Genetic diversity of benthic marine cyanophage as revealed by sequence analysis of various phage marker genes

Alec Vallota-Eastman[1], Kerry McPhail[2], Rebecca Vega-Thurber[3]
Bioresource Research[1], College of Pharmacy[2], Department of Microbiology[3], Oregon State University, Corvallis, Oregon, 97331

Abstract

Cyanophage are viruses that infect cyanobacteria, photosynthetic microorganisms which are known to produce several anticancer and antibacterial chemical compounds. Cyanophage are significant members of microbial communities in marine and freshwater environments. There has been significant difficulty inventorying marine and freshwater cyanophage, however, due to the difficulty in cultivating their hosts. Using a culture-independent single gene amplification approach, the diversity of phage infecting these microbes was analyzed within samples of benthic cyanobacterial mats collected from the Red Sea and Panama. PCR-based primers were used to amplify several conserved phage marker genes. Resulting PCR products were then sequenced using the Illumina platform and the resulting marker gene libraries were assembled. Sequences within these libraries were clustered by 97% similarity and differences in sequence. OTU grouping were used to preliminarily infer differences in cyanophage consortia between the two unique geographic regions.

1. Introduction

1.1. Marine Natural Products

Small molecule secondary metabolites of organisms, or natural products (NPs), have served a central role in developing many current anticancer and antibiotic medicines. The importance of discovering novel bioproducts is great, as new pharmaceuticals are needed to combat the development of drug resistance in both bacterial pathogens and cancers. However, many NPs have already been characterized, with new chemical entities becoming more difficult to isolate, particularly from terrestrial sources[1]. This has led to high rates of re-isolation of known compounds in bioproduct research.

The advent of SCUBA has enabled researchers access to the Earth's greatest source of biodiversity, the marine environment. Out of over 35 known phyla of bacteria to date, it is estimated that 34 inhabit the ocean, with 15 being exclusive to this environment[2]. Such great biological diversity is a promising source for additional chemical diversity. A study conducted by Kong and coworkers[3] demonstrates the marine environment is a superior source of chemical novelty. When comparing chemical scaffolds between marine and terrestrial natural product dictionaries, it was found that 71% of unique chemicals were exclusively produced by marine organisms. In a separate study done by the National Cancer Institute, marine and terrestrial samples were screened for cytotoxicity. The study found that 1% of the screened marine

samples showed anti-tumor potential compared to just 0.1% of the tested terrestrial samples, suggesting that marine natural products may be superior in terms of bioactivity in addition to chemical novelty[3].

1.2. Marine Cyanobacteria

Cyanobacteria is a metabolically diverse phylum of bacteria with the ability to obtain energy from photosynthesis. Eukaryotic chloroplasts originated from an ancient endosymbiont belonging to this phylum[4]. In addition to their ability to photosynthesize, some strains of cyanobacteria can fix nitrogen and are thus important providers for nitrogen fertilizer, used in the cultivation of several crops. Cyanobacteria are also known for their morphological diversity – some species are unicellular while others may form filamentous multicellular colonies. Members of the phylum inhabit many environments in nature, including terrestrial, aquatic, and marine environments. While many species of cyanobacteria have been found to produce interesting natural products, the majority of biologically active nitrogen-containing natural products have been reported from marine cyanobacteria[5] and, more specifically, those found in tropical climates[6–8]. One example includes dolastatin 10, an anti-microtubule agent currently in preclinical trials as an anticancer drug. The second and third generation analogues of this compound – TZT-1027 and dolastatin 15 – are currently undergoing phase I and phase II testing for tumor treatment, respectively[5].

1.3. Cyanophage

Cyanophage, or viruses that infect cyanobacteria, are members of the most abundant entities in the biosphere, with an estimated 10 bacteriophage (phage) for every 1 bacterium[9,10]. As a result of their abundance and host lysis, phage play a major role in the energy and nutrient cycles within earth's ecosystems. Phage have also been found to be drivers of evolution through the intra-species transfer and incorporation of genetic material to bacterial genomes[11]. Despite growing evidence of the dominance and immense diversity of these entities within marine environments, only an infinitesimal fraction of marine phage has been inventoried. This has largely been due to the difficulty of cultivating many of their hosts in laboratory culture conditions, specifically those found in the water column of both marine and freshwater environments. Methods for analyzing viral diversity have rapidly progressed in the era of genomics characterized by advances in molecular biology methodologies, sequencing technology, and computational analysis. Through DNA and RNA extraction from samples and the use of polymerase chain reaction (PCR), it is possible to analyze diversity within a sample population in a more high-throughput manner compared to culture-dependent methods[12].

