Inter-Research Science Publisher

Aquatic Microbial Ecology
Volume 91:1–14

Published February 6, 2025
https://doi.org/10.3354/ame02014

Check for updates

OPEN ACCESS

The hypersaline northwestern Arabian Gulf contains a phylogenetically diverse and highly uneven community of viruses related to cyanophages and pelagiphages

Awatef Almutairi[1,*], Curtis A. Suttle[2], Julia A. Gustavsen[3,4]

[1]Department of Biological Sciences, Faculty of Science, Kuwait University, Safat 13060, Kuwait
[2]Departments of Earth, Ocean and Atmospheric Sciences, Botany, Microbiology and Immunology, and the Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, BC, Canada
[3]Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[4]*Present address*: Agroscope, Rte de la Tioleyre 4, 1725 Posieux, Switzerland

ABSTRACT: Cyanobacteria are the dominant primary producers in many marine waters, and are intimately connected with the cyanophages that infect them. The most commonly isolated marine cyanophages form a monophyletic group based on several marker genes including g20, a gene that codes the capsid assembly protein. Based on morphology, these viruses are typically referred to as cyanomyoviruses. Here, we used g20 sequences to interrogate the diversity of cyanomyoviruses at 5 locations in the waters of the northwestern Arabian Gulf. The diversity and richness of g20 sequences varied among locations and were highest at the southernmost sites. Most sequences belonged to a small number of operational taxonomic units (OTUs), with the rest belonging to low abundance rare OTUs. Phylogenetic analysis revealed that the most abundant genotypes fell within the ubiquitous cyanomyovirus Cluster II, while others clustered with metagenome assembled pelagimyophage sequences and other environmental sequences across a broad diversity of clades. This study revealed a diverse community of cyanomyoviruses in the northwestern Arabian Gulf that is dominated by a few relatively abundant but phylogenetically diverse taxa.

KEY WORDS: Viruses · Cyanophage · Cyanomyovirus · Pelagimyophage · g20 · *Synechococcus* · Arabian Gulf · Kuwait

1. INTRODUCTION

Viruses are highly abundant in the marine environment and comprise a diverse and dynamic component of aquatic systems that infects autotrophic and heterotrophic prokaryotes and eukaryotes (Proctor & Fuhrman 1990, Breitbart & Rohwer 2005, Suttle 2005). Marine cyanophages are a morphologically and genotypically diverse group of tailed viruses that infect cyanobacteria in the genera *Synechococcus* and *Prochlorococcus* (Proctor & Fuhrman 1990, Suttle & Chan 1993, 1994, Waterbury & Valois 1993, Marston & Sallee 2003, Sullivan et al. 2003, Baran et al. 2022). Ubiquitously distributed in marine surface waters, cyanophages are significant agents of cyanobacterial mortality (Proctor & Fuhrman 1990, Suttle 1994, Suttle & Chan 1994, Mann 2003), and affect the density, distribution, and clonal composition of cyanobacterial communities (Wang et al. 2011, Marston et al. 2013, Chow & Suttle 2015, Ahlgren et al. 2019, Carlson et al. 2022, Dart et al. 2023).

Cyanophages are tailed double-stranded DNA viruses previously belonging to the order *Caudovirales*. Depending on tail morphotype, viruses of this

*Corresponding author: a.almutairi@ku.edu.kw

© The authors 2025

Publisher: Inter-Research · www.int-res.com

order were historically subdivided into 3 families *Myoviridae*, *Podoviridae*, and *Siphoviridae* (Suttle & Chan 1993, Waterbury & Valois 1993, Marston & Sallee 2003). Members of these families in aquatic environments include viruses that infect SAR11 bacteria and cyanobacteria. However, with the recent updates of phage classification by the International Committee on Taxonomy of Viruses (ICTV), all tailed bacterial and archaeal viruses with double-stranded DNA and icosahedral capsids are now members of the class *Caudoviricetes* (Walker et al. 2022).

The previous classification described *Myoviridae* as genetically related to T4-like phages with contractile tails. Currently, these viruses are designated as myoviruses based on their morphology. Historically, myoviruses infecting cyanobacteria have been termed cyanomyoviruses (Suttle & Chan 1993, Waterbury & Valois 1993, Marston & Sallee 2003, Sullivan et al. 2008). Cyanomyoviruses and cyanopodoviruses are both abundant in diverse marine environments (Huang et al. 2015) with cyanomyoviruses being in high abundance in coastal waters (Suttle & Chan 1993, 1994).

Cyanomyoviruses generally have relatively broad host ranges (Suttle & Chan 1993, Waterbury & Valois 1993, Sullivan et al. 2008) and high genetic diversity, often assessed based on genes coding the major capsid protein (g23) (Filée et al. 2005, Marston & Amrich 2009), DNA polymerase (g43) (Marston & Amrich 2009, Finke & Suttle 2019), and the capsid portal protein (g20) (Fuller et al. 1998, Zhong et al. 2002, Short & Suttle 2005, Sullivan et al. 2008).

Portal proteins can be used to classify abundant uncultured viruses (Lopes et al. 2014). In particular, analysis of g20 amplicon sequences has highlighted the high genetic diversity and prevalence of cyanomyoviruses in space and time (Fuller et al. 1998, Zhong et al. 2002, Frederickson et al. 2003, Marston & Sallee 2003, Dorigo et al. 2004, Short & Suttle 2005, Sandaa & Larsen 2006, Sullivan et al. 2008, Matteson et al. 2011, 2013, Wang et al. 2011, Zhong & Jacquet, 2013, Hanson et al. 2016). Moreover, studies have shown that g20 diversity varies with *Synechococcus* host abundance (Frederickson et al. 2003, Wang & Chen 2004, Wang et al. 2011, Hanson et al. 2016) and environmental changes, including temperature, salinity, and nutrient availability (Suttle & Chan, 1994, Lu et al. 2001, Wang & Chen 2004, Wang et al. 2011). Thus, amplicon sequencing of g20 gene fragments is a robust way to infer the diversity of cyanomyoviruses that can be compared to other studies and contextualized through phylogenetic analysis.

Although the potential for off-target amplification can be minimized by using modified g20 primers that specifically target cyanomyophage isolates (Sullivan et al. 2008), a criticism of using g20 amplicon sequences to assay cyanomyovirus diversity is the potential for the primers to amplify g20 sequences of other T4-like phages, or that some cyanomyoviruses may be missed (Short & Suttle 2005, Zhong & Jacquet 2013). To date, phylogenetic analyses have identified 4 clusters (I to IV) of g20 sequences with cultured cyanomyovirus representatives (Zhong et al. 2002, Sullivan et al. 2008), while other clusters are still known only from environmental sequences (Zhong et al. 2002, Frederickson et al. 2003, Short & Suttle 2005, Jameson et al. 2011, Zhong & Jacquet 2013). It has been speculated that g20 sequences of these environmental clusters belong to either undiscovered cyanomyophages or to myophages that infect other bacterial hosts (Short & Suttle 2005, Sullivan et al. 2008). Interestingly metagenomic studies of myophages infecting the globally abundant SAR11 clade (pelagimyophages) have shown a high similarity to cyanomyophages (Zaragoza-Solas et al. 2020). Metagenome-assembled genomes of putative pelagimyophages showed similar gene synteny with cyanophage isolates, and were closely related using a selection of concatenated genes (Zaragoza-Solas et al. 2020).

Despite extensive investigations of cyanophage diversity, the waters of the Arabian Gulf (also called the Persian Gulf and hereafter the Gulf) are unexplored, with the exception of a recent study of 3 areas along the coast of Kuwait that used denaturing gradient gel electrophoresis (DGGE) to show differences among months, and sequencing of 121 g20 amplicons to reveal the prevalence of cyanomyoviruses in Marine Cluster II (Almutairi et al. 2023).

The Gulf is a semi-enclosed shallow marine system in the northwestern Indian Ocean, surrounded by 8 countries and is a major region of oil production, refining, and shipping (Sheppard et al. 2010, Zhao et al. 2015). The northwest of the Gulf is affected by freshwater inflow from the Shatt Al-Arab River (Al-Yamani et al. 2004, Naser 2013, Devlin et al. 2015) that enters near the confluence of the borders of Kuwait, Iraq, and Iran, and introduces nutrient-rich water, as well as creating a seasonal north to south salinity gradient (Al-Yamani et al. 2004, Alosairi et al. 2011, Devlin et al. 2015). However, evaporation exceeds freshwater input, resulting in salinities in excess of 38 ppt for much of the Gulf (Pous et al. 2015).

