AEM Journals.ASM.org

MINIREVIEW

Using Signature Genes as Tools To Assess Environmental Viral Ecology and Diversity

Evelien M. Adriaenssens, Don A. Cowan

Centre for Microbial Ecology and Genomics, University of Pretoria, Pretoria, South Africa

Viruses (including bacteriophages) are the most abundant biological entities on the planet. As such, they are thought to have a major impact on all aspects of microbial community structure and function. Despite this critical role in ecosystem processes, the study of virus/phage diversity has lagged far behind parallel studies of the bacterial and eukaryotic kingdoms, largely due to the absence of any universal phylogenetic marker. Here we review the development and use of signature genes to investigate viral diversity, as a viable strategy for data sets of specific virus groups. Genes that have been used include those encoding structural proteins, such as portal protein, major capsid protein, and tail sheath protein, auxiliary metabolism genes, such as *psbA*, *psbB*, and *phoH*, and several polymerase genes. These marker genes have been used in combination with PCR-based fingerprinting and/or sequencing strategies to investigate spatial, temporal, and seasonal variations and diversity in a wide range of habitats.

Viruses of microbes are the most abundant biological entities on earth (1, 2). They play key roles in the shaping of microbial communities and are considered major factors in controlling nutrient cycling in a wide range of environments (3, 4). It is widely accepted that studies of viral abundance and diversity will lead (and have led) to novel insights into the functioning of the microbial biosphere. Several of the methods used in environmental virology, such as microscopy, genome/amplicon size fingerprinting, signature genes, and whole-genome or metagenome sequencing, have been recently reviewed, with the conclusion that all such methods have both benefits and drawbacks (5).

Viruses recovered in environmental studies can have a range of different hosts, from *Bacteria* and *Archaea* to amoeba, algae, plants, and animals. For viruses of prokaryotes, or phages, the most well-studied group is the order *Caudovirales*, comprising tailed bacteriophages and archaeal phages containing double-stranded DNA (dsDNA) and that can be further divided into three families: the *Myoviridae*, with long contractile tails, the *Siphoviridae*, with long noncontractile tails, and the *Podoviridae*, with short tails (6). Another group of dsDNA viruses that has been the focus of recent research is the nucleocytoplasmic large DNA viruses (NCLDVs), a grouping without taxonomic status containing, among others, the algae- and amoeba-infecting families *Phycodnaviridae* and *Mimiviridae* (7). Currently, the International Committee on the Taxonomy of Viruses (ICTV) recognizes 7 viral orders comprising 25 families and 71 families not assigned to any order (ICTV current taxonomy is available at www.ictvonline.org).

Here we focus on the use of viral signature genes to investigate the diversity of viral communities in different environments and provide an overview of both the genes and the techniques employed. Since there is no universal signature gene present in either prokaryotic or eukaryotic viruses (i.e., no homologue of the 16S rRNA gene in *Bacteria* and *Archaea*), many different genes were assessed as potential group-specific signature genes (Table 1) (8).

GENES ENCODING STRUCTURAL PROTEINS

T4-like g20 portal protein in the family *Myoviridae*. The application of the myovirus T4 portal protein gene, *g20*, as a signature gene originated from the identification of a conserved sequence in marine cyanophages that corresponded with *g20*, which led to the development of primer set CPS1-CPS2 (cyanophage specific) (Table 1) (9). Using additional CPS primer sets combined with denaturing gradient gel electrophoresis (DGGE) on an Atlantic Ocean north-south transect and depth profile, it was found that similar cyanophages were common along the transect but showed much greater diversity along the depth profile, reflecting nutrient and host bacterial abundance levels (10, 11). The same observations were made for cyanophage communities at the coast of British Columbia, Canada (12). A larger-scale Atlantic north-south transect revealed widely distributed similar cyanomyovirus signatures, but the variation present could not be linked to environmental factors or the host bacteria (13). g20 clone libraries of estuarine and open ocean environments revealed a high diversity in these two environments as well as high novelty in sequences (14).

Using *g20* PCR restriction fragment length polymorphism (RFLP) and sequence analysis, temporal changes in marine cyanophage community composition and abundance were found in coastal waters near Rhode Island, and several RIM (Rhode Island myovirus) clusters were defined (15). Several other studies found well-established seasonal changes in marine cyanophage diversity, linked to host cell density, but low spatial variation (16–18).