1.4. Metagenomic and Amplicon Analysis

The field of metagenomics strives to answer the question of what organisms are present through the direct sequencing of the collection of whole genomes present within the sample. In this approach, whole genomic DNA is extracted from the sample and is digested into short fragments. These fragments can then be sequenced using one of several available sequencing

technologies[13]. The sequencing data is processed to remove low-quality and low-abundance reads. Following this quality control, reads may be assembled based on overlapping sequences between two or more reads to form longer continuous sequences or 'contigs'. Assembly decreases computational power needed to analyze data downstream and can greatly assist in functional gene annotation; however, it is more likely that less abundant reads will be masked and that erroneous chimera sequences (i.e., sequences that result from incorrect joining of two or more unrelated sequences, either during assembly or PCR) will be produced[14]. Taxonomic members of the community can then be identified on assembled or non-assembled data by "binning" reads into groups based on inter-library sequence similarity or known gene similarity.

An alternative method for culture-independent community analysis is through single marker gene amplification, or amplicon analysis. In this method, PCR primers are designed to target a highly conserved universal gene widely distributed among organisms. This gene can be used as a genetic "fingerprint", or molecular marker, to detect the identity and abundance of organisms present within the sample. This is generally done using the small subunit ribosomal RNA (16S/18S) gene in bacteria and eukaryotes. Unlike cellular organisms, phage lack ribosomal RNA genes and there are no conserved genes that can be used as a universal molecular marker for phage diversity. Whole genome analyses, however, have identified functional and structural genes which are conserved among certain groups of phage. For example, DNA polymerase B (*polB*) and major capsid protein (MCP) genes have been found to be highly conserved among *Podoviridae* – a family of T7-like phage (which have short tails)[15,16] – while capsid assembly protein (*g20*) and portal head protein (*g23*) genes are conserved among T4-like *Myoviridae* (which have contractile tails)[17–19]. Therefore, several of these molecular markers can be used in parallel to capture a high-level snapshot of phage diversity, allowing a more targeted and less data-intensive approach compared to metagenomic analyses.

Here we report the results of using several modified analogues of previously designed primer sets[20] to investigate phage diversity present within samples of benthic cyanobacterial mats collected from the Red Sea and the coast of Panama. To our knowledge, this is the first example using single gene amplicon analysis to investigate cyanophage associated with raw mat-forming cyanobacterial aggregates. We hypothesized that the use of these selected primer sets would allow us to detect the presence of unique members of these viral families.

2. Methods

2.1. Sample Collection and DNA Extraction

Red Sea samples were collected off the Eastern coast near Al Ahlam Resort 24^{o}N latitude and 37^{o}E longitude) in May, 2013 and November, 2014. Panama samples were collected from the Gulf of Panama (8^{o}N latitude and 80^{o}W longitude) in November, 2003 and June 2004. Samples were filamentous algal-like cyanobacterial matts and were collected from various depths, ranging from 55 feet to the surface. All samples were stored in RNA*later* RNA Stabilization Reagent (Qiagen, Hilden, Germany) and kept frozen at -80^{o}C. A subset of fourteen samples

were selected for processing based on variable dive site coverage among the Red Sea and Gulf of Panama. Extraction was carried out using ~250 mg of each algal matt. In brief, DNA was extracted using an organic extraction protocol[21] on the raw sample, as previously described. Resulting extracts were purified using a DNA Clean & Concentrator kit, following manufacturers instructions (Zymo Research, Irvine, CA, USA).

2.2. Primer design

Primers were designed as previously described (Table 1)[13]. Adaptor and pad sequences were included on the 5' ends of each forward and reverse primer. The adaptor sequences were AF (5'-TCG TCG GCA GCG TC-3') and AR (5'-GTC TCG TGG GCT CGG-3'). The pad sequence, included between each adaptor and annealing region, was P (5'-AGA TGT GTA TAA GAG ACA G-3'). The final primer sequences were used in the first step in a two-step PCR amplification reaction protocol, further described below. Nextera primers specific to our adaptors were used to attach unique barcode sequences to the resulting amplicons in a second-step PCR reaction, further detailed by Fadrosh et al.[23]

