188) Arg (TCT) 72 bp Sc: 62.7	*Salmonella* phage 41	48.2	81	5E-004	80	KR296695.1
	trna14-MetCAT (111055-110983) Met (CAT) 73 bp Sc: 79.5	*Mycobacterium* phage Lukilu	180	91	6E-009	85	KX831080.1
	trna15-SupCTA (110856-110785) Sup (CTA) 72 bp Sc: 65.1	*Saccharopolyspora erythraea* NRRL2338	299	95	1E-011	87	AM420293.1
	trna16-SerGCT (108955-108865) Ser (GCT) 91 bp Sc: 24.9	no hits					
	trna17-AsnGTT (108792-108718) Asn (GTT) 75 bp Sc: 71.9	*Bacillus* phage Moonbeam	73.4	76	1E-011	89	KM236246.1
	trna18-TrpCCA (108711-108640) Trp (CCA) 72 bp Sc: 52.8	*Streptomyces* phage Brock	96.3	88	3E-007	85	KX925554.1
	trna19-PheGAA (108274-108202) Phe (GAA) 73 bp Sc: 68.6	no hits					
	trna20-CysGCA (108195-108119) Cys (GCA) 77 bp Sc: 51.4	*Lactococcus* phage P087	78.8	96	3E-013	88	FJ429185.1
Contig2	trna1-ProTGG (60517-60589) Pro (TGG) 73 bp Sc: 44.4	no hits					
	trna2-GluTTC (60595-60667) Glu (TTC) 73 bp Sc: 53.3	*Cellulophaga* phage phi4:1	62.6	65	2E-008	90	KC821632.1
	trna3-HisGTG (60836-60912) His (GTG) 77 bp Sc: 44.3	no hits					
	trna4-LeuTAG (60999-61083) Leu (TAG) 85 bp Sc: 25.4	*Cellulophaga* phage phi38:1	105	95	2E-021	89	KC821614.1
	trna5-LeuTAA (61087-61160) Leu (TAA) 74 bp Sc: 45.7	*Mycobacterium* phage MrMagoo	64.4	74	6E-009	85	KY223999.1
	trna6-ArgACG (61166-61239) Arg (ACG) 74 bp Sc: 49.1	no hits					
	trna7-LysTTT (61824-61896) Lys (TTT) 73 bp Sc: 36.6	*Enterobacteria* phage SPC35	51.8	86	4E-005	79	HQ406778.1
	trna8-IleGAT (61908-61981) Ile (GAT) 74 bp Sc: 45.3	no hits					
	trna9-GlyTCC (61987-62058) Gly (TCC) 72 bp Sc: 46.6	no hits					
	trna10-MetCAT (62607-62691) Met (CAT) 85 bp Sc: 48.7	no hits					
	trna11-AspGTC (62760-62831) Asp (GTC) 72 bp Sc: 36.6	*Bacillus cereus* phage vB_BceM_Bc431v3	53.6	88	1E-005	78	JX094431.1
	trna12-ThrTGT (63084-63157) Thr (TGT) 74 bp Sc: 43.6	no hits					
	trna13-ValTAC (63247-63322) Val (TAC) 76 bp Sc: 52.1	Halovirus HRTV-5	50	97	1E-004	78	KC292022.1
	trna14-UndetNNN (63467-63535) Undet (NNN) 69 bp Sc: 30.4	no hits					
	trna15-AsnGTT (27089-27013) Asn (GTT) 77 bp Sc: 48.7	no hits					
	trna16-CysGCA (26792-26722) Cys (GCA) 71 bp Sc: 20.8	*Cellulophaga* phage phi10:1	51.8	100	4E-005	76	KC821618.1



Appendix 4.4 Multiple genome alignments show genetic similarities between larger contigs 1-3 and respective reference genomes. (A) Contig_1 compared to *Cellulophaga* (phi4:1, phi17:2) and Cyanophage S-TIM5 references. (B) Contig_2 with respective references and (C) contig_3 comparisons. Coloured blocks indicate genetic similarities among sequences. Connecting lines between genomes indicate similar blocks. Coloured graphs lines within blocks shows the level of similarity. Open reading frames are displayed under the respective genome.

Appendix 4.5 Gene homologies with bacteria in cyanobacteria cultures and virus contigs.

Virus contig #	**Length**	**Coverage**	**GC-content**	Function	Matching taxon (bacteria)	Contig name (bacteria)	*E*-value
contig_1	26089.01	148900	34.24	tRNA-Glu-TTC	*R. reptotaenium* AO1	NODE_60_length_69290_cov_433.684_ID_119	3E-011
contig_101	13.38	15288	37.69	N/A	N/A	NODE_24_length_231210_cov_99.1565_ID_47	7E-028
contig_129	14.67	8335	39.74	N/A	N/A	NODE_13_length_404991_cov_102.715_ID_25	5E-006
contig_135	17.38	6128	39.12	Translation elongation factor G	Mixed bin, *Cytophagaceae* sp.	NODE_2_length_1495235_cov_100.329_ID_3	3E-043
contig_142	23.51	17576	44.00	lsuRNA; LSU rRNA	*R. reptotaenium* AO1	NODE_130_length_5159_cov_679.589_ID_259	3E-176
contig_142	23.51	17576	44.00	lsuRNA; LSU rRNA	*R. reptotaenium* AO1	NODE_130_length_5159_cov_679.589_ID_259	3E-097
contig_142	23.51	17576	44.00	lsuRNA; LSU rRNA	*R. reptotaenium* AO1	NODE_130_length_5159_cov_679.589_ID_259	2E-083
contig_142	23.51	17576	44.00	lsuRNA; LSU rRNA	*R. reptotaenium* AO1	NODE_130_length_5159_cov_679.589_ID_259	4E-016
contig_142	23.51	17576	44.00	N/A	N/A	NODE_385_length_332_cov_167.746_ID_769	1E-020
contig_142	23.51	17576	44.00	N/A	N/A	NODE_224_length_571_cov_89.5676_ID_447	1E-020
contig_142	23.51	17576	44.00	N/A	N/A	NODE_224_length_571_cov_89.5676_ID_447	1E-011
contig_142	23.51	17576	44.00	Contig 128, rRNA	N/A	NODE_128_length_5294_cov_214.689_ID_255	0
contig_142	23.51	17576	44.00	Contig 128, rRNA	N/A	NODE_128_length_5294_cov_214.689_ID_255	0
contig_142	23.51	17576	44.00	Contig 128, rRNA	N/A	NODE_128_length_5294_cov_214.689_ID_255	0
contig_142	23.51	17576	44.00	lsuRNA; LSU rRNA	*Alphaproteobacterium*	NODE_126_length_6254_cov_68.4601_ID_251	0
contig_142	23.51	17576	44.00	lsuRNA; LSU rRNA	*Alphaproteobacterium*	NODE_126_length_6254_cov_68.4601_ID_251	0
contig_165	14.59	7639	38.55	Na(+)-translocating NADH-quinone reductase	Mixed bin, *Cytophagaceae* sp.	NODE_7_length_885936_cov_106.412_ID_13	1E-028
contig_168	15.58	3097	39.36	Exconuclease ABC subunit	Mixed bin, *Cytophagaceae* sp.	NODE_7_length_885936_cov_106.412_ID_13	1E-013
contig_17	19.58	7957	40.63	N/A	N/A	NODE_8_length_882516_cov_107.531_ID_15	8E-025
contig_427	29.63	2296	40.07	tRNA-Glu-TTC	Mixed bin, *Cytophagaceae* sp.	NODE_7_length_885936_cov_106.412_ID_13	1E-037
contig_427	29.63	2296	40.07	Na(+)-translocating NADH-quinone reductasei	Mixed bin, *Cytophagaceae* sp.	NODE_7_length_885936_cov_106.412_ID_13	6E-031
contig_427	29.63	2296	40.07	tRNA-Glu-TTC	Mixed bin, *Cytophagaceae* sp.	NODE_7_length_885936_cov_106.412_ID_13	3E-024

contig_439	11.64	2786	55.35	N/A	N/A	NODE_4_length_1198380_cov_18.9772_ID_7	6E-022
contig_439	11.64	2786	55.35	N/A	N/A	NODE_107_length_18619_cov_20.8771_ID_213	6E-022
contig_439	11.64	2786	55.35	Translation elongation factor Tu	*Alphaproteobacterium*	NODE_176_length_1193_cov_44.074_ID_351	2E-155
contig_45	19	8230	38.96	Translation elongation factor Tu	Mixed bin, Cytophagaceae sp.	NODE_2_length_1495235_cov_100.329_ID_3	9E-139
contig_50	14.13	1311	41.27	N/A	N/A	NODE_8_length_882516_cov_107.531_ID_15	8E-018
contig_540	16.27	5836	38.47	N/A	N/A	NODE_1_length_2382106_cov_91.1957_ID_1	2E-020
contig_582	10.42	3375	52.33	N/A	N/A	NODE_4_length_1198380_cov_18.9772_ID_7	3E-015
contig_591	17.62	5041	41.08	N/A	N/A	NODE_5_length_1010206_cov_96.9797_ID_9	2E-083
contig_618	14.04	3715	34.40	N/A	N/A	NODE_12_length_520865_cov_97.2282_ID_23	6E-013
contig_619	9.54	4557	55.10	N/A	N/A	NODE_4_length_1198380_cov_18.9772_ID_7	5E-089
contig_619	9.54	4557	55.10	N/A	N/A	NODE_4_length_1198380_cov_18.9772_ID_7	1E-050
contig_652	6.95	2184	53.89	lsuRNA; LSU rRNA	*R. reptotaenium* AO1	NODE_130_length_5159_cov_679.589_ID_259	1E-111
contig_652	6.95	2184	53.89	lsuRNA; LSU rRNA	N/A	NODE_128_length_5294_cov_214.689_ID_255	7E-130
contig_652	6.95	2184	53.89	lsuRNA; LSU rRNA	N/A	NODE_128_length_5294_cov_214.689_ID_255	7E-055
contig_652	6.95	2184	53.89	lsuRNA; LSU rRNA	*Alphaproteobacterium*	NODE_126_length_6254_cov_68.4601_ID_251	0
contig_652	6.95	2184	53.89	lsuRNA; LSU rRNA	*Alphaproteobacterium*	NODE_126_length_6254_cov_68.4601_ID_251	4E-167
contig_68	20.21	8974	38.85	ATP synthase alpha chain	Mixed bin, *Cytophagaceae* sp.	NODE_7_length_885936_cov_106.412_ID_13	1E-102
contig_7	20.11	7225	40.14	ATP synthase alpha chain	Mixed bin, *Cytophagaceae* sp.	NODE_7_length_885936_cov_106.412_ID_13	1E-156
contig_740	8.86	2520	56.87	N/A	N/A	NODE_3_length_1457913_cov_20.7581_ID_5	9E-020
contig_783	4.71	1265	57.00	*R. reptotaenium* tRNA	*R. reptotaenium* AO1	NODE_65_length_62389_cov_459.575_ID_129	1E-015
contig_783	4.71	1265	57.00	sulfurtransferase, 8e-12763%, WP_015202277.1	*Geitlerinema* sp. BBD_1991	BBD_1001000 *Geitlerinema*	3E-016
contig_787	13.57	2538	40.23	Translation elongation factor Tu	Mixed bin, *Cytophagaceae* sp.	NODE_2_length_1495235_cov_100.329_ID_3	0

Appendix 4.6 Auxiliary metabolic genes (AMGs) characteristic for bacteriophages infecting cyanobacteria. Location of AMG associated genes are shown for respective contigs. Contigs are sorted according to their length [bp].

Contig #	Length	Coverage	GC-content	AMG	Taxon affiliation	E-value
Contig_760	1497	4.1	57.98	6-phosphogluconate dehydrogenase (**gnd)**	*Synechococcus* phage S-SM2	5E-010
Contig_444	1725	16.64	38.90	Transaldolase-like protein (**TalC)**	*Synechococcus* phage S-MbCM6	5E-050
Contig_767	2419	5.51	36.01	Heat shock protein **(HSP)**	*Prochlorococcus* phage P-SSM3	3E-006
Contig_113	3790	14.78	39.89	Glucose-6-phosphate dehydrogenase (**zwf)**	*Synechococcus* phage S-SM1	3E-101
Contig_213	5857	15	39.51	Cobalamin biosynthesis protein (**CobS)**	*Synechococcus* phage S-RIM2 R1_1999	0.007
Contig_292	11,579	14.2	39.04	Ferredoxin **(petF)**	*Synechococcus* phage S-MbCM7	3E-009
Contig_281	11,800	16.08	39.53	Phosphate starvation-inducible protein (**Phoh)**	Cyanophage S-TIM5	3E-024

Appendix 4.7 Gene similarities of virus contigs 1-3 and viral metagenomic reads from Sato et al. (2017). Virus contigs were analysed for gene similarities through a tBLASTx to the viral metagenome reads (hits are shown in column 'Annotation' of respective virus genes, Subject ID names the read that the gene belongs to, and the respective *e*-value of the comparison). The unknown reads of the metagenome were then identified with a BLASTp against the 'nr' database. The closest match of the BLASTp is shown in the column 'Subject ID (metagenome match) with the respective *e*-value and Genbank accession number. Highlighted in bold are the best blast matches for the respective genes.

Query	Annotation	Subject ID (contig match)	*E*-value tBLASTx	Subject ID (metagenome match)	*E*-value BLASTp	Accession Nr.
Contig_1	**ribonucleoside-diphosphate reductase alpha subunit**	**AGRF-22_0003:7:43:10556:18242#0**	**7E-005**	**hypothetical protein CPVG_00016 [Cyanophage KBS-S-1A]**	**1E-011**	**AET72825.1**
Contig_1	**HNH endonuclease**	**AGRF-22_0003:7:22:14792:2676#0**	**8E-007**	**hypothetical protein pVp-1_0037 [*Vibrio* phage pVp-1]**	**9E-009**	**YP_007007860.1**
Contig_1	HNH endonuclease	AGRF-22_0003:7:7:5028:15299#0	5E-005	hypothetical protein CPVG_00016 [Cyanophage KBS-S-1A]	1E-011	AET72825.1
Contig_1	HNH endonuclease	AGRF-22_0003:7:8:14599:6455#0	5E-005	putative endonuclease [uncultured Mediterranean phage uvMED]	9E-011	BAR34273.1
Contig_1	HNH endonuclease	AGRF-22_0003:7:103:7415:11009#0	4E-004	hypothetical protein PRRG_00010 [*Prochlorococcus* phage P-RSP2]	1E-004	AGF91521.1
Contig_1	HNH endonuclease	AGRF-22_0003:7:40:2946:16187#0	3E-005	putative endonuclease [*Prochlorococcus* phage P-SSM2]	3E-012	YP_214523.1
Contig_1	HNH endonuclease	AGRF-22_0003:7:43:7624:3776#0	3E-004	hypothetical protein CPVG_00016 [Cyanophage KBS-S-1A]	1E-012	AET72825.1
Contig_2	**putative ribonucleoside-diphosphate reductase subunit alpha**	**AGRF-22_0003:7:22:14792:2676#0**	**8E-004**	**ribonucleotide-diphosphate reductase subunit alpha [*Nitrincola* sp. A-D6]**	**4E-011**	**WP_052063864.1**
Contig_2	putative ribonucleoside-diphosphate reductase subunit alpha	AGRF-22_0003:7:103:7415:11009#0	6E-004	FAD-dependent thymidylate synthase [Cyanophage KBS-P-1A]	1E-008	YP_007676344.1
Contig_2	putative ribonucleoside-diphosphate reductase subunit alpha	AGRF-22_0003:7:43:7624:3776#0	1E-004	ribonucleotide-diphosphate reductase subunit alpha [*Pelagibacter* phage HTVC019P]	2E-009	YP_007517830.1
Contig_2	**thymidylate synthetase thyX**	**AGRF-22_0003:7:43:10556:18242#0**	**7E-006**	**FAD dependent thymidylate synthase [uncultured phage MedDCM-OCT-S04-C348]**	**4E-010**	**ADD95112.1**
Contig_2	thymidylate synthetase thyX	AGRF-22_0003:7:8:14599:6455#0	6E-004	hypothetical protein SSSM7_233 [*Synechococcus* phage S-SSM7]	3E-006	YP_004324285.1
Contig_2	thymidylate synthetase thyX	AGRF-22_0003:7:40:2946:16187#0	2E-005	thymidylate synthase [*Vibrio* phage VBP47]	2E-006	YP_007674123.1
Contig_2	**tail fiber like protein, *Synechococcus* phage S-RSM4**	**AGRF-22_0003:7:7:5028:15299#0**	**6E-004**	**hypothetical protein pVp-1_0037 [*Vibrio* phage pVp-1]**	**7E-008**	**YP_007007860.1**
Contig_3	**putative DNA methylase**	**AGRF-22_0003:7:75:1975:2807#0**	**4E-006**	**gp127 [*Mycobacterium* phage Omega]**	**2E-012**	**NP_818425.1**

Appendix - Chapter 5

CRISPR-Cas in BBD associated cyanobacteria

Appendix 5.1 Web access links for bioinformatic tools. Follow the links to access bioinformatic tools and genome data.

Tool	Web access
RAST	http://rast.nmpdr.org/ Access with guest account, username: guest; password: guest. ID: *Roseofilum reptotaenium* AO1 (564709.3), *Alphaproteobacterium* (28211.29), *Cytophagaceae* sp. (89373.4)
NCBI GenBank	www.ncbi.nlm.nih.gov, NCBI GenBank MLAW00000000. for *Roseofilum reptotaenium* AO1, *Alphaproteobacterium*, *Cytophagaceae* sp.
Artemis	www.sanger.ac.uk/science/tools/artemis
CG-View	http://stothard.afns.ualberta.ca/cgview_server/
CRISPRfinder, db	http://crispr.u-psud.fr/crispr/
CRISPRtarget	http://brownlabtools.otago.ac.nz/CRISPR_WEB/crispr_analysis.html
Bandage	Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualisation of *de novo* genome assemblies. Bioinformatics. 2015 Jun 22:btv383.
177PHAST	http://phast.wishartlab.com/
PHASTER	http://phaster.ca/
VIRsorter	https://de.iplantcollaborative.org/de/

Appendix 5.2 Functions of *R. reptotaenium* AO1 and *Geitlerinema* sp. BBD_1991 CRISPR-Cas type associated genes. CASCADE = CRISPR associated complex for antiviral defence. The gene Cas10d is representative for the CRISPR-Cas type I-D. Functions from Makarova et al. (2011a, b).

CRISPR type	Cas gene	Function
I-D	Cas2	Spacer acquisition, protospacer recognition
	Cas1	Spacer acquisition, protospacer recognition, also RNAse
	Cas4	Not sure, might be involved in spacer acquisition
	Cas6	RAMP family RNAse, for crRNA processing, snaps ssRNA and dsRNA
	Cas5	RAMP CASCADE subunit, RNA cleavage, might substitute for Cas6
	Cas7	RAMP CASCADE subunit, RNA cleavage
	Cas10d	Large CASCADE subunit
	Cas3 helicase	Cuts DNA during interference, promotes strand separation

CRISPR type	Cas gene	Function
I-D	Cas2	Spacer acquisition, protospacer recognition
	Cas1	Spacer acquisition, protospacer recognition, also RNAse
	Cas4	NA, might be involved in spacer acquisition
	Cas6	RAMP family RNAse, for crRNA processing, snaps ssRNA and dsRNA
	Cas5	RAMP CASCADE subunit, RNA cleavage, might substitute for Cas6
	Cas7	RAMP CASCADE subunit, RNA cleavage
	Cas10d	Large CASCADE subunit
	Cas3 helicase	Cuts DNA during interference, promotes strand separation
III-U	Csx3	Not yet assigned to a specific cas subtype
I-MYXAN	Cas4/Cas1	NA
	Cas5/Cmx5	NA
	Cas7/Cst2	NA
	Cmx8	NA
	Cas3	NA
	Cas6/Cmx6	NA
III-B	Cas1	Spacer acquisition, protospacer recognition, also RNAse
	Cas2	Spacer acquisition, protospacer recognition
	Cmr2	NA
	Cmr3	NA
	Cmr4 RAMP	NA
	Cas1	Spacer acquisition, protospacer recognition, also RNAse
Cas genes	Cas1	Spacer acquisition, protospacer recognition, also RNAse
Cas genes	Cas6	NA
Cas genes	Cas1	Spacer acquisition, protospacer recognition, also RNAse
Cas genes	Cas1	Spacer acquisition, protospacer recognition, also RNAse

Appendix 5.3 Direct repeat (DR) BLAST similarities. DR sequences were BLASTed against the CRISPRdb with default parameters and top 3 hits reported in the table.

R. reptotaenium AO1

DR sequence (array #)	Organism	RefSeq	Identity	*E*-value
(1) GTTTCAATCCACAGCAATCTCTATTAGATTTTGAAAC	*Tistrella mobilis* KA081020-065	NC_017958_2	0.938	3.20e-002
	Tistrella mobilis KA081020-065	NC_017957_2	0.938	3.20e-002
	Tistrella mobilis KA081020-065	NC_017957_1	0.938	3.20e-002
(2) GTTTCAATCCCTCATAGGGATTTATGTTGGTTTCAAC	*Rivularia* sp. PCC 7116	NC_019678_16	0.946	7.00e-011
	Rivularia sp. PCC 7116	NC_019678_4	0.919	1.00e-008
	Rivularia sp. PCC 7116	NC_019678_19	0.919	1.00e-008
(3) GTTGAAACCCATCTAAATCCCTATGAGGGATTGAAAC	*Cylindrospermum stagnale* PCC 7417	NC_019757_35	0.964	3.00e-008
	Nostoc sp. PCC 7120	NC_003272_20	0.912	3.00e-008
	Cyanothece sp. PCC 8802	NC_013160_1	1.000	4.00e-007
(4) TGTTTCCAACTAATCCGATTTAACCCAATCGGTAGGG	*Cyanothece* sp. PCC 7822	NC_014533_1	0.944	7.00e-011
	Rivularia sp. PCC 7116	NC_019678_24	0.917	3.00e-008
	Cyanothece sp. PCC 7424	NC_011738_1	0.917	3.00e-008
(5) GTTGAAATCGACCTAAATCCCTATTAGGGATTGAAAC	*Rivularia* sp. PCC 7116	NC_019678_19	0.946	7.00e-011
	Rivularia sp. PCC 7116	NC_019678_31	0.921	2.00e-010
	Cylindrospermum stagnale PCC 7417	NC_019757_35	1.000	1.00e-008
(6) GTTTCAATCCACAGCAATCTCTATTAGATTTTGAAAC	*Tistrella mobilis* KA081020-065	NC_017958_2	0.938	3.20e-002
	Tistrella mobilis KA081020-065	NC_017957_2	0.938	3.20e-002
	Tistrella mobilis KA081020-065	NC_017957_1	0.938	3.20e-002
(7) GTTGAAATGAACATAAATCCCTATTAGGGATTGAAAC	*Rivularia* sp. PCC 7116	NC_019678_16	0.947	2.00e-012
	Rivularia sp. PCC 7116	NC_019678_19	0.946	7.00e-011
	Calothrix sp. PCC 7507	NC_019682_29	0.919	1.00e-008

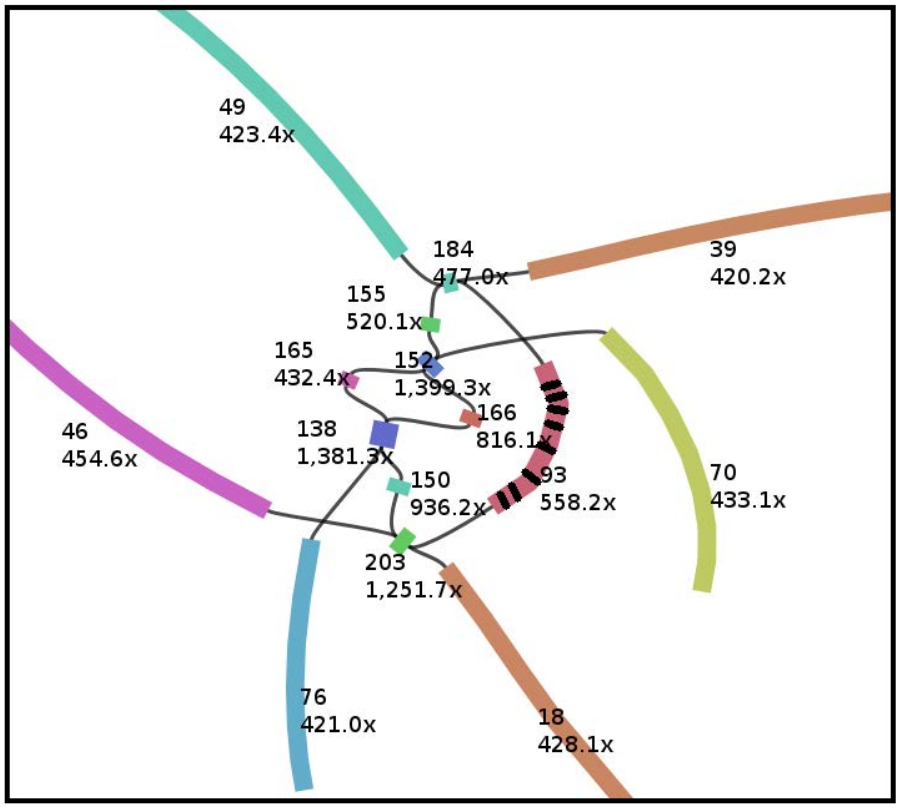
Geitlerinema sp. BBD_1991

DR sequence (array #)	Organism	RefSeq	Identity	*E*-value
(1) GTTTCAATCCCTAGAAGGGATTTTATCGGATTTAGAG	*Halothece* sp. PCC 7418	NC_019779_1	0.956	5.00e-06
	Calothrix sp. PCC 6303	NC_019751_6	0.923	5.00e-06
	Nostoc sp. PCC 7120	NC_003272_3	0.923	5.00e-06
(2) GTCTAAATCCGATAAAATCCCTTTTAGGGATTGAAAC	*Crinalium epipsammum* PCC 9333	NC_019753_5	0.961	1.00e-07
	Crinalium epipsammum PCC 9333	NC_019753_18	1.0	1.00e-07
	Crinalium epipsammum PCC 9333	NC_019753_17	0.961	1.00e-07
(3) GTTTCAATCCCTAAAAGGGATTTTATCGGATTTAGAC	*Crinalium epipsammum* PCC 9333	NC_019753_5	0.961	1.00e-07
	Crinalium epipsammum PCC 9333	NC_019753_18	1.0	1.00e-07
	Crinalium epipsammum PCC 9333	NC_019753_17	0.961	1.00e-07
(4) GTTTCAATCCCTAAAAGGGATTTTATCGGATTTAGAC	*Crinalium epipsammum* PCC 9333	NC_019753_5	0.961	1.00e-07
	Crinalium epipsammum PCC 9333	NC_019753_18	1.0	1.00e-07
	Crinalium epipsammum PCC 9333	NC_019753_17	0.961	1.00e-07
(5) GTGCTCAACGCCTAACGGCGATCGAAGGAAATTCAC	*Leptospira interrogans* L1-130	NC_005823_1	0.956	5.00e-06
	Stanieria cyanosphaera PCC 7437	NC_019748_5	0.916	2.00e-04
	Synechococcus sp. PCC 6312	NC_019680_1	0.892	2.00e-04
(6) GTTTCCATTCATTCCACTTCTCTAAAAAGAAGCGTC	*Cylindrospermum stagnale* PCC 7417	NC_019757_28	0.878	2.00e-05
	Stanieria cyanosphaera PCC 7437	NC_019748_1	0.818	0.009
	-	-	-	-
(7) CTGACAGCTTCTTTTGAAGCGGAATG	*Trichodesmium erythraeum* IMS101	NC_008312_20	0.952	2.00e-04
	-	-	-	-

AATGGAAAC	-	-	-	-
(8) GTTTCCGTCCCCTTGCGGGAAAAAG	*Crinalium epipsammum* PCC 9333	NC_019753_6	0.958	8.00e-07
	Synechococcus sp. JA-3-3Ab	NC_007775_2	0.958	8.00e-07
	Crinalium epipsammum PCC 9333	NC_019753_3	0.956	3.00e-06
(9) CATCCCCCGAAGGGGAAGTCGATCGAAAAG	*Synechococcus* sp. JA-3-3Ab	NC_007775_10	1.0	0.089
	-	-	-	-
	-	-	-	-
(10) ATTTCCATTCATTCCGCTTCAAAAGAAGCTGTCAG	*Trichodesmium erythraeum* IMS101	NC_008312_20	0.95	7.00e-04
	-	-	-	-
	-	-	-	-
(11) ATTTCCATTCATTCCGCTTCAAAAGAAG	*Trichodesmium erythraeum* IMS101	NC_008312_20	0.95	7.00e-04
	-	-	-	-
CTGTCAG	-	-	-	-

Appendix 5.4 Spacer matches to contig 93. A# = CRISPR array number. S# = spacer number out of total respective array. %Ident = percent identity of query to spacer target in BLASTn. M/G = number of mismatches or gaps in alignment of protospacer and spacer. *E-value* = BLASTn score of query match.