Kuwait is situated at the western edge of the northern Gulf, with a coastline of about 500 km that borders Iraq and is intersected by the Shatt Al-Arab River in the north, and borders Saudi Arabia in the south. The coastal waters of Kuwait are shallow, with a maximum

depth of about 30 m, and are well mixed year-round (Al-Yamani et al. 2004), but subject to wide shifts in temperature (15 to 19°C in winter and 30 to 32°C in summer) and salinity, and are affected by rapid coastal development and anthropogenic influence (Sheppard et al. 2010, Naser 2013, Devlin et al. 2015).

Proteobacteria constitute a major fraction of the bacterioplankton community in local waters, and the SAR11 clade is highly ubiquitous (Almutairi 2015, Ismail & Almutairi 2022). Phytoplankton biomass and primary productivity in the Gulf are similar to many other coastal areas (Al-Yamani et al. 2006, Sheppard et al. 2010), and seasonal phytoplankton blooms occur regularly in Kuwaiti waters (Al-Yamani et al. 2004, Polikarpov et al. 2009, Sheppard et al. 2010). In particular, the waters are rich in picocyanobacteria in the genus *Synechococcus*. Picocyanobacterial abundances range between ~1×10^5 and ~6×10^5 ml^{-1} (Al-Hasan et al. 2001) and compose 25 to 100% of the total picophytoplankton in Kuwaiti coastal waters (Al-Bader et al. 2011). Thus, *Synechococcus* spp. are major primary producers in local waters, and as cyanophage and *Synechococcus* abundances covary (Suttle & Chan 1993, 1994, Waterbury & Valois 1993, Marston & Sallee 2003), cyanomyoviruses would also be expected to be abundant in the waters of Kuwait.

Given the limited data on cyanophage diversity in environments analogous to those in the northern Arabian Gulf, we sequenced cyanomyovirus g20 amplicons across 17 stations in Kuwaiti coastal waters.

2. MATERIALS AND METHODS

2.1. Sample collection and processing

Seawater samples were collected during daylight hours from 2 m depth at 17 sites within the coastal waters of Kuwait (Fig. 1) from 12 to 22 March 2013. The sites were grouped by geographical location; 2 sites (KB2, KB5) were 18 km apart in northern Kuwait Bay; 4 sites near Messila (M1, M3, M4, M5) were 6 to 15 km apart; 5 sites near Fintas (F1, F2, F3, F4, F5) were <7 km apart; 4 sites (J1, J3, J4, J5) near Julaiah were 1 to 10 km apart; and 2 sites (KH1, KH2) near Khairan were 4 km apart.

At each site, temperature, pH, salinity, and dissolved oxygen were measured using a Horiba U-10 water-quality checker (Horiba), and 5 replicate 1 l samples were collected and kept on ice in the dark, and transported to the laboratory within 6 h. Replicate samples from each site were combined and

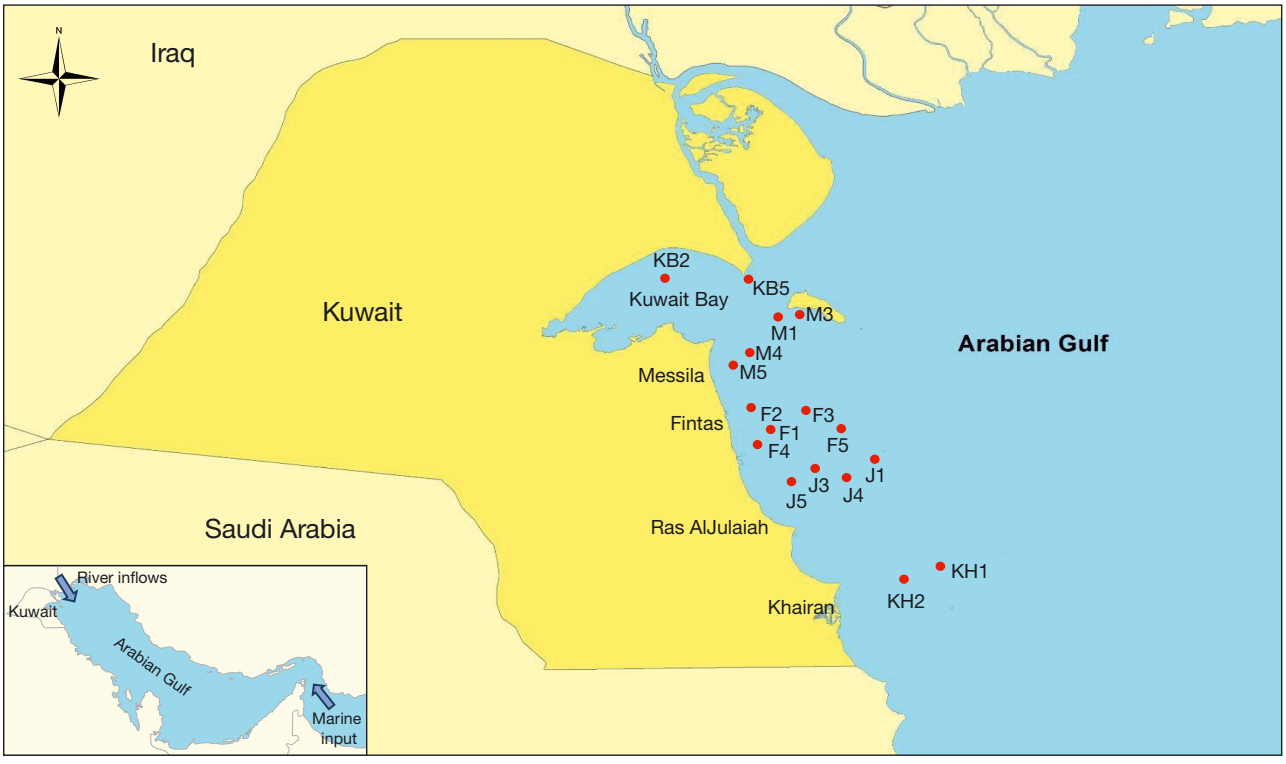

Fig. 1. Sampling sites. Sites are grouped and numbered by location. KB: Kuwait Bay; M: Messila; F: Fintas; J: Julaiah; KH: Khairan

sequentially filtered through 3 µm pore-size polycarbonate and 0.45 µm pore-size polyethersulfone filters (TPP) to remove most bacteria and larger particles. Tangential flow filtration (TFF) using a Vivaflow 200, 30 kDa molecular-weight-cutoff polyethersulfone membrane (Sartorius) was used to concentrate the viruses in the filtrate to a final volume of approximately 5 ml. The concentrates were stored at −80°C until use.

2.2. Viral DNA extraction and PCR

Viral-size particles were collected from the concentrate using aluminum oxide 0.02 µm pore-size syringe-tip filters (Anatop, Whatman) and the filters were stored at −80°C until processing. DNA was extracted from the filters using the MasterPure total nucleic acid extraction kit (MasterPure, Epicenter) following the method of Steward & Culley (2010). DNA was resuspended in autoclaved 0.02 µm filtered MilliQ water and stored at −20°C.

Fragments of the capsid assembly protein gene g20 were amplified from the concentrates using the CPS1.1 (5′-GTA GWA TWT TYT AYA TTG AYG TWG G-3′) and CPS8.1 (5′-ART AYT TDC CDA YRW AWG GWT C-3′) revised primers (Sullivan et al. 2008) with Ready-to-Go PCR beads (GE Healthcare). Each reaction mix (25 µl) contained 0.5 µM of each primer and 2 to 3 µl of DNA template. Triplicate PCR reactions were done for each sample using a Veriti 96-well thermocycler (Applied Biosystems). The PCR was performed as described by Zhong et al. (2002) using the following conditions: denaturation at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 15 s, annealing at 35°C for 1 min, ramping at 0.3°C s^{-1}, and elongation at 73°C for 1 min, followed by a final elongation step at 73°C for 4 min. All PCR reactions were visualized on 1% agarose gels, 1× Tris-acetate-EDTA (TAE) and stained with ethidium bromide. PCR products of each sample with the expected fragment size (~590 bp) were pooled and then purified using the MinElute PCR purification kit (Qiagen) according to the manufacturer's instructions and eluted in autoclaved 0.22 µm filtered MilliQ water.

2.3. Construction of g20 libraries

PCR products for each sample were purified using AMPure XP beads (Beckman Coulter) at a ratio of 1.2:1 beads:product. The purified products were resuspended in 30 µl EB buffer (Qiagen) and then quantified using the Picogreen dsDNA assay (Invitrogen) alongside the standardized Lambda DNA (Invitrogen). Concentrations were calculated using iQ5 and CFX96 Touch systems (Bio-Rad). Libraries for each sample were prepared using NxSeq DNA sample prep kit 2 following the manufacturer's recommendations with either NEXTflex 48 (BioO) or NEXTflex 96 HT barcodes (BioO). Libraries were purified using AMPure XP beads (Beckman Coulter) at a ratio of 0.9:1 beads:library.