The *g20* marker gene has also been successfully used for analysis of freshwater environments. In the mesotrophic Lake Bourget (France), a high diversity of cyanophages was identified, with significant similarity to clades of marine cyanophages (19). In accordance with results from the marine studies, phage abundance was lowest in winter months. Water samples from Lake Erie (Canada and United States) and a ship's ballast tank showed novel sequences and sequences related to marine cyanophages, suggesting a possible marine origin for the freshwater cyanophages or, alternatively, that they infect a different host in the lake environment

Published ahead of print 16 May 2014

Editor: T. K. Wood

Address correspondence to Don A. Cowan, don.cowan@up.ac.za.

doi:10.1128/AEM.00878-14

TABLE 1 List of signature genes described in this study and the associated primer sets

Signature gene	Gene product	Target virus group	Primer (5′–3′)	Reference(s)
g20	Portal protein	Cyanophages belonging to *Myoviridae* family	CPS1: GTAGWATTTTCTACATTGAYGTTGG	9
			CPS2: GGTARCCAGAAATCYTCMAGCAT	
			CPS4GC: CGCCCGGGGCGCGCCCCGGGCGGGGCGGGGGCA CGGGGGGGTAGAATTTTCTACATTGATGTTGG	11
			CPS5: GGTAACCAGAAATCTTCAAGCAT	
			CPS3: TGGTAYGTYGATGGMAGG	14
			CPS4: CATWTCWTCCCAHTCTTC	
			CPS8: AAATAYTTDCCAACAWATGGA	
			G20-2: SWRAAATAYTTICCRACRMAKGGATC	24
			CPS1.1: GTAGWATWTTYTAYATTGAYGTWGG	67
			CPS8.1: ARTAYTTDCCDAYRWAWGGWTC	
g23	Major capsid protein	T4-related members of *Myoviridae*	MZIA1: TGTTATIGGTATGGTICGICGTGCTAT	26
			CAP8: TGAAGTTACCTTCACCACGACCGG	
			MZIA1bis: GATATTTGIGGIGTTCAGCCIATGA	27
			MZIA6: CGCGGTTGATTTCCAGCATGATTTC	
			ScExoT-F: CWCGTCAAYTGAAAGCTCAA	28
			ScExoT-R: AWTTKMAYACCGTARCGAGT	
			T4superF1: [tetrachlorofluorescein]-GAYHTIKSIGGIGTICARCCIATG	30
			T4superR1: [6-carboxyfluorescein]–GCIYKIARRTCYTGIGCIARYTC	
			G23-For: ACWGGWCTKATYTTCGCAATG	90
			G23-Rev: AYTTYTCAACWGACCADCKACC	
mcp	Major capsid protein	Freshwater cyanophages of *Myoviridae* family	AN15 MCPF5: GTTCCTGGCACACCTGAAGCGT	56
			AN15 MCPR5: CTTACCATCGCTTGTGTCGGCATC	
mcp	Major capsid protein	*Phycodnaviridae*	mcp-F: GTCTTCGTACCAGAAGCACTCGCT	111
			mcp-R: ACGCCTCGGTGTACGCACCCTCA	
			mcp Fwd: GGYGGYCARCGYTTGA	57
			mcp Rev: TGIARYTGYTCRAYIAGGTA	
mcp	Major capsid protein	*Gokushovirinae* subfamily of *Microviridae* family of ssDNA viruses	MCPf: CCYKGKYYNCARAAAGG	59
			MCPr: AHCKYTCYTGRTADCC	
g91	Tail sheath protein	Cyanophages of *Myoviridae* family infecting *Microcystis aeruginosa*	SheathRTF: ACATCAGCGTTCGTTTCGG	60
			SheathRTR: CAATCTGGTTAGGTAGGTCG	
psbA	Photosynthesis protein D1	Cyanophages	58-VDIDGIREP-66: GTNGAYATHGAYGGNATHMGNGARCC[a]	112
			331-MHERNAHNFP-340: GGRAARTTRTGNGCRTTNCKYT CRTGCAT[a]	
			psbAF: GTNGAYATHGAYGGNATHMGNGARCC	63
			psbAR: GGRAARTTRTGNGC	
			Pro-psbA-F: AACATCATYTCWGGTGCWGT	67
			Pro-psbA-R: TCGTGCATTACTTCCATACC	
			psbA-93F: TAYCCNATYTGGGAAGC	69
			psbA-341R: TCRAGDGGGAARTTRTG	
psbD	Photosynthesis protein D2	Cyanophages	psbD-26Fa: TTYGTNTTYRTNGGNTGGAGYGG	67
			psbD-26Fb: TTYGTNTTYRTNGGNTGGTCNGG	
			psbD-54Fa: GTNACNAGYTGGTAYACNCAYGG	
			psbD-54Fb: GTNACNTCNTGGTAYACNCAYGG	
			psbD-308Ra: YTCYTGNGANACRAARTCRTANGC	
			psbD-308Rb: YTCYTGRCTNACRAARTCRTANGC	
			psbD-F: GGNTTYATGCTNMGNCARTT	68
			psbD-R: CKRTTNGCNGTVAYCAT	
phoH	Phosphate starvation protein	Cyanophages	vPhoHf: TGCRGGWACAGGTAARACAT	70
			vPhoHr: TCRCCRCAGAAAAYMATTTT	90
			phoH-For: GARATYGGDTTCYTDCCTGG	
			phoH-Rev: ACWARWCCAGADCKWACRATRTC	
polA	DNA polymerase	*Podoviridae*	T7DPol230F: ARGARMRIAAYGGIT	72
			T7DPol510R: GTRTGDATRTCICC	
			HECTORPol19F: GCAAGCAACTTTACTGTGG	
			HECTORPol711R: CGAGAGATACACCAACGAA	
			HECTORPol563F: CTTCTCAGTTTTCTGTT	
			HECTORPol800R: GCAAGCAACTTTACTGT	