A#	S#	Sequence	Target	Gene target	%Ident	M/G	*E*-value
2	4/14	GATCGGGCCCCCATCTGGGAATGCCCCGAGTTT	Contig_93	hypothetical protein	100	0	2e-011
2	7/14	CCAAAGGGTCCCCACTTACTGGATGTCTTGAAATATCCA	Contig_93	hypothetical protein	97.44	1	5e-013
3	12/12	CAAACCTATTTTTACTATTTCCGCAATTTATGGCAA	Contig_93	hypothetical protein	100	0	5e-013
5	6/11	GAAACATTCAGAATCAAATGGGAGGTTGATTGGGA	Contig_93	hypothetical protein	100	0	2e-012
5	7/11	CCGGCACCAGATTTTCCGGATGAATTCCTAGAAG	Contig_93	hypothetical protein	100	0	5e-012
5	9/11	GGGAAATTCTCCTTTCGGAGCGCTCGGTGAAATTCTTG	Contig_93	hypothetical protein	100	0	4e-014
6	11/27	CTGGGTTAAGAAGAGGATTCCATTCGCTATAATCAAAATCAT	Contig_93	hypothetical protein	97.3	1	3e-011
6	14/27	GCATAACAAAATACGATCGGGAAACATTGAAGAGCCCAAT	Contig_93	DnaB domain-containing helicase, VIRsorter: Phage_cluster _71 _PFAM-AAA_25	100	0	3e-015



Appendix 5.5 De Bruijn graph connection of contig 93. De Bruijn graph was visualised with the software bandage (Table S1). Contig 93 can be connected to several other contigs (contig number in figure with coverage). A potential prophage could be spread over the adjacent contigs. Black marks on contig 93 indicate CRISPR-Cas target loci.

Appendix 5.6 Short contigs of *R. reptotaenium* AO1 that are connected to self-target region contig_93. BLASTx matches (nr database) are indicated below the respective sequences.

>Contig_184_length_979_cov_476.981_ID_367

ATCGCCTACGACCCCAACCGAGAAGCGCGGGCCGAACAACTGCGGCTGAAGAACCAACCCACCATTCGGG
AACTGTGGGAGAAGTACAAAATCAAGCAGTTAAAACTGACCTCCAAAACCACCCAGAAAAGCGTTTGGGT
AGAGATTGACCGGGCACTGGATGCCTTATCCAGTAATGCCTTGAAGCTTGAGAGTCTGGATAATCTAGGC
GATGAGTATATGAAGCTCTACGCGATCGCCACCTGCCACCGTCATTTCGAGTCACTGCAACCAGCCATCC
GGTTGCACCTTCCCAACATTAAACTCAAGCCCCAACTTCCCAAGAGGGTAAAACGCCCTATCGAATGGTT
CCCCCCGGATGAAGTCAAAGTCATATTACAAGCGTTCAAAACCGACCGTTTTTCCTCTCCCTTTGCCCCA
GTTCCCCACAGCTACTACT ACCCCTATGTCTGTTTCCTCGCTCATACCGGGTGCAGACCCGAAGAAGCGA
TCGCCCTGCTGTGGTCAGATTTGTTCTGGCTGCGCGATCGGGCCGGTTGCGAAGCCTCCATCACCAAAGT
CTTCTCCAAAAAGGTCCTCAAACCCTACACCAAGAACCACCTCATCCGAAACATCCCCATCTCCCTAGCC
CTTCAAGACATATTAGAGGAGCGAAAGAAGCGATCGGGCTTAGTCTTCCCCTCTCCCCAAAAAAAGCACA
TCGATCAGAGCAACTTCTCCAGTCGCGTCTGGAACTGCGTCCTCCACTCCCTAGTTGCAGAGGGAGAAAT
CCGCAAGCGACTACGCCCCTACTGCCTACGTCACTCCTTCGTCACCAATATCCACCATGAGCATGGAGTT
CCCTTCCCCACCATTGCCCACCTGATTGGCGACAAAATTGAAACCGTCATCCGGTTTTACTCAGGAACCA
AACCCCTTACCACCCAAACCTTCCCCAACCTCTATTAAAACCCCTTCAAACTGGTACGATTTTGGATCG

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession
phage integrase family protein [*Stanieria* sp. NIES-3757]	169	169	88%	4e-46	39%	BAU63200.1

>Contig_156_length_1881_cov_520.087_ID_311

ATTGCCCACCTGATTGGCGACAAAATTGAAACCGTCATCCGGTTTTACTCAGGAACCAAACCCCTTACCA
CCCAAACCTTCCCCAACCTCTATTAAAACCCCTTCAAACTGGTACGATTTTGGATCGGAATTTCCGGCTC
CAATTGCCCACCTGGTAAAGGCTTGCCAGTTTGAGTAAGCGAAAAAAACTGCCCCAAAGTGGTACGATTT
TGGTACAGGGGTATTTTTAGAGAAAAATCTCAAACGGCTGAAACCCTTGCTGTGACTGGGATGGACGTAA
CGGGACTCGAACCTGTGAACCTTTAAGGGGTTGATGTCATTCCAAAGGCTAAAAGCCTGACCATGCATAC
ATTCAACGCATTAAACCTTAGAAATGTGTTGAAAATGTGTTGAATATGGACACCTACATAAAACAAGCAT
GGAGCCTGGTGAATGAGTACTTCTACTCTAACTCCATTGACATCTCCAAGCAGGTAGACCATGAGCTAGT
GAGGGCATACCTCAAAGCTTGCCAGAAGAGTACCCCCAAGGGTATTCGTATTGTCAGTAGCGGTAACCGA
CTGTATCTCCGATTCAAAACTACGACTAAGCCAGCAACGGTTAACAATTCCTGTAATGAGGACTTCACTC
GCGATGGTTGTATTAACGCTCTAGCTAAAGCCTTAGCTGTCTTTAACAAACTCAAGGAGACTGAAGCAGA
GTCTGAATTCTGGTCGTGGTATGAGTCAGAAATTAAAGGCATTGTACTCCTCAAGAATGACATTATTACC
ATTGGTCAAGCTATTGAGACTGTCAAATCAAACTACCTACTAGGTCGCGACAAGTGTGACCGAAATAGGA
ATGACGAAAAGCTGACGACTAACAGCTTGAGTATATACAATAAGACTTATGGCGGACATTACAAGAAACT
CAATCCATGTCTCAGACTTACCGGGGAAAACATAATCTCTGAAATCATGAGGAACTGGGGTCAATTAATT
ATTTCTACGACTGGCTCTCAAACCTTGTGTTCCAAAGGCTTTAAAAATGCCTATACAGCTTGCTGTAAGC
TTCTCCGGGATACAAAACTCTCATCAGAGCTGGACAAAGTTACCAGCTATTTTGAGAAATTGAAGGTCAC
TAGGAAAACAAAAATGCAAGCCATTGACCTTGAAGCCTTCCTAGACTTTAGGGCTAGAGTATTGGGGCTA
AATGGTTATGAGCTGACTAAAGCTCAATGGCGTAACATTGAGTCACGCAAATCATGGATGAAAGCTATTT
GCATTAATTTAATTTACGGCTTTCGTGCCTCTGAGTTCAAAGCAATTTTGAATTTTGATAAAGCAATTAC
TTTAGATGGTTACACCTTTTATGCACTTGATGACCCCAGTAACAACGAAAATATTGTAGTTATTGACGAA
GGCTTTTGGATAACTGATACTAGCGGAGAGTGTCATTACATCACAATTAAAACAGGCAAACGTATTGCAC
GTCCGATGATTCATCCCGATTACCCTAACTTGGTTGAATTATTGGGAATAAAAGACCCAAGAGTCAAGAT
ACCTGAATGTATCCCTAAAGCCAGTAGCAACCCGGACACGATTAAAGATATTTATACCCGTCAAATGGGG
CAGAGATTGGCTGATTACATTTCCCAAGTAGGAGGTCAAGGATTCACTCAAACTCATGCCCTACGCCACT
TAGCAAATTACCATGGCAAACTTGCCGGTTTAACGCGCGACCAACGTGCTTTGTCACTGGGACACTCTCA
AACAATGAATGACAAATACGACAAACATCAAACTACCAGGAATCAGGTGAACCTTCTGATGGCTGACATC
TCTGAGAAATCAGAAATCCAAAGGCTAAAAGATGAGCTTTCCCAAGCTCAGGAGACCATTA

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession
hypothetical protein [*Lyngbya aestuarii*]	323	323	56%	6e-102	46%	WP_023065230.1

>Contig_152_length_2049_cov_1399.3_ID_303

ATGAATGACAAATACGACAAACATCAAACTACCAGGAATCAGGTGAACCTTCTGATGGCTGACATCTCTG
AGAAATCAGAAATCCAAAGGCTAAAAGATGAGCTTTCCCAAGCTCAGGAGACCATTAAATCCCTAGAGGA
AACCATTCTCTTCCTGAAGAAGGAGAACGCCAGGCTGAATGAACTTTTAGGAGGTAACGACGACCTTCCC
AGAATTGATTAGAATGCAGGCGAATCCCCTTCTACTTGGAAGGGGTTAATTTACCTTAAATTTTTAAGGT
AAATATCTCATTACCTCCCTAACTTGGTGTTCGTCAAAACCCCATTCCTCCATCTGCCGACACAGTGTAG
AGAGTCGTCTAACGCTGAGTTTAAGAACACGAGTATTACCCTGATTATCTAAGTTACTAAAAACGACAGC
GTATGAATTAGAATTTTTAAGCTTCTTAATATTCAGAGTGCATTTTCCCCCCTCAACTACTCCACGCCAT
TCTGTTGAATTTTTATCTGAATTCCTTTTCTTAATCAATTCTCTGACAGCATCCTTCACTAGCCATGATT
CATTGATGGCTGTATACACTGCTAAATCGATACATTCCTCAGATTCTATACAATCCCCCCATTGATTGAT
TGTTGATGGGAATAAATCAATTATTCTCAACCACTCACTAACTTGCTCTAGCTTGCAGTCTAATTTTGAT
ATTCTGTCACACAACGCGTAAATCGTATCAGATGGTTTAGAGTAACGCTCTAATACGACCTCCCTAAAGT
CAATTGATAAAGCCTCTGCAATTTGGTCAATCAAATCAGTCTTGAGCCAACCACCGTCTACTAACGTAAT
AGTGGTATCCATTTTTTCGTTTAAAATCTCAGCCATTGCAACCTCATTCTTTCGTATAGAGAGTGTAATT
TACCTTAAAAATTTAAGGTAAAATTGCCATCTAAACCTCTATTCTACCGTTAGGTCTTTAATCTTAGGTC

TCTCCTGCCGTGTCCGGTTAAACGGCAGTGCATTATCTCCACTCGTTCGGATAGTAATTTGGCTACCATC
AGGAGACCTATAAATACCATCTGGCGGCAGATATCCCTCCTCAGTAAACTCTAGAAACTTCCGCCAATCT
TTCTGCTTGTACCTATCAGTATCTTCTAGACTTTCCGGCTCCTCTGTAATACCTTTAGCTTTCTTTACAG
CCCTCGTCCCCTGATTAACTTTGGTCTCGAAATCTTTAAGATACTTATTCAGATTCTCCTCCAAGTCTTC
TGAACCCCCTACAGACACAGTGAAACTATGTTTAATAATAGATATCGCCTGCTTTACATCAACCATCATA
TCCATTAAGTCTTCCGTCGAAGATTCTAATTCATAGGCTTTTACGGGGTTGGTGGATGGCTCCACAAATT
TATCTAATTGTTTGTCAGCGTAATCGTCAGTCGGTAAAGTATCTTTACCCATCTCCTCCTCCCATCCAGG
AATACTAAAAGCTGTAAAACCACCCCCACCAATTAAATTAGCTATATCTGGAGAAACTAAACCAGGGATA
GTAAAAGATGCCCTCAACTTTTTAGTGACTTGCTTGTATTTGAACCCTAGCCAATCGATTACAACTTCTA
TTCTATTTCTAGCCTTATATGCCTCCTCTCTGGCTTGGTAAGCTTCCAATACAGCCCTCATCCCTAGAGC
CTTAATCGTTCTTTCATCATCCTGCAAGTTCATGTTCAGCTTGATTATTTCAGTTATTGCGTCTTCCAAG
TTGTGTACAGTGACATTAACCGGTAAATTTTCGCCCGATTGTTCTGCTGCCTCTCTAATCCTTCTAGGGA
CTTTAATCGTGAAGGGAAACTGACCAAAAAAACCATCCAGGCAATCCATAAAATGCCAGATTAGCTCAGG
CTGAGAGCGAATAATCTTATTCTCCATAAAAGCGTTGAATCCAGTAGGCACAATCCCCTCTATACCATCG
TTGGAAATAAATGGGTTTCCAGGGACTCTAATCGCTGTGTTATTCGCCCCAGTCCACTGTTTAATCATCC
TCATATCACGCTGAATTTC

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession
no match						

>Contig_165_length_1517_cov_432.414_ID_329

CGTTGAATCCAGTAGGCACAATCCCCTCTATACCATCGTTGGAAATAAATGGGTTTCCAGGGACTCTAAT
CGCTGTGTTATTCGCCCCAGTCCACTGTTTAATCATCCTCATATCACGCTGAATTTCGTTAAGTTTTCTG
ATTACTGTTGGGCAGCAATTCCTCATATCATCCTCCTCTCCTCTTGGATTTCTTTTCAGTTTATGTAAGG
GGACAATTCCTTTCCTTGGGGGGAATTGAACAAGCCTATAAGGCACTCTAATTATAATTGATGCGCTGTC
AACATCTGTCCCCCAATCAGTAAATACTTCACCAGTTACGCTATTGATAGCCGTCCCCGGAGTCCCGAAA
GCGTTAACCCTATATTTACCATTTGGTCGTGGTTCTTTAGTTGGAGTTTTTCCTGCCCCATCCGTTGTGG
CACGCTCTCTATAATTAAGGAATTGCTCCATCGAGCTAGGTAGTCCCCATCTCGGACTGACTGGAGGGAC
TTCGTAGGGGTGACTCCACTCCATTCTTCCGTTTCGATATGCAGAATTATTGTAAACGTATATCTTTTCC
CCCAATCCATTTGTGTAACTTCCAATTACTCCCGAAGACCCAACTGGAAGAAGAAAACCTAATCTACTAC
CGTCTGGGTTTGAGGTTGGTCTGCCCGCTTCAAAATCTACGCCTCTAGCCTCTATTATGCCGTCAATGAT
TTCCGGGGTAATTCCTCCCAAAGATTCTGTATTTCCTGGAGGATAACTAATTGTTATTGTGTCTAAATTT
ATCGTTTCCTGAGTAACTTTATTTTCCTCTCTTATTTGTTGCTCTATGATATCCGGTACTTCGCTTGGAC
TGGTGTTAGAAATAATATCATATATTTCTCCTAGTGCTTGATATAAAGTGTAAACATCGTGTGCCCATAA
AGCTACGTTTAAAGCCGCTCCTGCTGGTCCGGTTGCGAGTGTAGTCAAGGCTAATCCTCGTTTTCCAGCC
CACTTTAGAAATTCTTTGACAATTCTTCCCCCTCCAATCTTACTGCCGCGCCTTAATAATCTAGCCGCTG
CACCAGTGAACCCACCTGAAGGCGCGGCATCTCTGAGTTTATTTCTAGGATCTAATAATTGTTTAGTTTT
TTCGGGGTCTCTTAACCATCGTTCTATAGGTTGTGCTGCCAATTCGCTATCGATTCTTCGACTAGGAGAT
AGAACTCTTGTAGCCCCTTCTCCCATGATGGCACTAGAGGTATTAGGCACTGGCGCTCTTAAGGATGTTG
ATTGATTGGGATTATATGGAGAAATACTTCCACCTCTTCCCGCTTTAATTCCATCAAAAGGATTGGGATT
TCTCTTGGGCATGATAACCTCATCGATTTGTACGGGTATGGCTACCGATGAGGTTCGTTGTGGTATTCCC
ATGACAGTTTTACCTTAAATATTTACCTTAAAATGGTCTGGGCATGGTCTCAATTGAATCGAAAAAGATT
TGGTCGAGCCACTCTTTTATCATGTTATCCATGGTTGGGTTATCTTC

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession
no match						

>Contig_166_length_1508_cov_816.137_ID_331

CGTTGAATCCAGTAGGCACAATCCCCTCTATACCATCGTTGGAAATAAATGGGTTTCCAGGGACTCTAAT
CGCTGTGTTATTCGCCCCAGTCCACTGTTTAATCATCCTCATATCACGCTGAATTTCATTTAATTTTCTA
AGTACCTTTGGGCAGCATTGCATGTTATCATCCTCCTCTCTATTTCTTTTTCTTTTTGCCTTTGGCTTGG
GTACAATTCCTGTCTTTGGCGGGGAATGTAGCACTATGTGAGGTATGCGGTCAACTCTCGCATATCCAGT
AACATCAACACCCGTACCGCCTGATATTGCATCGACTGTAATTGTTTCTCCATTTATCTCTATTTCAGTG
TTACGCCCTAACATCGAAGCAGGAGCTGTTACCGAACCCAGGGTATAAAGTCCATTTGCATCGGGTTCAC
CCTTAGTAAATTCTCCCCAATCTTTAAACTTATCTAATCTTTTTATATATTGCTCCATTGTGGAAACTAT
GTCCCGAAAGTTTAATGAACGCGGTTTTATTGTATCAAAGGGATAGCTCCATTTAATCTTTAATCCCGTA
CCGTAGGAAGTAAGTCCAACGCTACTAATAACTTCCCCGAACTCACCGCGAAGCCCGCCTAAAACAGCGA
CGTTATTAAAAGGAGTACGTCCTATAAAGGCACTATGTCCAATTGCCTTGACGTAACCTGATTTACTTCC
ATCACTTCTAGGGGTTGGGATATTTTCTATTTGCTCAATAGCTCCTGCCCCTTCTTGGATTCCTTCGACT
ATACCATCACCATATTCTTGTATGATTTCTTGAATAGATTCTGTGGAAATATTGCCCGCCGTTTCAATAG
CGCCTTCCAACCCAGAATACTTTTCTAGCAATTCTTTTCTCGCTCCTTGTTCCACGATATCTGGAACCTC
ACTTGGGCTAGTGTTCCCCATATTATCGAGAATATTCTCTAACGCTCTATATCCTTGGACGATAGCTATA
CCGCCAAGAACACCAGCGCCGATTAGCAGTCCCCCTTTGACTGTGATGGCACCACCAACAACCAATTTAG
TGCCCTTAATTGCTCCTCCAACAACCTTTTTCCCCGCAGCAGAGAATCCAGCCTTGCCGCCACCAATTCC
CCCAGCGAGAACCTTTGCGCCCTTACCTTGAAGGGCGCGTTGAGTTTTAGTAGGGTCTTTTAGCCAGTTT
TCTATTGGTTGTTTTGACAGATTCCAGTCTGATATTAGGCTACCTGCCGACGTTGGAAGCTTAACTGATG
CTCCCTTCAAAGTCCCTCCTGTCGGAAAGATTGCTGACCCTGTTGTAGGGGGCGATATTACCGTTGTTGG
CATCTTAGCCGCAACCTGTTCAGGCAGGTTAGACGGCATTGATATTGTACTTGGTATTCCCATGACAGTT
TTACCTTAAATATTTACCTTAAAATGGTCTGGGCATGGTCTCAATTGAATCGAAAAAGATTTGGTCGAGC
CACTCTTTTATCATGTTATCCATGGTTGGGTTATCTTC