2.4. Sequencing library quantification and quality control

Libraries were checked for small fragments using a 2100 Bionanalyzer (Agilent) with the High Sensitivity DNA kit (Agilent). The concentration of libraries was quantified using the Picogreen dsDNA assay as described in Section 2.3. The libraries were quantified and checked for amplifiable adapters using the Library Quantification DNA standards 1 to 6 (Kappa Biosystems) with the SsoFast EvaGreen qPCR supermix (Bio-Rad) using 10 µl EvaGreen master mix, 3 µl of the F primer (0.5 µM), 3 µl of the R primer (0.5 µM) and 4 µl each of 1:1000, 1:5000, and 1:10000 dilutions of the libraries in triplicate on either a iQ5 (Bio-Rad) or a CFX96 Touch qPCR machines. The following cycling parameters were used: 95°C for 30 s, then 35 cycles of 95°C for 5 s, 60°C for 30 s, and melt curve generation from 65 to 95°C in 0.5°C steps (10 s per step). The quantification values from the Picogreen assays and the qPCR assays were both used to evaluate the final pooling of the sequencing libraries. Libraries were sequenced using 2 × 250 bp PE MiSeq sequencer (Illumina) at the Génome Québec Innovation Centre at McGill University (Montreal, QC, Canada), and 2 × 300 bp PE MiSeq sequencer (Illumina) at Genoseq UCLA (Los Angeles, CA, USA).

2.5. Sequence analysis

Libraries were demultiplexed using CASAVA (Illumina). Initially, sequence quality was examined using FastQC (Andrews 2010). Any contaminating sequencing adapters were identified and removed using Trimmomatic version 0.32 (Bolger et al. 2014) using the default settings. The sequencing library quality was examined in more detail using fastx_quality (Gordon 2010). Since the expected amplicon can be up to 594 bp long (Zhong et al. 2002, Sullivan et al. 2008), the reads were not merged as they did not over-

lap. Reads from the forward primer were the longest and of the highest read quality and thus were used for further analysis. Sequence alignment of the libraries to sequences retrieved from NCBI's Conserved Domain Database (CDD, conserved protein domain family: PHA02531) was performed using align.seqs in mothur (Schloss et al. 2009). TBLASTX (Altschul et al. 1990) queries were performed using a custom blast database containing the g20 gene constructed from cyanomyovirus isolates available in NCBI (for accession numbers see Text S1 at www.int-res.com/articles/suppl/a091p001_supp.pdf). Reads matching the database with an E-value of less than 1×10^{-10} were kept. Reads were translated using FragGeneScan (Rho et al. 2010) using the Illumina_10 method. Sequences were aligned using Clustal Omega (Sievers et al. 2011), then trimmed to the same positions using USEARCH (Edgar 2010), and gaps introduced during the alignment step were removed using scripts in Biopython (Cock et al. 2009). Alignments were visualized using Aliview (Larsson 2014). Sequences aligning poorly or less than 66 amino acids long were removed. Chimera-checking of the sequences was performed using USEARCH denovo and reference methods (Edgar 2010). USEARCH was used to cluster sequences at 97% identity based on similarity to cyanomyovirus reference sequences. Several studies have shown that the performance of the amplicon sequence variant (ASV) method is affected by sequence depth (Joos et al. 2020, Jeske & Gallert 2022) and thus due to low sequencing depth in some samples, the operational taxonomic unit (OTU) picking method was used instead of constructing ASVs. To check the identities of the obtained OTUs, the top 20 OTUs were queried against GenBank's nr database using BLASTp. Raw sequence data are available in the NCBI BioProject database ID: PRJNA801482.

2.6. Statistical analysis

Significant variations in environmental parameters among sites were determined by 1-way analysis of variance (ANOVA) using Excel (Microsoft), after checking for normal distribution of the data using the descriptive measures of normality skewness and excess kurtosis. Relative abundances of OTUs were normalized by random resampling to address differences in sequencing depth among samples. Libraries were normalized to the library containing the fewest reads to generate equal sequence counts across samples. Rarefaction curves were generated using the vegan package (Oksanen et al. 2013) in R (R Core Team 2021). Non-metric dimensional scaling plots were generated and analysis of similarities (ANOSIM) was performed in vegan using Bray-Curtis-generated distance matrices. Alpha-diversity metrics and heatmaps were generated using the phyloseq (McMurdie & Holmes 2013) and ggplot2 (Wickham 2016) packages in R. After testing for normality via visual inspection and applying the Shapiro-Wilk's normality test as implemented in R, differences in alpha diversity metrics among sites were analyzed using ANOVA and pairwise comparisons with *t*-tests and corrected for multiple testing using Bonferroni correction in R (R Core Team 2021). Relationships among environmental and biotic variables were tested by canonical correspondence analysis in vegan.

2.7. Phylogenetic analysis

A reference alignment containing cyanomyovirus isolates (accession numbers are listed in Text S1) was created using Clustal Omega (Sievers et al. 2011) with default settings. Alignments were visualized using Aliview (Larsson 2014). Alignment masking was done using trimAl with the automatic heuristic (Capella-Gutiérrez et al. 2009) and then the alignments were edited manually. ProtTest 3.2 was used for amino-acid model selection (Darriba et al. 2011). An initial maximum likelihood phylogenetic tree was built with FastTree (Price et al. 2010). Environmental sequences from other studies (Zhong et al. 2002, Sullivan et al. 2008, Jameson et al. 2011, Zhong & Jacquet 2013, Jing et al. 2014) were downloaded from GenBank, and 28 pelagimyophage sequences (Table S4) were also included (Zaragoza-Solas et al. 2020). The unique sequences were clustered at 90% amino acid similarity (using centroid output) using USEARCH (Edgar 2010). These sequences were aligned to the reference alignment using Clustal Omega and then a reference tree was made using RAxML (Stamatakis et al. 2008) with rapid bootstrap analysis, using the amino-acid matrix LG with optimization for substitution rates and the Gamma model of rate heterogeneity using T4 virus (GI: NP 049782) as an outgroup. The top 50 OTUs at 97% sequence similarity found in the normalized dataset were aligned with the reference and environmental sequences using Clustal Omega and then the RAxML Evolutionary Placement algorithm (EPA) (Berger et al. 2011) was used to place the OTUs onto the reference tree using the amino-acid matrix LG, optimization for substitution rates, the Gamma model of rate heterogeneity and a threshold of 0.2. The resulting tree was viewed in Archaeopteryx (Han &

Zmasek 2009) and then the tree format was converted using biopython (Cock et al. 2009). Graphlan (Asnicar et al. 2015) was used to create the final tree with the placements and the relative abundance information for the top 50 OTUs.

3. RESULTS

3.1. Environmental parameters of study sites

The environmental data are provided in Table S1. Water temperatures ranged from 18.3 to 20.3°C and differed significantly among sites ($p < 0.001$); the highest temperature was at Site M5 near Messila (20.3°C) and the lowest at Site KB2 in (Kuwait Bay 18.3°C). At all sites, the waters were hypersaline, with salinity between 47 and 49 ppt and pH ranging between 9.0 and 9.2. Turbidity varied significantly among sites ($p < 0.05$) and was highest in Kuwait Bay at Site KB5 (8.2 nephelometer turbidity units [NTU]) and lowest near Messila at Site M3 (0.33 NTU). NMDS analysis of the environmental parameters clustered samples by location (Fig. S1).

3.2. Overview of sequence statistics

1 472 055 quality-filtered sequences were obtained from the 17 samples with an average of 457 119 sequences per sample with a range of 134 384 (J4) to 2 114 372 (M4). These were clustered at 97% sequence similarity into 11 735 OTUs with an average of 1399 OTUs per sample, with a range of 631 (KB4) to 6417 (M4). After normalization, for the forward primer, there were 1623 OTUs with an average of 495 OTUs per sample (range 389 [M5] to 581 [J1]). The sequences associated with the forward g20 primer (CPS1.1) were chosen for downstream analysis since quality processing of the sequences indicated that the forward primer recovered more reads than the reverse primer.