2.3. Marker gene PCR amplification

All PCR reactions were performed in a 25 uL volume of 1 uL of extracted DNA template or PCR product and 24 uL PCR mixture containing Tough-TaqDNA polymerase assay buffer (50 mM KCl/20 mM Tris•HCl, pH 8.4), 1.5 mM MgCl, 0.20 mM of each dNTP, 40 nmol each of the primers, and 1 unit Taq polymerase (AccuStartTM II PCR ToughMix, Quanta Biosciences, Gaithersburg, MD, USA). Negative controls consisted of 1 uL of nuclease-free water and 24 uL of PCR mixture. PCR commenced with an initial denaturation step at 94^{o}C for 3 min, followed by 35 cycles of denaturation at 94^{o}C for 45 seconds, annealing at variable temperatures for 1 min, extension at 72^{o}C for 1.5 min, and then a final extension step at 72^{o}C for 10 min. Primers were screened against several samples using a gradient PCR reaction[24] to identify the optimal annealing temperatures (Figure 1). For primers targeting g20, mcp, and DNA polymerase B genes (Table 1) an annealing temperature of 54.5^{o}C was used. For primers targeting the g23 gene, an annealing temperature of 41.9^{o}C was used. 5 microliters of PCR product was verified by agarose gel electrophoresis on a 1.0% (wt/vol) agarose gel in 1 × TBE (89 mM Tris-Cl, 89 mM boric acid, 2 mM EDTA, pH 8.0) and was stained with SYBR Green I Nucleic Acid Gel Stain (Invitrogen, Carlsbad, CA, USA) and viewed with an ultra violet transilluminator. Amplicons of the correct sizes were barcoded in a second PCR reaction as previously described[25]. Successfully barcoded amplicons were verified by agarose gel electrophoresis using unbarcoded amplicons as controls.

2.4. PCR sequencing

PCR products were purified using an Agencourt AMPure XP PCR cleanup kit, following manufacturer's instuctions (Beckman Coulter, Brea, CA, USA). The concentration of purified product was determined using using a Quibit Quantitation Kit as well as a Quibit Fluorometer

(Thermo Fisher Scientific, Waltham, MA, USA). All amplicons were pooled in equimolar ratios based on their predetermined DNA concentrations and were submitted to the Center for Genomics Research and Biocomputing (CGRB) at Oregon State University for Illumina sequencing (Illumina, San Diego, CA, USA).

2.5. Data Processing and Diversity Analysis

Sequencing data was processed through the QIIME pipeline to obtain a marker gene library[29]. First, adaptor, pad, and barcode sequences were trimmed from each read. Forward and reverse reads were then merged and sequences with low quality scores were discarded from the dataset. Chimeric sequences and sequences less than 100 nucleotides in length were also excluded from downstream analyses. Operational taxonomic unit (OTU) binning was performed at 97% sequence similarity threshold and without taxonomic assignment.

Microbial diversity was evaluated within-community (alpha diversity) using QIIME. Rarefaction, to a subsampling depth (determined by the minimum number of sequences in a sample) of 300 sequences/sample was performed on all samples to standardize the sequencing effort. Alpha diversity was measured with the Chao1 (richness) metric and alpha rarefaction curves were produced using QIIME. OTUs were visualized by heat maps, produced using the PhyloSeq package of R[30]. OTU data produced from QIIME was plotted in Excel to produce pie charts. OTUs with less than 10 reads were combined into a single category labeled “Other”. Major OTUs from each resulting pie chart were filtered from the complete OTU table produced by QIIME. The resulting filtered table was used to retrieve representative fasta sequences from each of these major OTUs. These sequences were then BLASTed using NCBI’s blastn suite against the “nr” database.

3. Results

3.1. Isolation of various marker genes

PCR products of approximately 540-660, 250, 610, 410, and 200 bp were amplified using the primers shown in the respective order that they are listed (Table 1).

Table 1.

Target Viral Group	Marker Gene	Primer Sequence*	Reference
T4 phage	g23/major capsid	MZIA1bis: GAT ATT TGI GGI GTT CAG CCI ATG A	Filee et al., 2005
		MZIA6: CGC GGT TGA TTT CCA GCA TGA TTT C	
T4 phage	g20/portal head protein	CPS1: GTA GWA TTT TCT ACA TTG AYG TTG G	Fuller et al., 1998
		CPS2: GGT ARC CAG AAA TCY TCM AGC AT	

	g20/portal head protein	CPS1: GTA GWA TTT TCT ACA TTG AYG TTG G	Zhong et al., 2002
		CPS8: AAA TAY TTD CCA ACA WAT GGA	
T7 phage	mcp	MCPF5: GTT CCT GGC ACA CCT GAA GCG T	Baker et al., 2006
		MCPR5: CTT ACC ATC GCT TGT GTC GGC ATC	
T7 phage	DNA polymerase B	AVS1: GARGGIGCIACIGTIYTIGAYGC AVS2: GCIGCRTAICKYTTYTTISWRTA	Chen and Suttle 1995

Primer sets and their associated gene targets used in the construction of marker gene libraries

* The codes for mixed bases are: R = A, G; Y = C, T; M = A, C; K = G, T; S = C, G; W = A, T; H = A, C, T; B = C, G, T; V = A, C, G; D = A, G, T; n = A, G, C, T; I = inosine.

Optimal annealing temperatures were identified by visual inspection of amplicon sizes. An example showing annealing temperature optimization for CPS1 and CPS8 is shown (Figure 1). Specific amplification of the greatest number of samples were used in the identification of the best temperature for downstream library preparations. For these reasons, 50.2^{o}C was selected for amplification with CPS1 and CPS8. The remainder of primer sets were evaluated in a similar manner.

