Unique viral capsid assembly protein gene (g20) of cyanophages in the floodwater of a Japanese paddy field. Biol Fertil Soils

ARTICLE *in* BIOLOGY AND FERTILITY OF SOILS · FEBRUARY 2009

Impact Factor: 3.4 · DOI: 10.1007/s00374-009-0410-y

Biol Fertil Soils (2010) 46:93–102
DOI 10.1007/s00374-009-0410-y

ORIGINAL PAPER

Unique viral capsid assembly protein gene (*g20*) of cyanophages in the floodwater of a Japanese paddy field

Guanghua Wang · Jun Murase · Susumu Asakawa · Makoto Kimura

Received: 1 June 2009 / Revised: 20 September 2009 / Accepted: 22 September 2009 / Published online: 20 October 2009

Abstract In order to evaluate the genetic diversity of cyanophage communities of rice fields, viral capsid assembly protein gene (*g20*) was amplified with primers CPS1 and CPS8. The DNA was extracted three times from viral concentrates obtained from floodwater samples collected in each of four different plots (no fertilizer; P and K chemical fertilizers; N, P, and K chemical fertilizers; and chemical fertilizers with compost). Denaturing gradient gel electrophoresis (DGGE) gave different *g20* clones. The sequencing of DGGE bands revealed that the *g20* genes of the floodwater were divergent and that the majority of clones formed several unique groups. However, they were more closely related to *g20* sequences from freshwaters than to those from marine waters, suggesting that *g20* genes in terrestrial aquatic environments are different from those in marine environments.

Keywords Capsid assembly protein · Cyanophage · DGGE · Floodwater · *g20* · Rice field

Introduction

Viruses are the most abundant biological entities on earth (Fuhrman 1999; Wommack and Colwell 2000; Kimura et al. 2008). Many studies have indicated the ecological importance of viruses as the greatest, though little known, genomic reservoirs in marine and freshwater environments due to their large numbers and high diversity (Weinbauer and Rassoulzadegan 2004; Frost et al. 2005; Paul and Sullivan 2005). Bacteriophages (phages) represent the majority of virus in the natural environments (Wommack and Colwell 2000; Kimura et al. 2008).

Cyanophages are viruses that infect cyanobacteria. Unicellular cyanobacteria of the genera *Synechococcus* and *Prochlorococcus* are among the most abundant forms of marine picoplankton (Waterbury et al. 1986; Partensky et al. 1999), whereas filamentous cyanobacteria such as *Anabaena*, *Cylindrospermum*, and *Phormidium* are dominant members in freshwaters (Canter-Lund and Lund 1995). Although several cyanophages of filamentous cyanobacteria have been isolated from freshwater (Singh 1973; Hu et al. 1981; Deng and Hayes 2008), the research on their genomic diversity is limited (Baker et al. 2006; Deng and Hayes 2008). At present, genetic diversity of cyanophages is mainly based on the phages infecting oceanic *Synechococcus* and *Prochlorococcus* (Marston and Sallee 2003; Hambly and Suttle 2005; Wang and Chen 2008). The majority of cyanophages are classified into three-tailed phage families, *Myoviridae*, *Podoviridae*, and *Siphoviridae*, among which cyanomyoviruses represent more than 80% of cyanophage isolates from marine environments (Suttle and Chan 1993; Waterbury and Valois 1993; Marston and Sallee 2003).

As only a small fraction of bacteria in environments are cultivable (Ward et al. 1990; Pace 1997), the diversity of phages, though more ubiquitous and abundant than their host, in the biosphere is still poorly known (Filée et al. 2005). Studying the diversity of phages has been proven difficult because no universal gene, analogous to ribosomal

G. Wang (✉) · J. Murase · S. Asakawa · M. Kimura
Graduate School of Bioagricultural Sciences, Nagoya University,
Furocho, Chikusa,
Nagoya 464-8601, Japan
e-mail: guanghuawang@hotmail.com

G. Wang
Northeast Institute of Geography and Agroecology,
Chinese Academy of Sciences,
Harbin 150081, China

RNA gene used for microbial communities, exists throughout all phage families (Paul et al. 2002). However, recently some family-specific genes have been proposed for the evaluation of phage diversity (Rohwer and Edwards 2002). With the polymerase chain reaction (PCR) primer set for capsid assembly protein gene *g20* of cyanomyoviruses, very diverse *g20* gene fragments were discovered in marine and freshwater environments. For examples, cloning and sequencing analysis of six natural virus concentrates from estuarine and oligotrophic offshore environments revealed nine phylogenetic groups (Zhong et al. 2002). The use of this primer set and its improved one showed further groupings of cyanophage *g20* genes in various seawaters (Wang and Chen 2004; Short and Suttle 2005) and freshwaters (Dorigo et al. 2004; Short and Suttle 2005; Wilhelm et al. 2006).

Paddy field is a unique, anthropogenic aquatic environment that is managed for better rice production. Cyanobacteria are one of the major microbial components in the floodwater and play an important role in maintaining soil fertility by fixing atmospheric N_2 to ammonia (Kimura 2005). Since the majority of cyanobacterial members in paddy fields are filamentous and taxonomically different from those found in marine environments (Kimura 2005), it is expected that the composition and diversity of cyanophages in the floodwater are also different from those in marine environments. In this study, we report the composition and diversity of cyanophage communities in the floodwater of a Japanese paddy field by the molecular analysis of *g20* sequences. This was the first study that elucidated the composition and diversity of *g20* sequences in the paddy floodwater (PFW) by using molecular techniques.

Materials and methods

Study site and floodwater sampling

The paddy field was located in the Aichi-ken Anjo Research and Extension Center (34°48′N, 137°30′E), Japan. The soil was an Anthraquic Yellow Soil (Oxyquic Dystrudept). The field had been subjected to a long-term fertilizer trial since 1925, and the following four plots were chosen for this study: a plot without fertilizers (NoF plot), a plot with P and K chemical fertilizers but no N (CF(−N) plot), a plot with N, P, and K chemical fertilizers (CF plot), and a plot amended with chemical fertilizers (N, P, and K) and rice straw compost (CM plot). The amounts of N, P_2O_5, K_2O, and compost applied were 80 kg, 86 kg, 56 kg, and 22.5 ton per hectare, respectively (Kimura et al. 2002). These different plots were chosen not for comparing the *g20* composition among the plots but for obtaining as many as possible *g20* sequences from PFW.