The diversity of OTUs was estimated across locations. Alpha-diversity based on the Chao1 richness index revealed no significant differences among locations; whereas, the Shannon and InvSimpson diversity indices showed significant differences among sites ($p \leq 0.0001$) (Fig. 2, Fig. S2, Table S2), with alpha diversity highest in Julaiah and lowest in Kuwait Bay. In addition, estimates of beta diversity across locations using Non-metric dimensional scaling (NMDS) ordination (Fig. 3) showed separation among the 5 locations (ANOSIM $R = 0.9591$, $p = 0.0001$), with more similarity among sites within the same location. However, the rarefaction curves did not saturate for any of the locations, indicating that sequencing depth was inadequate to capture the full diversity of g20 OTUs (Fig. S3). Relationships between biotic and abiotic parameters were not found to be significant (data not shown).

3.3. Cyanomyovirus distribution and phylogeny

Analysis of the rank abundance and distribution of the top 50 OTUs (⩾0.2% relative abundance) across locations showed that one to 3 OTUs were dominant at each location, followed by a long tail of low abundance OTUs (Fig. S4). OTU_1 was the most abundant at all locations except in Julaiah, where it was most abundant at Site J3 only. OTU_2 was the second most abundant at all sites except also in Julaiah, and OTU_9 had a higher abundance at all Kuwait Bay sites (15%) compared to the rest of the sampling locations.

The relative abundances and distribution of the top 20 OTUs (⩾0.8% relative abundance) varied across locations (Fig. 4). For example, the relative abundances of OTU_8 and OTU_9 were highest in Kuwait Bay, while OTU_7 had a higher relative abundance in Julaiah. Overall, the top 20 OTUs comprised 75% of the total relative abundance of the dataset, with more than 20% of the sequences belonging to OTU_1.

BLASTp analysis of the top 20 OTUs showed that most sequences were identified as gp20 portal proteins (average E-value $<4 \times 10^{-34}$), half of which were 80 to 98% similar to sequences in GenBank for cyanomyoviruses infecting *Synechococcus* (Table S3). Most had highest similarity to cyanomyoviruses isolated from the coastal waters of Rhode Island; 2 others were 98.4% similar to sequences for cyanomyoviruses infecting *Prochlorococcus*; 2 others were most similar to T4-like phages with unknown hosts; 5 were more similar to putative portal proteins from freshwater metagenomes.

Phylogenetic analysis of the top 50 OTUs and sequences from representative cyanomyovirus isolates and environmental samples (Zhong et al. 2002, Sullivan et al. 2008, Jameson et al. 2011, Zhong & Jacquet 2013, Jing et al. 2014) (Fig. 5) revealed that some sequences grouped with cyanomyophage isolates, others with pelagimyophages, and some with diverse environmental clades. A total of 53% of the OTUs fell within well-supported clades within cyanophage Marine Cluster II, and the remaining OTUs grouped with pelagimyophage sequences and various environmental clusters, including 8 OTUs that fell within Clusters A, B, C, D, and F defined by Zhong et al.

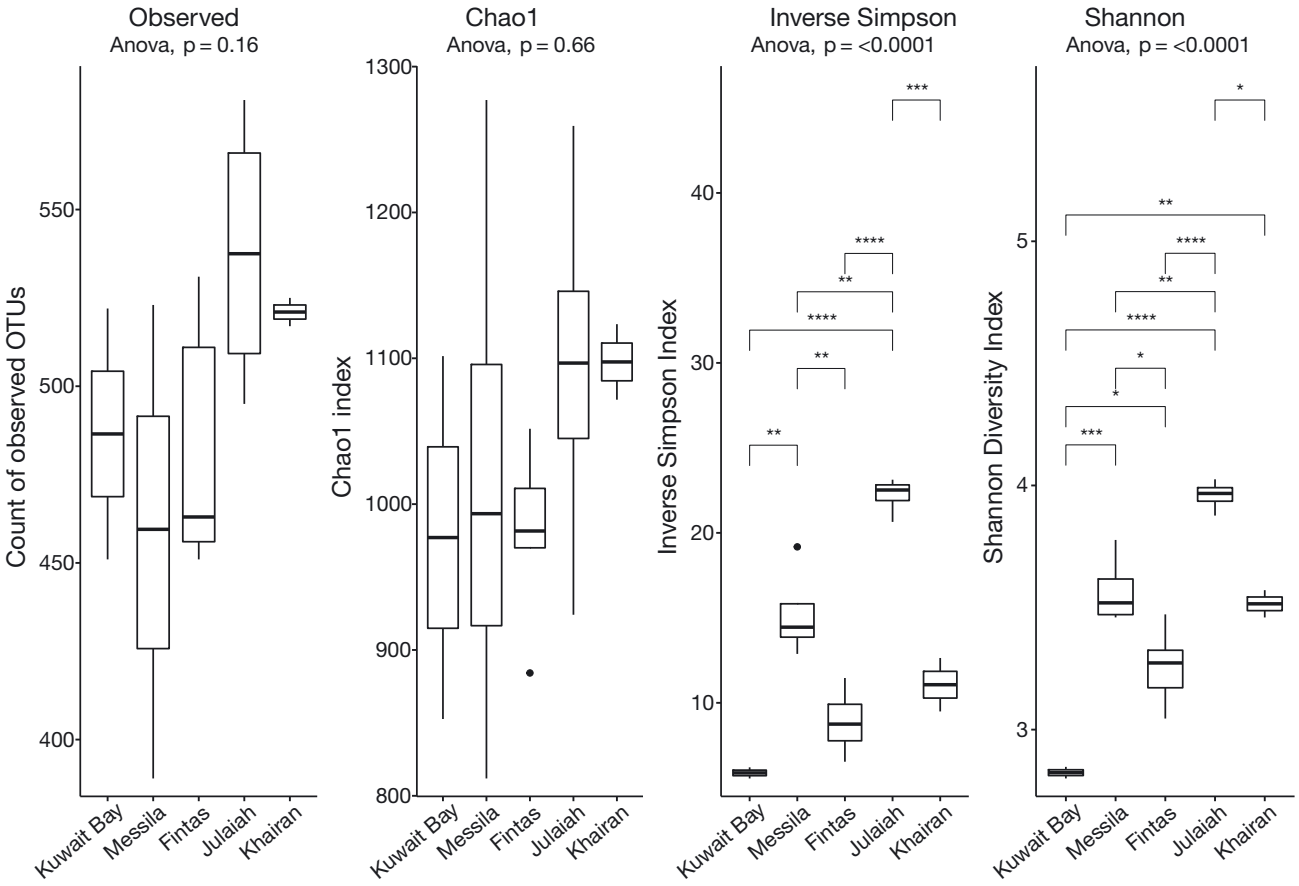

Fig. 2. Alpha diversity measures of reads recovered from the forward CPS primer at 97% sequence similarity. The plots represent the number of OTUs per location excluding singletons (Observed), the Chao1 estimate (Chao1), the Inverse Simpson Index (InvSimpson), and the Shannon Diversity Index (Shannon). Differences among sites were evaluated using ANOVA. To test for differences between sites, pairwise *t*-tests were performed with p-values adjusted for multiple testing using the Bonferroni method. ****p ≤ 0.0001; ***p ≤ 0.001; **p ≤ 0.01; *p ≤ 0.05; not significant: p > 0.05

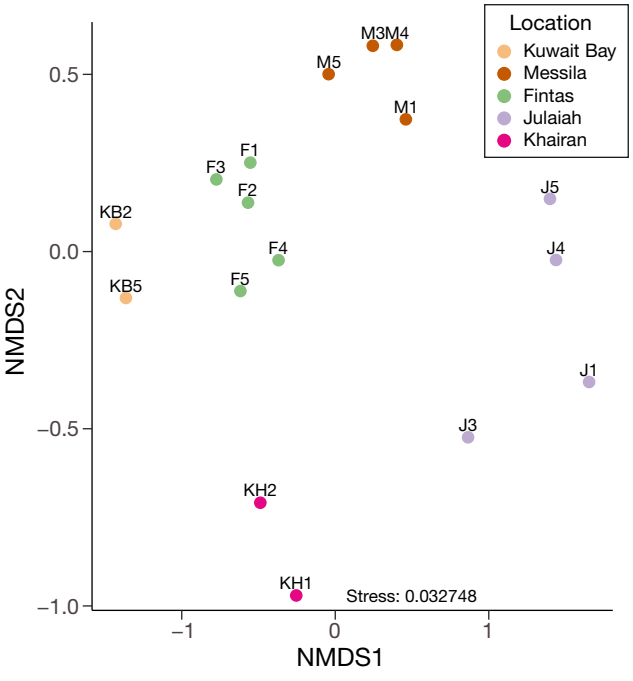

Fig. 3. Total forward CPS primer OTUs found in the 17 samples, based on Bray-Curtis dissimilarities. Each point represents 1 site and the points are colored by location

(2002). No OTUs grouped with isolates from Marine Clusters I or III. OTU_1, with the highest relative abundance at all sites, fell within Environmental Cluster A, which contains sequences from the surface waters of estuarine and perialpine lakes. The next most abundant OTU_2 fell within Marine Cluster II, while OTU_3 clustered with environmental sequences collected along a north–south transect in the Atlantic Ocean.