(Continued on following page)

TABLE 1 (Continued)

Signature gene	Gene product	Target virus group	Primers (5′–3′)	Reference(s)
			PARISPol25F: ATACTACACGCTACTCTGG	
			PARISPol701R: GAGTGGCAAGAGGAGTTAT	
			PARISPol480F: AAGTTGTGCTTCTGGTA	
			PARISPol786R: ATACTACACGCTACTCT	
			Podo-F: GACACHCTYRTVHTGTCWMGWYTG	74
			Podo-R2: MCKACCRTCYARDCCYTTMAK	
			CP-DNAP-349F: CCAAAYCTYGCMCARGT	75
			CP-DNAP-533Ra: CTCGTCRTGSACRAASGC	
			CP-DNAP-533Rb: CTCGTCRTGDATRAASGC	
			DPOL-341Fd: CCNAAYYTNGSNCARGTNCC	91
			DPOL-534Rd: TGNWRYTCRTCRTGNAYRAA	
			DPOL-349Fd: CCNAAYYTNGSNCARGT	
			DPOL533Rd: TCRTCRTGNAYRAANGC	
g43	DNA polymerase	T4-like members of *Myoviridae*	g43-For: GCWGGTGCWTATGTHAARGAACC	77, 90
			g43-Rev: CCWGASARAGTAATKGCYTCWGC	
polB	DNA polymerase	*Phycodnaviridae*	AVS1: GARGGIGCIACIGTIYTIGAYGC	78
			AVS2: GCIGCRTAICKYTTYTTISWRTA	
			POL: SWRTCIGTRTCICCRTA	
RdRp	RNA-dependent RNA polymerase	RNA viruses	RdRp1: GGRGAYTACASCIRWTTTGAT	87
			RdRp2: MACCCAACKMCKCTTSARRAA	
			Mpl.sc1F: TIGCIGGWGAYTWYARM	89
			Mpl.sc1R: YTCCTTWTCRGSCATKGTA	
			Mpl.sc2F: ITWGCIGGIGATTWCA	
			Mpl.sc2R: CKYTTCARRAAWTCAGCATC	
			Mpl.sc3F: TIATIGMKGGIGAYTA	
			Mpl.sc3R: TTMARGAAIKMAGCATCTT	
			Mpl.cdhF: GMIGGTGAYTAYAGCGCTTWYGAY	
			Mpl.cdhR: ATACCCAATGCCTYTTIARRAA	
Syn9_g101	Putative tail fiber	Cyanophages belonging to *Myoviridae*	Syn9_g101-For: GGTGGTAMATTAACTSTTGATACTG	90
			Syn9_g101-Rev: TCTAGAACCAACAATCTCRAAACC	
cobS	Putative porphyrin biosynthetic protein	Cyanophages belonging to *Myoviridae*	cobS-For: BACYGTWTGGCACAAYGG	90
			cobS-Rev: CTTRGTNTCMTCATCRAARCG	
T7*mcp*	Major capsid protein	T7-like members of *Podoviridae*	MCP-365AF: AARACNHTNGTNATGGAYGA	91
			MCP-1141AR: AYNANRTCNCCYTGRTA	
			MCP-190BF: CARTTYATHTWYACNGG	
			MCP-268BF: CCNCCNGTNGCNGARAARAC	
			MCP-631BR: GCRTARTAYTGNCKNGGRTT	
			MCP-736BR: ATNTKDATNCCNGCDAT	