Rice seedlings were transplanted on 9 June 2008 and managed with conventional practices. About 500 mL floodwater was collected three times from the middle part of each plot on 19 June, 10 July, and 30 July in 2008. The water samples were kept in a container with ice and transported to the laboratory within 8 h.

Concentration of viral communities

The PFW samples were centrifuged at 8,000×*g* for 20 min at 4°C to remove soil particles, plankton, and bacteria. Then, they were filtrated through a 0.2-μm cellulose filter to remove bacteria completely. Virus-size particles were collected on 0.02-μm filter (Anodisc 25; Whatman, Maidstone, England) by decompressed filtration. The filter was put into a 2-mL centrifuge tube with 700 μL 10 mM Tris–HCl buffer (pH 7.5).

DNA extraction and PCR amplification

The filter in the tube was treated with DNase and RNase (40 μg mL^{-1}) for 5 h at 37°C to decompose free DNA and RNA; then 38 μL 10% SDS, 7.5 μL 1 M Tris−HCl, 15 μL 0.5 M EDTA, and 2 μL proteinase K (10 mg ml^{-1}) were added to the tube, which was voltexed for 2 min and incubated for 30 min at 55°C with gentle shaking by hand every 10 min. At the end of incubation, we added 140 μL 5 M NaCl and 150 μL CTAB/NaCl solution, and the tube was incubated for 10 min at 65°C (Casas and Rohwer 2007). Viral DNA was extracted twice with PCI solution (phenol:chloroform:isoamyl alcohol=25:24:1, *v*/*v*) and once with CIA solution (chloroform:isoamyl alcohol= 24:1, *v*/*v*). The aqueous phase was treated with 0.6 volume of isopropanol and centrifuged at 18,000×*g* for 20 min at 4°C to obtain DNA pellet. Precipitated DNA was washed with 70% ethanol, dried, and resuspended in TE buffer.

The capsid assembly protein gene, *g20*, was amplified with primers CPS1 (5′-GTA GWA TTT TCT ACA TTG AYG TTG G-3′) and CPS8 (5′-AAA TAY TTD CCA ACA WAT GGA-3′) (Zhong et al. 2002). Fifty microliter of PCR mixture contained 0.4 μL of forward and reverse primers (50 pmol each), 1~2 μL of DNA template, 5 μL of dNTP (2.5 mM each; TaKaRa, Otsu, Japan), 0.5 μL of Ex-*Taq* polymerse (TaKaRa, Otsu, Japan), and 5 μL of Ex-*Taq* buffer (TaKaRa, Otsu, Japan) and was filled up to the required volume (36.7–37.7 μL) with MilliQ water. Negative control contained all reagents and sterile MilliQ water without the template. PCR amplification was performed by a thermal cycle PCR machine (TaKaRa, Otsu, Japan) at 94°C for 5 min (an initial denaturation), followed by 35 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min, with a final extension at 72°C for 5 min. One microliter of PCR product was used as template to conduct

the second PCR with GC-CPS1 and CPS8 primers under the same conditions as described above, except that PCR cycle was reduced to 25 times.

Denaturing gradient gel electrophoresis, cloning, and sequencing of *g20*

About 5~10 μL of PCR products (determined from PCR product concentration) were applied on an 8% (*w*/*v*) acrylamide gel with a 20% to 40% denaturant gradient (where a 100% denaturing gradient was defined as 7 M urea and 40% (*v*/*v*) deionized formamide). Denaturing gradient gel electrophoresis (DGGE) was run in 1×TAE buffer for 16 h at 60°C and 80 V with a Dcode Universal Mutation Detection System (Bio-Rad Laboratories, Hercules, CA, USA). After the electrophoresis, the gel was stained in 1:10,000 (*v*/*v*) SYBR™ Green I nucleic acid staining solution (CAMBREX, Rockland, ME, USA) for 30 min and photographed under ultraviolet light.

Using a sterile pipette tip, plugs were excised from DGGE bands. Each plug was placed in a sterilized microcentrifuge tube with 30 μL of sterilized TE buffer and kept at 4°C for more than 12 h to elute DNA from the excised plug. One microliter of the eluate was used as a template for PCR amplification with CPS1/CPS8 primers under the same PCR conditions described above. PCR products were purified with the QIAquick Gel Extraction kit (QIAGEN, Tokyo, Japan), and the purified DNA was cloned into pT7 Blue Vector (TaKaRa, Otsu, Japan). One positive clone from each transformation was chosen from white colonies and PCR-checked with the same primers for the correct insertion. Plasmid DNA was harvested from overnight cultures, and about 500 ng were subjected to sequencing reactions by using Ampli-Taq Fs BIGDYE Terminator Cycle Sequencing Chemistry (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's recommendations. Excess dye-terminators were removed from the completed sequencing products by ethanol precipitation using EDTA buffer, and the products were run in ABI 3130 automated genetic sequencers (Applied Biosystems, Foster City, CA, USA).

Phylogenetic analysis

Sequences obtained from the clones of respective DGGE bands were translated to deduced amino acid sequences by EMBOSS Transeq program at European Bioinformatics Institute web site (http://www.ebi.ac.uk/). Closest relatives of respective *g20* clones were examined by using the basic local alignment search tool (BLAST) search program at the National Center for Biotechnology Information (NCBI) web site (http://www.ncbi.nlm.nih.gov/) at the amino acid level. The closest relatives of respective *g20* fragments and three outgroup *g20* sequences of Coliphage T4 (AF158101), Vibriophage KVP40 (AB020525), and *Aeromonas* phage Aeh1 (AY266303) were retrieved from GenBank for constructing the phylogenetic tree (Wang and Chen 2008). Amino acid sequences were aligned with ClustalX 1.81 (Thompson et al. 1997). Phylogenetic tree using the minimum evolution approach was constructed by Molecular Evolutionary Genetic Analysis software (MEGA 3.0) (Kumar et al. 2004) with 1,000-fold bootstrap support. Nucleotide sequences determined in this study have been deposited in the DNA Data Bank of Japan as accession numbers ranging from AB471562 to AB471638.

Results

DGGE band patterns

About 600-bp PCR products were obtained from all the floodwater samples. Many DGGE bands appeared in every floodwater sample (Fig. 1). DGGE band patterns of floodwaters of different plots were different on 19 June and 10 July but similar on 30 July (Fig. 1).