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession

no match

>Contig_138_length_3886_cov_1381.31_ID_275

TGGTATTCCCATGACAGTTTTACCTTAAATATTTACCTTAAAATGGTCTGGGCATGGTCTCAATTGAATC
GAAAAAGATTTGGTCGAGCCACTCTTTTATCATGTTATCCATGGTTGGGTTATCTTCTCGGTAGTATAAC
TCAAATCTTTTGCTCATCTTCCTAGAAGACTTCTCTCTGTTTGGGTCATTTAATCGGTATCTACCTCGTC
TATCTTGACCCGTGCGAACAATCCAAAAGCGAATAGGGTAAACCGTCATCTCTTCTGGAAACACTCCCGA
TTGAATGTTGTCAAGGTTTTTGACTTTTTCGGGTAACGTCATATTCTTGATCTGGTCATCAGAAACTAAG
TTAAATAAATCTCGAACAAATCGCTCAGATTTATTTTTGTCATCTGTTGAAAAACAGGTATTAGTCCCGC
TTCGCAGGTCAACCCAACACTTATATTTACCTAGCCGGATGCCACTACTTCCACCGCAAGCCGCCTTTAA
TTGCTTCCACGTCCAGCCAGATCGTTTGACCATAGGAATAGTCATCGTAACCGTTTTATATGGTTCACCG
GCTCTCCTTAATGGGGGTCTAGGCTGGTTGCACCATAAAATTTTCATACTGATTGTATGAGCCGCTTTGA
GTTTGTATGCTGGGGTACTTCCTTCAATTATTTCCCCTACATCCTTCTCATACATCATCTGGTAATTAGC
TAAGAAGTTTGTGATGTCATTTCTACCTATGTTTTCTCGCGTAGAAAAGTAATAGGGAGTTCCTTCTTCT
CTATTGGGAATTACTTCAATGAATTTTTGCTCACCCAATACGATATCGGGTTCCTTGACTCTATTCAAAG
CCATCTGGCTATCGATAGCATCCGCTACTACCATGAATCCTTCTTTGCAGGACTGCCAAAGCTCTTCTGA
AAATTCTTCTACAAATTCTCTAACATTTGGATTTTCAATATCTTCTATTCTGTCTCTCTGCCATTGTGCT
AGAGACAGGGTTTTACTCCCTTCTTCTCCCCATTTATCAAGAGCTTCCTGTCCGAACACCCTCTGGGCTA
TTCCATTCTTTTTAATCGCTTTTCTAGCATTTTGAAAAGCAAAATAGAATGCATGTTCTTGTCCTAATTG
ATAAATAGCCCTACCAGTTTGCACCCATTCCTGCTTGATTTCTTCGTATATCTCTTCCGAAACTTCATCC
CTAATTGCTAGTGCCATGGATGGGTCTATCTTAAAGGTCAAGGTATTGGCTATTTGACCGCACAATACCT
GCCCAACCGTTACTCCTAGCGCACCGCCTAACATACCAGCCCCTTGCTGGATAAGTTGCTGCCTCTGCAT
GTCAATTTGGGTATCTGTAGTGTTCCAATCAAAGTTCCACAAGAATTCAGCAGACTGCAATAATCCGCCA
ACGATTGTAGATACTCCAAGAAACTTTAATACATTCCCGGCGACTGCTCGAAGTAACCCGGCAATCCCCT
TCTCCAAAAAGGATTGAGCCAGACCCGTTCCTACACCAATAAGTTCAGCTTGCTTTTGTCGGTAAAAATC
TCTTAGTTCGTTCATGAGCCGTGCCCATAAACTTTTAGTAAAAACCTTTATACCGCCACTGCCGCCGCCT
CTATTTCCTCTTAATTGTGGTGGTAGCGCTAAAGAGCCAGGCATGACAAAAATCCTTTACTAATCTACTA
AAGAGTACCATAACTTAATAGTATATATAATTTCTTCCTTCGGAATTCTACCCATCCATCCTTTATAAAT
TAATAAACTACTTTCCTCTAAATCATTTTTGATAAAGACGGGTATTCGTGGCGGTCTAGTTATCGTTTTA
TTTATTGTGTTGGGAATGGTTTCTACCCAACGAGTAAAAATGATTCCATTATTGAATGTCTTCAAGCGAA
GCGCTAACCATAATCGACTTCTTACCAGCTTTGGGTCAGCNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
AGGTTCACTAAGAATTTCAATAGCATCATCGTCTACCAAAATTTTCAGTCCATCTATACCATTGGGCATA
ACTTCCACTGGAAATCGTGCAGATACTTGATAGATGTCTTCGTTGCCCGATTGAGTGAGCCTGCCTTGGA
CGGTTGTCCCCTCTTTCAAAACCAATCGCCCTTCCTTTCTGGCAACTTCAATGTCGCCAAATAAATTAAC
GATATCAAGTGCTTGTCCCATGTGTCCTTCTCCTGGTCGCTACATCTCAGTTATAGCGCGACTTTTCGCT
ATTTTCGTGATAAAATGGACTATCAGAACCCTATAGGAGTGTGATAGGACAGTATGAGTGTAATCGTCAT
TCAGAAAAAGGTAATTTGCGATAGTACAGAGCGGCAAATTATCCTAGAACCTCAAGAGCAAGACATACTT
TTTGAAGAGGACTTGGTTTTGCTTTATCGATTCCACTATGACGCTAACGATGGTGGCATTAAGATAACGA
CTCCCTTCGTAGAGCAAAGACCAGAGGATACAGTGGAGTGGGCAGCAGTAAAGGAATATTACAATTCTTG
GGCAATACCCCATATACTCCTTTATTTGTTCGTTTTCGGGCAAGAAGTAGACGATGACGCACCTCGACAG
ACTCTCGCGTTACCTAATTATGGTGTGCCTTTTCCGTTACCGCTCAACAATCCTATTCGCCTAAAAGCAG
GGCAAGGATTAGCCATTGAGTTAGTCAATAATGGGTATGATTTTCCTGATAGCAATAGACTTGATAATTT
CATTACTATTTATGGGGATGCTTGGCTGGGTAGGAATGAATTTAATGCACAAACGGTTAACGTGGATGCT
ACTAGCCAGACAACGGTTAACGCAGCAGTGCAAGAAGGAACACCACCAGAAATAGACCTTAGTCAATTAA
GGAAAGTTGGCATGGGTGATATCAGTTCGATTAAAGTTAAATTCAGAATGCCTTCTGGAAGTACCGGAGT
GGCTCTTTATGAACCATCGCCAACCATGAATAATGCAATTCTTTTTGATGATGATGAAGTGGGTTTATGG
CAGCCAGTCGCCAATAACGTGCGAGGAAATAATTCAGATTTAGATATCTTTACAACCCTTAGAAGTTGCG
GTCACGTTATCGAGGTTGAGTTTAATCAAACATTAACGATTGGTCAAATAAATATTGTTACTCCTGATTT
TCCAGATAAAGCCATGGGCAATATTAGGGCAACAGAAGATTATAGAGGATGGAAGCAAATACATACCGGT
TATTTGTCGGGAAATATTTCCCTACCTAAGCCAGAAAGGGCTAGAAAAATAGAAATAGTAGCTCCCTTAG
AAGATAATTTAGTTGATGGAGTTTTTAAGATTTATGAAATAAATCTATGGGTAGCAGAAGAAATATTAGA
GGTGAATAATGACCCGCTAATTATTCAACAGGCAACGTTTAGATTGAACCCAGAGAAGGTAAACACTTCA
TCTTTAGAAACTTATGAATCATTTACTCTCGCAGATGGCTCTACAGCGCCTTTGGTAAAAACAGAGGATA
ATCGAAAATATATTGAATGGAGGTTTACAGGAAATAATGATAAGAAACAGCTTGTTTCTTCTGAGCAATT
TGTAGTTAAGTCGTTATTAATAGAGTTACGCTCTCCTACTACCTCTTGGGAGGTAGATGGAAGCTTTGAT
AACGCTAATGGAGGTATGATAACTGGATTGAATCCACAAAGTTCTCCTTTATTCACTGCCCTCGGTGGCA
CTTCTTCAATAAATTGGTATACAGCCTACAAGAAGATATCAAGAAATAACATAACGTTAACAGGAGACAA
TGTCAGAGGAGTTGATTTATCTCCAATTGATAATTG

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession
no match						

>Contig_151_length_2086_cov_936.202_ID_301

TATTCACTGCCCTCGGTGGCACTTCTTCAATAAATTGGTATACAGCCTACAAGAAGATATCAAGAAATAA
CATAACGTTAACAGGAGACAATGTCAGAGGAGTTGATTTATCTCCAATTGATAATTGGTTCTGGGTATAT
ATCGAACTAAACGAAAGCTTTACCGACCAACTTATGTTAGGGTTGGGGTCTTATTCCTGGCATAGGGTTG
ATGTGGATGTGCGGCAAATATTAACGTATGATTCCTCACTTTCTGAGAGGAATATAAATGATGTAATAGC

TTATTTCGGTACTTCCCCATCGCCAACCCCTACTCCAGTACCTGTTCCCACGCCAACACCATCAGAGCCA
TCGGCGTACTTGGATAGCAACAACATCATTCTTAAGCCAGACACTTTAGCTGATGGTAATTTATCGAATT
GGGGCGGATATGTTCAACGTGGTTCTACGCAGCCAACAGTGACTAGCCATGGCAATTATAAGCTGGTCAG
TTTTTCCGGTTACTCCGAATTGGCGAAATCTCTTTCTAGTAAAGCAAATTTTTATGGCGTAACCTGCATA
ATAAAATCGCCCAATCCCTCTTGGAATAATTGGGGAAATGTTTGGTCTCATAATGTTCCCAATCATTCCT
GGACATTTGCACTTGATAAAGACGGGACAGCCATCGGGAACAATCCGCCTTTATCTCGTGTACGAGTAAA
TAAATCAGAGATTAATCTAATTTCCAACTGGGAACTGGGAACTATAACTAATCTCTTTTGCCTGACGATT
TTCTGGGATGCTTCAATGATTACCCATTCATCTGGGGATTTAGTTATTCACAAGAGCAGTATCTCAGATT
GGCAATGTGAGTATTCATTAGGGGATGTTTTAGTGTGGGGAAATATGCCATCTGATGCCGATTTAATTGG
CGTAGAGGATTGGATGATGGATAAATATTCAATTCTTTGACAGTACAATAAGGAAGAAGAACTAAACAGA
AGGAGAAGCAAGGAAGATGGACGAATTGTTACGAGGAATAACGTTGAGTGAGACTCTAGATTTCCTGGTT
TTATTGGGGGAGTTTTTTCTCGCTGTCATTATGAAGAAATTGAATGACAAATTAGAGAGGATGACTAGCA
AAGAGGGTGTCCAGATGGATGATGTGCTAAGTAAGTTGAACAAGATACAGCTTTGGGTGGATAGTAATTA
GTAGGTCTACCATGAACTTTTACCTTAAAAATTTAAGGTAAATTAACTTTCTCCTCATCGTCTTCTTGAT
GGGGTGAAGAATTTACCTTAAATTTTTACCTTAAATATCGCACTACACCCTCATCGAAGCATGTAGACGA
TGGCTGTCAATATCACCACAGAAGGTATCAACACGTTGAACACATCGATAACATCACGTGCTTTCAGCTA
CTTCTCAGCTTTCCTGAGGTATCTCTGTATTCTATTCATAGCTCTTAAATAGGTCGCAATACTCATCGAA
GGTGATATCAAGGATTCTGCAAAGCGTCTGAATCTGTTTGGGTTCTAATCTTGGAGGTTTATTGCTAGAT
ACACACCTTCTATATGTTGTCGCACTCATTTTGCATTTTTTTACAAAATCCTCTTGAATCAAATCCGTTC
TTTCTTTACGCAATCGCTCCAGAGGGTAAATATCATCAATTTCAACTGGTGCAAACAGTAGCTTTCCTTT
TTCTGTTCGTTGATTCATAAGTCAATTATAGACAAAAATCTATTGTATTTATAAGATTACAATGCTGTCA
ACCAAAACGACCGGTGCTACACTAAGGACATCGGAGCCAAGCACTCCTGGTCTAGAGCCTCTAAGCACTA
CTATCGCTTAACTGAAGGCAATCACCAAGTGTTGACAACCGATACTATCTTTTTGGGGTAAAACCTTTTC
ACGTTCTGTTATGCACCATTTTCTGCGTGTATATGCAGAGATGTCGTAACAGTCAAAAAAGGCTTACTCT
CCATTGTACCTTGACAATTTGGTTCACCTGTTGACCCTTCTTTCACTTACTCAATCTGAAGAAAGGAGTC
AACAAATGGTTAAACAACAACGCATTCCTGAAGTAGCAGGAGGGTTTAACGTTTTG

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession
transcriptional regulator [*Nodosilinea nodulosa*]	45.8	45.8	10%	0.007	33%	WP_017301714.1

>Contig_203_length_708_cov_1251.67_ID_405

TCAGCCGAACATTTATCACGCCAGTTTTTACTCTTGGCACGACGTTCGTCGGCATCTGAAACGGCAAGGT
TGGAAACGCTTCGATTGATGTTGCATAAACCTCGGACACCGTGTCCGCTATCAAACGTGGTTTTCTTTAG
GTCTTGATAAGACTTGCGATAATCACGTTTGTCTTTTTTGCTTGCCTTCCGATGCTTAGTGAGTCCTTTC
TGGAGGGTTGCACCTGCATCAAGGTACTCTTTACGGCTTCGGTTGATTCCAAGTGCTTCATCAATCGTTT
GCTGCAAGATGACTGCAAATGATGCTTCCTGCATCGACTTGGCTACTGGGTAACCTTTGTCGGATGCAAG
TTTCACTAGGAGGGTTAAGTCGGTTTGGGTGATGATGGAGATGGGCTTAGAGTTGACTTCGGTCAAAAGG
TTGCCCCTCTTGGAAGTCTCGCCAGCTAAGGCTTTGAAGGCTTCCGAGTTAAGTAAGCGCCCCACGACTC
TTTTATCAATTTCCAATCCTCTGGACACGCCACGCAGGCTAAAAACCACCTGATTGGTGCTTGGGTTAAG
GAAACCCTGAAACACCACACCCAAAACGTTAAACCCTCCTGCTACTTCAGGAATGCGTTGTTGTTTAACC
ATTTGTTGACTCCTTTCTTCAGATTGAGTAAGTGAAAGAAGGGTCAACAGGTGAACCAAATTGTCAAGGT
ACAATGGA

BLASTx	M.score	T.score	Q.cover	*E*-value	Ident	Accession
Nuclear factor, kappa-B-binding prot. [*Eufriesea mexicana*]	40.8	40.8	27%	1.5	33%	OAD62547.1

Appendix 5.7 Spacer details and protospacers. A = CRISPR array number. S# = spacer number. Length = basepairs of spacer sequence. S= Score cut-off at default value of 20, however, values of 18 and 19 were considered if protospacer originated from BBD relevant match (e.g. cyanophage).

A	S#	Sequence	[bp]	S	Potential target	Category
1	Spacer1	CCAGAAAACCTTATCTCCCAAAAACCAGTAGGAGGCAGGA	40	NA	NA	NA
	Spacer2	TTGTCCATTGGGAAAAGGAAGAATGCATTGTAGA	34	20	*Synechococcus* phage S-MbCM6 (NC_019444)	phage
	Spacer3	AGTACCCATTCCTTCCGCAGAACCGCACTAACGCAA	36	23	*Salmonella* phage FSL SP-101 (KC139511)	phage
	Spacer4	AAGCCTTTGAAATGGGCATGACCGAGCTAGAGGA	34	NA	NA	NA
	Spacer5	GCAGGAGAGCGCAGATATCGAGCGGCTGTAGAGC	34	24	*Ruegeria pomeroyi* DSS-3 megaplasmid (NC_006569)	plasmid
	Spacer6	GAGGTATAAATCCAGATACGATCCTGAAGGGCTGAGGGTTGC	42	NA	NA	NA
	Spacer7	ACAAGGCAATTCCCAGAGATATAGATTTCAAG	32	NA	NA	NA
	Spacer8	TTGATTAATCGGTCGATGGGATGAGCATTAGGATT	35	NA	NA	NA
	Spacer9	GATAGATACCTGAAATGGTTGGAGGGGAAATCATGA	36	25	*Bifidobacterium* phage Bbif-1 (GQ141189)	phage
	Spacer10	TTCAGGCAAAAACCTGTGATATTCAGGCTTGCCATTGGG	39	25	*Granulicella tundricola* plasmid pACIX905 (NC_015060)	plasmid
	Spacer11	CGGGGGGATTGATGGGGTAAACGATCGCGTTTGCCAGTCGAT	42	NA	NA	NA
	Spacer12	GGCAAGAAATATAATCACGGGTACTCTCACATGAGTTGAAA	41	NA	NA	NA
	Spacer13	CCCCAAAAAACACAATCAGAAGAATCCCCAGAAGAATCCCAA	42	20	*Apocheima cinerarium* nucleopolyhedrovirus (NC_018504)	virus
	Spacer14	CGTGCGAATTTCGTGCGAATTTCGTGCGAATTTCGC	36	NA	NA	NA
	Spacer15	TTGCAGGATTAACAATTAATGGACGAAGCAAGGGCGATCGC	41	NA	NA	NA
	Spacer16	CTCTTCTTTAACCCCGGACTGTTCCTCACCTTCTT	35	NA	NA	NA
	Spacer17	TGGTCTTTTGCAGCCAATGCGCGCCAATTGCGAGCGTA	38	NA	NA	NA
	Spacer18	AACAGTGATCGCGCTAGAGAATTGCAAGAATTAGA	35	24	*Stanieria cyanosphaera* PCC 7437 plasmid pSTA7437.02 (NC_019749)	plasmid
	Spacer19	AATCTTTCCGACTTAACTTATGAGCTGAATTCTC	34	NA	NA	NA
	Spacer20	TGAGCTAACCGCAGATCGGCTGGCAAAAATGATCAA	36	22	*Azospirillum lipoferum* 4B plasmid AZO_p1 (NC_016585)	plasmid
2	Spacer1	TTTACTGTAGCTCCATAATTCGAGTTATTGACGATATGCTT	41	27	*Anabaena cylindrica* PCC 7122 plasmid pANACY.01 (NC_019772)	plasmid
	Spacer2	CTTAAACGCCAGCCGATTCATTTCGATAAATTTTTTGCT	39	NA	NA	NA
	Spacer3	ATAACATTTCCGTCCGCTTATCTCGATGGTGCTGATTGGC	40	27	*Ralstonia eutropha* JMP134 megaplasmid (NC_007336)	plasmid
	Spacer4	GATCGGGCCCCCATCTGGGAATGCCCCGAGTTT	33	NA	Contig_93, hypothetical protein	self-target
	Spacer5	AATTTTATTTGGATCGGGTTTTGGGTGAGGATTTGCCTCT	40	21	*Vibrio* phage 11895-B1 genomic sequence (NC_020843)	phage
	Spacer6	TAAAGGCAGTGTGCTGATCAAAGCTTTGTAATCAGC	36	NA	NA	NA
	Spacer7	CCAAAGGGTCCCCACTTACTGGATGTCTTGAAATATCCA	39	NA	Contig_93, hypothetical protein	self-target
	Spacer8	TCACCTGGTACGTCCAGCCCTTATCCTCTGTAAA	34	21	*Meiothermus silvanus* DSM 9946 plasmid pMESIL01 (NC_014213)	plasmid
	Spacer9	TATTGGCTTGTAGCCTACCTCCTGCCCTTCATTACCGCCAAT	42	NA	NA	NA
	Spacer10	TTTCCAATAAGGGGGTAATTGCCTCTTATGCCAA	34	NA	NA	NA
	Spacer11	ATGATTGAATAGCTGGATAATTCTAATTGCCTTA	34	NA	NA	NA
	Spacer12	GCTGCGGTGAGGCTGGCTTTTGTGCTATGGGTT	33	22	Beet curly top virus – California (NC_001412)	virus
	Spacer13	TTCCGTATTCTTGCGAAATACATATTTCTTTTTTGTT	37	21	*Bacillus thuringiensis* Bt407 plasmid BTB_502p (NC_018878)	plasmid
	Spacer14	CCGCGTATTTGGAATCCAACGATGACCCGATATGA	35	NA	NA	NA
3	Spacer1	GAGAAAGTGAAACAACGAAGCTTTTCAGCGATTAGCCCCC	40	NA	NA	NA
	Spacer2	AAGAAAGCCTTGGACGATACTGAGGCAACTGTGGAT	36	20	*Escherichia coli* K-12 plasmid F DNA (NC_002483)	plasmid
	Spacer3	TATGAAGAAAAGATGGAGCATCTCAACGCAGCAAT	35	24	*Spodoptera frugiperda* MNPV	virus

					virus (NC_009011)	
	Spacer4	AGAGATTTTAAGAAATGCAAAGCAACCAATGACAGCAAACTTG	43	NA	NA	NA
	Spacer5	GTTTTCTCAGGACTATGGGGGCGAAAGTTTTCACTTAT	38	23	*Acaryochloris marina* MBIC11017 plasmid pREB2 (NC_009927)	plasmid
	Spacer6	TTTTATTTTGCACTGAAGTACTCCCCTTCTCGTTCGTATT	40	23	*Leptospira* phage LnoZ_CZ214 (KF114877)	phage
	Spacer7	TTTAATGGCGCTTACGCCATTTGCAATTTATATCTCATT	39	NA	NA	NA
	Spacer8	AGGGACTTCCTGCACGATGCCTTCTTCAAGCATTTC	36	NA	NA	NA
	Spacer9	TGCAACGCCAAATTGAGAGAGATGTGGCCCGCCAGT	36	NA	NA	NA
	Spacer10	ATCCAACATCCTCTGAAGGCTGTACCACGGGATGTATGA	39	NA	NA	NA
	Spacer11	CAAATTTTCTACTTGCAAGCGCTCCATACTTTCAA	35	28	*Sclerotinia sclerotiorum* dsRNA mycovirus-L (NC_017915)	virus
	Spacer12	CAAACCTATTTTTACTATTTCCGCAATTTATGGCAA	36	NA	Contig_93, hypothetical protein	self-target
4	Spacer1	GCTACTTTTGACATTACTGTGTCCGGCACCCC	32	19	*Cyanobacterium aponinum* PCC 10605 plasmid pCYAN10605.01	plasmid
	Spacer2	CTAGAGACAGAAAATTACTCGAGGAAAAAATCATGTCT	38	22	*Burkholderia phymatum* STM815 plasmid pBPHY01 (NC_010625)	plasmid
	Spacer3	TTGAAGTGACTGGAAGGGAACGTATCGTCACTGG	34	NA	NA	NA
	Spacer4	TTACCGCCTACCTTACTAGGACGGTATATAGAGA	34	NA	NA	NA
	Spacer5	CCAACACACCCAAGCGGGTGCAATTCTGTGAGGAGATGCGGGGC	44	18	*Cyanophage* KBS-M-1A genomic sequence (NC_020836)	phage
	Spacer6	GGCACCGCGTACAAAATTTCCGTGGACATCTTTA	34	NA	NA	NA
	Spacer7	TTACCTCCAGATTGCCCCTGATTTGGTAAAACCATCTCTTAA	42	NA	NA	NA
5	Spacer1	TCAGGTAGACAAAGTGAGGTTAAGCGCCTCCCATTA	36	NA	NA	NA
	Spacer2	TCTTGCCATAAGAGATTTAATGATGTCAAATCAAGA	36	NA	NA	NA
	Spacer3	ACTAACAAGCTTTTATCCTCATGCTTATGAAGAA	34	22	*Loktanella* phage pCB2051-A genomic sequence (NC_020853)	phage
	Spacer4	TGCCCCCACCAGCCGAAGCGAACAGCAAACCCAC	34	24	*Micrococcus* sp. V7 plasmid pLMV7 (NC_022599)	plasmid
	Spacer5	AGATTATGATAAGTCAAATGCGGGCAAAGCCCAACCCCT	39	NA	NA	NA
	Spacer6	GAAACATTCAGAATCAAATGGGAGGTTGATTGGGA	35	NA	Contig_93, hypothetical protein	self-target
	Spacer7	CCGGCACCAGATTTTCCGGATGAATTCCTAGAAG	34	NA	Contig_93, hypothetical protein	self-target
	Spacer8	GCGTATGGACACCTTTTCATCACCATTGATATGTT	35	20	*Acidianus* two-tailed virus complete viral genome (AJ888457)	virus
	Spacer9	GGGAAATTCTCCTTTCGGAGCGCTCGGTGAAATTCTTG	38	NA	Contig_93, hypothetical protein	self-target
	Spacer10	ATTAAATTTTATTTCCATTTCAAAGTTCCACAACCCAT	38	21	*Staphylococcus* phage StauST398-3 (NC_021332)	phage, temperate
	Spacer11	CAACATTATTTGGGACATAGAAATATCAGGAACACCTTGATA	42	30	*Cyanothece* sp. PCC 7424 plasmid pP742401 (NC_011738)	plasmid
6	Spacer1	CAAAAGATTCGCCAGTCACCAATGCCTGAGTTAC	34	NA	NA	NA
	Spacer2	GCAATAAGAATCAGGATGACAACCGCCCAAAATCCGATAA	40	24	*Sinorhizobium fredii* HH103 plasmid pSfHH103e (NC_016815)	plasmid
	Spacer3	CTCCTCCGACTAGAACACTACAATGCCCATCAACTGCTGGA	41	NA	NA	NA
	Spacer4	CCAAGCCCTCACCGCAGCCGCCTATAGTGGTC	32	21	*Paracoccus aminophilus* JCM 7686 plasmid pAMI5 (NC_022043)	plasmid
	Spacer5	AAGTAATTACCAGAAAGACATATTGGAGTGGGTGAA	36	21	*Erwinia amylovora* CFBP 2585 plasmid pEA3 (NC_020920)	plasmid
	Spacer6	TGAAAAACTAGAAACGGCAACGGCTCGTGCTATGCAAA	38	20	Enterobacteria phage EK99P-1 (KM233151)	phage, virulent
	Spacer7	AGTGGATCTGGTGGCGATCGCTAAAATAATGGTAGA	36	NA	NA	NA
	Spacer8	GGCAGAACCTACTACATCCGTAAGAAAATCGCCCTG	36	NA	NA	NA
	Spacer9	CTTGAGCAATTAGGCTTGCCCAAGGTTGCCATAAT	35	NA	NA	NA
	Spacer10	TAATCACCACAACATCAGGACTGACGATTCTAA	33	NA	NA	NA
	Spacer11	CTGGGTTAAGAAGAGGATTCCATTCGCTATAATCAAAATCAT	42	NA	Contig_93, hypothetical protein	self-target