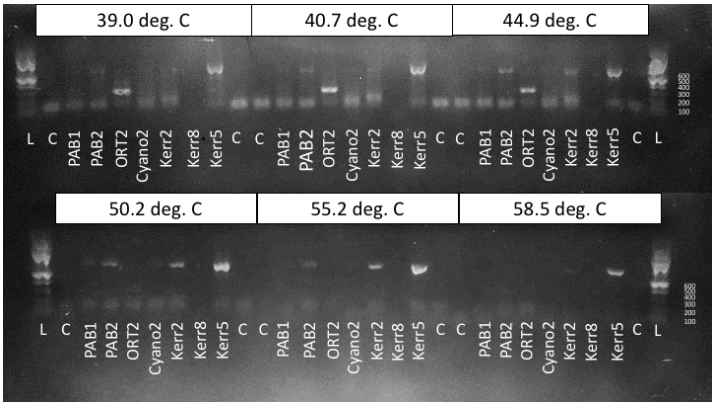


Figure 1.
Gel electrophoresis of PCR product amplified using CPS1 and CPS8 primers under various annealing temperatures (shown above). Sample names are shown under their

3.2. Diversity analyses

Resulting OTUs clustered closely together for a given marker gene used. When looking at OTU diversity sorted by gene target (Figure 2a), primers targeted to the DNA Polymerase B gene were specific to a small number of closely related OTUs compared to all other primers. Primers amplifying the g23 gene detected the largest OTU diversity. Chao1 rarefaction curves (Figure 2b) were produced with a subsampling depth of 300 sequences/sample, appropriate for the diversity present. These measures confirm patterns seen within the heatmap, with the highest alpha diversity seen in g23 amplicons and lowest average diversity seen in DNA polymerase B amplicons. Additionally, major OTUs detected between different marker genes were dissimilar, as expected (Figure 3). One exception to this pattern is OTU111, which was found by using DNA polymerase B and mcp genes as markers. BLASTing the representative sequence from OTU111

returned Xanthomonas campestris as a hit with 88% identity and an E-value of 0.003. This is a bacterial species that causes a variety of plant diseases in cruciferous vegetables.

When sorting by location (Figure 2c), we see OTUs that are somewhat universal across samples from both locations, although they are of low abundance. We also can see unique OTU groupings between the two locations. OTUs higher on the y-axis were unique to samples collected from Panama. A cluster of OTUs near the center also shows unique diversity found from Red Sea samples. Additionally, the heat map and alpha diversity rarefaction curves for sample location (Figure 2d) show Panama samples to have a higher average diversity compared to samples from the Red Sea.

Major OTUs detected from all marker genes between the two locations were somewhat dissimilar with a few exceptions. OTU1 was found in high abundance by marker gene g20 in both Panama and Red Sea samples. BLASTing a representative sequence from this cluster returned the following hit: “Uncultured cyanomyovirus clone T4-like portal protein (g20) gene” with 82% identity and an E-value of 3e-23. OTUs 67 and 119 were also BLASTed which returned a hit for *Ovis Canadensis* (bighorn sheep) with 76% identity and an E-value of 0.13 and *Spirometra erinaceieuropaei* (a species of tapeworm) with 81% identity and an E-value of 1.6, respectively.

When using mcp as a marker, OTU0 was found to be a major component within Red Sea samples. BLASTing this OTU returned Uncultured Myoviridae clone major capsid protein gene as a hit with 76% identity and an E-value of 2e-11. The major OTU from Panama samples using this genetic marker, OTU 190, was identified as a cyanobacteria belonging to the genus *Cyanothece* with 75% identity and an E-value of 2e-30.

OTU2 was found in both Red Sea and Panama samples using the g23 gene. BLAST results for this OTU returned *Moorea producens* (a species of filamentous toxin-producing cyanobacteria) with 100% identity and an E-value of 1e-140. OTU 5, found in Red Sea samples using the g23 gene, returned uncultured cyanophage T4-like portal protein (g20) gene as a hit with 94% identity and an E-value of 9e-42.