4. DISCUSSION

We analyzed amplicon sequences of the capsid portal protein (g20) to examine cyanomyovirus diversity in Kuwait's coastal waters of the northwestern

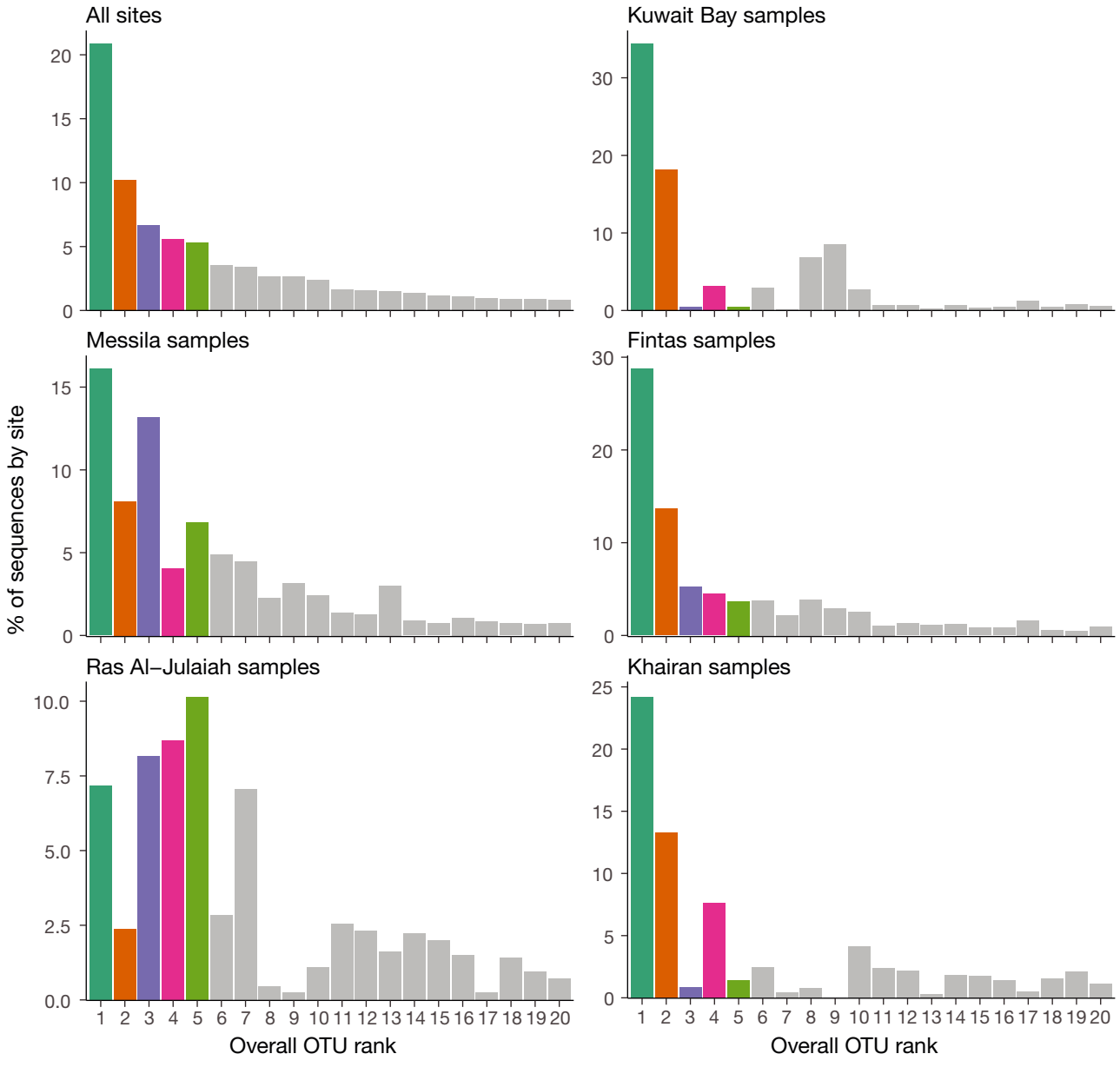

Fig. 4. Top 20 OTUs by location. OTU rank is based on the overall ranking when considering all sites together, and the relative abundance is considered by location. The top 5 most abundant OTUs bars are colored in each plot. The list of the top 20 OTUs is available in Text S1

Arabian Gulf. This semi-enclosed marine system forms an unusual negative estuary (where evaporation exceeds freshwater inflow) with freshwater inflow in the northwest, setting up a strong north-to-south salinity gradient (Al-Yamani et al. 2004, Devlin et al. 2015). This environmentally sensitive region is surrounded by 8 countries that are responsible for nearly a third of global oil production; yet, little is known about the dominant cyanobacteria in the region or the viruses that infect them. We used the g20 gene as a proxy (Zhong et al. 2002, Short & Suttle 2005, Sullivan et al. 2008) to survey the genetic diversity of cyanomyovirus populations in this region. Deep sequencing of g20 allowed us to sample genetic diversity and cyanomyovirus community composition from 5 different locations in Kuwait's coastal waters.

This first detailed analysis of the cyanomyovirus communities of the Arabian Gulf leads to several striking observations. The relative abundance of OTUs was highly uneven, with only a few dominant but phylogenetically dispersed OTUs. Furthermore, despite the hypersaline conditions of the Gulf, most OTUs belong