[a] This primer was originally developed for cyanobacteria, but it has been used for cyanophages in subsequent research.

(20). A quantitative PCR approach to assess the abundance of cyanomyoviruses in this lake yielded 1.3×10^5 to 4.3×10^6 *g20* copies per ml of water over the stations, depths, and seasons, showing a significantly higher viral density in summer (21). Analysis of floodwater and soil of rice paddy fields in Japan revealed that these communities differed significantly from each other, with the soil sequences grouping into more distinct clades whereas the floodwater community contained more sequences similar to marine, freshwater, and isolated cyanophages (22, 23).

In a study that spanned a wide range of environments, including marine (from tropical to polar and over 3,000 m in depth) and freshwater environments (lakes and ponds), highly similar sequences were found in highly variable environments (24).

When investigating the prevalence of the *g20* gene in cyanophages with a confirmed myovirus morphology, it was found that only two-thirds of the population carried the gene, and it was concluded that the full diversity of cyanomyoviruses cannot be explored using this marker (25). In addition, the primer sets described for this marker (Table 1) are specifically targeted to cyanophages, making them relevant only to habitats where cyanobacteria make up the majority of the bacterial population.

Major capsid protein gene (*g23* of T4) in the family *Myoviridae*. The major capsid gene of phage T4, *g23*, was one of the conserved virion genes used to distinguish between T-even-, Schizo-T-even-, Pseudo-T-even-, and Exo-T-even-type phages (26). Based on this phylogenetic analysis, a new degenerate primer set for *g23* (MZIA1bis and MZI6) was developed as an alternative for the *g20* signature gene described above (Table 1) (27). In a wide range of marine habitats, this primer set could identify five new groups of previously unidentified T4-type phages, revealing that some of the groups were present from the Gulf of Mexico to the Arctic Ocean while others were more geographically distinct. Combining all isolated T4-type phages and known environmental *g23* sequences up to that time with data from the Global Ocean

Survey (GOS), 1,399 sequences were aligned, leading to the creation of the Far-T4 phylogenetic group, comprised almost exclusively of uncultured sequences, and the Cyano-T4 group, which includes all the cultured cyanomyoviruses and half of the environmental sequences described by Filée and colleagues in 2005 (28).

Recently, *g23* TRFLP (terminal restriction fragment length polymorphism) analyses have been used to assess temporal variations in marine viral communities. A study conducted over 78 days at one location off the coast of California showed significant fluctuations of the bacterial and T4-type myovirus populations from days to weeks but a more resilient population structure over longer time periods (29). A 3-year study off the California coast (San Pedro Ocean Time-Series [SPOT]) found seasonable variability, with spring/summer and autumn/winter T4-like operational taxonomic units (OTUs) but also persistent OTUs throughout the 3 years, in contrast with the Bank model and the "Killing the Winner" hypotheses, which predict much more dynamic virus-host fluctuations (1, 30–32). At the same site and over the same time period, the interactions between protists, bacteria, and T4-like phages were charted, revealing a complex microbial network in which viruses are thought to be mainly controlled by host availability (33). A similar seasonability (winter/spring, summer, autumn) combined with persistent OTUs throughout the seasons was found in a fjord system in Norway, albeit at a lower diversity (34). Using *g23* amplicon sequencing on samples from the Chesapeake and Delaware Bays, a phylogenetic analysis showed no associations between *g23* polymorphism, phage genome size, sampling time, or location (35).