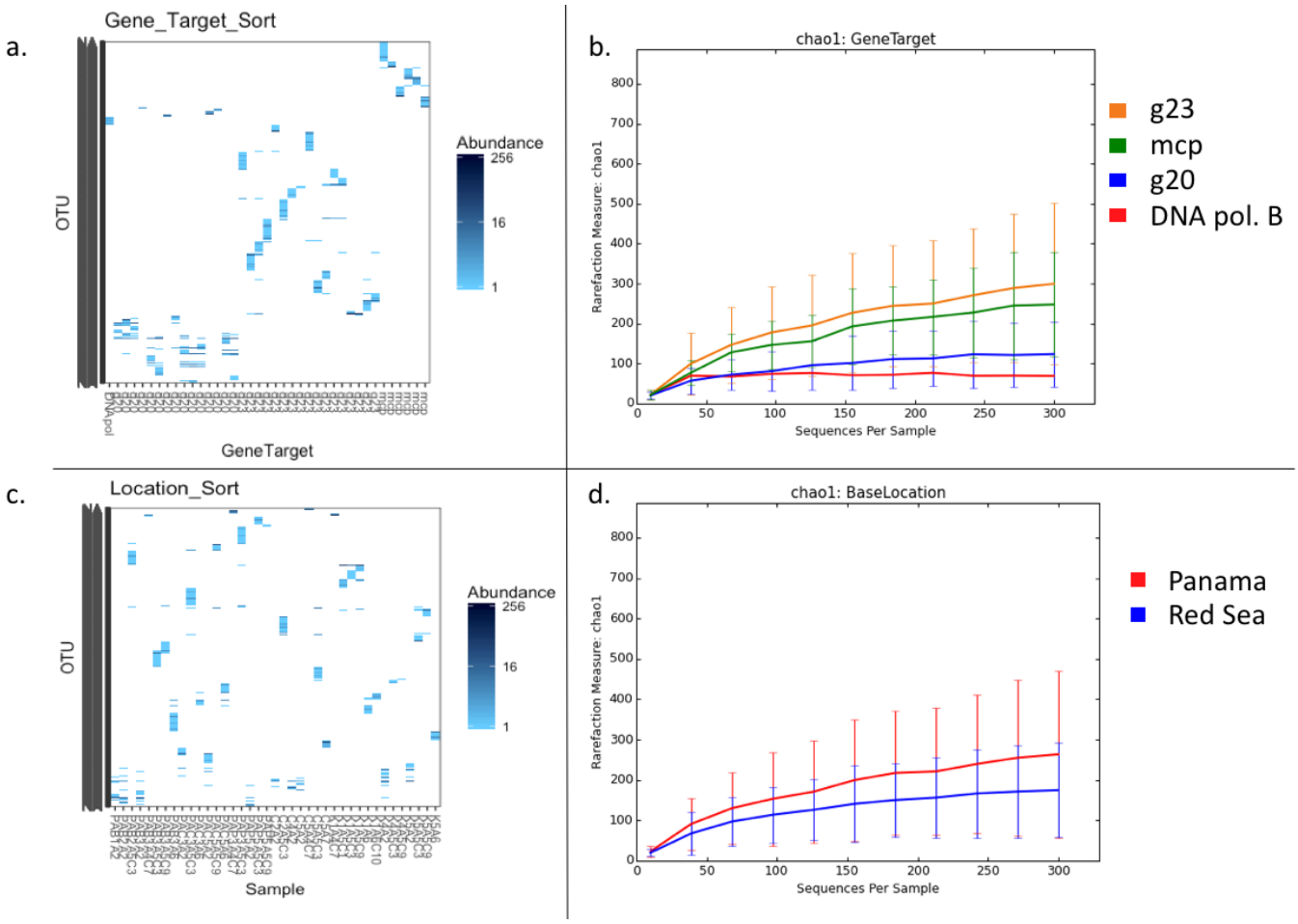