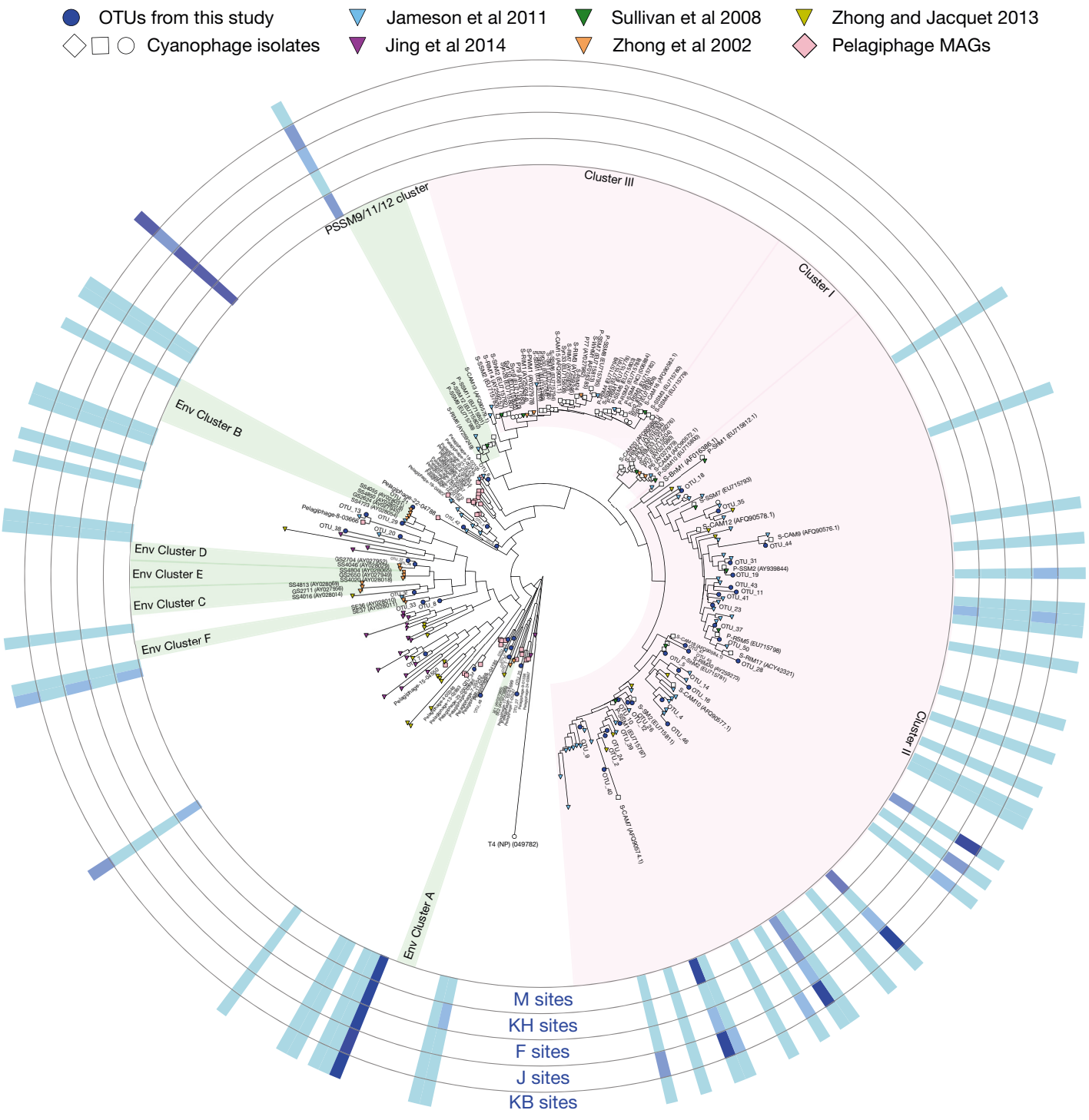

Fig. 5. Maximum likelihood g20 tree (RAxML) of cyanomyovirus isolates, sequences from environmental surveys with the evolutionary placement (EPA) of the top 50 OTUs in this study. The heatmaps in the outer rings represent the relative abundance by location of the high-throughput environmental sequencing reads of the top 50 OTUs (97% sequence similarity) retrieved in this study. Phylogenetics groups indicated are based on Sullivan et al. (2008)

to Marine Cluster II, a phylogenetic group that includes both isolates and representatives from a wide range of marine and freshwater habitats. These observations and their implications are discussed in detail below.

Cyanomyovirus diversity varied across locations. In particular, the lower alpha diversity and richness in the semi-enclosed, highly productive Kuwait Bay coincides with urbanization, pollution, and other anthropogenic

activities that have affected water quality (Al-Abdulghani et al. 2013, Devlin et al. 2015), shifted the phytoplankton community composition, and decreased primary productivity (Devlin et al. 2015, Ahmed et al. 2022). Turbidity was also higher in Kuwait Bay, following the typical pattern for Kuwait coastal waters of higher water clarity from north to south (Al-Yamani et al. 2004). This pattern is due to the effect of winds and tidal currents and discharges of plumes of suspended sediments released from Shatt Al-Arab (Al-Yamani et al. 2004, Al-Ghadban & El-Sammak 2005). Turbidity not only reduces primary productivity (Grobbelaar 1985, Huisman et al. 1999, Bouman et al. 2011), but can affect the abundance of picocyanobacteria and their community structure (Stomp et al. 2007, Buzzani et al. 2022). Along the coast of Kuwait, *Synechococcus* abundance and diversity increase with increasing water clarity (Al-Bader et al. 2011), as seen in other studies (Stomp et al. 2007, Haverkamp et al. 2008, Xia et al. 2018). High turbidity levels were also found to favor mostly filamentous cyanobacteria (Al-Bader et al. 2011). Given that cyanophage and picocyanobacteria abundances co-vary (Waterbury & Valois 1993, Suttle & Chan 1994, Wang et al. 2011), the lower diversity of g20 sequences in Kuwait Bay is likely the result of lower abundance and diversity of *Synechococcus*.

The samples were collected over 10 d and the top 20 OTUs in relative abundance occurred at all 5 locations. This could be because well-mixed water columns usually have stable viral communities (Wang & Chen 2004, Goldsmith et al. 2015), and Kuwait's coastal waters are typically well-mixed. In addition, the prevailing counterclockwise circulation pattern in Gulf waters (Sheppard et al. 2010) might also lead to a wide distribution of cyanomyoviruses and their hosts.

Although the top 20 OTUs occurred at all locations, their relative abundances varied markedly. Given that environmental variation can affect viral abundance and diversity (Chow & Fuhrman 2012, Brum et al. 2015) and that cyanomyovirus communities are reflected by the physical structure of the water column (Frederickson et al. 2003), we measured temperature, salinity, and turbidity, and found significant variations among sites. However, unlike other studies that have demonstrated relationships between environmental variables and viral dynamics (Weinbauer et al. 2009, Mojica & Brussaard 2014), we found no significant associations using canonical correspondence analysis with the environmental data and the OTUs (data not shown). Potentially, more in-depth spatial and temporal sampling could disentangle the influence of abiotic and biotic factors on viral diversity and abundance patterns.

Despite the variation among sites, more than half of the g20 sequences at any location belonged to the top 5 OTUs. Similarly, highly uneven OTU rank-abundance curves are consistent with other cyanomyovirus diversity studies (Jameson et al. 2011, Chow & Fuhrman 2012, Marston et al. 2013, Needham et al. 2013, Pagarete et al. 2013, Goldsmith et al. 2015), in which most environmental viral sequences belonged to a few common and relatively abundant OTUs, whereas most OTUs were less abundant and rare. Such highly uneven rank abundance curves have been interpreted as evidence for a 'seed-bank' model (Breitbart & Rohwer 2005) in which widely dispersed virus genotypes come to prominence when suitable hosts are available. Similarly, in the 'kill the winner' model, specific viruses dominate in response to rapid host growth (Thingstad 2000). Both the 'seed-bank' and 'kill the winner' models are supported by observations of the release of free ribosomal RNA (rRNA) that demonstrate that within a population of potential hosts, lysis is restricted to very few genotypes (Zhong et al. 2023). For example, of 138 different sequence variants (taxa) of rRNA assigned to the genus *Synechococcus*, cell lysis was only detected in 9 relatively rare genotypes (Zhong et al. 2023). Given that lysis of a specific taxon is likely the result of infection by a single phage type genotype, this would explain the highly uneven rank-abundance curves. Moreover, the high Bray-Curtis dissimilarities among the 5 locations imply high local diversity, which is consistent with the 'seed-bank' model (Brum et al. 2015), as well as 'kill the winner' (Thingstad 2000) and taxon-specific cell lysis (Zhong et al. 2023). Furthermore, the 'royal family' framework proposed by Breitbart et al. (2018) characterizes a scenario where dominant bacterial and phage populations, termed the 'royal family', persist over time, exhibiting 'kill the winner' dynamics among them. This framework sheds light on the long-term dominance observed in our study.

Highly uneven rank-abundance curves are a feature of aquatic virus communities and the populations they infect (Pedrós-Alió 2006, Suttle 2007, Breitbart et al. 2018). Rare OTUs and the rank-distribution of taxa in viral communities can be driven by many factors including dispersal, host range, burst size, host distribution, and a range of other host, virus, and environmental factors (Chow & Suttle 2015). For example, cyanomyoviruses can have broad host ranges (Sullivan et al. 2008) and may respond rapidly to increases in population abundances of multiple host genotypes (Doron et al. 2016). Cyanomyoviruses that infect genetically diverse hosts are also expected to prevail when host abundances are low or variable

(Suttle 2007). Hence rare OTUs could reflect rare hosts, wide dispersal, or remnant viruses from previous infection events.

The cyanomyovirus g20 isolate-containing phylogenetic Clusters I, II, and III have been frequently recovered from oceanic, freshwater, and coastal marine environments (Zhong et al. 2002, Marston & Sallee 2003, Short & Suttle 2005, Wilhelm et al. 2006, Sullivan et al. 2008, Jameson et al. 2011, Marston et al. 2013, Jing et al. 2014, Hanson et al. 2016). In this study, over half of the top 50 relatively most abundant OTUs grouped with the isolate-containing Cluster II (Zhong et al. 2002), while none grouped with culture-containing Clusters I, III (Zhong et al. 2002), or IV (Sullivan et al. 2008), despite the cosmopolitan distribution of members of these groups (Zhong et al. 2002, Short & Suttle 2005, Sullivan et al. 2008). Although a high diversity of g20 OTUs was recovered in our study, the rarefaction curves (Fig. S3) did not saturate for any of the samples, indicating that the full range of OTUs was not recovered.

Phylogenetic analysis showed that 9 of the top 20 OTUs, including those with high relative abundance (OTU-1 and OTU-3), grouped within environmental sequence clusters. Similarly, g20 diversity along an Atlantic Ocean transect and in alpine lakes revealed that many g20 sequences fell into clades with no cultured representatives, which are often referred to as environmental sequence clusters (Jameson et al. 2011, Zhong & Jacquet 2013). It should be noted that when using g20 as a proxy, Short & Suttle (2005) were able to amplify g20 sequences from depths of over 3000 m, where picocyanobacterial abundance was minimal. Thus, these environmental sequence clusters may contain myophages with diverse host ranges, including pelagiphages. Although, members of the genus *Synechococcus* dominate the picophytoplankton community in Kuwait waters (Al-Hasan et al. 2001, Ahmed et al. 2022, Ismail & Almutairi 2022), few strains have been isolated and identified (Al-Hasan et al. 2001), and representative cyanomyovirus isolates have yet to be described; hence, there is an opportunity and a need to explore the local picophytoplankton community and the viruses that infect them.