	Spacer12	TTGTCATCAAGATATCCCCACGCCTAGCGAAATCTT	36	NA	NA	NA
	Spacer13	CTCTTCCTCGCCTATTGGGGAAAACCAGACCCCGGT	36	NA	NA	NA
	Spacer14	GCATAACAAAATACGATCGGGAAACATTGAAGAGCCCAAT	40	NA	Contig_93, DnaB domain-containing helicase, VIRsorter: Phage_cluster _71 _PFAM-AAA_25	self-target
	Spacer15	CGCACTGGCCGATAGAAATCATTTTCATCGCCAAGGG	37	26	*Cyanothece* sp. PCC 7822 plasmid Cy782202 (NC_014534)	plasmid
	Spacer16	TGGCTATAATTCTCTCTGGATAAGGGTTTCAAGCTTT	37	23	Uncultured bacterium plasmid pEFC36a (NC_025088)	plasmid
	Spacer17	CACAAAAATGGGGGATAGCTTACTTCCCCCTGCGA	35	NA	NA	NA
	Spacer18	TCTGGGAACGAGGTGGAGTCGCTGGGGATGGTTTCGAGT	39	NA	NA	NA
	Spacer19	AGACGGATTAATCGAACTCAGTAACGACATCCAA	34	38	*Natrialba* phage PhiCh1 (NC_004084)	phage, temperate
	Spacer20	GTGTAGCTTTACAGGTCGCACTTATGGAATGGAAA	35	NA	NA	NA
	Spacer21	CAAGCGCTTTTGGGGGATGGATGCCGTGGGGAGTT	35	NA	NA	NA
	Spacer22	GAAGAATTGTCAACAGAAGATGCCGAATATGTGGC	35	19	*Synechococcus* phage S-ShM2 (NC_015281)	phage, virulent
	Spacer23	TTGCCCTAAATTGTTTTTTAGTCTTCTGCGATAATC	36	26	*Staphylococcus epidermidis* plasmid pSWS47 (NC_022618)	plasmid
	Spacer24	CAAAAAGCGATCATTTCCCATCGCTCTATGGAGCGATCGCAC	42	NA	NA	NA
	Spacer25	AATTTAGAAACCTTGATGCCTTGATCGCGCCCTACCG	37	NA	NA	NA
	Spacer26	GTGTCAAAATGGCAAGTGCGCCCACTTCCTCAA	33	NA	NA	NA
	Spacer27	AGACAAAGAATACAGCGAAGGCGAACCGAGCGAACCA	37	NA	NA	NA
7	Spacer1	TGGAGGATTTGGACTGGCATCGAGAAAAAGTACA	34	23	*Sinorhizobium fredii* NGR234 plasmid pNGR234b (NC_012586)	plasmid
	Spacer2	TTACATTGCCTATTCTTTTGAGCAAGATGATGACT	35	24	Influenza A virus 107399 (H9N2) (NC_004911)	virus
	Spacer3	TATCTATGAGCCAGGTATTTCCGATGAAGAATGTTATA	38	NA	NA	NA
	Spacer4	GTATATTTTCCTCCCAGATTGGCTAAGGAGTAACT	35	20	*Methanobacterium formicicum* DSM 3637 (NZ_AMPO01000001)	NA
	Spacer5	TGCAGGAGGAAAATAGGGGAAGGACATGGTTGATAGA	37	19	*Synechococcus* phage Syn19 (NC_015286)	phage, virulent
	Spacer6	CTTGAGATTCGTTGATGTAGGTTTGAAGCCTACCG	35	NA	NA	NA
	Spacer7	ATGAAAAGTACTGACAACTATTAATACTGTACAAA	35	NA	NA	NA
	Spacer8	ATCATATTAAATTCACTCATGCCATTATTTCTGCTCAG	38	NA	NA	NA
	Spacer9	GAAGATAGAAATCTAAATCAGTCTGCAAGCTTTGAGT	37	NA	NA	NA

Geitlerinema sp. BBD_1991

A	S#	Sequence	Length [bp]	Score	Potential target	Category
1	Spacer_NoG1_1	GTCTTCGTCGTTCTGAGAAGAGTGCCAGAGGTAG	34	NA	NA	NA
	Spacer_NoG1_2	CTGCGGCGCTTCTTCGTTCCCGAGAAGTTGCGATCGGACG	40	22	*Cupriavidus metallidurans* CH34 megaplasmid	plasmid
	Spacer_NoG1_3	CTTCAATCGTTACGTCCCCCCCCCTCAGCTTCCACGA	37	NA	NA	NA
	Spacer_NoG1_4	CTGCCTTGATGGGGGACTCAAACGCCATCACCCG	34	NA	NA	NA
	Spacer_NoG1_5	CGGAGATGCTCGCCGTATTGGCTCTGCAAAACCG	34	NA	NA	NA
	Spacer_NoG1_6	AAGTCGAGCCCATTGTTGTTGAATGCACGGAGGAGGA	37	NA	NA	NA
	Spacer_NoG1_7	TCGAAAATCTCCGAATTCAGTTTTGATGATAAATTAGTA	39	NA	NA	NA
	Spacer_NoG1_8	GTTACGTCAGAACTCCCAGATCGGCGCATTCAGG	34	NA	NA	NA
	Spacer_NoG1_9	TTTATCGGATTTAGAGCCGAGTGCCTCGCATTGTTCCAATATGTTCTG	48	20	KT895374 *Bacillus* phage vB_BpuM-BpSp	phage
	Spacer_NoG1_10	TCCTTATCGACTTCTACCCCAAATCCCCACCC	32	NA	NA	NA
	Spacer_NoG1_11	CTGGGATAATCGGGGGCGGAAACCCGCAAAACAA	34	NA	NA	NA
	Spacer_NoG1_12	TAAAGAAAAAAACCTCGAAACCTATGGGAACTAAC	35	NA	NA	NA
	Spacer_NoG1_13	CTTGAGGTCGGCGAAAGGTAAACGATCGCGAATCCCG	37	NA	NA	NA
	Spacer_NoG1_14	CCCCCTGATATCTTCCGTATCAGGGGGGTAAGGAG	35	NA	NA	NA
	Spacer_NoG1_15	ACATCGGCACGTGACGGCGAAAAATATCTCAAAC	34	NA	NA	NA

Spacer_NoG1_16	CAGACAACAATCGGCAATTTCCAGTCGAGGCAGATTTTC	39	NA	NA	NA
Spacer_NoG1_17	AACAGTTGAGCTTCGACACCGAGTCAATCGATCTGTC	37	NA	NA	NA
Spacer_NoG1_18	GTTGTGTTCGGTGCAGCAACACTACACCGAACAC	34	NA	NA	NA
Spacer_NoG1_19	TCTGATGCTGAACTTCAGGCCGAAATTCATGCGATG	36	NA	NA	NA
Spacer_NoG1_20	GGCTCTAATCGAGTCTCATACCATCGCGCCAAATCCAA	38	NA	NA	NA
Spacer_NoG1_21	ACCTTCGATATGATGCTCGTCATCATGAAA	30	NA	NA	NA
Spacer_NoG1_22	GGTCGGACGGACAAGCCAGTAGTCTCGCCCTCCTC	35	NA	NA	NA
Spacer_NoG1_23	GCGTTGTTTCCAGCCTATCGCAGTTTGCAACAGAT	35	NA	NA	NA
Spacer_NoG1_24	CCCGTTGCTGAAGTCGGGCGCGATGTAAGCCGCCC	35	NA	NA	NA
Spacer_NoG1_25	TGCCCTTATTTCGCCTCCTACCACCGCTCTGCCACCATCCA	41	21	*Halobacterium* sp. JI20-1 plasmid: II	plasmid
Spacer_NoG1_26	GGGGACAAAGAGCGGGTTCGGGATTACTGCTTT	33	NA	NA	NA
Spacer_NoG1_27	ATTCGCCAAATTCGCCACCTGTCACGGCATCACG	34	NA	NA	NA
Spacer_NoG1_28	TCTTTATCACGCATGGATAAGTGACGATGTTCACC	35	NA	NA	NA
Spacer_NoG1_29	GAGGCATGGCGATCGTTGTACTGACCGATGGCGATTC	37	NA	NA	NA
Spacer_NoG1_30	GGGTCGGCCAGGATTGAGGTTTGCTTTCACCTCAAATTT	39	NA	NA	NA
Spacer_NoG1_31	TCGGTTAGCGTTACGGCTCCATCCGCAGATTTCACCAC	38	NA	NA	NA
Spacer_NoG1_32	CTTCGAGGAAGATAAGATTTTCCAAGTCGATCTCCA	36	NA	NA	NA
Spacer_NoG1_33	TCGAAGCCCTTGCCAAAGGGGCAGGCTTTGACCACCTGGAG	41	NA	NA	NA
Spacer_NoG1_34	CAGTCTTTCACATGGCTATTACGCCAGCGATCGAC	35	NA	NA	NA
Spacer_NoG1_35	TTGCAAGCCTTGGCACTGGACCGGACGGGTACC	33	NA	NA	NA
Spacer_NoG1_36	CTTCACACGGAAGCGGGTCTGAGGAGTCGCCAAGCTTCAG	40	NA	NA	NA
Spacer_NoG1_37	AGTCAATACACAAATATAGTCTATGAACAGATTAA	35	NA	NA	NA
Spacer_NoG1_38	TCTAATTAGTTCTAAAGCTCTATTAGATCCTTTGGC	36	NA	NA	NA
Spacer_NoG1_39	TATAATGGTCATCGGGAAAGAGTGCGGGTATAGCAA	36	NA	NA	NA
Spacer_NoG1_40	TACTACTTCTTATGCTCTAGAGATTCAAAGAATCAA	36	NA	NA	NA
Spacer_NoG1_41	TGTAACAGGGAAGAACAGCCGGATCAATAATGTAAAGATCC	41	NA	NA	NA
Spacer_NoG1_42	TAATGCATGATTTTTGTATTAAGGTAATCTTTTCT	35	NA	NA	NA
Spacer_NoG1_43	ATAGAAGAGAGAGAGGAGATTTTAAAAGCTATT	33	NA	NA	NA
Spacer_NoG1_44	CTTAATCCTATAGTTTCAAATGTTACGTTAAGCGCTATTCC	41	NA	NA	NA
Spacer_NoG1_45	AGGCATTAAATACTTGAAAGGCTAAATAGCCTTG	34	NA	NA	NA
Spacer_NoG1_46	GAAAACTGGTTAAGAAATAATGGCTAAAGTTTTT	34	20	C2PVCG_L48605 Bacteriophage c2	phage
Spacer_NoG1_47	GATTTATTACTTACATCCCCGGCATCAGGCATATG	35	NA	NA	NA
Spacer_NoG1_48	CTTAGGGGAGGCTAACTGAGATGACTAGATACGTTTTA	38	NA	NA	NA
Spacer_NoG1_49	TGCATTCGCCCAAGCTTGGGAATCACCGAAGTCGAT	36	NA	NA	NA
Spacer_NoG1_50	AACAAGCTTTCCTTAAGACTTTAGAAGATGCCCCTA	36	NA	NA	NA
Spacer_NoG1_51	CTTATACGAAGGGGTATTGCAATCTTACAGAG	32	NA	NA	NA
Spacer_NoG1_52	CAAAAGTCTTTACATTTCCAGAGAAATCTAATAT	34	NA	NA	NA
Spacer_NoG1_53	GCGAGTAAGCAGGTAAAGTTACCAGATTATGGAAATTT	38	NA	NA	NA
Spacer_NoG1_54	GGGAGTTTCGGCAACGCCCCTGTCGAAGTTGTTCTTGGT	39	NA	NA	NA
Spacer_NoG1_55	AGTGAATATTTGAAGCTTCCACCGTTGGAAAAAACT	36	NA	NA	NA
Spacer_NoG1_56	GGGAACTAACTGGCGATACGCCGGGTACGAGCC	33	NA	NA	NA
Spacer_NoG1_57	AGCATCTACATAGTCAGATTGTGCGGTTAGGTACAG	36	NA	NA	NA
Spacer_NoG1_58	TTTAGTTTTTATGTCCGGCTCATTAGTATCCTACTT	36	NA	NA	NA
Spacer_NoG1_59	CCACCCTCATATATGGCCCGCCTGGTTCGGGGAAATC	37	NA	NA	NA
Spacer_NoG1_60	TCGCGATCTGTGACGAGGGTGTTGGGTTGAAATGG	35	NA	NA	NA
Spacer_NoG1_61	TAAAGGTGAGCTATTATAGTGTACCTTTCCACCCT	35	NA	NA	NA
Spacer_NoG1_62	GATTAAAGCCATCACTATTATAAGCATTGAAAGTCTC	37	NA	NA	NA
Spacer_NoG1_63	TCAAAATCGTTGAAAGACTTTGCTGGTGGTAGTTT	35	NA	NA	NA
Spacer_NoG1_64	ATATATATTGGTAATCCGTCAGCGTCATTTACTTT	35	NA	NA	NA
Spacer_NoG1_65	CCTTAAATTTTTTTGTGTTAAAATAGTTAATATAT	35	21	*Staphylococcus aureus* NCTC8532, plasmid: 2	plasmid

Spacer_NoG1_66	TTAAGTCAACCTATAACTTCAGCACCTAAAATAAACCA	38	NA	NA	NA
Spacer_NoG1_67	GTAAGATTTTGAGCTAATACCCTGTAAGTCTTAGA	35	NA	NA	NA
Spacer_NoG1_68	ATTTGGTTAGAATCGTTCTCGACGTAGTAAAATTT	35	NA	NA	NA
Spacer_NoG1_69	ATTAGAGCTAGTTGGTAAATCATCCAATCCACACATCAG	39	NA	NA	NA
Spacer_NoG1_70	GAAAGGCTCTTTTGCCCCTAAATCAGTCTTGTTTGTT	37	NA	NA	NA
Spacer_NoG1_71	AAGTATAAGCATTTTTGGTATTGACACCTTAATCTCA	37	NA	NA	NA
Spacer_NoG1_72	TAGATTCCACCACTCGGGTAGTTTACAACAAATCT	35	NA	NA	NA
Spacer_NoG1_73	TTTCTATAATTTTTGAGTTCTTTAGAATAATAATTTTT	38	22	JN258408_JN258408 Megavirus chiliensis	virus
Spacer_NoG1_74	AGTGCCGCACCAAATCCCCAAGGTTCACCAGAAA	34	NA	NA	NA
Spacer_NoG1_75	TCAGAAGCCGTTTGGCTCCCTTGGTGCCTTTCGCTTG	37	21	*Meiothermus silvanus* DSM 9946 plasmid pMESIL01	plasmid
Spacer_NoG1_76	TTAATATCTATCGCGAAGCGATCGCACAATGCCTTAAATCT	41	NA	NA	NA
Spacer_NoG1_77	CTCCAGCGGCAGAGGGTTGGCAATCGGTTCGTCA	34	20	*Escherichia coli* ST131 EC958 plasmid pEC958	plasmid
Spacer_NoG1_78	AATTCGGGGCTGTACTCCTCTAGCAGTAATCCTG	34	NA	NA	NA
Spacer_NoG1_79	CTTTTAAGATTGGGTTCACGGCTACCCCTTACCG	34	NA	NA	NA
Spacer_NoG1_80	ACGAAGAAGTATCGCGCTGGGTGACGGGATTTTGG	35	NA	NA	NA
Spacer_NoG1_81	TTTGCCGCTCCAATTCCTCCAGCTTGCGTTGGAGTGCCAG	40	NA	NA	NA
Spacer_NoG1_82	GTCTGCAAAGATGCACTCAGGGTTATTTACTAACTG	36	NA	NA	NA
Spacer_NoG1_83	GATTGTACAAAATCAACGGGAAGAATTAGAAAATAGAA	38	NA	NA	NA
Spacer_NoG1_84	TTACTTTAATCAATCCCTCCCAATTACCTCCAATAGATG	39	NA	NA	NA
Spacer_NoG1_85	TTTTGTCCGATTCGATCGAGAGACTCTTTTGCCTCCCGATCGACT	45	NA	NA	NA
Spacer_NoG1_86	AAAATTCCTCCGATCGCCAAGCCGAAACCGACACC	35	NA	NA	NA
Spacer_NoG1_87	TTGTGCTTCCGATTGTGGCTTCTTACTTACTTCAATCGCCTC	42	NA	NA	NA
Spacer_NoG1_88	TTGACGGAAGGGTTGAAATCGCTAAAGGCGTCTGCGA	37	NA	NA	NA
Spacer_NoG1_89	TCGATCTCCGTCACGCAAGGTATATCCATTTGTGA	35	NA	NA	NA
Spacer_NoG1_90	TTTGTTTGGTACAGATTGATTATTAGAACCTGG	33	21	HM595733 *Spodoptera frugiperda* MNPV isolate Nicaraguan	plasmid
Spacer_NoG1_91	TTCATAGCCAGGCATGATTGACCAGAGTGATTTGCCATT	39	NA	NA	NA
Spacer_NoG1_92	TCCTTTTCTACTCCACAATGTTGAATTAAAGTCTC	35	NA	NA	NA
Spacer_NoG1_93	CCTTAAGAGAACTGCAATCTGGACACTTAAGTCCAGA	37	NA	NA	NA
Spacer_NoG1_94	TTTATTTGTTCATAAATTACATTAGTACTAAGGGTT	36	NA	NA	NA
Spacer_NoG1_95	GAGGTCGTAACTTATGATAGAATTCTAGATGAATATTT	38	NA	NA	NA
Spacer_NoG1_96	CATCTTTATATCCGCTGGCAGCTTTGCCACCTAGACCG	38	NA	NA	NA
Spacer_NoG1_97	GTTCTGTCTGTTTTTGAAGCAGCCGCGATAAGAA	34	NA	NA	NA
Spacer_NoG1_98	GTAGCAAAGGAAGCTTCCTTTGCTATCGAGACAA	34	NA	NA	NA
Spacer_NoG1_99	GCAACGGGAAAAAGGAAAAATGCAATCATCCCTA	34	NA	NA	NA
Spacer_NoG1_100	CGGTTCTCGACGATCGCGGGTTCCCCCAGTTGCAT	35	NA	NA	NA
Spacer_NoG1_101	GAAAGAGGACTATCAGATCCTCTCCAACGGGCAACGG	37	NA	*Sphingobium* sp. EP60837 plasmid pEP2	plasmid
Spacer_NoG1_102	TCCACGCCGTCCAGCTCCCCACGGTGGCTGTAAATCC	37	NA	NA	NA
Spacer_NoG1_103	TTGATGCTTGAGTGCGAAATTTTCTATCCCTTTGA	35	NA	NA	NA
Spacer_NoG1_104	GTTATAGGGACGACGACTTATCCGAGGACTGGAA	34	NA	NA	NA
Spacer_NoG1_105	TGCAAAGTGCGATTGGCTGTGGGAGGCAATTCAA	34	NA	NA	NA
Spacer_NoG1_106	CTTATTGGTATCAGCGAGCTAGGTTGGCTCTCACTGATTTG	41	NA	NA	NA
Spacer_NoG1_107	TCTGTTACCCAATTTGGAGCGGGATCGAGGACGATCTC	38	NA	NA	NA
Spacer_NoG1_108	CCGTACCATAGAGATAGCGCGATCGAGGCAAAACTC	36	NA	NA	NA
Spacer_NoG1_109	GCCGAGTTGAAGGCGATCGCCGCCAAACTGATGG	34	20	*Rhizobium* sp. N541 plasmid pRspN541e	plasmid
Spacer_NoG1_110	TCGAACACGGGCAAGGCAGCCCCCTCAACGTCTTC	35	NA	NA	NA
Spacer_NoG1_111	GAGAAGAAGTTCGTCACGGGCGGCGCAGTTGGCGAGACT	39	NA	NA	NA
Spacer_NoG1_112	ACCAACAGACGGAGCCACGCCAGACACCCCAGACGGA	37	NA	NA	NA
Spacer_NoG1_113	GGATAAAGGACAGATGAATATCGATAACTTTCTT	34	NA	NA	NA
Spacer_NoG1_114	TCTAAAAATACATCCCACCAGTGACCGAAATCAA	34	NA	NA	NA

Spacer_NoG1_115	CCAGCGCCGTGGCGCTTGCCCCCGTGTCCGTTA	33	NA	NA	NA
Spacer_NoG1_116	CTTTGCTATCGCTGTAGAGCAAGCAAAGGATACG	34	NA	NA	NA
Spacer_NoG1_117	GAAGGTAACGCTCGGCCGGGGCGATTGTTTTGGCTCG	37	NA	NA	NA
Spacer_NoG1_118	TTGTAGGTAAACACACTCCTCAATCAGATCTAGGATCT	38	NA	NA	NA
Spacer_NoG1_119	TTGGAACAACTTGGGGACTAGAAATCATTGGATTAGGTAA	40	20	*Clostridium botulinum* 202F plasmid pCBI	plasmid
Spacer_NoG1_120	TTAAAGTCCCCCATCCTAGAACTGGAAATTAAG	33	NA	NA	NA
Spacer_NoG1_121	GAGAAACAAATACCCAATTTAAGGGCGCAAGTTGCTCAG	39	NA	NA	NA
Spacer_NoG1_122	CTGAAGGTTATGCACCTTATTATGTTAGGTACGGAGTATTT	41	NA	NA	NA
Spacer_NoG1_123	TCGGAATGATACTCGTCAAGCCTTCCTGTCCCGT	34	NA	NA	NA
Spacer_NoG1_124	TTGATTCAAAGAATCATTTACGGTAGTATCAAAG	34	NA	NA	NA
Spacer_NoG1_125	GGGTTCAAGCTTTAGAGTACATTCTCAAGTCTGCTACC	38	NA	NA	NA
Spacer_NoG1_126	ACGAGATTTTCATTAAATTTTATAATGGAGATATTTTCTTA	41	21	*Peptoclostridium difficile* NCTC13307, plasmid: 2	plasmid
Spacer_NoG1_127	ATATTTTCGATTAGTTCATACATCGCTAAATTCAGATG	38	NA	NA	NA
Spacer_NoG1_128	GGAGATATGTGGAGTGATTTAATCCACGCCCACATC	36	NA	NA	NA
Spacer_NoG1_129	TTTGCGAATGAGACTACGGGTTTCATGGTAGATA	34	NA	NA	NA
Spacer_NoG1_130	TCTGCCCCGATCTGGAGGGGTATCTTTTAGAGAATCCA	38	20	*Mycobacterium chubuense* NBB4 plasmid pMYCCH.01	plasmid
Spacer_NoG1_131	TTCTACGCCTCCACCCCGTCGTCGGAAAACTACGCCAAATTC	42	22	*Granulicella tundricola* MP5ACTX9 plasmid pACIX902	plasmid
Spacer_NoG1_132	CTCAAAAGTTTTATCTGAGAGAGGGGAGCCTTTT	34	NA	NA	NA
Spacer_NoG1_133	GGAAATTGGCTTCGGGGAACCAGCAGACAACACT	34	NA	NA	NA
Spacer_NoG1_134	TAACCTATTTAAAAGAGGATGATTAATATGAT	32	22	*Clostridium botulinum* plasmid pCB111 DNA strain: 111	plasmid
Spacer_NoG1_135	GCTTAACAAAGCCATCTAAGTTAGTAGTGTCTTTGTAAGA	40	NA	NA	NA
Spacer_NoG1_136	ATTGATGGCTGTTCCGACTATAGATAGGTCGTAGTTATCTAA	42	NA	NA	NA
Spacer_NoG1_137	TTGAAACTGAATCTTATAGGTTCACAGGAAATTAT	35	NA	NA	NA
Spacer_NoG1_138	TCGGTAAGACCCATCCTCGAAAAGTTCAAGTGCGT	35	21	KU568494 *Mycobacterium* phage Bactobuster	phage
Spacer_NoG1_139	TATGCCTGGATGGACAGCCCGTATTGCACGCACTTCT	37	NA	NA	NA
Spacer_NoG1_140	GATGAGCCACTAGATAGTCCATATTAGCTAGAAAA	35	NA	NA	NA
Spacer_NoG1_141	TCGTTTTTACCTCAATGCTATTGAGGAGATGGATTGT	37	NA	NA	NA
Spacer_NoG1_142	GTATCCCTAGATCTCGTTAATCTAGGTGATGTCA	34	NA	NA	NA
Spacer_NoG1_143	GATGGAGAAGAAATTTACTCTGCATCGTTCGCAGGG	36	NA	NA	NA
Spacer_NoG1_144	TAGATGTAGAGAATCTTATAGCCCAAACTTAAAAATA	37	NA	NA	NA
Spacer_NoG1_145	TTAGACGCGAGGCAATGTAGAATATGCGATAAAAACTAGTT	41	NA	NA	NA
Spacer_NoG1_146	AGCTGTTACGACAACGCTATTGAGGAGATGGATTGT	36	NA	NA	NA
Spacer_NoG1_147	TGGCGATCGATTTCGTCAAAGTAGCGATCGTATTCGTCTTGGCT	44	22	*Haloterrigena turkmenica* DSM 5511 plasmid pHTUR01	plasmid
Spacer_NoG1_148	CTCACGCCAGCGATTGGGAGGCGGGGGGGACGATC	35	NA	NA	NA
Spacer_NoG1_149	CGGGTACGACTTCGGTTCGTATTCGTAGTCGTCGGG	36	NA	NA	NA
Spacer_NoG1_150	AATAGAGGGTTCTATACGACCATTAGCTACTTATAGA	37	NA	NA	NA
Spacer_NoG1_151	ATATTGACGCCCCTTAGAAATATTGGGGTTTTCAAGTCT	39	NA	NA	NA
Spacer_NoG1_152	GAGCTAGTCGGTAAATCATCTAAACCACACATCAG	35	NA	NA	NA
Spacer_NoG1_153	ATATAGAAGAGGGTAAAAAGAAGGCTAGAAGGTTAGGT	38	20	JQ340389_JQ340389 *Vibrio* phage pVp-1	phage
Spacer_NoG1_154	CTCTTAAGGTTAAATGAATCTACACTGAGGGAGGTTTT	38	NA	NA	NA
Spacer_NoG1_155	CTCAGCCAATCTTAGTGTCTCAGCTCGTTTAGATTCA	37	NA	NA	NA
Spacer_NoG1_156	ATATTTTTACATCACATCCACTCGAACATATTCTC	35	NA	NA	NA
Spacer_NoG1_157	TGGAAAATAGATGAAATGGATTGATTCCTCCTGG	34	NA	NA	NA
Spacer_NoG1_158	CTTAGATCAGTTGAGGGGGTATGTTATAGCCTCAG	35	NA	NA	NA
Spacer_NoG1_159	GAAGAATCAATGGCAGATTCCACGATATGACTTTGG	36	NA	NA	NA
Spacer_NoG1_160	GACCTGAATCTAGATCTTCAAAATTGCTAGACCACACA	38	NA	NA	NA
Spacer_NoG1_161	AAGCACCGATCCTATGCACGGTAAAGTTACCAGGGG	36	NA	NA	NA
Spacer_NoG1_162	GCGGCTATAAAGGAGGTTCGTCCGAACCTCCTTTATT	37	NA	NA	NA