The closest relatives of *g20* genes in the floodwater

All the DGGE bands were excised for sequencing on 19 June. The new DGGE bands (those appearing in a different position with respect to those of the profile of 19 June) of the following sampling days were excised for sequencing (Fig. 1). In total, 77 bands with different nucleotide sequences were identified as *g20* gene fragments, among which 22, 10, 12, and 33 clones were obtained from NoF, CF(−N), CF, and CM plots, respectively. The length of PCR products (between the primers) varied among clones: 546 bp for 66 clones (86%), 549 bp for 8 clones (10%), and 555 bp for 3 clones (4%). BLAST search within the NCBI for the closest relatives at the amino acid level revealed that the clones obtained in this study had the highest identities from 67% to 98% with the cyanophage clones or isolates obtained from marine and freshwater environments (Table 1). The identity within the clones in this study at the amino acid level was from 48% (PFW-CF10 and PFW-CF(−N)16) to 100% (PFW-NoF2, PFW-CF9 and PFW-CM28; PFW-CF(-N)5, PFW-CF(−N)7 and PFW-CM3; PFW-CF1 and PFW-CM1; PFW-CF12 and PFW-NoF21; PFW-CM12 and PFW-CM13; PFW-NoF13, PFW-NoF15 and PFW-NoF16).

Phylogeny of *g20* genes in the floodwater

Phylogenetic analysis revealed that *g20* clones could be grouped at least in the five major clusters (α to ε)

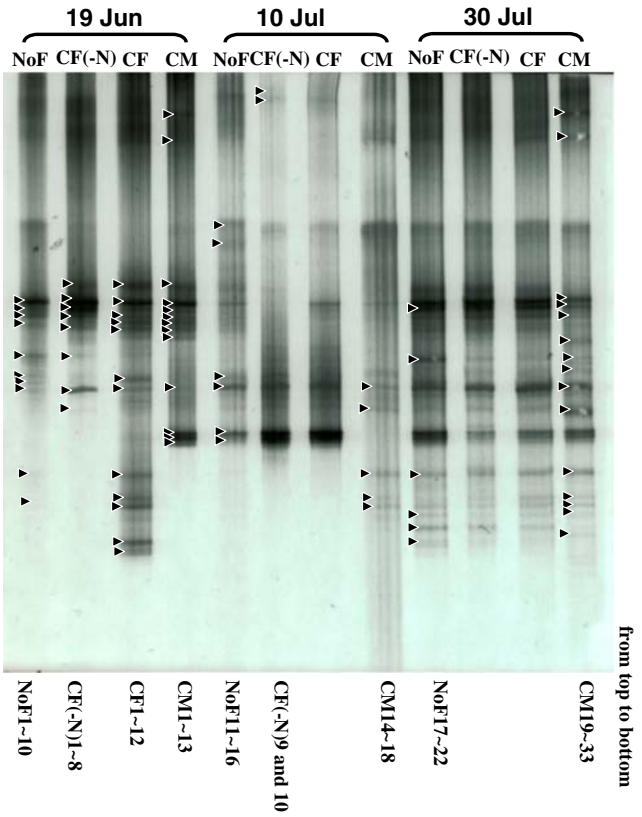


Fig. 1 DGGE band patterns of PCR products resulting from the amplification of *g20* fragment with primers GC-CPS1/CPS8 in a Japanese paddy floodwater. *NoF*, *CF(-N)*, *CF*, and *CM* represent the plots treated with no fertilizer, P and K chemical fertilizers, N, P, and K chemical fertilizers, and N, P, and K chemical fertilizers with compost, respectively

corresponding to clones and phage isolates reported previously (Fig. 2) (Fuller et al. 1998; Zhong et al. 2002; Marston and Sallee 2003; Dorigo et al. 2004; Wang and Chen 2004; Mann et al. 2005; Short and Suttle 2005; Wilhelm et al. 2006).

Cluster α was a big and strongly bootstrap-supported (85%) cluster that contained our *g20* clones, *g20* clones obtained from several lakes in Canada, Germany, France and USA (Dorigo et al. 2004; Short and Suttle 2005; Wilhelm et al. 2006) and two clones (BES02B-27 and BES02B-28) from the Arctic Ocean of Beaufort Sea (Short and Suttle 2005). Clones obtained from PFW formed four independent subclusters, designated as subclusters PFW-I, PFW-II, PFW-III, and PFW-IV and consisted of 16, 14, 4, and 16 clones, respectively (Fig. 2). The clones in groups PFW-I and PFW-II and group PFW-IV showed the highest identity of about 78% and 70% with clones CUL02H19 and CUL02H16, respectively, obtained from Cultus Lake (Short and Suttle 2005), and the clones in group PFW-III had the highest identity of 93% with clone MC33 obtained from Laurentian Great Lake (Wilhelm et al. 2006; Table 1).

Cluster β was a weakly bootstrap-supported (58%) cluster included nine clones of this study, seven clones from freshwater environments (Dorigo et al. 2004; Wilhelm et al. 2006), and three clones (OTU1, OTU7, and BES02A-4) from marine environments (Wang and Chen 2004; Short and Suttle 2005). Within this cluster, two clones (PFW-CM17 and PFW-CM33) formed an independent remote subcluster (subcluster PFW-V) and showed the highest identity of about 67% with clone OTU1 from the surface water of Chesapeake Bay (Wang and Chen 2004; Table 1).

Cluster δ was strongly bootstrap-supported (87%) and included marine and freshwater clones, and *Synechococcus* and *Prochlorococcus* phages. All clones obtained from PFW belonged to two independent subclusters, and no clone or phage isolate from other environments fell into these subclusters. As cluster δ coincided with cluster cultured *Synechococcus* phage (CSP) proposed by Short and Suttle (2005), the PFW clones of this cluster were named as CSP-PFW1 and CSP-PFW2 members. BLAST search clones of CSP-PFW1 and CSP-PFW2 subclusters showed an identity of 91% with clone MC4 from Laurentian Great Lake (Wilhelm et al. 2006), clone SAI99-44 from Salmon Inlet (Short and Suttle 2005), and cyanophage S-RIM21 (Marston and Sallee 2003; Table 1).