5. CONCLUSION

This study examined the previously unexplored diversity of cyanomyoviruses in Kuwait's coastal waters. Along a south–north gradient, our results revealed higher cyanomyovirus diversity in the southern Kuwait coastal waters. Cyanomyovirus community composition was variable among sites, with a high relative abundance of sequences in cyanomyovirus Marine Cluster II, as well as novel environmental pelagimyophage-like sequences with no cultured representatives that may be endemic to these waters. Similar to other marine environments, a few diverse OTUs were dominant, consistent with the 'seed-bank', 'kill the winner', and 'royal family' models. This investigation advances our understanding of cyanomyovirus diversity, laying the groundwork for future research into the intricate dynamics of these viruses in the northwestern Arabian Gulf.

Acknowledgements. We thank Dr. Mohamad AlKhalidi from the Civil Engineering Department at Kuwait University for providing the Kuwait map. We are grateful to the Faculty of Science boat skipper Mr. Katta Sudhanandham for his assistance in sampling. This research was supported by Kuwait University Grant SL08/12 and partially through Grants SL01/10 and SL08/14 to A.A., and an NSERC Discovery grant to C.A.S.

LITERATURE CITED

Ahlgren NA, Perelman JN, Yeh YC, Fuhrman JA (2019) Multi-year dynamics of fine-scale marine cyanobacterial populations are more strongly explained by phage interactions than abiotic, bottom-up factors. Environ Microbiol 21:2948–2963

Ahmed A, Madhusoodhanan R, Yamamoto T, Fernandes L and others (2022) Analysis of phytoplankton variations and community structure in Kuwait Bay, Northwestern Arabian Gulf. J Sea Res 180:102163

Al-Abdulghani E, El-Sammak A, Sarawi M (2013) Environmental assessment of Kuwait Bay: an integrated approach. J Coast Conserv 17:445–462

Al-Bader D, Rayan R, Mahmood H, Al-Hasan R, Bobby L, Eliyas M (2011) Comparison of the seasonal diversity of planktonic cyanobacteria in clear and turbid shallow coastal waters of Kuwait using culture-independent techniques. Kuwait J Sci Eng 38:163–183

Al-Ghadban A, El-Sammak A (2005) Sources, distribution and composition of the suspended sediments, Kuwait Bay, Northern Arabian Gulf. J Arid Environ 60:647–661

Al-Hasan RH, Khanafer M, Eliyas M, Radwan SS (2001) Hydrocarbon accumulation by picocyanobacteria from the Arabian Gulf. J Appl Microbiol 91:533–540

Al-Yamani F, Bishop J, Ramadhan E, Al-Husaini M, Al-Ghadban A (2004) Oceanographic Atlas of Kuwait's waters. Kuwait Institute for Scientific Research, Safat

Al-Yamani F, Rao DVS, Mharzi A, Ismail W, Al-Rifaie K (2006) Primary production off Kuwait, an arid zone environment, Arabian Gulf. Int J Oceans Oceanogr 1:67–85

Almutairi A (2015) Spatial-temporal variations and diversity of the bacterioplankton communities in the coastal waters of Kuwait. Mar Pollut Bull 100:699–709

Almutairi A, Al-Bader D, Al-Mutairi M (2023) Seasonal variations in cyanophage communities in the coastal waters of Kuwait. Kuwait J Sci 50:31–39

Alosairi Y, Imberger J, Falconer RA (2011) Mixing and flushing in the Persian Gulf (Arabian Gulf). J Geophys Res Oceans 116:C03029

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Andrews S (2010) FastQC. www.bioinformatics.babraham.ac.uk/projects/fastqc

Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N (2015) Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ 3:e1029

Baran N, Carlson MC, Sabehi G, Peleg M, Kondratyeva K, Pekarski I, Lindell D (2022) Widespread yet persistent low abundance of TIM5-like cyanophages in the oceans. Environ Microbiol 24:6476–6492

Berger SA, Krompass D, Stamatakis A (2011) Performance accuracy and web server for evolutionary placement of short sequence reads under maximum likelihood. Syst Biol 60:291–302

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Bouman HA, Ulloa O, Barlow R, Li WK and others (2011) Water-column stratification governs the community structure of subtropical marine picophytoplankton. Environ Microbiol Rep 3:473–482

Breitbart M, Rohwer F (2005) Here a virus, there a virus, everywhere the same virus? Trends Microbiol 13: 278–284

Breitbart M, Bonnain C, Malki K, Sawaya N (2018) Phage puppet masters of the marine microbial realm. Nat Microbiol 3:754–766

Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G and others (2015) Patterns and ecological drivers of ocean viral communities. Science 348:1261498

Buzzani A, Paranhos R, Mello MP, Abreu F and others (2022) Rainfall governs picocyanobacterial ecology in a tropical estuary (Guanabara Bay, Brazil). Hydrobiologia 849: 175–196

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973

Carlson MCG, Ribalet F, Maidanik I, Durham BP and others (2022) Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. Nat Microbiol 7:570–580

Chow CET, Fuhrman JA (2012) Seasonality and monthly dynamics of marine myovirus communities. Environ Microbiol 14:2171–2183

Chow CET, Suttle CA (2015) Biogeography of viruses in the sea. Annu Rev Virol 2:41–66

Cock PJA, Antao T, Chang JT, Chapman BA and others (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423

Darriba D, Taboada GL, Doallo R, Bangor D (2011) ProtTest3: fast selection of best fit models of protein evolution. Bioinformatics 27:1164–1165

Dart E, Fuhrman JA, Ahlgren NA (2023) Diverse marine T4-like cyanophage communities are primarily comprised of low-abundance species including species with distinct seasonal, persistent, occasional, or sporadic dynamics. Viruses 15:581

Devlin MJ, Massoud MS, Hamid SA, Al-Zaidan A and others (2015) Changes in the water quality conditions of Kuwait's marine waters: long term impacts of nutrient enrichment. Mar Pollut Bull 100:607–620

Dorigo U, Jacquet S, Humbert JF (2004) Cyanophage diversity inferred from g20 gene analyses in the largest natural lake in France, Lake Bourget. Appl Environ Microbiol 70: 1017–1022

Doron S, Fedida A, Hernández-Prieto MA, Sabehi G and others (2016) Transcriptome dynamics of a broad host-range cyanophage and its hosts. ISME J 10:1437–1455

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461

Filée J, Tétart F, Suttle CA, Krisch HM (2005) Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc Natl Acad Sci USA 102: 12471–12476

Finke JF, Suttle CA (2019) The environment and cyanophage diversity: insights from environmental sequencing of DNA polymerase. Front Microbiol 10:167

Frederickson CM, Short SM, Suttle CA (2003) The physical environment affects cyanophage communities in British Columbia inlets. Microb Ecol 46:348–357

Fuller NJ, Wilson WH, Joint IR, Mann NH (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

Goldsmith DB, Parsons RJ, Beyene D (2015) Deep sequencing of the viral *phoH* gene reveals temporal variation, depth-specific composition, and persistent dominance of the same viral *phoH* genes in the Sargasso Sea. PeerJ 3: e997

Gordon A (2010) fastx_toolkit. Github repository. https://github.com/agordon/fastx_toolkit

Grobbelaar JU (1985) Phytoplankton productivity in turbid waters. J Plankton Res 7:653–663

Han MV, Zmasek CM (2009) PhyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics 10:356

Hanson CA, Marston MF, Martiny JBH (2016) Biogeographic variation in host range phenotypes and taxonomic composition of marine cyanophage isolates. Front Microbiol 7:983

Haverkamp T, Acinas SG, Doeleman M, Stomp M, Huisman J, Stal L (2008) Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. Environ Microbiol 10:174–188

Huang S, Zhang S, Jiao N, Feng C (2015) Marine cyanophages demonstrate biogeographic patterns throughout the global ocean. Appl Environ Microbiol 81:441–452

Huisman J, van Oostveen P, Weissing FJ (1999) Species dynamics in phytoplankton blooms: incomplete mixing and competition for light. Am Nat 154:46–68

Ismail N, Almutairi A (2022) Bacterioplankton community profiling of the surface waters of Kuwait. Front Mar Sci 9: 838101