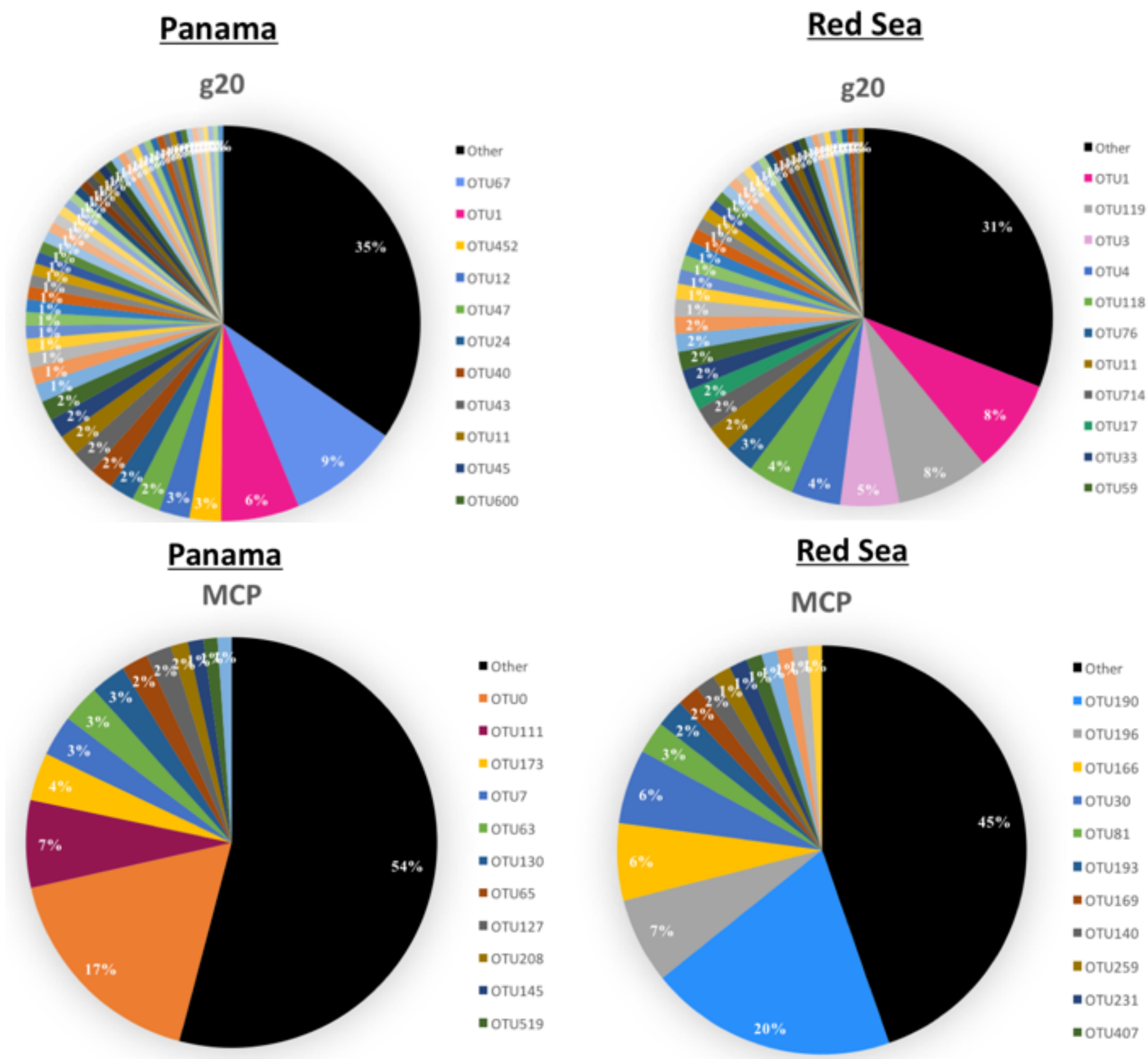