Clusters γ and ε with very weak bootstrap supports (<25%) contained environmental *g20* clones obtained from marine, brackishwater, freshwater, and paddy floodwater environments. Five *g20* clones (PFW-CM12, PFW-CM13, PFW-CF(−N)9, PFW-CF(−N)10, and PFW-CM11) of cluster γ formed a small subcluster with strong bootstrap support (98%), and it was designated as PFW-VI. These clones showed about 69% (SE13) or 72% (SS4714) identity with marine clones (Zhong et al. 2002; Table 1). The phylogenetic positions of clone PFW-CM14 was remote.

Discussion

PCR and DGGE

Many primer sets had been designed to amplify cyanophage *g20* gene (CPS1/CPS2, CPS1/CPS4, and CPS1/CPS8 Fuller et al. 1998; Zhong et al. 2002; Marston and Sallee 2003). A recent study by Sullivan et al. (2008) revealed that primers CPS1/CPS8 only amplified *g20* genes from cyanophages but did not amplify *g20* genes of other 12 tested noncyanophages. The specificity of CPS1/CPS8 pairs to *g20* genes was also supported in this study from the finding that every clone in DGGE bands belonged to *g20* gene in BLAST search (Table 1).

The DGGE fingerprints (Fig. 1) indicated that *g20* assembly structure in the floodwater consisted of diverse

Table 1 The closest relative of sequenced *g20* clones of paddy floodwater at the amino acid level

Clone name (PFW-)	Length[a]	Alignment	Closet relatives	Accession number[b]	Identity (%)	Remarks
NoF1, 5	183	177/182	Clone MC24	ABC49816	97	Laurentian Great Lake (Wilhelm et al. 2006)
NoF2, 8, 9, 19	182	127/180	Clone CUL02H-16	AAW48781	70	Cultus Lake (Short and Suttle 2005)
NoF3, 7	182	143/179	Clone j04	AAR10331	79	Lake Bourget (Dorigo et al. 2004)
NoF4	182	144/179	Clone j04	AAR10331	79	Lake Bourget (Dorigo et al. 2004)
NoF6	183	180/182	Clone MC24	ABC49816	98	Laurentian Great Lake (Wilhelm et al. 2006)
NoF7	182	143/179	Clone j04	AAR10331	79	Lake Bourget (Dorigo et al. 2004)
NoF10	182	125/180	Clone CUL02H-16	AAW48781	69	Cultus Lake (Short and Suttle 2005)
NoF11	182	141/180	Clone CUL02H19	AAW48778	77	Cultus Lake (Short and Suttle 2005)
NoF12	182	139/180	Clone CUL02H19	AAW48778	76	Cultus Lake (Short and Suttle 2005)
NoF13~16	182	170/181	Clone MC33	ABC49822	93	Laurentian Great Lake (Wilhelm et al. 2006)
NoF17	182	126/180	Clone CUL02H-16	AAW48781	69	Cultus Lake (Short and Suttle 2005)
NoF18	182	165/182	Clone MC4	ABC49801	91	Laurentian Great Lake (Wilhelm et al. 2006)
NoF20	183	167/182	S-RIM21	AAP83556	91	Cyanophage of *Synechococcus* (Marston and Sallee 2003)
NoF21	185	168/185	Clone MC28	ABC49818	91	Laurentian Great Lake (Wilhelm et al. 2006)
NoF22	185	167/185	Clone MC28	ABC49818	90	Laurentian Great Lake (Wilhelm et al. 2006)
CF(−N)1, 8	182	142/179	Clone j04	AAR10331	78	Bourget Lake (Dorigo et al. 2004)
CF(−N)2	182	144/180	Clone CUL02H19	AAW48778	79	Cultus Lake (Short and Suttle 2005)
CF(−N)3~7	182	143/179	Clone j04	AAR10331	79	Bourget Lake (Dorigo et al. 2004)
CF(−N)9	182	132/182	Clone SS4714	AAK31745	73	Sargasso Sea (Zhong et al. 2002)
CF(−N)10	182	131/182	Clone SS4714	AAK31745	72	Sargasso Sea (Zhong et al. 2002)
CF1	182	144/179	Clone j04	AAR10331	79	Bourget Lake (Dorigo et al. 2004)
CF2	182	167/182	Clone MC4	ABC49801	92	Laurentian Great Lake (Wilhelm et al. 2006)
CF3	182	143/180	Clone CUL02H19	AAW48778	79	Cultus Lake (Short and Suttle 2005)
CF4	182	165/182	Clone MC4	ABC49801	91	Laurentian Great Lake (Wilhelm et al. 2006)
CF5, 7, 8	182	166/182	Clone MC4	ABC49801	91	Laurentian Great Lake (Wilhelm et al. 2006)
CF6	183	167/181	Clone SAI99-44	AAW48744	91	Salmon Inlet (Short and Suttle 2005)
CF9	182	127/180	Clone CUL02H-16	AAW48781	70	Cultus Lake (Short and Suttle 2005)
CF10	182	125/180	Clone CUL02H-16	AAW48781	69	Cultus Lake (Short and Suttle, 2005)
CF11	183	166/182	S-RIM21	AAP83556	91	Cyanophage of *Synechococcus* (Marston and Sallee 2003)
CF12	185	168/185	Clone MC28	ABC49818	91	Laurentian Great Lake (Wilhelm et al. 2006)
CM1,5~7, 10, 22	182	142/179	Clone j04	AAR10331	78	Bourget Lake (Dorigo et al. 2004)
CM2, 3, 21, 23	182	143/179	Clone j04	AAR10331	79	Bourget Lake (Dorigo et al. 2004)
CM4, 9, 19, 20	182	141/179	Clone j04	AAR10331	77	Bourget Lake (Dorigo et al. 2004)
CM8	182	144/179	Clone j04	AAR10331	79	Bourget Lake (Dorigo et al. 2004)
CM11	182	125/181	S-SM1	ACD93421	69	Cyanophage of *Synechococcus* (Sullivan et al. 2008)
CM12, 13	182	126/181	Clone SE13	AAK31688	69	Surface water of Savannah estuary (Zhong et al. 2002)
CM14	182	122/182	Clone OTU14	AAO13212	67	Surface water of Chesapeake Bay (Wang and Chen 2004)
CM15	182	128/182	Clone SS4051	AAK31722	70	Sargasso Sea (Zhong et al. 2002)
CM16	183	125/183	Clone SE29	AAK31699	68	Surface water of Savannah estuary (Zhong et al. 2002)

Table 1 (continued)