Spacer_NoG1_163	TTATTGTTTTCTATCTTATAAAACTACTTGATCTAA	36	20	*Ralstonia solanacearum* CMR15 plasmid CMR15_mp	plasmid
Spacer_NoG1_164	ATGTCGTAGCATTAGCCAAATTAGACGCTTCTAA	34	NA	NA	NA
Spacer_NoG1_165	CTCAATATTAGTGTAATTGCAATCACGCACTCATT	35	NA	NA	NA
Spacer_NoG1_166	TCGCCAACAGCGCCGGAGGGATGCCGTTCTTCTCG	35	21	*Sinorhizobium fredii* NGR234 plasmid pNGR234b	plasmid
Spacer_NoG1_167	TATGCGGTATCGTTGGTGTCGTTGACTTCTATGGC	35	NA	NA	NA
Spacer_NoG1_168	GAGGGACTAGATCCTCTACATGGTGAGGTCGTTTTGGG	38	NA	NA	NA
Spacer_NoG1_169	AAGTTTAACACCAACAGGTAGTAAAAAAGGTATTGAAA	38	20	*Bacillus cereus* strain CMCC P0021 plasmid pRML04	plasmid
Spacer_NoG1_170	ATTAACAGAATCTTTCATTTCCTCAGTTACATTTCCACCTTGT	43	21	*Campylobacter* iguaniorum strain 1485E plasmid pCIG1485E	plasmid
Spacer_NoG1_171	TCTTAATTCTATAGAGATTTGTTAGCCAATGCCAAAAAATT	41	NA	NA	NA
Spacer_NoG1_172	GATTGAATTAATCGAAAAAAGCAAGGATTTGGGG	34	22	DQ092789 *Enterovirus* E isolate Jena 3802 polyprotein gene, partial cds	virus
Spacer_NoG1_173	TTTTCTCTGTATCTAAGTCAGAAGTCGTAAAAGTTGAAAAA	41	NA	NA	NA
Spacer_NoG1_174	TAGGTAAATTTTCATTTTTTTTCTTCCATTGACTAGATTCAATAT	45	19	*Cyanothece* sp. PCC 8801 plasmid pP880101	plasmid
Spacer_NoG1_175	CCTGAAAATATTACGGCCGTTGCTACAGGTAAATC	35	NA	NA	NA
Spacer_NoG1_176	CTGAGGAAGATATTTATAATGCCATAGACAAGTGGAACGA	40	20	KT968831 *Pseudomonas* phage YMC1102R656	phage
Spacer_NoG1_177	GTGGAAAAGCAAGTACCGGGGTCAGTGGAGATCA	34	NA	NA	NA
Spacer_NoG1_178	GAGAATAGGACTCATGCGTCCTTTAGCCAATTGAGGACTAA	41	NA	NA	NA
Spacer_NoG1_179	CTGAAGACAAATTAGATAATATTAGTCAGACAAA	34	20	*Cronobacter turicensis* z3032 plasmid pCTU1	plasmid
Spacer_NoG1_180	GATACGGTGACAAACTTGGCGGCGGCTACCGGGGGAACC	39	NA	NA	NA
Spacer_NoG1_181	CTGTGAAGCATCCTAGCCATCATGGAAACCTGTTAGCAGA	40	NA	NA	NA
Spacer_NoG1_182	GGTGGCGTCCTTCCCATCCCTGTAAGGCTGTAGGGTG	37	NA	NA	NA
Spacer_NoG1_183	TTGCCGGATTGTCCGCCCCTCTACCCGCGCTCGT	34	NA	NA	NA
Spacer_NoG1_184	TTAAAATCTATACCCAAAAACTTAACTATGGATCAGGTAGT	41	19	*Anabaena cylindrica* PCC 7122 plasmid pANACY.04	plasmid
Spacer_NoG1_185	TTTCTGCCGCCCTCGATGCCGAACTGGAGATCCTCGAAG	39	NA	NA	NA
Spacer_NoG1_186	TTTAGAATTTACCGAAGCTGATTACAAAGAGTTAAGAGAA	40	20	*Jeotgalibacillus* sp. D5 plasmid	plasmid
Spacer_NoG1_187	GGAAAAATGTGTAACTATTTCCCACAAATGGCTGA	35	NA	NA	NA
Spacer_NoG1_188	TGCCAGGAGAGGATCTCCTCTCCGTTTTTAAAGATAG	37	NA	NA	NA
Spacer_NoG1_189	GGGAAATATCCCTCTGTCCACAAGATCCCGCAC	33	NA	NA	NA
Spacer_NoG1_190	AGTAGCTTTCGCGTCGGATGGTGCCGAAGCGATCG	35	NA	NA	NA
Spacer_NoG1_191	ATGTTTTGTTATCCGAGTCATAATCAAACAACATAG	36	20	*Rhizobium* sp. N541 plasmid pRspN541e	plasmid
Spacer_NoG1_192	AGGAAGAAGTCGAAACGCTGGGGCGACTAATTCAC	35	NA	NA	NA
Spacer_NoG1_193	GTACACCAACCCCGGATCGTCGGCGAACAGCAAG	34	NA	NA	NA
Spacer_NoG1_194	TTTCTCTATTTTTAATTTTTAGTTAACGGTTGAATAG	37	21	KU556803 *Faustovirus* strain D3	virus
Spacer_NoG1_195	CCCAGGGGTTCGAGTCGCCATCGAAAATCATGAATATT	38	NA	NA	NA
Spacer_NoG1_196	GCCGGAAAGGTTTTTGGCGGTGGCACATCTAGCAATGA	38	NA	NA	NA
Spacer_NoG1_197	GGAACGAAAGCGTACCAACAGGCGCGCGAGGAATTC	36	NA	NA	NA
Spacer_NoG1_198	TTAGACTTCCAAAAAAAAGTCTAACAAAAAGTCT	34	20	*Borrelia valaisiana* VS116 plasmid VS116_cp32-2-7	plasmid
Spacer_NoG1_199	AGCTTTCTCGCGACCTGGGCGCAAGTCAACCCTCG	35	NA	NA	NA
Spacer_NoG1_200	ATTAATCGCGATGTCCGTATATTGGCGTGAGTATTCGC	38	NA	NA	NA
Spacer_NoG1_201	AATATGGTCAGCTCGACCATATCTCCCGTGTCATG	35	NA	NA	NA
Spacer_NoG1_202	AATAGGAATCCCTCCACTCTCTGATGAGGAATGGGAT	37	20	VSVGLYPW_M35229 Vesicular stomatitis Indiana virus strain 85-GM-B glycoprotein gene	virus
Spacer_NoG1_203	CCTACTGAGACACAATTAGAATTCTTATACTCTAAATTCTCA	42	NA	NA	NA
Spacer_NoG1_204	TTTAGATCTAAATCGTAATAAGGAGAGGGCTGAA	34	NA	NA	NA
Spacer_NoG1_205	CATTTTTCTTTCTCCCATCGATTAGAATAAATAT	34	NA	NA	NA
Spacer_NoG1_206	TAGTTACCAGGGGGGCAGAATCCTACTCGTGGAA	34	NA	NA	NA
Spacer_NoG1_207	CTGTGCAAGGTAAATTTGCATGAGGATCTCCTTCACA	37	NA	NA	NA

Spacer_NoG1_208	TAGAGCTACCATGGTGTTTATGTCTCCTATTTAGACTCTG	40	NA	NA	NA
Spacer_NoG1_209	ACCATTATGAGGAGGGTTCCTCATATGAACTCAAG	35	NA	NA	NA
Spacer_NoG1_210	ACATTCAAAATCATCATAGGTTTGATTTTTAATC	34	NA	NA	NA
Spacer_NoG1_211	CTCAAAATTAACCGATAATCCAAAGCTTATAGAA	34	NA	NA	NA
Spacer_NoG1_212	ACTAGTCCGCCAATATTCTATAATGCAGTAGTTTTC	36	NA	NA	NA
Spacer_NoG1_213	CCCAGTAGCCATCGGGGTTCGGTGTTAGAGGTACAA	36	NA	NA	NA
Spacer_NoG1_214	TAGTGTGACTTCTTCTTCGTTACCACACTGATGGGGTATATTT	43	NA	NA	NA
Spacer_NoG1_215	TTTACCTTGGGGAAACGACAATATGATGACCCCAGAA	37	NA	NA	NA
Spacer_NoG1_216	CTTCACCCTCTCCCGCAGTCGCGAAATCTAATCGCTTCCCTAAAGCTTCAA	51	NA	NA	NA
Spacer_NoG1_217	TTTAAAGATTCTTCTATTTCTGCTTTTATTCTCT	34	20	*Ornithobacterium rhinotracheale* plasmid pOR1	plasmid
Spacer_NoG1_218	TTGGAAGCCCCGTCGCTTTAGCGCGGGGATTGGTTACACAACTTGTCAGATA	52	NA	NA	NA
Spacer_NoG1_219	TGTAAACCCAGTTGTGCCAAAATTGGTTCCG	31	NA	NA	NA
Spacer_NoG1_220	ATTCTGGCGGATTGACATAGGCATTGCTACTGCAC	35	NA	NA	NA
Spacer_NoG1_221	CATTGCCCAAAGCGTTCGGCGTCTGTTCGGCG	32	NA	NA	NA
Spacer_NoG1_222	GAGGGTGTGGAGAGTCACGCTGAGCGCCTCGTCGAT	36	22	DQ184476 Orf virus strain NZ2	virus
Spacer_NoG1_223	TGGAGGGGAGGTGGTAGTACCTCCCCGGTTGTGTTA	36	NA	NA	NA
Spacer_NoG1_224	GCTACTCGTCCTCGCCGTCTTCGAGCCACAACTC	34	NA	NA	NA
Spacer_NoG1_225	TCCAGATTTTGGGCGATACTCGGAGGAAGTAGGCGCG	37	NA	NA	NA
Spacer_NoG1_226	TGTATAATGGATACAAACAGACACAGCAGAACCT	34	NA	NA	NA
Spacer_NoG1_227	CTTATTATCAAAATTAGAATCGGGGCAAAGAAATCT	36	NA	NA	NA
Spacer_NoG1_228	GTAAGATAAGATGCCCGAAATCACAAATCCTCGCGCTGT	39	NA	NA	NA
Spacer_NoG1_229	CGGCGTTCCGATCGAAGGCGATCGTCGTTGCATCGA	36	NA	NA	NA
Spacer_NoG1_230	GCATTCGTTGCAAGGTTCTCCATCAATTGGGTTCA	35	NA	NA	NA
Spacer_NoG1_231	ATTCATTTAATTAAATCGTAATACCTATAGAAAAT	35	NA	NA	NA
Spacer_NoG1_232	ACTCCTTTACTTAAGAGGTAATCTGATAATTGGT	34	NA	NA	NA
Spacer_NoG1_233	CTGAGTTGCCGTAATTGAATTTGTGTGCCGATTATAAGT	39	NA	NA	NA
Spacer_NoG1_234	TTGTGGCATTCGTTGCAAGGTTCTCCATCAATTGGGTTCA	40	NA	NA	NA
Spacer_NoG1_235	ATTATTGGGACGATACAGATTGGATCTAATTTTTG	35	NA	NA	NA
Spacer_NoG1_236	TAAAGAGGATCTATATTGGGTACAGAAAGAAAAAG	35	NA	NA	NA
Spacer_NoG1_237	GTTAAATTTTTCGGTTTTGAATGTGGCATGAAGGA	35	NA	NA	NA
Spacer_NoG1_238	TTTACGCCAAGGACGTAAACTAGTTTTAGCCGCT	34	NA	NA	NA
Spacer_NoG1_239	CGGCGTCAGGCATATGTACCAGTGTCTGTGGATTGAAC	38	NA	NA	NA
Spacer_NoG1_240	TCGGACACAGATTAAAGAACATGAGGCCTCAATG	34	NA	NA	NA
Spacer_NoG1_241	GGGCTTGGCTAGCATCCTCGGGAGTTGTCGTCACTCAG	38	NA	NA	NA
Spacer_NoG1_242	TCGGCTCCGTCAAGTTAGCCCGATCGCGCTTCGACAT	37	NA	NA	NA
Spacer_NoG1_243	TTTGGCAAAGGCATCCCGCGCCTCTTGAGTGTTGA	35	NA	NA	NA
Spacer_NoG1_244	TTGTCAGTGTTCCACCCGGATAAGTATAAACTAA	34	NA	NA	NA
Spacer_NoG1_245	TTTAAAAGGCTTTGGTGGTGTTAGTAACCCCGTTGGTCTAA	41	19	AB626962 *Staphylococcus* phage S24-1 DNA	phage
Spacer_NoG1_246	ATTTTTGCAAGAGTGCCTTCCTGTCTTAGTCCTC	34	NA	NA	NA
Spacer_NoG1_247	GCTGATTTAATTTAGGCTTTAGATCGTCTAATTCGATTTCAA	42	NA	NA	NA
Spacer_NoG1_248	GGTTCCGCCCATGCCTTTGGCGTACTCCGTCCAGTAAGG	39	NA	NA	NA
Spacer_NoG1_249	ATTAGATCCCGGTAGGCCAGCTACAAAAGTACTGA	35	NA	NA	NA
Spacer_NoG1_250	ATAGTAGGGTTAGGATCCACTTACTATGGCTTCC	34	NA	NA	NA
Spacer_NoG1_251	GAAGAAAAATTCTTTAGTTTTTACTTTAGGTTCCTTAG	38	NA	NA	NA
Spacer_NoG1_252	GTGGTGATTTTCGGGTGGATCTTTCACCCTATCGAG	36	NA	NA	NA
Spacer_NoG1_253	TTGTCGTACCTTCAGATAATTGTGATTGACGACTTTTGAC	40	NA	NA	NA
Spacer_NoG1_254	CATATATTTTAAATATCGACACTTTACTTTTGAGTTAG	38	NA	NA	NA
Spacer_NoG1_255	GCTATGAGCCGGACATCCAGCCTGGTGTACCTGATGA	37	NA	NA	NA
Spacer_NoG1_256	CCCAATTCTACTAATTCTCTAATAGAAGCACTTTT	35	NA	NA	NA
Spacer_NoG1_257	ATATCAGGTTCCCCCTCCACGGTCTACTCTTGCCC	35	NA	NA	NA

	Spacer_NoG1_258	GATAGCCCAATTCTACTAATTCTCTAATAGAAGCACTTTT	40	NA	NA	NA
	Spacer_NoG1_259	TTATTGGATTAGGTAGTACAAAAGTAGCTAGAAG	34	NA	NA	NA
	Spacer_NoG1_260	AATAGTAGCCATATAAAACTATAAAATAAAAATTAC	36	20	*Helicobacter cetorum* MIT 00-7128 plasmid pHCW	plasmid
2	Spacer_NoG2_01	CTTTTTAAATTAGGTCAGCAAGCCTCTGACCTC	33	NA	NA	NA
	Spacer_NoG2_02	ATCACTGGAGACTTCTCCCTCTATTGGGTTCCTA	34	NA	NA	NA
	Spacer_NoG2_03	AAAAGATTTTTAGATGAGGAATAGTCAGCGAGTTGATCTGCTGGAGC	47	NA	NA	NA
	Spacer_NoG2_04	ATCTTCATACCAATTAATATAAACATTTAGGATACG	36	NA	NA	NA
	Spacer_NoG2_05	GCAGTTTCTTATGGAAGAAAGGGTGATTCAATCTATC	37	NA	NA	NA
	Spacer_NoG2_06	ATTTGAGAAGCTATGTCAGGGATTTCGACATCTAATT	37	NA	NA	NA
	Spacer_NoG2_07	CTGCAACCTAGAAAAATAAATGTCTTATTTAGAA	34	NA	NA	NA
	Spacer_NoG2_08	TTCGCCGGGAGAGCCGTTGGCGGTGTGGCGGACGAA	36	NA	NA	NA
	Spacer_NoG2_09	GACCGACCGACAGGCGATCGCGATTCAACTCAGC	34	NA	NA	NA
	Spacer_NoG2_10	TTGTGTTTTTACTTTTGTTAGATTAGGGTTTATAACTC	38	NA	NA	NA
	Spacer_NoG2_11	TGGACATCTCGGCTTTAAGGTAATCATTGACTCTT	35	NA	NA	NA
	Spacer_NoG2_12	TGCAAGATAAAGGACATTAAAGATCCTCAAAACCGC	36	NA	NA	NA
	Spacer_NoG2_13	GAAAAACCGTGGTTCAATCATCACTTTCCTCTTCTT	36	NA	NA	NA
	Spacer_NoG2_14	TGTGCTATGACATGTGCTATCTTTTTCTTACCCTTGAAAAGCTGATT	47	NA	NA	NA
	Spacer_NoG2_15	TGTGCTATGACACCACTTCTCCTTA	25	NA	NA	NA
	Spacer_NoG2_16	GTGAATAATAAAATATCATCCCATCAATGGTATCAC	36	NA	NA	NA
	Spacer_NoG2_17	AGTGTAATTACCGGGGGCGTTAGTAAGTACTGCCCCCT	38	NA	NA	NA
	Spacer_NoG2_18	ATGTGGTTGCAGTCTCCCATGATTCCAACTACAAA	35	NA	NA	NA
	Spacer_NoG2_19	AAAAATGTAATTATCAGTGACGCTACGGCTAAAG	34	NA	NA	NA
	Spacer_NoG2_20	ACCCGGTTAGGTTGTGTCAAACTATGCTTGTCACAA	36	NA	NA	NA
	Spacer_NoG2_21	ATTACAACAGGTAGAGGATGGCCTGAGGGGGGATACGG	38	NA	NA	NA
	Spacer_NoG2_22	TTCTAAATCAATCCCACTTGTACGCCAATCCATGA	35	NA	NA	NA
	Spacer_NoG2_23	GGCAAACAACCTGTCAATACAGGGGAGCGACGCAAGGTAAG	41	NA	NA	NA
	Spacer_NoG2_24	TAGCATAATTTCTACTAACATATACATAAGAAC	33	NA	NA	NA
	Spacer_NoG2_25	CTCTAGGCGGATAATGCACATCAGAGATCAAATCCC	36	NA	NA	NA
	Spacer_NoG2_26	CTCAGGAGATGGGGGCTCAGCCTGTGTATTTGA	33	NA	NA	NA
	Spacer_NoG2_27	CTTCCCTTTGTTTCTATACCCAATATAGATCGGT	34	NA	NA	NA
	Spacer_NoG2_28	GCTCTACCTTAGAATCAAAAACTAATCCTTGTGAGGAAGA	40	NA	NA	NA
	Spacer_NoG2_29	TTAGGTAATATTAAGAAGCTCTGCCATCATTGGAAT	36	NA	NA	NA
	Spacer_NoG2_30	TTTAGTTGCATCCATTTTTAGTCCTTCCCATTTA	34	NA	NA	NA
	Spacer_NoG2_31	CCCACTTCCCGGTATTAAGGAGGGCATTCTTTACTTC	37	NA	NA	NA
	Spacer_NoG2_32	TGGCAAGATTTCGATCGGGGATACACTCGGAACCTC	36	NA	NA	NA
	Spacer_NoG2_33	TACAAATACTATAAGACTGTGTTACATTATATGTGT	36	NA	NA	NA
	Spacer_NoG2_34	ACTACAGTGGTAGGCAATTAAATATCATGGTTAAA	35	NA	NA	NA
	Spacer_NoG2_35	AACGCCTTTTGGAAGGCGGCGGCAAAAGTAAACGAGT	37	NA	NA	NA
	Spacer_NoG2_36	AAACCTTGGAAGGCTACACGATAACTTAAACCTTTA	36	NA	NA	NA
	Spacer_NoG2_37	GTTAGACAGCTTGTCCTTGTTTCTTGTTTTTGCCC	35	NA	NA	NA
	Spacer_NoG2_38	GAAAGATGCTCCAACTGTTCCTACTGACATCCCGCCT	37	NA	NA	NA
	Spacer_NoG2_39	TTATAGTTATAAAAACTTATCTAACCCATAGGATAAA	37	NA	NA	NA
	Spacer_NoG2_40	CTCGGATTTCACCGTTGTCTTTAAGTTTCACGAAACT	37	NA	NA	NA
	Spacer_NoG2_41	ACGTCAAGCATCACCATGTATTTCTTCTTACTGCTA	36	NA	NA	NA
	Spacer_NoG2_42	ATATTCCTAAAGGATACATAAAATTAAACTATAAA	35	NA	NA	NA
	Spacer_NoG2_43	TAGTAGCTACGAAGATTTAGATTCAGCTTATGAAA	35	NA	NA	NA
	Spacer_NoG2_44	TGTTGCCTGATAATCTTCAATAGCCTTTTTAATCCT	36	NA	NA	NA
	Spacer_NoG2_45	CTTTTTCTGTTTCACCTTCCACCCATCTTTTCCT	34	NA	NA	NA
	Spacer_NoG2_46	GAGCCATAGCGTGGCTGTGGGCGACTCCATTGAAA	35	NA	NA	NA
	Spacer_NoG2_47	CCTTTATTTTGGTTCTAGAAATAATACCTATTGCATTAGT	40	NA	NA	NA
	Spacer_NoG2_48	TCATTTTCTGGAATTAAGTACATATTGGATTACT	34	NA	NA	NA

Spacer_NoG2_49	ATGTATTTAATAAGTGTGGCTAATAGGTTAAATAAA	36	NA	NA	NA
Spacer_NoG2_50	TCAATAGTTGGATCGTAACTGTAATCTATTCTATAGT	37	NA	NA	NA
Spacer_NoG2_51	ACGGGAAGTTGCATTTTTTCCGCCCATTCGTTTAC	35	NA	NA	NA
Spacer_NoG2_52	ATGATGAGGTGGATTTTCAATCATTAGATGGTGCGGGTT	39	NA	NA	NA
Spacer_NoG2_53	CTGACGGGGTAGCTAACAGGTTATTCAATGCAAA	34	NA	NA	NA
Spacer_NoG2_54	ATTATTGCCCTATTTCTAGGTAAAATATCTTGTAGA	36	NA	NA	NA
Spacer_NoG2_55	CAATAGACTAATTATTGTTATTGTAACATTGACAA	35	NA	NA	NA
Spacer_NoG2_56	GGCATCGTCAGCTACTACTTCTCTCCAAAGGTATC	35	NA	NA	NA
Spacer_NoG2_57	CTCAAGACCTAGTATAGTTAGATTAACCGATTCAGT	36	NA	NA	NA
Spacer_NoG2_58	ACTATCCCCAAGCATCCGAAATACAGAGAGCATA	34	NA	NA	NA
Spacer_NoG2_59	AGCTGTGATAGTATAATCTTCACCCAGAGGATACCAA	37	NA	NA	NA
Spacer_NoG2_60	TGAATCAAGTTGGAATTCTTGCCAAAACCCTGTATA	36	NA	NA	NA
Spacer_NoG2_61	CTAGATTTACGAATGCTTCCTCTTGGTAGACGT	33	NA	NA	NA
Spacer_NoG2_62	TTTGCATCGCGAAGTTAAGAGATACTCTGACGAT	34	NA	NA	NA
Spacer_NoG2_63	CTCTTAGAGTTAATAGATACTGATGAAAAATTGG	34	NA	NA	NA
Spacer_NoG2_64	CTGGTGGAGGTTGCGGATAGTCGATATCAATCCCT	35	NA	NA	NA
Spacer_NoG2_65	ATAAGCGATCATGTCCTCTAAGCTACTGTAAATGGCTT	38	NA	NA	NA
Spacer_NoG2_66	TTGAGTCAATTCCTTAAGAAAGTTCAAACACCTGGAAC	38	NA	NA	NA
Spacer_NoG2_67	TATAAAGATGATAGATACTTTAATGATGTCTTAAA	35	NA	NA	NA
Spacer_NoG2_68	ATCGACGATGTCGATCCGTGGACTTTGGAAGTGTCGGT	38	NA	NA	NA
Spacer_NoG2_69	CTCTCTCGCCACCACACTTGGGGACGCTTACATCGTTCCTCAA	43	NA	NA	NA
Spacer_NoG2_70	CCTTCAATTGACGAAGACCCCTCCCAACAGACCCC	35	NA	NA	NA
Spacer_NoG2_71	CGATACTTTTTCGAGGTTCTCCCCGAAAGCGTTG	34	NA	NA	NA
Spacer_NoG2_72	ATGTAAAGTTGGGTTGACACCCTGATGAGGGCTTGA	36	NA	NA	NA
Spacer_NoG2_73	TTTGTATCCAAAATAGATCGGTATAGGGAGGTTGTCAA	38	NA	NA	NA
Spacer_NoG2_74	ATCTATGTCCATTTTTTGTACCCGTCTCCTTTGTTTA	37	NA	NA	NA
Spacer_NoG2_75	AAGTTAGAAATACTTTATGGTTCTTTGACTTAAAC	35	NA	NA	NA
Spacer_NoG2_76	GTGATGATTTTCTTAGTCATGGTTGTTTCCTTTTCTT	37	NA	NA	NA
Spacer_NoG2_77	AAACATCTGTTTAACAAAAATTTAATATTAATCAT	35	NA	NA	NA
Spacer_NoG2_78	ATGGATAGATAAGATAAACTCTTCATCCATATTT	34	NA	NA	NA
Spacer_NoG2_79	TGGCGGTAAGTATTTCGATTCTTTTCTTAGGAGTAA	36	NA	NA	NA
Spacer_NoG2_80	TCCCACAGGTTCCGGCAAGACGTTTATGTTTTGCCAAA	38	NA	NA	NA
Spacer_NoG2_81	ATTCGTGAATCTCAGATTTACTATAGCAAGTATAAGGAA	39	NA	NA	NA
Spacer_NoG2_82	TTATATGGCTACTATTGGACACAAATCAGAATTGT	35	NA	NA	NA
Spacer_NoG2_83	CCCGTCTCTTCACGGTGACAGATTGGGTCTAGA	33	NA	NA	NA
Spacer_NoG2_84	GGGTCTGGTCTCCTCTACCCCGATGACTGATTAAAT	36	NA	NA	NA
Spacer_NoG2_85	ACCTTCCCCTACATCTGGGATATCTTCTAATAGT	34	NA	NA	NA
Spacer_NoG2_86	AGAAAAGGAAAAATACTCACCTACTTCTAAATCCTC	36	NA	NA	NA
Spacer_NoG2_87	AACAATTAGAATCATGGCGACTGCGATACTCATA	34	NA	NA	NA
Spacer_NoG2_88	GTGGTGTCGTCGTCGTGGTTTCCGGTGTGTTTCCA	35	NA	NA	NA
Spacer_NoG2_89	CGATCACCCTATCGCTCAAGGGCGGTGTGGTGT	33	NA	NA	NA
Spacer_NoG2_90	AATAAGCACGAAGTCGCCATCGGGGAGTCGGGTAACAA	38	NA	NA	NA
Spacer_NoG2_91	TTAAATAACCGTAAGCGCGCCAGCACTCACGATAAG	36	NA	NA	NA
Spacer_NoG2_92	GGCGATCGAGATACGGCAGATTGGCATGATTAGGTCT	37	NA	NA	NA
Spacer_NoG2_93	AGTTCGAGCTTTTACCTTCCGAACTCGCCTAACAACC	37	NA	NA	NA
Spacer_NoG2_94	TCCGATCTTCATCTTAATAGCAATCAGACTCGCGTAATG	39	NA	NA	NA
Spacer_NoG2_95	CGAATTGGGCTGAAGCAAAATAGCGGTTGAAACCC	35	NA	NA	NA
Spacer_NoG2_96	ACAATGAAGGTTTGAAGCCCGGTGGCTGGTGCGGGAG	37	NA	NA	NA
Spacer_NoG2_97	TCCTGCCTATGTAATCGGTACGAAAAGACGCTTAA	35	NA	NA	NA
Spacer_NoG2_98	GGAAACTTTCGGTTAGCCCTCGGTGCTGATACCACCGA	38	NA	NA	NA
Spacer_NoG2_99	TTTTGAAAGGCTTTAACAGTCGCCTTCCCCGAAACA	36	NA	NA	NA