The *g23* marker gene has also been extensively used to investigate phage diversity in Japanese and Chinese rice paddy field-associated niche environments. Phylogenetic analyses revealed distinct clades associated with paddy soil, rice straw, and paddy flood water (paddy groups I to IX), as well as a small number of sequences similar to environmental marine samples from previous studies (36–40). When comparing manganese nodules in Japanese paddy soil with plow soil and subsoil layers, no differences in phage community structures were found, nor were differences found when comparing soil depth profiles or soils in different regions of Japan or different rice plant parts (40–43). In an investigation of decomposing straw, a reduction in richness in T4-type phage signatures was found in the late stages of decomposition, the opposite of what was observed for the local bacterial population (44). During the decomposition of root callus cells, a proliferation of T4-type phages belonging to a limited number of paddy groups was observed in aerobic soils; no signal was found in anaerobic soils (45).

In a comparison of the *g23* sequences from five different rice field soil types in northeast China, there was only a small difference between the soil types and the phylogenetic analysis showed most sequences clustered with the Japanese paddy groups and one novel China-specific group (46). In several dry upland black soils, 40 to 50% of the *g23* clones grouped with the paddy groups in a phylogenetic analysis, while the rest formed separate groups or clustered separately, indicating a relatedness with rice paddy soils but a distinct community structure, which varied according to sampling location (47, 48). T4-type community structure in flooded paddy field soils and natural wetland soils in northeast China was found to be very diverse, and while the compositions were similar, they were significantly different from paddy soils in Japan and from marine, lake, and upland soil environments (49, 50).

In the freshwater Lake Baikal (Russia), *g23* clones were most closely related to uncultured or cultured (ExoT-even) marine groups, with a few clones clustering with paddy group VII (51). The eutrophic Lake Kokotel, which is situated less than 5 km from the eastern shore of Lake Baikal, comprised a T4-like community that was different from that of Lake Baikal and more closely related to that of Lake Donghu, a lake with a similar trophic status (52, 53). The Antarctic Lake Limnopolar, which was dominated by eukaryotic viruses, still had a wide diversity of *g23* signatures, mainly clustering in the cyano-T4 clade (54).

g23 clone analysis of water and sediment from cryoconite holes in an Arctic glacier (Svalbard, Norway) led to the creation of three unique T4-type clades, polar I to III, containing sequences from only Arctic or Antarctic regions, and three novel environmental clades (Env I to III) with sequences from a variety of habitats (55). An interesting finding in this study was a cluster of sequences in polar group III which shared up to 99% amino acid sequence identity with sequences from Lake Limnopolar (54).

The *g23* signature gene appears to span a much greater diversity of bacteriophage groups within the *Myoviridae* family than the *g20* gene, encompassing both cyanophages and noncyanophages and making it a more desirable marker gene for investigating myovirus diversity. Nevertheless, the prevalence of *g23* in myoviruses has not yet been studied, leaving the question of how much of this family's diversity remains unstudied.

Alternative major capsid protein genes (*mcp*). To investigate a specific group of non-T4-related cyanophages that infect freshwater filamentous cyanobacteria, a primer set was developed that amplified neither the *mcp* gene of T4-related marine cyanophage S-PM2, T4 itself (*g23*), nor that of several phycodnaviruses (56). The use of this novel primer pair revealed a heterologous group of previously uninvestigated cyanomyoviruses which were genetically very different from their marine counterparts that infect unicellular cyanobacteria. This study emphasized the limitation of surveying cyanophage diversity with the *g20* and *g23* marker genes.

The *mcp* gene of the heterogenous group of NCLDVs that infect unicellular eukaryotes was used as an alternative marker to the family B polymerase marker (see below) (57). The genus subdivisions of the *Phycodnaviridae* family were largely supported in phylogenetic analyses of the *mcp* marker gene. However, a close relationship with sequences of the family *Mimiviridae* was observed, suggesting some form of evolutionary relationship. For *Emiliana huxleyi* viruses, sequence analysis of environmental *mcp* amplicons significantly increased the known species richness of this viral group (58).

Very recently, the first single-stranded DNA (ssDNA) virus signature gene study was published, in which the major capsid protein of the subfamily *Gokushovirinae* in the *Microviridae* family was investigated (59). *mcp* signatures were found in all the environments tested (estuarine, freshwater, sediments, sewage) and were only distantly related to the cultured isolates of this subfamily. The findings of this paper suggested that the ssDNA viruses are much more cosmopolitan than previously believed, and when generalizing we can hypothesize that there are many more groups of uninvestigated viruses that may have such a global spread.