Clone name (PFW-)	Length[a]	Alignment	Closet relatives	Accession number[b]	Identity (%)	Remarks
CM17	182	121/182	Clone OTU1	AAO13199	66	Surface water of Chesapeake Bay (Wang and Chen 2004)
CM18	182	127/182	Clone SS4051	AAK31722	70	Sargasso Sea (Zhong et al. 2002)
CM24, 26, 29, 30, 32	182	126/180	Clone CUL02H-16	AAW48781	69	Cultus Lake (Short and Suttle 2005)
CM25	183	175/182	Clone MC24	ABC49816	96	Laurentian Great Lake (Wilhelm et al. 2006)
CM27, 28, 31	182	127/180	Clone CUL02H-16	AAW48781	70	Cultus Lake (Short and Suttle 2005)
CM33	182	122/182	Clone OTU1	AAO13199	67	Surface water of Chesapeake Bay (Wang and Chen 2004)

[a] The length of amino acid residues

[b] Accession number of amino acid sequences

members. The remarkable difference in DGGE band patterns observed among different fertilizer treatments on 19 June decreased among the treatments along with the sampling time and indicated the change of *g20* assembly structure in PFW with growth stage and fertilizer treatment (Fig. 1). Cyanobacterial communities in the floodwater were not examined in this study, but several studies have shown that the cyanobacterial communities of paddy fields change through season (Song et al. 2005) and depend on fertilizer treatments (Jha et al. 2001; Irisarri et al. 2001) and application of organic materials (Jha et al. 2004; Asari et al. 2008). Therefore, the DGGE band patterns in Fig. 1 might indicate that cyanophage communities changed with the change of cyanobacterial communities in PFW.

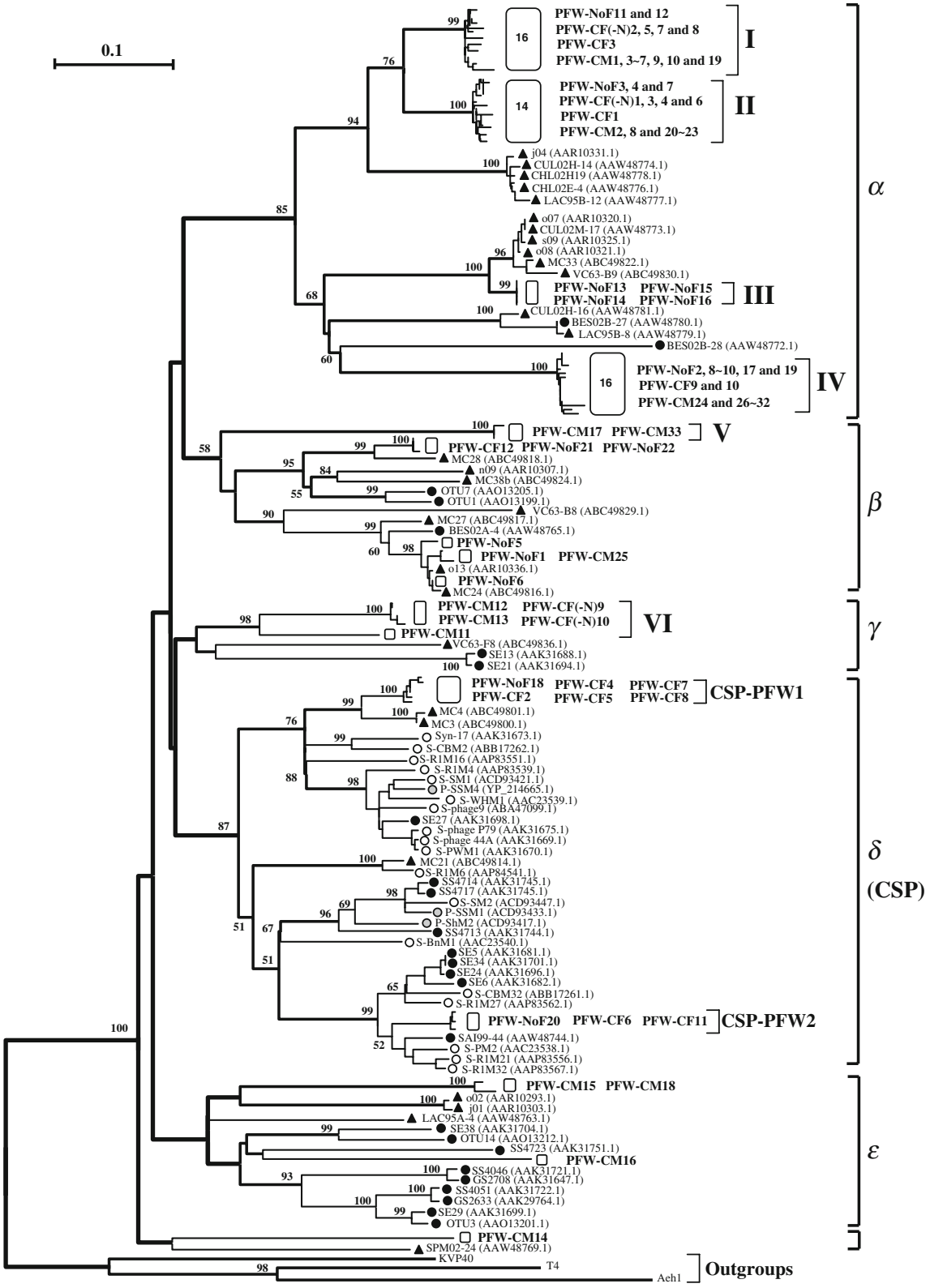
Phylogenetic position of *g20* genes in the floodwater

The *g20* clones obtained from PFW were distributed among all the five major clusters of Fig. 2. Even within a single cluster, clones/phages demonstrated their divergent nature by the wide range of sequence identity, and the pairwise sequence identity levels ranged from 41% to 100% in cluster α, from 43% to 100% in cluster β, from 54% to 100% in cluster γ, from 51% to 100% in cluster δ, and from 50% to 99% in cluster ε. When the pairwise sequence identity of the clones/phages belonging to different clusters were compared, it ranged between 27% and 61% for cluster α clones, between 29% and 67% for cluster β clones, between 36% and 68% for cluster γ clones, between 33% and 68% for cluster δ clones, and between 28% and 67% for cluster ε clones. Thus, the diversity of cyanophage communities was large in PFW.