	Spacer_NoG2_100	GCTCTGAATTCGCTCCTGCTTTTGGCGAAGACGAGT	36	NA	NA	NA
3	Spacer_NoG3_01	TCTCCTGCACCACAGTAGGGGTGGAACACTGTTGAGC	37	NA	NA	NA
	Spacer_NoG3_02	TCTGCGATGTCCGCCCACGGCATTCGCGTAGGAATCTC	38	NA	NA	NA
	Spacer_NoG3_3	GCGGACTTTGTGAAGATCTCGATAAAGAGATTCTTCGAAAAA	42	NA	NA	NA
	Spacer_NoG3_4	ATCGAGCGCATACCGTCGAGCTGCTCGCCCTCTTC	35	NA	NA	NA
	Spacer_NoG3_5	GGGGGATTTCAGGCAAGAGATGTTTTCAAATTCTTGGTTTA	41	NA	NA	NA
	Spacer_NoG3_6	AAGAAGTGCTCGCTCAAGAGAGCGATATAGCGCT	34	NA	NA	NA
	Spacer_NoG3_7	TTTCTACATTGGGAGCGATCCAGAAAACCCA	31	NA	NA	NA
	Spacer_NoG3_8	CCTCATACCTTAGCTGGAAATCGTTAGTCACCGCCGTCCGA	41	NA	NA	NA
	Spacer_NoG3_9	CCAACGCCCCGGAAACCGTACGCGGTGCGATCG	33	NA	NA	NA
	Spacer_NoG3_10	TTCACTTTAGGGGAAGAGGGTTTAACATCACCTAGATT	38	NA	NA	NA
	Spacer_NoG3_11	CTGTCGGCGGTTTTCGGGGATGCCGTTTTTAAGA	34	NA	NA	NA
	Spacer_NoG3_12	TGACCAATTTAACGCCCTCGAGAGTGAAGCGTT	33	NA	NA	NA
	Spacer_NoG3_13	CACCGCCCGGCGGGGCGCAAACACAGCAACTCCTTC	36	NA	NA	NA
	Spacer_NoG3_14	GGTAACAAATCAGGGCGGTGTCGAGGCGGGATTCAA	36	NA	NA	NA
	Spacer_NoG3_15	AAGATCGACCCCCCGACGATGAGGAGTCCGCCGATCGCACC	41	NA	NA	NA
	Spacer_NoG3_16	GTTCGCCCCTGGCACTTGTTGCGATCGGGCTGTTCTCAAA	40	NA	NA	NA
	Spacer_NoG3_17	GCCTTATCGAACGCGGTTTCGATCGGCGGAAGCGTTGCC	39	NA	NA	NA
	Spacer_NoG3_18	TTGATGATGGCATCTATTCGGAGAAAGCTGAAAG	34	20	*Rhizobium* sp. N941 plasmid pRspN941a	plasmid
	Spacer_NoG3_19	TGGCGATCGCTCCCACCGAGTCAATTAGGACA	32	NA	NA	NA
	Spacer_NoG3_20	CTGTAACAGTGTAAAGTCCGTTGAAGCCATAAAAA	35	NA	NA	NA
	Spacer_NoG3_21	TTAAGGGATTTCGGAGAGCAATTGTTTGCTACTACGTC	38	NA	NA	NA
	Spacer_NoG3_22	GTCGGGTGTTGTTGACTCCGATTCAATAACATTG	34	NA	NA	NA
	Spacer_NoG3_23	CTACCTTCCTCAAGATTAGAGTCTTCATTTTCCAT	35	NA	NA	NA
	Spacer_NoG3_24	AAGATTTATGGAGAAGGGTATAGAATCTTGATTA	34	NA	NA	NA
	Spacer_NoG3_25	AGTTTCAGGGCAAATTAATATTAATATTCTACAT	34	NA	NA	NA
	Spacer_NoG3_26	ACTGAATAAAGCTAAAGCTAGTATCAAAGAGGCAGA	36	NA	NA	NA
	Spacer_NoG3_27	AATTTCTCTTCTGATAAACCTTGTCTAGATCCATACTCCAA	41	NA	NA	NA
	Spacer_NoG3_28	AGTGGAAAAGCAAGTTCCGGGGACAATGGATCTCG	35	NA	NA	NA
	Spacer_NoG3_29	GATCCGGTTATAGCTTTTTTAGCCTGTTTAGCATCG	36	NA	NA	NA
	Spacer_NoG3_30	CTACGACTACCTCTGTTAAAGCGTCGATATCATCATCATT	40	NA	NA	NA
	Spacer_NoG3_31	GAGGGGTTTCCGAAAATCGAATTTCTAGGGTTTGAA	36	22	*Bacillus thuringiensis* strain KNU-07 plasmid pBTKNU07-01	plasmid
	Spacer_NoG3_32	ATATTTGGCAAAAGGGCAACAATGTAATCGATTTA	35	NA	NA	NA
	Spacer_NoG3_33	AATGGGTGAATTTTCCCATTTTAGAGATGCAATCTAA	37	NA	NA	NA
	Spacer_NoG3_34	CTAGGGCTTTGCAAAAGCCCTCATTCTCTTGTAGGTA	37	NA	NA	NA
	Spacer_NoG3_35	TTTATTCGGCTTCCTTCGTAGGAAATGTAGTTTGCCCC	38	NA	NA	NA
	Spacer_NoG3_36	GATTTGTAAAGCTAGTTGTCCCGAAATTAGTGCCA	35	NA	NA	NA
	Spacer_NoG3_37	GAATTTTAAAGGTAGTAGGCTCCAATGAAATCAC	34	NA	NA	NA
	Spacer_NoG3_38	AAATCCCCTCTTGAGAAGCATAATTTAGAATTTT	34	NA	NA	NA
	Spacer_NoG3_39	GAGGTAGCACAGCCAGGCGAGCGGCGTCGAGTTGGGATGCCGT	43	NA	NA	NA
	Spacer_NoG3_40	GCTTCCCAACTGTCGATCGTCAAACTCGAACAGTTG	36	NA	NA	NA
	Spacer_NoG3_41	CTCTAGCAAAGATACTACTCCACTCGAGAAGGAAGT	36	NA	NA	NA
	Spacer_NoG3_42	TCCTCAAAGGATTCCCCCTCCACTTCTTTCCTAG	34	NA	NA	NA
	Spacer_NoG3_43	GTAATTCTTGCTCAATTGCGGGTTGGAGCAATCG	34	NA	NA	NA
	Spacer_NoG3_44	TCTTCTATTAGGCAGCCGTGATCCTTTCCGCAGATAG	37	NA	NA	NA
	Spacer_NoG3_45	TATCTGAAATTGAGCAAGGTATTTCCAGACTTAGAGCTA	39	21	KX507046 *Vibrio* phage S4-7	phage
	Spacer_NoG3_46	TAGAGGCACTAGAGGCTGAAAAAGCATTAGATAATC	36	NA	NA	NA
	Spacer_NoG3_47	TCCTCCGCGATCAATTGAGCAATTTCCCCATTCGA	35	NA	NA	NA
	Spacer_NoG3_48	CGTAGTTGTTTGGGGGAAAGAGTCCTACCAGTGGA	35	NA	NA	NA
	Spacer_NoG3_49	CTTCGAAGAGTATCCCCCTTTCTTAAGGGATACCCTTTTGG	41	NA	NA	NA

Spacer_NoG3_50	TCAAATGTTACTTTAAGTGCAATTCCAACATATGG	35	NA	NA	NA
Spacer_NoG3_51	CTGGTGTACCCTCTGATACACCAGGAAGTGGGAGCCTCTTCCTAG	45	NA	NA	NA
Spacer_NoG3_52	TTATTGCCCTGTCTATAGACATTAGGTACTACGCT	35	NA	NA	NA
Spacer_NoG3_53	CTAGAGAATTGAGAAATTTTACGACATTTCCTTCTCT	37	NA	NA	NA
Spacer_NoG3_54	TACCCCCTTTGATTCGGAAAGACTCATTGTCTTTC	35	NA	NA	NA
Spacer_NoG3_55	ATAGTTATATTGGTAGCTGTAGTCACTTATGCCAT	35	NA	NA	NA
Spacer_NoG3_56	CGTACCCAATACTAGTAATAATATGGCTAGGCAAA	35	NA	NA	NA
Spacer_NoG3_57	CTAATTTTTTGTCCTCTTCCGGCCAACTTGGTGGTGTAA	39	NA	NA	NA
Spacer_NoG3_58	TTCGCCAGAATAAAAATGATGATTACGGTCGTATCCT	37	NA	NA	NA
Spacer_NoG3_59	TCGTCAATGTAGACGACATTACCTCCGTATTAGA	34	NA	NA	NA
Spacer_NoG3_60	TCTATGGGTGTTGGCGCAAATGGTGCTTTAATTTTCTCC	39	NA	NA	NA
Spacer_NoG3_61	TAGATTCTTGCCTAGTCAGTACACAAATATAGTCTACG	38	NA	NA	NA
Spacer_NoG3_62	GATTTGTGATTATATGTGATGTCCTCTTATAGAA	34	NA	NA	NA
Spacer_NoG3_63	ACCTTAGATAAGTTAGGATCTTCTCAGTTCCCTC	34	NA	NA	NA
Spacer_NoG3_64	TGTCGAACACCTTATCTTCTCCCTCCCCGCCGTTGCG	37	NA	NA	NA
Spacer_NoG3_65	GATATAAGTCTGACTGTCTACCCAGTAGACACAGCA	36	NA	NA	NA
Spacer_NoG3_66	TTGAAGCAGTGGAAATGGACGTGTACACGGCTA	33	NA	NA	NA
Spacer_NoG3_67	TTTATGGATCATCTATATATCCTGGAAGGCAATCAT	36	NA	NA	NA
Spacer_NoG3_68	GTCCATCTATTTGCGGCTATAAAGGAGGTGCGGACGA	37	NA	NA	NA
Spacer_NoG3_69	TCGAGGTGGTATTTGAAGGTAGATACAGGTACTCCAT	37	NA	NA	NA
Spacer_NoG3_70	TCTCCTCTTGATAAGAGGGAGAATGGAGTCATTCT	35	NA	NA	NA
Spacer_NoG3_71	AGCCCCTTAAAATAATGTTTTTTAATAAAATCATGATGTCA	41	NA	NA	NA
Spacer_NoG3_72	TAGGATCTGTCTCAATCTCTTTTGTAGACTCCACTA	36	NA	NA	NA
Spacer_NoG3_73	GCAACAGCCACGGGTAGGCTTGCACGTCCGTTTAACAA	38	NA	NA	NA
Spacer_NoG3_74	CAGGAAATAAGTCCGAATTAGAATTGATACAGGAGC	36	NA	NA	NA
Spacer_NoG3_75	GACCTACTGATGTTTTGGCAACAGAGGATGGCACGAT	37	NA	NA	NA
Spacer_NoG3_76	GCACCATACTTAGATCGTATTCAAATCTTTGGGTACTTG	39	NA	NA	NA
Spacer_NoG3_77	AGTGTAGATTACACTGAAAATGGTGCAAATACCGCCA	37	NA	NA	NA
Spacer_NoG3_78	TTTGTACTTGAAATTCAAACCTTTTTCCTCACTG	34	NA	NA	NA
Spacer_NoG3_79	AGAAGAAGTCGGATCCTGTGATTGAA	26	NA	NA	NA
Spacer_NoG3_80	CGGTTTTACTACGGATAAGCTGATAGGAGAGTTTA	35	NA	NA	NA
Spacer_NoG3_81	CACGGGGTGAAATACCCCGATGAGATGTAGTTTGAATG	38	NA	NA	NA
Spacer_NoG3_82	TCGTATGAACTACGGAGCAACGAGGATGTCTACGAATG	38	NA	NA	NA
Spacer_NoG3_83	TTTAGGATCAATCCTTGTATTAGATAATACATTTTTT	37	21	AB620173 Influenza A virus PB1 gene for polymerase PB1	virus
Spacer_NoG3_84	TCGGGGGTAACGACTTTTTTGCAACGAGTGACGCCGTAG	39	NA	NA	NA
Spacer_NoG3_85	CCTATAGAAATTTAATTTCTGTTTAACATGGATGT	35	21	*Clostridium perfringens* strain JP838 plasmid pJFP838A	plasmid
Spacer_NoG3_86	AGGGAGGGTAATATTTTATATGAGAGGTTGGGCTGT	36	NA	NA	NA
Spacer_NoG3_87	TTCATTTGGGTAGCTAAATTCTTTTGTAAGCTCTTTA	37	NA	NA	NA
Spacer_NoG3_88	AAACCCCTTTATCAACGTGCAAATAATAATGTTTT	35	NA	NA	NA
Spacer_NoG3_89	ATAATAGTAAAGATAGGTGATAGTTCCCCTAATGGA	36	NA	NA	NA
Spacer_NoG3_90	TCGGAAACTGGAGAATTATCTTAGCCTTTATAACTCT	37	NA	NA	NA
Spacer_NoG3_91	GGGATTCTTTAATGATTCCGGACATTACCAGAAA	34	NA	NA	NA
Spacer_NoG3_92	GGTAATATCTAAATTAGAATTGATTAATTCCTCTA	35	NA	NA	NA
Spacer_NoG3_93	TGTTGCCGCTTGGTGGCTGGGGCGTAAGAGCGGACA	36	NA	NA	NA
Spacer_NoG3_94	ACTGCGACGCCAAACCGGCCGCCACGTCAGACTCG	35	NA	NA	NA
Spacer_NoG3_95	CGAGGTTTGCGCCGCAGTGCGAGATCTACGAAAGCTGCAACCT	43	NA	NA	NA
Spacer_NoG3_96	ACGCCACAAATCCGGCGTTCGTCCGAGCAACGATTTTTGAA	41	NA	NA	NA
Spacer_NoG3_97	CCCTACGCTGTTACAACAATGCGGTGGAGGAGATG	35	NA	NA	NA
Spacer_NoG3_98	CCTAAAGTGTAGTTAGTATTATAAAAGCGTTTCGAC	36	NA	NA	NA
Spacer_NoG3_99	GCGGTGTTTGGCAGTCGCCACGAACGATCGCCCGTCTGCTAATTC	45	NA	NA	NA

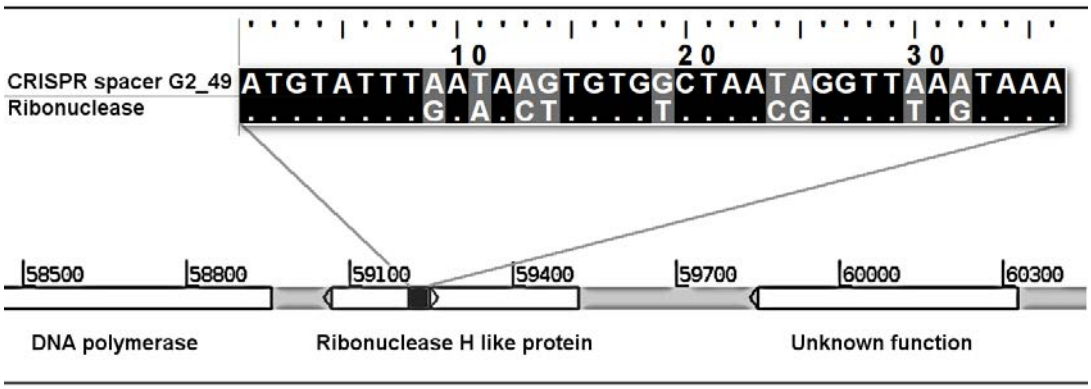
	Spacer_NoG3_100	AACAGATGAGGCTTTTAATCAAAGAGAACTACCAT	35	NA	NA	NA
	Spacer_NoG3_101	GAGCTGAATTCTTTGGATTTGAATGTGGCATTAAGGA	37	NA	NA	NA
4	Spacer_NoG4_001	GGGGCGTCTGCGGGAATTTCGCCCTACCCGCACTGAAGAA	40	NA	NA	NA
	Spacer_NoG4_002	AGCAGAGATAATCTCGTGGGCATTACAGATGGACAAA	37	NA	NA	NA
	Spacer_NoG4_003	TCGATCTCGAGGATGATAGCGGCGACGAAGCC	32	NA	NA	NA
	Spacer_NoG4_004	CACAATTGTTTTTGTCAGAGTAGTCCCCGAATTGAT	36	NA	NA	NA
	Spacer_NoG4_005	TTTCTGTTACTAAGATTGCCCTTCAGCAACGGGGA	35	NA	NA	NA
	Spacer_NoG4_006	AATTCTTTTTCCAATATTATCATTTTGATACTGCTGTT	38	NA	NA	NA
	Spacer_NoG4_007	TAAAGTTAAGTTCTCTTTTTTCTTACCAATACTGA	35	NA	NA	NA
	Spacer_NoG4_008	CTGATATAGAAACAGCTAGACGGCAACTGTTACTAATAATCACTAGATC	49	NA	NA	NA
	Spacer_NoG4_009	GCGATTCCGATGTCTCCGAGTCGCTCGAAGAACT	34	NA	NA	NA
	Spacer_NoG4_010	GTTCTTCAATTTAATTTCTGTTTTACAGGGATTA	34	NA	NA	NA
	Spacer_NoG4_011	GGGTGTTGGCGCAAACGGTGCTTTAATTTTCTCC	34	NA	NA	NA
	Spacer_NoG4_012	TCAAAAAAAATATTTACCTCTACAACGATTTGAG	34	NA	NA	NA
	Spacer_NoG4_013	AAAAAATTTCGTGGGTGCGATTCAGTCCGCGTATG	35	NA	NA	NA
	Spacer_NoG4_014	TAGATTCAGCCAGCACTTTATCCTGGCTACGTTGACGCGTATG	43	NA	NA	NA
	Spacer_NoG4_015	TTATTTTTTTCGTCTAGCTTACTTAGCGTACCTA	34	NA	NA	NA
	Spacer_NoG4_016	AAGACAAAATAGTTGAATTACAGTCCCTTTTGGCATCTC	39	NA	NA	NA
	Spacer_NoG4_017	CCTTCTTCGGCGTAAACCACTGTCCCCGAATGCGTCGC	38	NA	NA	NA
	Spacer_NoG4_018	ATGAAGAGTACGCTGAAACTCTTCGGTACGTCTTC	35	NA	NA	NA
	Spacer_NoG4_019	GAGGGAATAGAGCGACTTGGGAAGAGGTTGCTCAAGTACGTCTTC	45	NA	NA	NA
	Spacer_NoG4_020	GACATTTGTTCCTTGGTGAAGCTTTCCTTCCGCTTG	36	NA	NA	NA
	Spacer_NoG4_021	TCGTATGAACTACGGAGTGGGGTGGACGTCTACCAGTG	38	NA	NA	NA
	Spacer_NoG4_022	TTTAGATAAGGATCTACGTCCCGTTAAACCAGAA	34	NA	NA	NA
	Spacer_NoG4_023	GTTCTATCACATTTAGAGGCAAACTCTGACTGTGTTA	37	NA	NA	NA
	Spacer_NoG4_024	GGGAGCGACCTACAAGCTGGGATGTGTTACTATTACCA	38	NA	NA	NA
	Spacer_NoG4_025	AGTATGAACTAGGCGTTTCCCTTCCTGGTTGTAGATCTTCAGATCTAC	48	NA	NA	NA
	Spacer_NoG4_026	ATCTCTCTACTTTACTCGGGTTAAATTGTTCCGCTAT	37	NA	NA	NA
	Spacer_NoG4_027	GAGAAGGGTTTGGGTCGTTAAAGGTATATTGTAGATACAG	40	NA	NA	NA
	Spacer_NoG4_028	CAGGCACTTTCTTTTCCCTCCGTGTTATTATAAA	34	NA	NA	NA
	Spacer_NoG4_029	CCTAAAACCCTTGTCCCGTTTAATTCCTGTAA	32	NA	NA	NA
	Spacer_NoG4_030	TTGAATGCTGTGCCTTTTTGGCGACGCTCAAGG	33	NA	NA	NA
	Spacer_NoG4_031	TTAAAAAAGGAGGATACTCTGAGAAGTATGGATTT	35	NA	NA	NA
	Spacer_NoG4_032	ATATTTTATATGAGAGGTTGGGCTGCAACTATAG	34	NA	NA	NA
	Spacer_NoG4_033	GGTATATAAGTCTCTATGGTCATGTTAGAGGCCTCTATAAG	41	NA	NA	NA
	Spacer_NoG4_034	CTGAGGGCGTAAATTCTTCTTGACCATTTAAACCT	35	NA	NA	NA
	Spacer_NoG4_035	AAGGGTAGTGACGAGTCCGTTAAGATTAAAGATCTT	36	NA	NA	NA
	Spacer_NoG4_036	TTTGATTCCCGAAGCGATCGCGAACCATCAACA	33	NA	NA	NA
	Spacer_NoG4_037	TTCATCTCAGCACCCCTCGTCCACAGCAACATGC	34	NA	NA	NA
	Spacer_NoG4_038	TTTAAGGTAGTTCTACCTTGGGGAAATGACAATATGCT	38	NA	NA	NA
	Spacer_NoG4_039	TCAAATCCGGTCTGTGGCTCGTCCAGCGTGATGATG	36	NA	NA	NA
	Spacer_NoG4_040	ATAATTTACTTAACAGATTTAGTACTTTTTTCT	33	NA	NA	NA
	Spacer_NoG4_041	ATTTTATTTGCATAGTTGAAGGCTGGAATAAAGA	34	NA	NA	NA
	Spacer_NoG4_042	GAGGATCGCTTGGGCGATCGTCACCGAGGCCGG	33	NA	NA	NA
	Spacer_NoG4_043	TTGAGTTTGGTGAACAGACGATTGACCCGTCATT	34	NA	NA	NA
	Spacer_NoG4_044	GAAATAATATAGTTATCATGTTATCACATAAAAAT	35	NA	NA	NA
	Spacer_NoG4_045	TCTCCTTCTCTAATTCAATTAAAATCATCTCAGCTAG	37	NA	NA	NA
	Spacer_NoG4_046	TGGTCGCGATGCTCGTAGTCGACTGCGATCGC	32	NA	NA	NA
	Spacer_NoG4_047	GAGCCGCTATCTAAGGGATGAAAGAGGTATAAGCCCT	37	NA	NA	NA
	Spacer_NoG4_048	CCGTCATCCAGTGGATCGCATCTAATCTCGATTG	34	NA	NA	NA
	Spacer_NoG4_049	TATAAAAACTACATGATCCGTGAAGCACACTATAAAG	37	NA	NA	NA