Jameson E, Mann NH, Joint I, Sambles C, Muhling M (2011) The diversity of cyanomyovirus populations along a North–South Atlantic Ocean transect. ISME J 5: 1713–1721

Jeske JT, Gallert C (2022) Microbiome analysis via OTU and ASV-based pipelines — a comparative interpretation of ecological data in WWTP systems. Bioengineering (Basel) 9:146

Jing R, Liu J, Yu Z, Liu X, Wang G (2014) Phylogenetic distribution of the capsid assembly protein gene (g20) of cya-

nophages in paddy floodwaters in Northeast China. PLOS ONE 9:e88634
Joos L, Beirinckx S, Haegeman A, Debode J and others (2020) Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. BMC Genomics 21:733
Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinformatics 30: 3276–3278
Lopes A, Tavares P, Petit MA, Guérois R, Zinn-Justin S (2014) Automated classification of tailed bacteriophages according to their neck organization. BMC Genomics 15: 1027
Lu J, Chen F, Hodson RE (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. Appl Environ Microbiol 67:3285–3290
Mann NH (2003) Phages of the marine cyanobacterial picophytoplankton. FEMS Microbiol Rev 27:17–34
Marston MF, Amrich CG (2009) Recombination and microdiversity in coastal marine cyanophages. Environ Microbiol 11:2893–2903
Marston MF, Sallee JL (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. Appl Environ Microbiol 69:4639–4647
Marston MF, Taylor S, Sme N, Parsons RJ, Noyes TJ, Martiny JB (2013) Marine cyanophages exhibit local and regional biogeography. Environ Microbiol 15: 1452–1463
Matteson AR, Loar SN, Bourbonniere RA, Wilhelm SW (2011) Molecular enumeration of an ecologically important cyanophage in a Laurentian Great Lake. Appl Environ Microbiol 77:6772–6779
Matteson AR, Rowe JM, Ponsero AJ, Pimentel TM, Boyd PW, Wilhelm SW (2013) High abundances of cyanomyoviruses in marine ecosystems demonstrate ecological relevance. FEMS Microbiol Ecol 84:223–234
McMurdie PJ, Holmes S (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLOS ONE 8:e61217
Mojica KDA, Brussaard CPD (2014) Factors affecting virus dynamics and microbial host–virus interactions in marine environments. FEMS Microbiol Ecol 89:495–515
Naser HA (2013) Assessment and management of heavy metal pollution in the marine environment of the Arabian Gulf: a review. Mar Pollut Bull 72:6–13
Needham DM, Chow CET, Cram JA, Sachdeva R, Parada A, Fuhrman JA (2013) Short-term observations of marine bacterial and viral communities: patterns connections and resilience. ISME J 7:1274–1285
Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P and others (2013) vegan: community ecology package. R package version 2.0-9, https://vegandevs.github.io/vegan/
Pagarete A, Chow CET, Johannessen T, Fuhrman JA, Thingstad TF, Sandaa RA (2013) Strong seasonality and interannual recurrence in marine myovirus communities. Appl Environ Microbiol 79:6253–6259
Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? Trends Microbiol 14:257–263
Polikarpov I, Al-Yamani F, Saburova M (2009) Space-time variability of phytoplankton structure and diversity in the north-western part of the Arabian Gulf (Kuwait). BioRisk 3:86–96
Pous S, Lazure P, Carton X (2015) A model of the general circulation in the Persian Gulf and in the Strait of Hormuz: intraseasonal to interannual variability. Cont Shelf Res 94:55–70
Price MN, Dehal PS, Arkin AP (2010) FastTree2 — approximately maximum-likelihood trees for large alignments. PLOS ONE 5:e9490
Proctor L, Fuhrman J (1990) Viral mortality of marine bacteria and cyanobacteria. Nature 343:60–62
R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
Rho M, Tang H, Ye Y (2010) FragGeneScan: predicting genes in short and error-prone reads. Nucleic Acids Res 38:e191
Sandaa RA, Larsen A (2006) Seasonal variations in virus–host populations in Norwegian coastal waters: focusing on the cyanophage community infecting marine *Synechococcus* spp. Appl Environ Microbiol 72:4610–4618
Schloss PD, Westcott SL, Ryabin T, Hall JR and others (2009) Introducing mothur: open-source platform-independent community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541
Sheppard C, Al-Husiani M, Al-Jamali F, Al-Yamani F and others (2010) The Gulf: a young sea in decline. Mar Pollut Bull 60:13–38
Short CM, Suttle CA (2005) Nearly identical bacteriophage structural gene sequences are widely distributed in both marine and freshwater environments. Appl Environ Microbiol 71:480–486
Sievers F, Wilm A, Dineen D, Gibson TJ and others (2011) Fast scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539
Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web servers. Syst Biol 57: 758–771
Steward GF, Culley AI (2010) Extraction and purification of nucleic acids from viruses. In: Wilhelm SW, Weinbauer MG, Suttle CA (eds) Manual of aquatic virology ecology. American Society of Limnology and Oceanography, Waco, TX, p 154–165
Stomp M, Huisman J, Vörös L, Pick FR, Laamanen M, Haverkamp T, Stal LJ (2007) Colourful coexistence of red and green picocyanobacteria in lakes and seas. Ecol Lett 10: 290–298
Sullivan MB, Waterbury JB, Chisholm SW (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus.* Nature 424:1047–1051
Sullivan MB, Coleman ML, Quinlivan V, Rosenkrantz JE and others (2008) Portal protein diversity and phage ecology. Environ Microbiol 10:2810–2823
Suttle CA (1994) The significance of viruses to mortality in aquatic microbial communities. Microb Ecol 28:237–243
Suttle CA (2005) Viruses in the sea. Nature 437:356–361
Suttle CA (2007) Marine viruses—major players in the global ecosystem. Nat Rev Microbiol 5:801–812
Suttle CA, Chan AM (1993) Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. Mar Ecol Prog Ser 92:99–109
Suttle CA, Chan AM (1994) Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. Appl Environ Microbiol 60:3167–3174

Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol Oceanogr 45:1320–1328

Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR and others (2022) Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses (2022). Arch Virol 167:2429–2440

Wang K, Chen F (2004) Genetic diversity and population dynamics of cyanophage communities in the Chesapeake Bay. Aquat Microb Ecol 34:105–116

Wang K, Wommack KE, Chen F (2011) Abundance and distribution of *Synechococcus* spp. and cyanophages in the Chesapeake Bay. Appl Environ Microbiol 77: 7459–7468

Waterbury JB, Valois FW (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. Appl Environ Microbiol 59:3393–3399

Weinbauer MG, Bettarel Y, Cattaneo R, Luef B and others (2009) Viral ecology of organic and inorganic particles in aquatic systems: avenues for further research. Aquat Microb Ecol 57:321–341

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY

Wilhelm SW, Carberry MJ, Eldridge ML, Poorvin L, Saxton MA, Doblin MA (2006) Marine and freshwater cyanophages in a Laurentian Great Lake: evidence from infectivity assays and molecular analyses of g20 genes. Appl Environ Microbiol 72:4957–4963

Xia X, Liu H, Choi D, Noh JH (2018) Variation of *Synechococcus* pigment genetic diversity along two turbidity gradients in the China Seas. Microb Ecol 75:10–21

Zaragoza-Solas A, Rodriguez-Valera F, López-Pérez M (2020) Metagenome mining reveals hidden genomic diversity of pelagimyophages in aquatic environments. mSystems 5:e00905-19

Zhao J, Temimi M, Al Azhar M, Ghedira H (2015) Satellite-based tracking of oil pollution in the Arabian Gulf and the Sea of Oman. Can J Rem Sens 41:113–125

Zhong X, Jacquet S (2013) Prevalence of viral photosynthetic and capsid protein genes from cyanophages in two large and deep perialpine lakes. Appl Environ Microbiol 79:7169–7178

Zhong Y, Chen F, Wilhelm S, Poorvin L, Hodson R (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

Zhong KX, Wirth JF, Chan AM, Suttle CA (2023) Mortality by ribosomal sequencing (MoRS) provides a window into taxon-specific cell lysis. ISME J 17:105–116

Editorial responsibility: Mya Breitbart, St. Petersburg, Florida, USA
Reviewed by: T.V. Butina and 1 anonymous referee
Submitted: November 13, 2023; Accepted: November 27, 2024
Proofs received from author(s): January 24, 2025

This article is Open Access under the Creative Commons by Attribution (CC-BY) 4.0 License, https://creativecommons.org/licenses/by/4.0/deed.en. Use, distribution and reproduction are unrestricted provided the authors and original publication are credited, and indicate if changes were made