Cluster CSP, first designated by Short and Suttle (2005), contained all of the *g20* sequences from cyanophage isolates infecting *Synechococcus* and *Prochlorococcus*, and about 40.7% of environmental clones collected from marine and freshwater environments. Only 12% (nine clones) of the PFW clones fell into cluster δ (Fig. 2). Clones from the floodwater formed two small but unique subclusters CSP-PFW1 and CSP-PSW2 in Fig. 2. Since *Prochlorococcus* is restricted in marine environments (Partensky et al. 1999), *Synechococcus* seems to be the host for those clones. Indeed, 16S ribosomal DNA sequences related to *Synechococcus* were isolated from paddy fields (Song et al. 2005). Only a low proportion of *g20* clones belonging to cluster CPS in PFW indicated that *Synechococcus* are not predominant in paddy environments (Kimura 2005; Song et al. 2005). In addition, PFW clones formed two unique small clades (Fig. 2), suggesting that *Synechococcus* communities of paddy freshwaters are different from those in freshwaters (lakes) and oceans. *Synechococcus* spp. are known to be genetically diversified in different environments (Fuller et al. 2003).

Cluster α of Fig. 2 exclusively included PFW clones and clones from freshwater environments except for two clones of marine origin (BES02B-27 and BES02B-28). The majority of *g20* clones (50 of 77 clones) from PFW fell into this cluster. This finding may indicate that cluster α is specific to *g20* of freshwater environments including the floodwater of paddy fields.

Fig. 2 Neighbor-joining phylogenetic tree of *g20* sequences showing the relationships of *g20* amino acid sequences from the paddy floodwater with those from freshwater lakes (Dorigo et al. 2004; Short and Suttle 2005; Wilhelm et al. 2006) and marine environments (Fuller et al. 1998; Mann et al. 2005; Marston and Sallee 2003; Short and Suttle 2005; Wang and Chen 2004; Zhong et al. 2002). *Black triangles and circles* indicated *g20* clones obtained from freshwater and marine environments, respectively; *White and gray circles* indicated the cyanophage isolates infecting marine *Synechococcus* and *Prochlorococcus*, respectively. Clones obtained in this study are presented with *square boxes*, and the *numbers in the box* indicate the clone number of the clusters. The *numbers in the parenthesis* are the accession numbers of amino acid sequences in NCBI web site. Bootstrap values less than 50 are not shown. The *scale bar* represents the abundance of amino acid substitutions per residue


0.1
16
PFW-NoF11 and 12
PFW-CF(-N)2, 5, 7 and 8
PFW-CF3
PFW-CM1, 3~7, 9, 10 and 19
I
14
PFW-NoF3, 4 and 7
PFW-CF(-N)1, 3, 4 and 6
PFW-CF1
PFW-CM2, 8 and 20~23
II
j04 (AAR10331.1)
CUL02H-14 (AAW48774.1)
CHL02H19 (AAW48778.1)
CHL02E-4 (AAW48776.1)
LAC95B-12 (AAW48777.1)
o07 (AAR10320.1)
CUL02M-17 (AAW48773.1)
s09 (AAR10325.1)
o08 (AAR10321.1)
MC33 (ABC49822.1)
VC63-B9 (ABC49830.1)
PFW-NoF13 PFW-NoF15
PFW-NoF14 PFW-NoF16
III
CUL02H-16 (AAW48781.1)
BES02B-27 (AAW48780.1)
LAC95B-8 (AAW48779.1)
BES02B-28 (AAW48772.1)
16
PFW-NoF2, 8~10, 17 and 19
PFW-CF9 and 10
PFW-CM24 and 26~32
IV
α
PFW-CM17 PFW-CM33
V
PFW-CF12 PFW-NoF21 PFW-NoF22
MC28 (ABC49818.1)
n09 (AAR10307.1)
MC38b (ABC49824.1)
OTU7 (AAO13205.1)
OTU1 (AAO13199.1)
VC63-B8 (ABC49829.1)
MC27 (ABC49817.1)
BES02A-4 (AAW48765.1)
PFW-NoF5
PFW-NoF1 PFW-CM25
o13 (AAR10336.1)
PFW-NoF6
MC24 (ABC49816.1)
β
PFW-CM12 PFW-CF(-N)9
PFW-CM13 PFW-CF(-N)10
VI
PFW-CM11
VC63-F8 (ABC49836.1)
SE13 (AAK31688.1)
SE21 (AAK31694.1)
γ
PFW-NoF18 PFW-CF4 PFW-CF7
PFW-CF2 PFW-CF5 PFW-CF8
CSP-PFW1
MC4 (ABC49801.1)
MC3 (ABC49800.1)
Syn-17 (AAK31673.1)
S-CBM2 (ABB17262.1)
S-R1M16 (AAP83551.1)
S-R1M4 (AAP83539.1)
S-SM1 (ACD93421.1)
P-SSM4 (YP_214665.1)
S-WHM1 (AAC23539.1)
S-phage9 (ABA47099.1)
SE27 (AAK31698.1)
S-phage P79 (AAK31675.1)
S-phage 44A (AAK31669.1)
S-PWM1 (AAK31670.1)
MC21 (ABC49814.1)
S-R1M6 (AAP84541.1)
SS4714 (AAK31745.1)
SS4717 (AAK31745.1)
S-SM2 (ACD93447.1)
P-SSM1 (ACD93433.1)
P-ShM2 (ACD93417.1)
SS4713 (AAK31744.1)
S-BnM1 (AAC23540.1)
SE5 (AAK31681.1)
SE34 (AAK31701.1)
SE24 (AAK31696.1)
SE6 (AAK31682.1)
S-CBM32 (ABB17261.1)
S-R1M27 (AAP83562.1)
PFW-NoF20 PFW-CF6 PFW-CF11
CSP-PFW2
SAI99-44 (AAW48744.1)
S-PM2 (AAC23538.1)
S-R1M21 (AAP83556.1)
S-R1M32 (AAP83567.1)
δ
(CSP)
PFW-CM15 PFW-CM18
o02 (AAR10293.1)
j01 (AAR10303.1)
LAC95A-4 (AAW48763.1)
SE38 (AAK31704.1)
OTU14 (AAO13212.1)
SS4723 (AAK31751.1)
PFW-CM16
SS4046 (AAK31721.1)
GS2708 (AAK31647.1)
SS4051 (AAK31722.1)
GS2633 (AAK29764.1)
SE29 (AAK31699.1)
OTU3 (AAO13201.1)
ε
PFW-CM14
SPM02-24 (AAW48769.1)
KVP40
T4
Aeh1
Outgroups