	Spacer_NoG4_050	CACTAATTTAATCATTGACATATTTAATAGTTCCATC	37	NA	NA	NA
	Spacer_NoG4_051	CAGGCAAATTACTAAGGGATAAAGGTTTAATATTG	35	NA	NA	NA
	Spacer_NoG4_052	AAAATCATTTATATAGGTGATATCCTTTACATCACC	36	NA	NA	NA
	Spacer_NoG4_053	GATCGGGGCTGTTGACTCTGCCAACAAATGATCCGTTC	38	NA	NA	NA
	Spacer_NoG4_054	AGGGCTAAAAGATTGCGACGATATCGATCTTCTAC	35	NA	NA	NA
	Spacer_NoG4_055	TCGCGGACGATCCCTATCGATGTTTGACTGTTGAAGGTTTTTAT	44	NA	NA	NA
	Spacer_NoG4_056	CAGATCCCGATGGCGTAATTGCATCTTATCTAGATCCG	38	NA	NA	NA
	Spacer_NoG4_057	TGGATAGTAGGCTCGAAATCCGTGTACTATGGCTTCC	37	NA	NA	NA
	Spacer_NoG4_058	ATGAAGATGTTATCACCCCAGGTGAACAGGAACCTG	36	NA	NA	NA
	Spacer_NoG4_059	AGGATTTTAAGTCCTCCCTATTAAATAATGTCAAT	35	NA	NA	NA
	Spacer_NoG4_060	CGGTGGAGTGCTACTATATCGGGGTATAAAGCGCCCAC	38	NA	NA	NA
	Spacer_NoG4_061	CCTCAATCTTTGGTGCTTCTCAATTAAGCTTCGA	34	NA	NA	NA
	Spacer_NoG4_062	AGTTATTCACTCACTTCAATGACACAAAGGAAAG	34	NA	NA	NA
5	Spacer_NoG5_01	ACAGGAGTGGGAAATCGCACGATTTCTAGTCACATC	36	NA	NA	NA
	Spacer_NoG5_02	ACTCGGGGAACTCGGCTATATCTCTCGCTTCCACC	35	NA	NA	NA
	Spacer_NoG5_03	TGTCGGCGGTGCAGTACGCCGAAACACCGTTCACCA	36	NA	NA	NA
	Spacer_NoG5_04	ACTTTAGCGGCGATGGAATTAGTGTCGGTGGGCATCGC	38	NA	NA	NA
	Spacer_NoG5_05	TCGGCTTTTCCAAAGAGTGATAGTTGTTGTTGTTT	35	NA	NA	NA
	Spacer_NoG5_06	CTTTCGCTCCCGTCACTCGATAGACCTTTGTCAC	34	NA	NA	NA
	Spacer_NoG5_07	AAATCGGCGCGGCGGCGTACCGCGACGAAAACCAGCC	37	NA	NA	NA
	Spacer_NoG5_08	ATTAAACGACCTAGACTGCCGGAGCCTATCGCGGA	35	NA	NA	NA
	Spacer_NoG5_09	TTTGATATCGAATTCAGGGGCAATTCCTACCCTATCCT	38	NA	NA	NA
	Spacer_NoG5_10	GACACTCCGGTAACAGTCAGCCGCTCGGCGACGATT	36	20	*Sphingobium* sp. EP60837 plasmid pEP2	plasmid
	Spacer_NoG5_11	ATTAATCCCGCCTTCCCCATCTCGATCGCCCAGATT	36	NA	NA	NA
	Spacer_NoG5_12	GAATAATTCGCCCCAAATTCCCACCATCTGGGCGTT	36	NA	NA	NA
	Spacer_NoG5_13	CTTACAGCCCGCTGTATAGCACGCCGAAGGGTTTTT	36	NA	NA	NA
	Spacer_NoG5_14	AAAACTACCGCCCCTGCAGGGGAATTAATCTACCT	35	NA	NA	NA
	Spacer_NoG5_15	TTAATACATACAGGTATTGTACCATGTTTTCGCCAA	36	NA	NA	NA
	Spacer_NoG5_16	GTTCCCCTTGCAATCGTCGCTGTTCGTCAGCCCAAT	36	NA	NA	NA
	Spacer_NoG5_17	GGTTTGGAGGATTATCGTTAAATGACTTCATTAC	34	NA	NA	NA
	Spacer_NoG5_18	ATCTCAAGGAACTTGAATGTCTGTAACAACATAACA	36	NA	NA	NA
	Spacer_NoG5_19	AACCGCACTCGGTTTTGCGTCCTGATGTTTGGTTT	35	NA	NA	NA
	Spacer_NoG5_20	GTTGCAGACGACGATCGCCTCGGAAGTCGTCCCCCT	36	NA	NA	NA
	Spacer_NoG5_21	CGGCATCAGCAACCAAGCAGGCATCCTCTACGGGTTC	37	NA	NA	NA
	Spacer_NoG5_22	CGTCCGCAAGCTGACGGGCGATCGCAGCGTTCC	33	NA	NA	NA
	Spacer_NoG5_23	TGTACCAGGGGGTTTCGCCCTGGGGAGAAAGGAATA	36	NA	NA	NA
	Spacer_NoG5_24	AGACGCCCGCAACAGGACGAAGACGAAGAAGAGA	34	NA	NA	NA
	Spacer_NoG5_25	GCGTTGGGGACGACGACGGTGTTGGCGTCGGCCG	34	24	*Halomicrobium mukohataei* DSM 12286 plasmid pHmuk01	plasmid
	Spacer_NoG5_26	AGTGAACCGTTCGCGATACAAGGGGCGGCGATACC	35	NA	NA	NA
	Spacer_NoG5_27	CGACGCTGCCAGTGCCGCAGCCCTGCATCGCCAAGTC	37	27	GU936714_GU936714 *Synechococcus* phage S-CBS2	phage
	Spacer_NoG5_28	TCCACTGAAACCGTTGCCTCGGGATACATTCCCAA	35	NA	NA	NA
	Spacer_NoG5_29	CGAGTAAAAACGGAACTCGAAAAAATTGCCGAAAAGTA	38	20	*Escherichia coli* UMN026 plasmid p1ESCUM	plasmid
6	Spacer_NoG6_02	GTCCGAAGATGGTGAGGTTTTTATCTCGCTTTAGTACCC	39	NA	NA	NA
	Spacer_NoG6_03	TAAATATGAATCCAGAACTTGAATTTTTGTTGCAGCTA	38	20	KF148616 *Campylobacter* phage CP8	phage
	Spacer_NoG6_04	ATTACGCTATCTTGGTTAATCGTGGCGGACGGTTTCGA	38	NA	NA	NA
	Spacer_NoG6_05	GAAAATCCCCTGAGTTGGTTGGCGCTCGATTTGACCAAAAACTGAA	46	NA	NA	NA
	Spacer_NoG6_06	TTATGCACTTGTCGATCTATCGTTCGGTGATCCTGCACCG	40	NA	NA	NA
	Spacer_NoG6_07	GGGTATATAGCCCCCGCGGTCGTACACCTCAACACAA	37	NA	NA	NA

Spacer_NoG6_08	TTGGTGATAACCTATACCCCGGAGTAACAACAATACTAAA	40	NA	NA	NA
Spacer_NoG6_09	CCTCGAAAGAGGGCGTTATTTCTGTAGCCGAAATCAAAAA	40	NA	NA	NA
Spacer_NoG6_10	AAAGTTCGGTATCGGCTATTCGCCTCGAGGCAAGCTCCT	39	NA	NA	NA
Spacer_NoG6_11	ATGCCGTCGGGATGAACCTGTCAGCCGAGACGCTGAGGACT	41	NA	NA	NA
Spacer_NoG6_12	TTTATCGACTCCAAAACGATCCCCCTGGCAACAGATA	37	NA	NA	NA
Spacer_NoG6_13	AGAAAGATCGTGAATTTAAAGATGGAGCCAAAATAAA	37	21	AP008983 *Clostridium* phage c-st	phage
Spacer_NoG6_14	ATGGACGCAGGGGCTACGGTTCCCAGATACGCCGTCGA	38	NA	NA	NA
Spacer_NoG6_15	GACGCCGTAGACTTCTACACTCCCAATCTGGATGCAATATTTG	43	NA	NA	NA
Spacer_NoG6_16	AGGATTTATGCAATTGGAGTCGCACTCTCAACTAGGGAT	39	NA	NA	NA
Spacer_NoG6_17	GAGAGAAGTAGCGATGAATTGTATTGATATTGTGCGGGA	39	NA	NA	NA
Spacer_NoG6_18	TCAACGTATATTTATTAGTACCGATATCAAAATTAGTATG	40	20	HF679131 *Adoxophyes honmai* enomopoxvirus 'L'	virus
Spacer_NoG6_19	CCTACTTGCCACCCAGTAGAATTGTCGGTATAAATTAAA	39	NA	NA	NA
Spacer_NoG6_20	GATCTCTAAGAGTTTACACCCGTTGCGGAGTTGTTGG	37	NA	NA	NA
Spacer_NoG6_21	ACGACGTCAATCGAGCGCTACGACTCGTTCGTACGT	36	NA	NA	NA
Spacer_NoG6_22	GCTTTGGGGATGTTGCCCAAGGACATAACGATTTTTT	37	NA	NA	NA
Spacer_NoG6_23	TTTATTGTACGAGCGGGAGGGTACCTCCCAGACTTGGAG	39	NA	NA	NA
Spacer_NoG6_24	TTTAGCGGCTCCAGACCAATTGATTTCATTATTGTTAT	38	20	AF020713 Bacteriophage SPBc2 complete genome	phage
Spacer_NoG6_25	TGGCTATCAGACATCAACCCCAATCTGGCGCTGACGTAT	39	NA	NA	NA
Spacer_NoG6_26	GTGGACGGGATTATTTAGTATACGATCCATGGGATTAC	38	NA	NA	NA
Spacer_NoG6_27	AGGGCTTATGGTTTAAGTATTCAGGGTCGCTTTGATTGG	39	NA	NA	NA
Spacer_NoG6_28	AAATTACGCAAAAGGCTGCCTGGCAAATTGTGTATCGACTGG	42	NA	NA	NA
Spacer_NoG6_29	TGCTTGGTGAGCGGTTTTTGTCCACTGATTTAAT	34	NA	NA	NA
Spacer_NoG6_30	TTTACCGCAACGGACACGCCGACGAACGTGCGGGAT	36	NA	NA	NA
Spacer_NoG6_31	GTAACAGCGCAATCAGTTTCTATGGCTGCGTTAAAGCTGTA	41	NA	NA	NA
Spacer_NoG6_32	GAGGAATATTTTCCGGAAGTGTCGGCGCGTCACTCCC	37	NA	NA	NA
Spacer_NoG6_33	TCTTCTACCGATAGAGTTACGTCTTATTTTGTAGATTGAGT	41	NA	NA	NA
Spacer_NoG6_34	CCCGCTACGGCTGTCGGACTCCAAATCGTATTTAATA	37	NA	NA	NA
Spacer_NoG6_35	TATTAGTATTAGTAAATAGAGAGCCTCTGATCGAACAAAC	40	NA	NA	NA
Spacer_NoG6_36	CTTTGGGATATAACCCCAACGAGTCGTTTCCCGCAGACA	39	NA	NA	NA
Spacer_NoG6_37	TTGACCGTCGTGCAACAGTTGGGTGGAAGCTTGGGG	36	NA	NA	NA
Spacer_NoG6_38	TTTTCGTTAAGGGCAACGATTTCTTCATCGAATCCCT	37	NA	NA	NA
Spacer_NoG6_39	AAGCAAGGCAAGTATTACGTCGAGGACTCCATCAAGCA	38	NA	NA	NA
Spacer_NoG6_40	TTGTCGATCTTCAGATGGAATATTTCGAGCTTGGCA	36	NA	NA	NA
Spacer_NoG6_41	TTTAATTTTTATGGACACTTTGACTTCGGGACGTT	35	NA	NA	NA
Spacer_NoG6_42	ACTGTCGGACGTGGGAAGGATATCCCCGCGCTCACAA	37	NA	NA	NA
Spacer_NoG6_43	GGTAACACGACCTATCGCTACCGTATTAAAGAAGATTT	38	NA	NA	NA
Spacer_NoG6_44	AACCCTACAACGCCTTCAACAAGCCAGAGGACAACGC	37	NA	NA	NA
Spacer_NoG6_45	AAAACGGCGAATCGCTTGCTCCCTACCTCCGTCAGAT	37	NA	NA	NA
Spacer_NoG6_46	GCGAATTACCTGCGTGGGTCGTCAGCACCTCGCTCATTTT	40	NA	NA	NA
Spacer_NoG6_47	ATCCCGTACCTCGTAAGGGGCCGCTCGACCTCGAT	35	NA	NA	NA
Spacer_NoG6_48	ACCGTTCGACCTCGAGATAGAGCGGAACGAATACGATTC	39	21	KU760999 Bluetongue virus isolate BTV-27FRA2014v03	virus
Spacer_NoG6_49	AATTTAAGCTTGGTCTTTTCGCGGTTTTTACCTCCTTT	38	NA	NA	NA
Spacer_NoG6_50	TGTCTGTTTATCTCCACCGGGTTGTCCAACCCCATAGAG	39	NA	NA	NA
Spacer_NoG6_51	TGACTTCGGCGTTGATTATAAGCCGTTAACGTGGCGAA	38	NA	NA	NA
Spacer_NoG6_52	GTCTGCAAACCGTTCCCGTCATCGAAACATCCTTTTACTT	40	NA	NA	NA
Spacer_NoG6_53	CTGACCAAAGCGCTGCCGTTGGAGCTATCCGAAGGAG	37	NA	NA	NA
Spacer_NoG6_54	ATGCGATCGGGGACTTGGGATTTCTCAATCCCAAGTGTG	39	NA	NA	NA
Spacer_NoG6_55	ATCCCGGCATGGTCGCACCCCAACGTTGCCGATAAATA	38	NA	NA	NA
Spacer_NoG6_56	TCGAACACGCTTACGCCGTGGCCGAAGCGTTCCGCGCC	38	NA	NA	NA
Spacer_NoG6_57	TTGGCGTTCTTCCAATCCCTTCTAATCTATCCACGTATT	39	NA	NA	NA

	Spacer_NoG6_58	GATTTCGGCGTTTACCAGATATTGCACGTCACGAAC	36	NA	NA	NA
	Spacer_NoG6_59	ATCCAGAGAGAGAAAACGGTAGCGACGATTCGCTTTTT	38	NA	NA	NA
	Spacer_NoG6_60	TTGCTCGTACTGCTTCCTAGAGAATAACGACACTCCGACAG	41	NA	NA	NA
	Spacer_NoG6_61	TTTTAGGGGAAAGATAGCCGATCTTTTCCGCCACCCT	37	NA	NA	NA
	Spacer_NoG6_62	CGATATGAACACCAACGGCTTTTTTCAGCCAGGGAC	36	NA	NA	NA
	Spacer_NoG6_63	TCAGCGTGATTCACAAAAGTCGCTTCAATAAAAGCCTCGAAAT	43	NA	NA	NA
	Spacer_NoG6_64	CTCAGCCGTCTCGGTTGCCTCAGGTTCCTCAGGTTCCGTCTT	42	20	KF056323 *Haloarcula hispanica* pleomorphic virus 2	virus
	Spacer_NoG6_65	CTTACGCCGTGGGCGTTGCTCTCGATCGCTACGAAAAA	38	NA	NA	NA
	Spacer_NoG6_66	CTTTGAACGCCTTTAACCGCTTATCGGGCTTCGTTGTTA	39	NA	NA	NA
	Spacer_NoG6_67	TGAAATACCTGTAATTGGTTTAGAGGGATTGAAAGGAAC	39	NA	NA	NA
	Spacer_NoG6_68	TACTGGAGGTATCGGACGGTACAGCTCGCATTGAACAA	38	NA	NA	NA
	Spacer_NoG6_69	AAATTTTTATTTGTAGTCATGGGATTAAATCTAGTTTTT	39	NA	NA	NA
	Spacer_NoG6_70	TGAACTGGAGTTTAAAAGATTAATCAACTCCTCATTTTTT	40	NA	NA	NA
	Spacer_NoG6_71	ATGGGGCGAAATGCGAACCTGTTACTATCAGTTA	34	NA	NA	NA
	Spacer_NoG6_72	TGTCATAATGCCAGATGGGTTACAAATTATCAATGTTA	38	NA	NA	NA
7	Spacer_NoG7_02	GCTGTTCGATGGATGCGGAGGTGGGGTCGGCGATTC	36	NA	NA	NA
	Spacer_NoG7_03	CAACTTCCACACCCACTCCAGACGTAGTCGACGTCGACGA	40	NA	NA	NA
	Spacer_NoG7_04	TACAATAATACCTAAGAGTAGGCTGCTGAGAAGGCTAATT	40	NA	NA	NA
	Spacer_NoG7_05	GGGGTGGTTTCCCGGACGGGGGTTGCAGAGACAA	34	NA	NA	NA
	Spacer_NoG7_06	CTCTTCCTCCTCGCCTTCGGCCCGGCTCGAAATCTCCGTC	40	24	JF974315 *Rhizobium* phage RR1-B genomic sequence	phage
	Spacer_NoG7_07	CGTTGTTCTCGCTCAGCTTGTTGTTTCCAACAAAGAACA	39	NA	NA	NA
	Spacer_NoG7_08	AAGACGACCCCGATTATTAAAATTCCTCCAGCGTATGAAG	40	NA	NA	NA
	Spacer_NoG7_09	AATCTAGACTAGTTTTTTTAAAAAAACTAGCCCAATCAAAACA	43	NA	NA	NA
	Spacer_NoG7_10	ATTGCTCGTGCAGGAGTGGTGGAAGAATTGACCAAACC	38	NA	NA	NA
	Spacer_NoG7_11	TCCATCATTGGAAACTTCATCAACGCCTTTCGG	33	NA	NA	NA
	Spacer_NoG7_12	CACCACGCCCAAGTACACCTTGGGTGTAATCCAAGA	36	NA	NA	NA
	Spacer_NoG7_13	ATTCTACCACATCTTCTTTGGAACGAAGGATATTAAAAT	39	NA	NA	NA
	Spacer_NoG7_14	ATTTCCATCGTCGCTGTCATCTTTGAAAGCGTTGCCATAT	40	20	HM144385 *Brochothrix* phage NF5	phage
	Spacer_NoG7_15	ATTTCGAACTATCGCTACGACGAGAGCCTACCTCTTT	37	NA	NA	NA
	Spacer_NoG7_16	TACTCGCCTAAGTTTGTAGCCTCCCCTGTCTCATCATCGA	40	NA	NA	NA
	Spacer_NoG7_17	ATATAGATCTAGGGGGTTTGCGTTTCCCTCAAACCTTAAGA	41	NA	NA	NA
	Spacer_NoG7_18	AGTTCTTTGCAAGAACTCGCCACAATATTGGCAGTAGGT	39	NA	NA	NA
	Spacer_NoG7_19	TCGTACCGGTAAGCCCGGCCGTACCGGCGAGCCTTTGCC	39	NA	NA	NA
	Spacer_NoG7_20	CGTGTAGTCTTTAGCGAGAACCTCAGCGGACTCTTTAGCG	40	NA	NA	NA
	Spacer_NoG7_21	CCCTCAATTAAGTGAGGGAGTCGGGACTAACTAATA	36	NA	NA	NA
	Spacer_NoG7_22	CAGATTGTATACTCCCCGTCCTTCTTCATGGACTGGGA	38	NA	NA	NA
	Spacer_NoG7_23	AATATTGAATACTTCTTCTAGGGAATTGAAATTCCCGAACG	41	NA	NA	NA
	Spacer_NoG7_24	CTTCCACCGCTACTGCCAAGGGGCTAGTGATAGCGAA	37	NA	NA	NA
8	Spacer_NoG8_02	GTCGGAAAAAAACCCAGAGCTATAAAAATGTACTTGTCCCGATAATTG CTTT	51	NA	NA	NA
	Spacer_NoG8_03	TTGAAAAAAAAACTCCAACAATATTACCCCAAAAAACGAGAAAATTCTT GA	50	22	*Ilyobacter polytropus* DSM 2926 plasmid pILYOP01	plasmid
	Spacer_NoG8_04	GTCAAAAAAAAAACGAAATTTGACAAATGCTAAAAGATAACTGAATT	46	20	*Lactobacillus salivarius* UCC118 plasmid pMP118	plasmid
9	Spacer_NoG9_01	TATTCTATTGGGTCAGGACAACTTTAATCTGTTGACCCGTTTT	43	NA	NA	NA
	Spacer_NoG9_02	CTCTGGCTCATCAGCTCTCCCTAGAGCGAATAGACAGTTTC	41	NA	NA	NA
	Spacer_NoG9_03	TCTCGAAACGCCCGTGAAACTTACCGTCACAAAATGTTTT	40	NA	NA	NA
10	Spacer_NoG10_01	GGGCCAAGAAAGCCCCGATCCCAATCCCAATCCCATG	37	NA	NA	NA
	Spacer_NoG10_02	CGGGAACTTACCTCACTGCCGAAGACGCCATCCATG	36	NA	NA	NA
	Spacer_NoG10_03	AAAACGTTTACGGGCTAACCCGTTTCCCCGGAACTTA	37	NA	NA	NA
	Spacer_NoG10_04	TATTTGTCTCCCGAAGACACAATTCATGCTGGAT	34	NA	NA	NA
	Spacer_NoG10_05	TCAAAGAGTTCCTGTGAGGAAGCGCTAGCCCGTAGCCGAGAGG	43	NA	NA	NA

	Spacer_NoG10_06	GTTTGGGACTTACGAGTCCGCCGAAGCCGCCATTTACGATG	41	NA	NA	NA
	Spacer_NoG10_07	CGTTCTCGGTATCGCTAACGCTACGTTAGCGACGCCTAT	39	NA	NA	NA
	Spacer_NoG10_08	GTTCTTGCAATCAACACCGAGCAAGGCTCTTGTTTCTTAGA	41	NA	NA	NA
	Spacer_NoG10_09	CGTCGTTGCTATCAACAGCAACCAGGGGCTATGTTTT	37	NA	NA	NA
	Spacer_NoG10_10	TATACTGACTTCATTACCGACCCAAAGACTGGAAAACAA	39	NA	NA	NA
	Spacer_NoG10_11	GGCTACTGTAGCCCGGGACCGGTCTATGAGACTGAT	36	NA	NA	NA
	Spacer_NoG10_12	CCGCAATTTACAATGGGGGGGTATGTCCCTCCTCAC	36	NA	NA	NA
	Spacer_NoG10_13	AGTACGTCATCGCCATCAGTGACGGCCGATCATGCTATT	39	NA	NA	NA
11	Spacer_NoG11_15	CCGTCGCAATTAAAGATGAGGAGCTGCACGACATCGC	37	NA	NA	NA
	Spacer_NoG11_16	CTCTTGCTATTTAGCTTCAAGGGGTATCGTATGCCATAA	39	NA	NA	NA
	Spacer_NoG11_17	TCTTTGCTACCTCGACGGTAGCAAGCCCAGTTCA	34	NA	NA	NA
	Spacer_NoG11_18	CTTTTCATCGCTGTCGTAATCTTTTTCGTTCGGAA	35	NA	NA	NA
	Spacer_NoG11_19	TTATGTTGTTTGATGATGATGATGACGACGGGGAGAAAA	39	NA	NA	NA
	Spacer_NoG11_20	TGGAAACGTGAGGAGGGACATACCCCCCATT	31	NA	NA	NA
	Spacer_NoG11_21	CCCATTCCAAGCTCGTTAGAAGGGGAAAACCCCGTCTT	38	NA	NA	NA
	Spacer_NoG11_22	TCAGCCCCCATACCTCGTCTTGCGACTTCGCGGAGA	36	NA	NA	NA
	Spacer_NoG11_23	CCCTTCGCTTCGGGGTAGAATATGAAAGCTGCGATGATTT	40	NA	NA	NA
	Spacer_NoG11_24	TTTTGGGTTCAGTTTCTTACTGAACTTTCCAAGTTAT	37	NA	NA	NA
	Spacer_NoG11_25	GAAGCCCATCGCGGGTCGGGTCGTATCGGGTAATTAGC	38	NA	NA	NA
	Spacer_NoG11_26	TGTTCGCAACGGCGTGCTTGACTGTATTGCTGTCTACGACAACGA	45	NA	NA	NA
	Spacer_NoG11_27	AAGAGGAGTCCGACGAGGGTCTAGATCCTACGATTA	36	NA	NA	NA
	Spacer_NoG11_28	TCCTTCGCCACCCCCACGATTGTAGCTAAAGAATACACCCCAG	43	NA	NA	NA
	Spacer_NoG11_29	TCTGTAGCAGGCGTGCCCTCTGTAGCTAAAGAATACATCCCG	42	NA	NA	NA
	Spacer_NoG11_30	ATTTTCATCCTCTTTCTAGGAGGGGGTGTGATTACCA	37	NA	NA	NA
	Spacer_NoG11_31	TGACAAATTATTTGAAGAAGGAGGTGAGGAATGTGAC	37	NA	NA	NA
	Spacer_NoG11_32	CTCTGAAATATCTGAAATAGAGAAGAGAGTCCCCCGAAC	39	NA	NA	NA
	Spacer_NoG11_33	CGTCGTCGGTCTCCATGGTCCTCCCCGGACAAGAAGGTGAT	41	NA	NA	NA
	Spacer_NoG11_34	CAGCCCCACTCCCTCGAGTGTATGAAGGAGATATGGGTATG	41	NA	NA	NA
	Spacer_NoG11_35	ACGGAAGTTGTGGAATTTCAAACCCCTCTTCAAA	34	NA	NA	NA



Appendix 5.8 CRISPR-Cas spacer target on genome contig_1. CRISPR spacer-NoG2_49 of BBD cyanobacterium *Geitlerinema* sp. BBD_1991 (Buerger et al. 2016; Den Uyl et al. 2016) showed similarities (score: 18) to a gene on virus genome contig_1 (*Chapter 4*), which encodes for a putative ribonuclease (RNase) H like protein. Alignment is shown for the spacer and the target gene, as well as an overview for the genome region with adjacent open reading frames on contig_1. Dots indicate the same nucleotide is present in both sequences.