Tail sheath protein. Typically, a tail sheath protein is only present in myoviruses, making this a potential marker for this

specific viral family, but a general myovirus tail sheath primer set has not yet been developed. A real-time PCR of the tail sheath protein gene specifically for *Microcystis aeruginosa* Ma-Lmm01-type cyanophages in a freshwater lake allowed successful quantification of the virus particles (60). However, these quantitative values could not be reliably linked to host blooms, potentially because of inactivation of the virus particles. Additionally, a combination of this real-time PCR with a real-time PCR analysis of the host cells gave a negative correlation between phage quantity and host cell number, indicating that cyanomyoviruses are an important factor in *M. aeruginosa* population dynamics (61).

AUXILIARY METABOLISM GENES AS SIGNATURE GENES

Photosynthesis-related genes *psbA* and *psbD* targeting cyanophages. The use of the *psbA-psbD* signature genes originated from the discovery that the cyanophage S-PM2 carries the genes for the core proteins of photosystem II, D1 and D2 (genes *psbA* and *psbD*, respectively), which were later found to be widespread in cyamyoviruses and cyanopodoviruses (62, 63). Using the *psbA* gene, significant differences were found between viral (cultured, environmental, and prophage) and host gene phylogenies, possibly reflecting different evolutionary stresses (64). Phylogenetic analysis of *psbA* in marine and freshwaters revealed that this marker can distinguish between these two environments as well as between *Synechococcus* and *Prochlorococcus* hosts and myoviruses and podoviruses, but not between different geographical locations ranging from the Mediterranean Sea to the Arctic Ocean (65). Analysis of Japanese rice paddy field floodwaters showed novel viral *psbA* signatures (66). In a coastal environment in Hawaii where *Prochlorococcus* dominated, environmental *psbA* sequences clustered with either *Prochlorococcus* podoviruses or myoviruses and with two "unrepresented" clusters that were more distantly related, but not with *Synechococcus* phages, while the *psbD* sequences all grouped into one cluster with the cultured cyanophage P-SSM4 (67). The *psbD* gene sequence was used on a group of isolated phages from the Indian Ocean and revealed no distinct community structure at different depths (68).

The *psbA* marker has been found in isolated myoviruses and podoviruses infecting both *Synechococcus* and *Prochlorococcus* cyanobacterial hosts (67, 69), potentially making it a more complete marker than *g20* for cyanophages. However, the full cyanophage richness remains unexplored, since the signature gene was found neither in siphoviruses nor in the full complement of myo- and podoviruses.

phoH. The *phoH* gene, the expression of which was linked to phosphate starvation conditions but with no identified specific function, has been recently developed as a novel marker gene (70, 71). This gene was found in 40% of cultured marine phages but in only 4% of nonmarine phages, making it a good signature gene for assessment of marine phage diversity. Phylogenetic analysis showed that phages clustered separately from their hosts, that phages infecting heterotrophic bacteria were more diverse than cyanophages, that eukaryotic viruses formed a distinct cluster, and that the phage community composition differed with depth and geographical location (70).

phoH has some advantages as a marker gene compared with some of the other signature genes described in this review. The gene has been found in phages infecting autotrophs and heterotrophs, is not restricted to one phage family, and has also been detected in viruses of photosynthetic green algae (70). Therefore, despite only including about 40% of marine viruses, the use of this single signature gene will produce a more complete picture of marine phage diversity in a chosen habitat than any other single marker gene. On the other hand, caution is needed when using this marker, since several enterobacteriophages also contain a copy of the *phoH* gene, which could lead to an increased signal in habitats contaminated with human or animal waste.

POLYMERASE GENES

T7-like DNA polymerase *polA* in the family *Podoviridae*. The family A DNA polymerase of the T7-like "supergroup," which comprises the genera belonging to the subfamily *Autographivirinae* and several cyanopodophages, contains several conserved regions which can serve as targets for degenerate primers in this subfamily (Table 1) (72, 73). Two partial polymerase sequences, named HECTOR and PARIS, have been found in biomes around the world (marine, freshwater, estuarine, extreme, terrestrial, and metazoan associated), and while PARIS was less abundant than HECTOR, both were more abundant in marine environments than other environments (72). With a different primer set for the DNA polymerase gene (Podo-F/Podo-R2) (Table 1), three environmental groups were found in marine samples clustering separately from the HECTOR and PARIS sequences and from previously cultured phage groups, such as the T7 group, the P60 group, and the SI-01 group (74). In Chesapeake Bay, seasonal changes of the podovirus community were found between winter and summer based on use of a 550- to 600-bp fragment of the *polA* gene (75). Compared with the estuarine Chesapeake Bay community, the open ocean podoviruses were less diverse, yet globally ubiquitous (76). A bioinformatics analysis of the GOS data revealed that those viruses that were present all fell into the preexisting clades (74), suggesting that the currently used primer sets (Table 1) span the full diversity of *polA*-containing podoviruses. However, the metagenomic data suggested that there are many cyanopodovirus groups in which the *polA* gene is not detected (76).