Previous researches showed that only some of the *g20* sequences of freshwater formed separated clusters from these of marine sequences (Dorigo et al. 2004; Short and Suttle 2005; Wilhelm et al. 2006). In addition, *g20* sequences in the estuary (Zhong et al. 2002) and Chesapeake Bay (Wang and Chen 2004) differed from those in the open oceans. These findings were consistent with what observed in our study; indeed, clusters β, γ, and ε contained *g20* clones from PFW, freshwaters, and sea waters. To compare the phylogenetic position of *g20* sequences in different environments, an unrooted phylogenetic tree was constructed with all *g20* sequence data available in GenBank (Fig. 3). The *g20* genes were divided into three large groups: the first group was CSP which contained cultured cyanophages and *g20* clones from various environments including PFW; the second group consisted of *g20* clones from PFW and freshwaters with the exception of two clones from the Arctic Ocean; and the third group was much more diversified including *g20* clones from various environments. The PFW clones were nearer to those from freshwaters than those from marine waters even in the third diversified group. Thus, PFW sequences formed several unique branches and were more closely related to *g20* sequences of freshwaters than to those from marine environments.

Cyanophage hosts estimated from *g20* diversity

Short and Suttle (2005) obtained various *g20* clones by using newly designed CPS4/G20-2 primers which amplify the same region of *g20* gene by CPS1/CPS8 primers, and they suggested that the *g20* sequences outside the CSP group may not be from cyanophages by considering that three *g20* clones outside the CSP group were found in >3,000-m-deep water in the Arctic Chukchi Sea and in the Arctic Beaufort Sea where the picocyanobacterial abundance was very low ($<10^2$ cells per milliliter). As describe before, two clones (BES02B-27 and BES02B-28) from the Arctic Beaufort Sea (Short and Suttle 2005) were in cluster α, which might indicate that cluster α consists of non-cyanophages. However, other clones belonging to cluster α such as j04, o07, o08 s09, MC33, and VC63-B9 were cyanophages from lake waters, and the relative extracted DNA was amplified by primers CPS1/CPS8 (Dorigo et al. 2004; Wilhelm et al. 2006). As the degeneracy of CPS4/G20-2 primers (1,024-fold) is much greater than that of primers CPS1/CPS8 (48-fold) and Sullivan et al. (2008) revealed the primers CPS1/CPS8 to be cyanophage-specific as mentioned before, most of PFW clones, if not all, could be regarded as cyanophage genes. Therefore, the wide distributions of PFW clones in Fig. 2 suggested that various

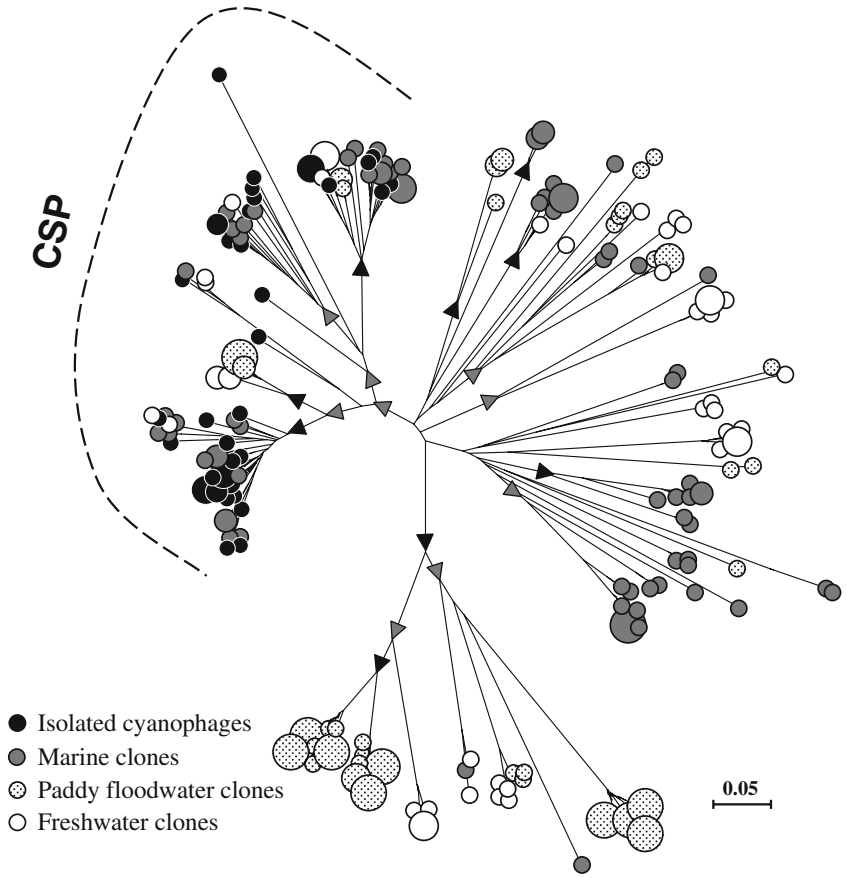


Fig. 3 Unrooted phylogenetic tree comparing *g20* amino acid sequences from environmental clones from marine, freshwater, and paddy floodwater, as well as from isolated cyanophages of marine *Synechococcus* and *Prochlorococcus*. The *black and gray triangles* indicate internal nodes with at least 90% and 50% bootstrap supports, respectively. The *size of circles at the end of branches* is proportional to the number of clones/phages, and the *smallest and largest circles* represent one and four clones/phages, respectively

cyanobacteria including *Synechococcus* were the hosts of phages in PFW, although most of these host cyanobacteria are unknown.

Since the first isolation of cyanophage LPP-1 which infects three genera of filamentous cyanobacteria (*Lyngbya*, *Plectonema*, and *Phormidium*; Safferman and Morris 1963), dozens of cyanophages of filamentous cyanobacteria had been obtained from freshwater and PFW (Singh 1973; Hu et al. 1981; Deng and Hayes 2008). However, information on their *g20* genes are still very limited (Baker et al. 2006; Deng and Hayes 2008). Baker et al. (2006) found that primer sets CPS1/CPS2 and CPS1/CPS8 were unfitted for the amplification of *Anabaena* phages AN-15, A-1 (L), and N-1. On the contrary, Deng and Hayes (2008) successfully amplified *g20* gene of phage P-Z1 infecting *Planktothrix rubescens* BC9307 by CPS1/CPS2 among 24 tested phages of filamentous cyanobacteria; however, they did not test the amplification of *g20* gene of P-Z1 with CPS1/CPS8. Thus, it was not clear whether PFW clones included the *g20* of cyanophages infecting filamentous cyanobacteria.