Appendix 5.9 Phage related genes in *Roseofilum reptotaenium* AO1. Annotations were taken from RAST and BLASTp.

Feature ID	Contig	RAST annotations	BLAST similarity
fig|564709.3.peg.63	Contig_102_length_22140_cov_415.14_ID_2031875719653897	phage protein	putative lipoprotein [*Vibrio* phage CKB-S1], *E*-value 2e-45
fig|564709.3.peg.309	Contig_123_length_7114_cov_1067.01_ID_24544845416933	phage protein	**1)** Multi-domain Transposase [Mobilome: prophages, transposons], COG3415, *E*-value 2.47e-08. **2)** Gp20, *Siphoviridae*, *E*-value: 8.2e-30
fig|564709.3.peg.623	Contig_15_length_308535_cov_441.995_ID_292752362770261791	phage tail sheath protein	Multi-species: tail protein [*Arthrospira*], *E*-value 0.0
fig|564709.3.peg.624	Contig_15_length_308535_cov_441.995_ID_292752362770261791	hypothetical protein	T4-like virus tail tube protein gp19, pfam06841, *E*-value 1.10e-17
fig|564709.3.peg.625	Contig_15_length_308535_cov_441.995_ID_292752362770261791	hypothetical protein	T4-like virus tail tube protein gp19, pfam06841, *E*-value 3.64e-16
fig|564709.3.peg.626	Contig_15_length_308535_cov_441.995_ID_292752362770261791	hypothetical protein	T4-like virus tail tube protein gp19, pfam06841, *E*-value 1.05e-28
fig|564709.3.peg.2506	Contig_37_length_117591_cov_438.095_ID_73116563114182238	DNA primase, phage associated # P4-type	Phage- or plasmid-associated DNA primase [Mobilome: prophages, transposons]; *E*-value 3.24e-32
fig|564709.3.peg.3086	Contig_45_length_94947_cov_442.235_ID_893854739005459	phage protein	Nucleoside Triphosphate Pyrophosphohydrolase (EC 3.6.1.8) MazG-like domain *E*-value 3.12e-05
fig|564709.3.peg.3307	Contig_49_length_86236_cov_423.446_ID_	phage endolysin,	Glycoside hydrolase family 19

	97612250581065	Phage lysis modules	chitinase domain. *E*-value 2.17e-04
fig|564709.3.peg.3537	Contig_51_length_84369_cov_463.726_ID_1015790356944960	putative prophage protein (ps3)	Uncharacterized phage-associated protein [Mobilome: prophages, transposons]. *E*-value 1.06e-20
fig|564709.3.peg.4364	Contig_68_length_58146_cov_430.42_ID_13554321531251197	hypothetical protein	Phage- or plasmid-associated DNA primase [Mobilome: prophages, transposons] *E*-value 3.94e-35
fig|564709.3.peg.4366	Contig_68_length_58146_cov_430.42_ID_13554321531251197	phage integrase	Shufflon-specific DNA recombinase Rci and Bacteriophage Hp1_like integrase. *E*-value 1.31e-17
fig|564709.3.peg.5051	Contig_86_length_35161_cov_369.366_ID_17179687234735	phage shock protein A	phage shock protein PspA; Provisional. *E*-value 7.35e-34
fig|564709.3.peg.5052	Contig_86_length_35161_cov_369.366_ID_17179687234735	phage shock protein A	phage shock protein PspA; Provisional. *E*-value 7.26e-24
fig|564709.3.peg.5237	Contig_91_length_31916_cov_504.263_ID_18115950149101041	phage protein	RNA ligase, DRB0094 family. *E*-value 1.81e-64

Appendix 5.10 PHAST, PHASTER and VIRsorter prophage detection in *R. reptotaenium* AO1. The detected prophage is considered as questionable or incomplete due to for example the lack of essential virus related genes, such as capsid, head, and tail genes.

#	*R. reptotaenium* AO1, R1 **CDS position (contig 41)**	**PHASTER - BLAST hits**	**E-value**
1	71377..72384	*PHAGE_Synech_ACG_2014f_NC_026927*: NAD-dependent epimerase/dehydratase PP_02287 phage (gi815854730)	3.37e-029
2	72417..73613	*PHAGE_Synech_ACG_2014f_NC_026927*: hypothetical protein PP_02288 phage (gi815854731)	7.15e-086
3	73622..74668	*PHAGE_ProchI_P_TIM68_NC_028955*: putative transketolase central region-containing protein PP_02289 phage (gi971760363)	1.78e-013
4	74671..75612	*PHAGE_Synech_S_SM2_NC_015279*: transketolase central region-containing protein PP_02290 phage (gi326781943)	1.52e-009
5	75648..76571	hypothetical PP_02291	0
6	76629..77624	*PHAGE_Synech_ACG_2014f_NC_026927*: GDP-D-mannose 4,6-dehydratase PP_02292 phage (gi815854729)	5.64e-040
7	77682..78776	*PHAGE_Caulob_Cr30_NC_025422*: D,D-heptose 7-phosphate kinase PP_02293 phage (gi725949173)	5.36e-044
8	78817..79218	*PHAGE_Caulob_Cr30_NC_025422*: phosphoheptose isomerase PP_02294 phage (gi725949171)	1.69e-015
9	79397..79978	hypothetical PP_02295	0
10	79981..80703	*PHAGE_Entero_phi92_NC_023693*: Phi92_gp066 PP_02296 phage (gi726646999)	1.68e-014
11	80715..81653	*PHAGE_Synech_ACG_2014f_NC_026927*: ADP-L-glycero-D-mannoheptose-6-epimerase	5.82e-011

		PP_02297 phage (gi815854739)	
12	81674..81751	hypothetical PP_02298	0
13	81784..82971	hypothetical PP_02299	0
14	83072..83188	hypothetical PP_02300	0
15	83400..84692	PHAGE_Plankt_PaV_LD_NC_016564: ABC transporter PP_02301 phage (gi371496158)	8.98e-010

#	*R. reptotaenium* AO1, R2 **CDS position (contig 72)**	**PHAST - BLAST hits**	**E-value**
1	7385..7849	*PHAGE_Pseudo_F116_NC_006552*: DNA adenine methyltransferase PP_00007; phage (gi56692911)	7.00e-018
2	7869..8273	*PHAGE_Pseudo_F116_NC_006552*: DNA adenine methyltransferase PP_00008; phage (gi56692911)	1.00e-020
3	8280..9293	*PHAGE_Microc_Ma_LMM01_NC_008562*: lysozyme/metalloendopeptidase PP_00009; phage (gi117530266)	7.00e-017
4	9309..11027	*PHAGE_Parame_bursaria_Chlorella_virus_FR483_NC_008603*: hypothetical protein FR483_N733R PP_00010; phage (gi155370831)	9.00e-032
5	11328..12539	PHAGE_Microc_Ma_LMM01_NC_008562: transposase PP_00011; phage (gi117530306)	3.00e-104
6	12554..12685	hypothetical PP_00012	0
7	12936..13088	hypothetical PP_00013	0
8	13192..14520	*PHAGE_Cardio_polyomavirus_NC_020067*: major structural protein VP1 PP_00014; phage (gi440285304)	2.00e-007

#	*R. reptotaenium* AO1, R3 **CDS position (contig 93)**	**VIRsorter - BLAST hits**	**E-value**
1	1..358	hypothetical protein	-
2	358..2963	Phage_cluster_71; PFAM-AAA_25 (DNA repair protein)	5.37e-25
3	3009..3455	hypothetical protein	-
4	3458..3919	hypothetical protein	-
5	4421..4942	hypothetical protein	-
6	4935..5204	hypothetical protein	-
7	5311..5814	hypothetical protein	-
8	6163..6423	hypothetical protein	-

Appendix 5.11 PHAST, PHASTER and VIRsorter prophage detection in *Geitlerinema* sp. BBD_1991. The detected prophage is considered as questionable or incomplete due to for example the lack of essential virus related genes, such as capsid, head, and tail genes.

#	*Geitlerinema* sp. BBD_1991, G1 **CDS pos. (BBD_1000996)**	**PHAST - BLAST hits**	**E-value**
	337736..337747	AttL AAGTGGCGTTTT	0
1	338024..338470	*PHAGE_Lactob_phiPYB5_NC_027982*: putative integrase; PP_00353; phage (gi937456185)	9.00e-005
2	338504..338869	*PHAGE_Dinoro_IMEphi4_NC_024367*: putative host-like protein; PP_00354; phage (gi658311036)	7.00e-006
3	339100..340053	*PHAGE_Synech_S_SKS1_NC_020851*: GDP-L-fucose synthase; PP_00355; phage (gi472340899)	5.00e-062
4	340160..341236	*PHAGE_Synech_S_SKS1_NC_020851*: GDP-D-mannose 4,6-dehydratase; PP_00356; phage (gi472340900)	5.00e-109
5	341422..342153	glycosyl transferase [*Oscillatoria acuminata* PCC 6304]. gi|428211603|ref|YP_007084747.1|; PP_00357	9.00e-098
6	342220..343476	*PHAGE_Synech_ACG_2014f_NC_026927*: group 1 glycosyl transferase; PP_00358; phage (gi815854524)	5.00e-009
7	343579..344703	*PHAGE_Bathyc_BpV1_NC_014765*: hypothetical protein; PP_00359; phage (gi313768026)	6.00e-049
	357614..357625	AttR AAGTGGCGTTTT	0
#	*Geitlerinema* sp. BBD_1991, G2 **CDS pos. (BBD_1001009)**	**PHAST - BLAST hits**	**E-value**
1	2477808..2478164	*PHAGE_Helico_phiHP33_NC_016568*: transposase; PP_02553; phage (gi371671361)	3.00e-015
2	2478257..2479516	*PHAGE_Microc_Ma_LMM01_NC_008562*: transposase; PP_02554; phage (gi117530202)	2.00e-045
3	2479573..2482482	*PHAGE_Ectoca_siliculosus_virus_1_NC_002687*: EsV-1-65; PP_02555; phage (gi13242537)	5.00e-040
4	2482470..2482589	hypothetical; PP_02556	0
5	2482597..2482731	hypothetical; PP_02557	0
6	2482789..2482908	hypothetical; PP_02558	0
7	2483196..2483342	hypothetical; PP_02559	0
8	2483336..2484139	*PHAGE_Synech_S_SKS1_NC_020851*: cyanobacterial phosphoribosylglycinamide formyltransferase; PP_02560; phage (gi472340960	2.00e-036
9	2484180..2484779	*PHAGE_Prochl_P_SSM7_NC_015290*: orotate phosphoribosyltransferase; PP_02561; phage (gi326784523)	3.00e-032
10	2485011..2486813	*PHAGE_Cyanop_S_TIM5_NC_019516*: virion structural protein; PP_02562; phage (gi422936314)	3.00e-016
#	*Geitlerinema* sp. BBD_1991, G3 **CDS pos. (BBD_1001028)**	**PHAST - BLAST hits**	**E-value**
1	4406431..4408458	*PHAGE_Cyanop_NATL2A_133_NC_016659*: hypothetical protein; PP_04556; phage (gi372217849)	2.00e-030
2	4408486..4409214	phage tail protein [*Calothrix* sp. PCC 7507]. gi|427718458|ref|YP_007066452.1|; PP_04557	8.00e-030
3	4409211..4410185	hypothetical protein Anacy_0193 [*Anabaena cylindrica* PCC 7122]. gi|440679915|ref|YP_007154710.1|; PP_04558	8.00e-065
4	4410236..4413910	*PHAGE_Bacill_BCD7_NC_019515*: putative baseplate J family protein; PP_04559; phage (gi422936037)	5.00e-016
5	4413999..4414385	*PHAGE_Synech_ACG_2014e_NC_026928*: base plate wedge subunit; PP_04560; phage (gi815854880)	1.00e-010
6	4414418..4414843	PAAR motif protein [*Teredinibacter turnerae* T7901]. gi|254787612|ref|YP_003075041.1|; PP_04561	3.00e-046
7	4414853..4415503	*PHAGE_Campyl_CP30A_NC_018861*: putative baseplate hub and tail lysozyme; PP_04562; phage (gi410493030)	9.00e-008
8	4415500..4416558	*PHAGE_Bacill_BCD7_NC_019515*: hypothetical protein; PP_04563; phage (gi422936041)	2.00e-010
9	4416577..4417227	*PHAGE_Salmon_Vil_NC_015296*: conserved uncharacterised protein; PP_04564; phage (gi326804610)	4.00e-005
10	4417224..4418816	*PHAGE_Molliv_sibericum_NC_027867*: hypothetical protein; PP_04565; phage (gi927594325)	3.00e-009
11	4418829..4419368	*PHAGE_Bacill_BCD7_NC_019515*: putative tail tube protein 2; PP_04566; phage (gi422936047)	4.00e-012
12	4419384..4419521	hypothetical protein *Anacy_0201* [*Anabaena cylindrica* PCC 7122]. gi|440679923|ref|YP_007154718.1|; PP_04567	4.00e-010
13	4419527..4419904	hypothetical protein *Lepto7376_0625* [*Leptolyngbya* sp. PCC 7376]. gi|427722594|ref|YP_007069871.1|; PP_04568	1.00e-045
14	4419904..4420365	*PHAGE_Bacill_BCD7_NC_019515*: putative tail tube protein 2; PP_04569; phage (gi422936047)	2.00e-018

15	4420381..4421694	*PHAGE_Bacill_BCD7_NC_019515*: putative tail sheath protein; PP_04570; phage (gi422936048)	2.00e-026
16	4421802..4423052	*PHAGE_Bacill_BCD7_NC_019515*: putative tail sheath protein; PP_04571; phage (gi422936048)	9.00e-038
#	***Geitlerinema* sp. BBD_1991, G4 CDS pos. (BBD_1001072)**	**PHAST - BLAST hits**	**E-value**
1	5824500..5825465 BBD_1001072	*PHAGE_Chryso_virus_NC_028094*: putative glycosyltransferase; PP_06057; phage (gi939177431)	1.00e-009
2	5825458..5825685	hypothetical protein *Cal7507_2917* [*Calothrix* sp. PCC 7507]. gi|427718171|ref| YP_007066165.1|; PP_06058	9.00e-014
3	5825691..5825807	hypothetical; PP_06059	0
4	5826905..5827384	glycosyl transferase family 2 [*Oscillatoria nigro-viridis* PCC 7112]. gi|428319715|ref| YP_007117597.1|; PP_06060	5.00e-026
5	5827381..5827593 BBD_1001073	hypothetical; PP_06061	0
6	5828156..5829235	*PHAGE_Ostreo_2_NC_028091*: hypothetical protein; PP_06062; phage (gi939177229)	8.00e-011
7	5829191..5829349	hypothetical; PP_06063	0
8	5829364..5829681	*PHAGE_Megavi_chiliensis_NC_016072*: thioredoxin-like protein; PP_06064; phage (gi363540574)	2.00e-008
9	5829758..5829895	hypothetical; PP_06065	0
10	5829878..5830357	*PHAGE_Pandor_salinus_NC_022098*: guanine deaminase; PP_06066; phage (gi531037005	4.00e-025
11	5830479..5831102	*PHAGE_Microm_12T_NC_020864*: hypothetical protein; PP_06067; phage (gi472342811	7.00e-007
12	5831263..5832063	*PHAGE_Entero_ST104_NC_005841*: ORF19; PP_06068; phage (gi46358666)	8.00e-018
13	5832192..5833109	type 11 *methyltransferase* [*Calothrix* sp. PCC 7507]. gi|427718977|ref|YP_007066971.1|; PP_06069	4.00e-106
14	5833106..5834035	iron permease FTR1 [*Oscillatoria* nigro-viridis PCC 7112]. gi|428315250|ref| YP_007113132.1|; PP_06070	3.00e-102
15	5834481..5835026	*PHAGE_Rubell_virus_NC_001545*: non-structural polyprotein; PP_06071; phage (gi336284683)	2.00e-012
16	5835195..5835266	tRNA	0
17	5835298..5835510	periplasmic solute binding protein [*Geitlerinema* sp. PCC 7407]. gi|428224031|ref| YP_007108128.1|; PP_06072	4.00e-010
18	5835859..5836659	hypothetical protein PCC7424_4846 [*Cyanothece* sp. PCC 7424]. gi|218441742|ref| YP_002380071.1|; PP_06073	5.00e-084
19	5836794..5837648	biotin-(acetyl-CoA-carboxylase) ligase BirA [*Oscillatoria acuminata* PCC 6304]. gi| 428211981|ref|YP_007085125.1|; PP_06074	2.00e-057
20	5837759..5838928	*PHAGE_Lactob_phig1e_NC_004305*: minor capsid protein; PP_06075; phage (gi23455811)	1.00e-016
	5837696..5838928	BLASTp: Peptidase, M23/M37 family [*Phormidium* sp. OSCR], Den Uyl., 2016: BBD_100107313, Membrane proteins related to metalloendopeptidases COG0739	2.00e-124
21	5839022..5839315	*PROPHAGE_Escher_Sakai*: putative transposase TnA; PP_06076; phage (gi15832531	8.00e-020
22	5839572..5839844	*PHAGE_Geobac_E3_NC_029073*: transposase; PP_06077; phage (gi985758480)	4.00e-007
23	5839852..5840331	*PHAGE_Geobac_E3_NC_029073*: transposase; PP_06078; phage (gi985758480)	2.00e-027
24	5840636..5840941 BBD_1001074	*PROPHAGE_Deinoc_R1*: serine protease; PP_06079; phage (gi15807944)	5.00e-007
25	5840913..5841986	*PROPHAGE_Deinoc_R1*: serine protease; PP_06080; phage (gi15807944)	0.00007
26	5842143..5843258	subtilase family protease [*Dactylococcopsis salina* PCC 8305]. gi|428780017|ref| YP_007171803.1|; PP_06081	4.00E-082
27	5843285..5843407	hypothetical; PP_06082	0
28	5843519..5843752	hypothetical; PP_06083	0
29	5843783..5843911	hypothetical; PP_06084	0
30	5844106..5845404	*PHAGE_Strept_20617_NC_023503*: enolase; PP_06085; phage (gi588295080)	8.00e-136
#	***Geitlerinema* sp. BBD_1991, G1 CDS pos. (BBD_1000996)**	**PHASTER - BLAST hits**	**E-value**
1	136544..136555	attL	0.0
2	136832..137278	*PHAGE_Bacill_Fah_NC_007814*: site-specific serine recombinase; PP_00294; phage (gi89152504)	9.48e-06
3	137312..137677	*PHAGE_Dinoro_DFL12phi1_NC_024367*: putative host-like protein; PP_00295; phage (gi658311036)	6.82e-09
4	137908..138861	*PHAGE_Synech_S_SKS1_NC_020851*: GDP-L-fucose synthase; PP_00296; phage	2.59e-76

		(gi472340899)	
5	138968..140044	*PHAGE_Synech_S_SKS1_NC_020851*: GDP-D-mannose 4,6-dehydratase; PP_00297; phage (gi472340900)	5.17e-139
6	140230..140961	hypothetical; PP_00298	0.0
7	141028..142284	*PHAGE_Synech_ACG_2014f_NC_026927*: N/A; PP_00299; phage (gi815854524)	7.99e-10
8	142366..143511	*PHAGE_Synech_ACG_2014f_NC_026927*: N/A; PP_00300; phage (gi815854731)	3.21e-19
9	156422..156433	attR	0.0
#	*Geitlerinema* sp. BBD_1991, G3 **CDS pos. (BBD_1001028)**	**PHAST - BLAST hits**	**E-value**
1	59826..61865	PHAGE_Cyanop_NATL2A_133_NC_016659: hypothetical protein; PP_03742; phage (gi372217849)	8.97e-27
2	61881..62588	hypothetical; PP_03743	0.0
3	62606..63580	hypothetical; PP_03744	0.0
4	63631..67305	*PHAGE_BacilI_BCD7_NC_019515*: putative baseplate J family protein; PP_03745; phage (gi422936037)	1.49e-17
5	67394..67780	*PHAGE_Synech_ACG_2014e_NC_026928*: base plate wedge subunit; PP_03746; phage (gi815854880)	8.45e-14
6	67813..68214	hypothetical; PP_03747	0.0
7	68248..68829	PHAGE_Campyl_NCTC12673_NC_015464: gp5 baseplate hub subunit and tail lysozyme; PP_03748; phage (gi332672341)	7.30e-08
8	68895..69953	*PHAGE_BacilI_BCD7_NC_019515*: hypothetical protein; PP_03749; phage (gi422936041)	3.37e-11
9	69972..70571	*PHAGE_Salmon_VilI_NC_015296*: conserved uncharacterised protein; PP_03750; phage (gi326804610)	4.24e-07
10	70619..72187	hypothetical; PP_03751	0.0
11	72224..72763	PHAGE_BacilI_BCD7_NC_019515: putative tail tube protein 2; PP_03752; phage (gi422936047)	2.74e-14
12	72779..72859	hypothetical; PP_03753	0.0
13	72922..73221	hypothetical; PP_03754	0.0
14	73299..73724	*PHAGE_BacilI_BCD7_NC_019515*: putative tail tube protein 2; PP_03755; phage (gi422936047)	2.24e-22
15	73776..75089	*PHAGE_BacilI_BCD7_NC_019515*: putative tail sheath protein; PP_03756; phage (gi422936048)	9.98e-30
16	75197..76447	*PHAGE_BacilI_BCD7_NC_019515*: putative tail sheath protein; PP_03757; phage (gi422936048)	3.98e-42
#	*Geitlerinema* sp. BBD_1991, G5 **CDS pos. (BBD_1001065)**	**VIRsorter - BLAST hits**	**E-value**
1	369..650	PFAM-DUF4258	5.80e-22
2	656..883	hypothetical protein	-
3	946..1206	hypothetical protein	-
4	1417..1635	hypothetical protein	-
5	1727..3124	hypothetical protein	-
6	3114..4043	hypothetical protein BLASTp: metalloendopeptidase-like membrane protein [*Oscillatoria acuminata*], WP_015152078.1 Den Uyl., 2016: BBD_10010657, membrane proteins related to metalloendopeptidases, COG0739	- 2.00e-19
7	4084..4608	hypothetical protein	-
8	4613..4909	hypothetical protein	-
9	4909..5232	hypothetical protein	-
10	5196..6245	hypothetical protein	-
11	6245..7012	hypothetical protein	-

Appendix 6

Details of publication prepared during the time of candidature

Chapter 2

Buerger P, Weynberg KD, Wood-Charlson EM, Willis B, van Oppen MJH (**near final version, in prep**). T4-bacteriophage communities associated with coral black band disease at Orpheus Island, Australia. Target: Environmental Microbiology.

Chapter 3

Buerger P, Alvarez-Roa C, Weynberg KD, Baekelandt S, Oppen MJH (**2016**) Genetic, morphological and growth characterisation of a new *Roseofilum* strain (Oscillatoriales, cyanobacteria) associated with coral black band disease. PeerJ 4:e2110 doi.org/10.7717/peerj.2110.

Chapter 4

Buerger P, Laffy P, Wood-Charlson EM, Weynberg KD, Willis B, van Oppen MJH (**in prep**). Phage therapy of BBD - implications for future experiments. Target: Coral Reefs.

Chapter 5

Buerger P, Wood-Charlson EM, Weynberg KD, Willis B, van Oppen MJH (**2016**). CRISPR-Cas defense system in *Roseofilum reptotaenium*: evidence of a bacteriophage-cyanobacterium arms race in the coral black band disease. Frontiers in Microbiology (Aquatic Microbiology) 7:2077.

Other relevant publications

Weynberg KD, Voolstra CR, Neave MJ, **Buerger P**, van Oppen MJH (**2015**). From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen. Nature Scientific Reports 5, 17889.

Montalvo Proano J, **Buerger P**, Weynberg KD, van Oppen MJH (**near final version, in prep**). Diversity of a putative *Symbiodinium* ssRNA virus isolated from corals on the Great Barrier Reef. Target: Frontiers in Microbiology (Virology).

Buerger P, Laffy P, Botte E, Bell SC, Cantin NE, van Oppen MJH, Webster NS (**in prep**). Virome analysis of bleached and healthy coral *Pocillopora damicornis* following the 2016 mass bleaching event. Target: Virology.

Buerger P, Montalvo Proano J, Bell SC, van Oppen MJH (**in prep**). Amplicon sequencing of a dinorna-like virus which infects *Symbiodinium* and potentially contributes to coral bleaching. Target: Nature Scientific Reports.