Figure 2.
Heat map of samples, labeled and sorted by marker gene used with OTU IDs on the y-axis and samples on the x-axis (a.). Rarefaction curves for Chao1 alpha diversity between gene markers (b.). Heat map of samples sorted by location (i.e. Panama or Red Sea). Panama samples are coded by a “P”. Red Sea samples are coded by “C” or “K” (c.). Rarefaction curves between sample location (d.).


Panama
g20
Other
OTU67
OTU1
OTU452
OTU12
OTU47
OTU24
OTU40
OTU43
OTU11
OTU45
OTU600
Red Sea
g20
Other
OTU1
OTU119
OTU3
OTU4
OTU118
OTU76
OTU11
OTU714
OTU17
OTU33
OTU59
Panama
MCP
Other
OTU0
OTU111
OTU173
OTU7
OTU63
OTU130
OTU65
OTU127
OTU208
OTU145
OTU519
Red Sea
MCP
Other
OTU190
OTU196
OTU166
OTU30
OTU81
OTU193
OTU169
OTU140
OTU259
OTU231
OTU407

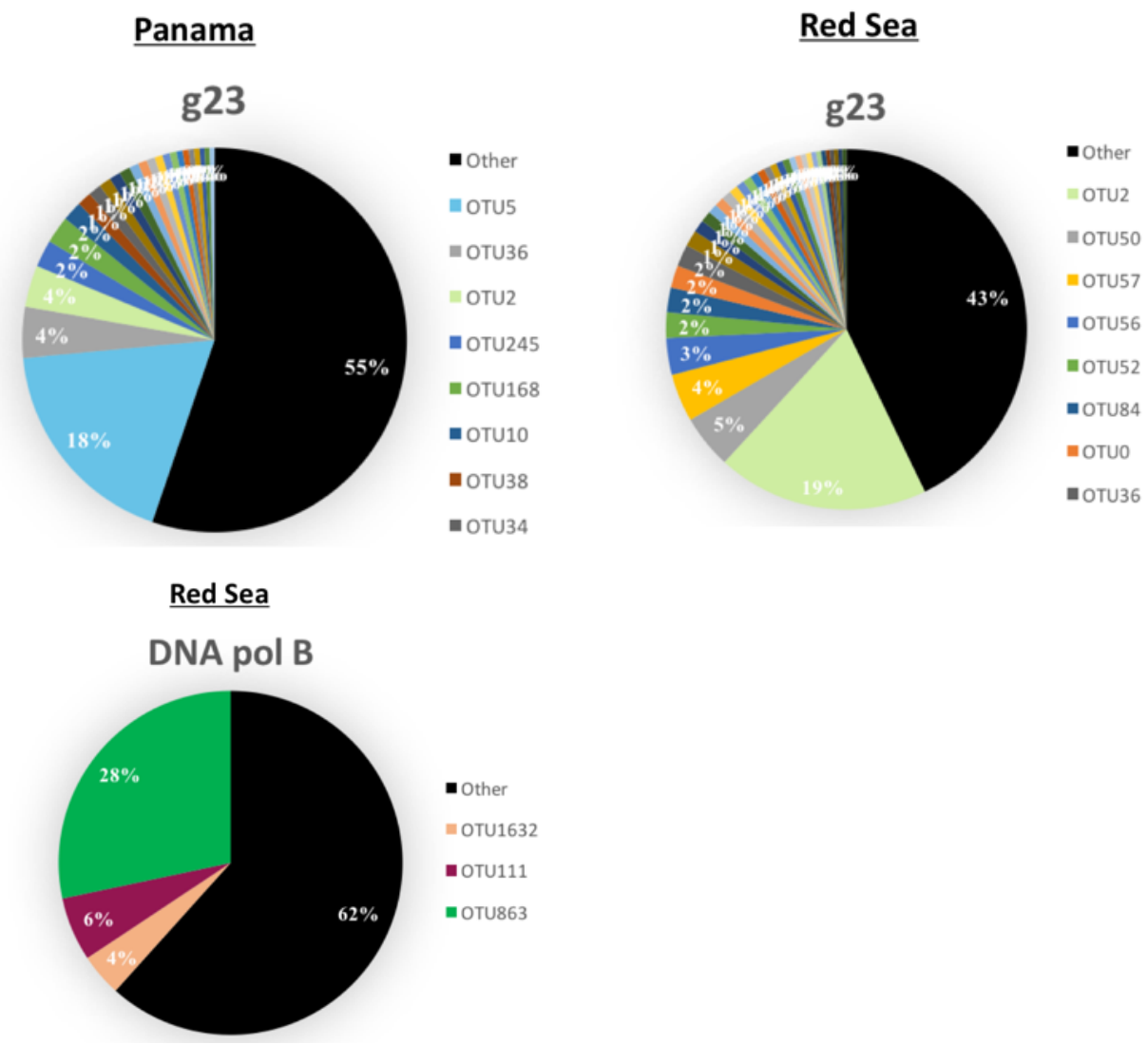


Figure 3.
OTU diversity within marker gene library grouped by marker gene type and sample location. OTUs with less than 10 reads were combined into a single category, labeled “Other”.

4. Discussion

Current findings on the immense ubiquity and diversity of marine cyanophage suggests that they play major roles in energy and nutrient and cycles within cyanobacterial communities. Their known ability to transfer genetic material to hosts begs the question of whether they may be associated with the evolution of novel biosynthetic gene clusters and, thus, chemical diversity. Since phage genomes do not contain universal genetic markers, there is a need to utilize several primer sets when evaluating natural diversity through single marker gene amplification. Several primers were identified from literature, which have proved useful for the construction of viral marker gene libraries[13]. In this study, we employed several of these genes to investigate the genetic diversity of natural cyanophage communities associated with

benthic marine mat-forming cyanobacteria, sampled from the Red Sea and Panama. Although these markers have been used individually for separate analyses of aquatic cyanophage isolates, this will be the first study, to our knowledge, utilizing several of these genes, in parallel, for the analysis of cyanophage associated with raw tropical cyanobacterial mats from marine environments. Using a variety of primers allowed us to capture a large spread of genetic diversity, as clustered OTUs were generally not shared between different genes used. This points to the importance of using markers specific to different phage groups.

Future work on this project will develop a method for taxonomic assignment to the major OTUs identified by each marker gene and within each geographical location. Following taxonomic assignment, cyanophage diversity may be compared to our previously collected liquid chromatography-mass spectrometry chemical signature data. These comparisons may lead to interesting hypotheses about the correlation between cyanophage and chemical diversity in nature.

Acknowledgements

We thank all supporters of this study, Rebecca Vega-Thurber for her constant guidance and project development, Kerry McPhail for providing samples and associated metadata, Ryan McMinds for assistance with sequencing preparation and technical expertise, Mark Dasenko of the Center for Genome Research and Biocomputing for assistance with Illumina sequencing, Thomas Sharpton and Grace Klinges for assistance with QIIME and R/PhyloSeq, Ian Munoz for graphical assistance, and Wanda Crannell and Kate Field for their assistance with editing and general advising on the project.