In conclusion, viral capsid assembly protein genes (*g20*) of the floodwater of a Japanese paddy field were PCR amplified with CPS1/CPS8 primers and separated by DGGE to evaluate the genetic diversity of cyanophage communities. The DGGE profile showed that cyanophage communities of the floodwater were affected by both fertilizer treatments and sampling time. In total, 77 bands with different nucleotide sequences were identified as *g20* clones. The *g20* clones obtained from the floodwater were grouped in five major clusters (α to ε), and the majority of clones from the floodwater formed several unique groups within each cluster. They were, in general, more closely related to *g20* sequences from freshwaters than those from marine waters, indicating that *g20* genes in terrestrial aquatic environments are different than those from marine.

Acknowledgments Guanghua Wang extends his gratitude to Japan Society for the Promotion of Science (JSPS) for financial support to conduct this study.

References

Asari N, Ishihara R, Nakajima Y, Kimura M, Asakawa S (2008) Cyanobacterial communities of rice straw left on the soil surface of a paddy field. Biol Fertil Soils 44:605–612

Baker AC, Goddard VJ, Davy J, Schroeder DC, Adams DG, Wilson WH (2006) Identification of a diagnostic marker to detect freshwater cyanophages of filamentous cyanobacteria. Appl Environ Microbiol 72:5713–5719

Canter-Lund H, Lund JWG (1995) Freshwater algae: their microscopic world explored. Biopress Ltd, Bristol, England, pp 194–237

Casas V, Rohwer F (2007) Phage metagenomics. Methods Enzymol 421:259–268

Deng L, Hayes PK (2008) Evidence for cyanophages active against bloom-forming freshwater cyanobacteria. Freshwater Biol 53:1240–1252

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

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

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3:722–732

Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. Nature 399:541–548

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

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. Appl Environ Microbiol 69:2430–2443

Hambly E, Suttle CA (2005) The viriosphere, diversity, and genetic exchange with phage communities. Curr Opin Microbiol 8:444–450

Hu NT, Thiel T, Giddings TH, Wolk CP (1981) New *Anabaena* and *Nostoc* cyanophages from sewage settling ponds. Virology 114:236–246

Irisarri P, Gonnet S, Monza J (2001) Cyanobacteria in Uruguayan rice fields: diversity, nitrogen fixing ability and tolerance to herbicides and combined nitrogen. J Biotechnol 91:95–103

Jha MN, Presad AN, Sharma SG, Bharati RC (2001) Effects of fertilization rate and crop rotation on diazotrophic cyanobacteria in paddy field. World J Microbiol Biotechnol 17:463–468

Jha MN, Presad AN, Misra SK (2004) Influence of source of organics and soil organic matter content on cyanobacterial nitrogen fixation and distributional pattern under different water regimes. World J Microbiol Biotechnol 20:673–677

Kimura M (2005) Populations, community composition and biomass of aquatic organisms in the floodwater of rice fields and effects of field management. Soil Sci Plant Nutr 51:159–181

Kimura M, Shibagaki T, Nakajima Y, Matsuya K, Ikenaga M (2002) Community structure of the microbiota in the floodwater of a Japanese paddy field estimated by restriction fragment length polymorphism and denaturing gradient gel electrophoresis pattern analysis. Biol Fertil Soils 36:306–312

Kimura M, Jia Z, Nakayama N, Asakawa S (2008) Viral ecology in soil: past, present, and future perspectives. Soil Sci Plant Nutr 54:1–32

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Mann NH, Clokie MR, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A, Krisch HM (2005) The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. J Bacteriol 187:3188–3200

Marston M, Sallee J (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

Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734–740

Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynethetic prokaryote of global significance. Microbiol Mol Biol Rev 63:106–127

Paul JH, Sullivan MS (2005) Marine phage genomics: what have we learned? Curr Opin Biotechnol 16:299–307

Paul JH, Sullivan MB, Segall AM, Rohwer F (2002) Marine phage genomics. Comp Biochem Phys B 133:463–476

Rohwer F, Edwards R (2002) The phage proteomic tree: a genome based taxonomy for phage. J Bacteriol 184:4529–4535

Safferman RS, Morris ME (1963) Algal virus isolation. Science 140:679–680

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

Singh PK (1973) Occurrence and distribution of cyanophages in ponds, sewage and rice fields. Arch Mikrobiol 89:169–172

Song T, Martensson L, Eriksson T, Zheng W, Rasmessen U (2005) Biodiversity and seasonal variation of the cyanobacterial assemblage in rice field in Fujian, China. FEMS Microbiol Ecol 54:131–140

Sullivan MB, Coleman ML, Quinlivan V, Rosenkrantz JE, De Francesco AS, Tan G, Fu R, Lee JA, Waterbury JB, Bielawski JP, Chisholm SW (2008) Portal protein diversity and phage ecology. Environ Microbiol 10:2810–2823

Suttle CA, Chan AM (1993) Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-reactivity and growth characteristics. Mar Ecol Prog Ser 92:99–109

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

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, Chen F (2008) Prevalence of highly host-specific cyanophages in the estuarine environment. Environ Microbiol 10:300–312

Ward DM, Weller R, Bateson MM (1990) 16S rRNA sequences reveal numerous uncultured microorganisms in a nature community. Nature 344:63–65

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

Waterbury JB, Watson SW, Valois FW, Franks DG (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. Can Bull Fish Aquat Sci 214:71–120

Weinbauer MG, Rassoulzadegan F (2004) Are viruses driving microbial diversification and diversity? Environ Microbiol 6:1–11

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 assay and molecular analyses of *g20* genes. Appl Environ Microbiol 72:4957–4963

Wommack KE, Colwell RR (2000) Virioplankton: viruses in aquatic ecosystems. Microbiol Mol Biol Rev 64:69–114

Zhong Y, Chen F, Wilhelm SW, Poorvin L, Hodson RE (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