Funding Statement

This work was supported by the Undergraduate Research, Innovation, Scholarship, & Creativity (URISC) grant, Louis Stokes Alliance for Minority Participation (LSAMP), Oregon State University College of Pharmacy, The E.R. Jackman Internship Support Program, the Bioenergy Scholarship, and the Ben and Phyllis Newell Scholarship.

References

1 Dias, D.A. *et al.* (2012) A Historical Overview of Natural Products in Drug Discovery. *Metabolites* 2, 303–336
2 Malve, H. (2016) Exploring the ocean for new drug developments: Marine pharmacology. *J. Pharm. Bioallied Sci.* 8, 83–91
3 Munro, M.H. *et al.* (1999) The discovery and development of marine compounds with pharmaceutical potential. *J. Biotechnol.* 70, 15–25
4 Raven, J.A. and Allen, J.F. (2003) Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol.* 4, 209
5 Uzair, B. *et al.* (2012) Exploring Marine Cyanobacteria for Lead Compounds of Pharmaceutical Importance. *Sci. World J.* 2012,

6 McPhail, K.L. *et al.* (2007) Antimalarial linear lipopeptides from a Panamanian strain of the marine cyanobacterium Lyngbya majuscula. *J. Nat. Prod.* 70, 984–988
7 Vining, O.B. *et al.* (2015) Depsipeptide Companeramides from a Panamanian Marine Cyanobacterium Associated with the Coibamide Producer. *J. Nat. Prod.* 78, 413–420
8 Medina, R.A. *et al.* (2008) Coibamide A, a Potent Antiproliferative Cyclic Depsipeptide from the Panamanian Marine Cyanobacterium Leptolyngbya sp. *J. Am. Chem. Soc.* 130, 6324–6325
9 Suttle, C.A. (2005) Viruses in the sea. *Nature* 437, 356–361
10 Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* 399, 541–548
11 Clokie, M.R. *et al.* (2011) Phages in nature. *Bacteriophage* 1, 31–45
12 Perez Sepulveda, B. *et al.* (2016) Marine phage genomics: the tip of the iceberg. *FEMS Microbiol. Lett.* 363,
13 Rebecca L. Vega Thurber *et al.* Marine Viral Metagenomics with Emphasis on Coral Microbiomes.
14 Sharpton, T.J. (2014) An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5,
15 Baker, A.C. *et al.* (2006) Identification of a Diagnostic Marker To Detect Freshwater Cyanophages of Filamentous Cyanobacteria. *Appl. Environ. Microbiol.* 72, 5713–5719
16 Takashima, Y. *et al.* (2007) Development and Application of Quantitative Detection of Cyanophages Phylogenetically Related to Cyanophage Ma-LMM01 Infecting *Microcystis aeruginosa* in Fresh Water. *Microbes Environ.* 22, 207–213
17 Zhong, Y. *et al.* (2002) Phylogenetic Diversity of Marine Cyanophage Isolates and Natural Virus Communities as Revealed by Sequences of Viral Capsid Assembly Protein Gene g20. *Appl. Environ. Microbiol.* 68, 1576–1584
18 Filée, J. *et al.* (2005) Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12471–12476
19 Fuller, N.J. *et al.* (1998) Occurrence of a sequence in marine cyanophages similar to that of T4 g20 and its application to PCR-based detection and quantification techniques. *Appl. Environ. Microbiol.* 64, 2051–2060
20 Short, S.M. *et al.* (2010) The construction and analysis of marker gene libraries. *Man. Aquat. Viral Ecol.*
21 Vega Thurber, R. *et al.* (2009) Metagenomic analysis of stressed coral holobionts. *Environ. Microbiol.* 11, 2148–2163
22 Zymo Research Corp. DNA Clean & Concentrator. , D4005
23 Fadrosh, D.W. *et al.* (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2, 6
24 Vincent R. Prezioso and Axel Jahns Using Gradient PCR to Determine the Optimum Annealing Temperature. , *Drug Discovery Online*. [Online]. Available: https://www.drugdiscoveryonline.com/doc/using-gradient-pcr-to-determine-the-optimum-a-0002. [Accessed: 09-Jun-2017]
25 Wu, L. *et al.* (2015) Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol.* 15, 125
28 Mitra, A. *et al.* (2015) Strategies for Achieving High Sequencing Accuracy for Low Diversity Samples and Avoiding Sample Bleeding Using Illumina Platform. *PLoS ONE* 10,

29 Caporaso, J.G. *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336
30 McMurdie, P.J. and Holmes, S. (2012) Phyloseq: a bioconductor package for handling and analysis of high-throughput phylogenetic sequence data. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.*