T4-like DNA polymerase *g43* in the family *Myoviridae*. To date, only one study has used the T4-like DNA polymerase gene *g43* for diversity studies. Isolated cyanophages infecting *Synechococcus* sp. WH7803 and seawater viral clone libraries were investigated in a 6-year study (77). Clear spatial (southern New England coast [United States], Long Island coast [United States], and Bermuda inshore waters) and temporal community compositional variations were found, although the overall richness and evenness did not change over the seasons at one location.

The *g43* signature gene targets the same group of phages (cyanomyoviruses) as the *g23* marker, but no comparative studies to evaluate the performance of these signature genes have yet been reported.

Family B DNA polymerase *polB* of NCLDVs. Primers for the family B DNA polymerase gene *polB* were first designed for viruses infecting microalgae and were shown to amplify DNA from environmental viral communities, particularly viruses belonging to the *Phycodnaviridae* family (78, 79). When comparing the *polB* marker sequences to polymerase sequences of other virus groups, it was shown that the algal viruses were related to the *Herpesviridae* and that the *Herpesviridae*, *Poxviridae*, *Baculoviridae*, and African swine fever virus each clustered separately (79). PCR-RFLP analysis of a natural viral community in the Gulf of Mexico revealed five different OTUs which, after sequencing, were all grouped into

the *Phycodnaviridae* family, in which *Micromonas pusilla* viruses seemed the most abundant (80).

PCR-DGGE of a *polB* fragment revealed that the algal virus population in coastal waters was much more stable than the eukaryotic population and, at times, correlated with tide height, salinity, and chlorophyll *a* content (81, 82). Variance in community composition through the seasons was found in the subtropical coastal water of Hawaii, with phylotypes specific for winter or summer and only two phylotypes present throughout the year and at high abundance (83). In freshwater lakes in North America, *polB* phylogenetic analysis showed that the freshwater *Phycodnaviridae* sequences clustered together but were still related to the marine sequences (84). In another freshwater study, no clear temporal and spacial patterns of *Phycodnaviridae* were found but, based on analysis of the monophyletic groups, 13 different freshwater hosts were inferred (85). A comparison of marine and freshwater *polB* sequences, with addition of sequences from Amazon River systems, suggested that there is no significant gene flow of phycodnaviruses between freshwater and oceanic aquatic systems (86).

The AVS primer set does not amplify certain marine *Phycodnaviridae*, such as *Emiliana huxleyi* viruses and *Herterosigma akashiwo* viruses, and is suspected to preferentially amplify *Micromonas pusilla* virus isolates (84, 85), implying that abundance studies are skewed toward the latter and that the full phycodnavirus richness is not captured with these primers.

RNA-dependent RNA polymerase in the order *Picornavirales*. The RNA-dependent RNA polymerase, present in all RNA viruses except retroviruses, was targeted to study the diversity of marine picorna-like viruses (87). This study found that this virus type is widespread and persistent in ocean environments with sequences related to viruses infecting algae, shrimp, and mammals, making this group extremely diverse. The use of this marker gene for investigation of marine RNA virus metagenomes revealed a distinct clade of marine picorna-like (Mpl) virus sequences (88). Extending the environmental range to subtropical waters increased the diversity within the Mpl clade, which was subsequently assumed to be a protist-infecting clade (89).

MULTI-SIGNATURE GENE STUDIES

Based on multilocus sequence typing (MLST) of four core viral genes (*g20*, *g43*, *g23*, and the tail fiber gene of cyanophage Syn9) and four bacteria-derived virus-encoded genes (*psbA*, *psbD*, *phoH*, and *cobS*, a putative porphyrin synthesis protein) in 60 isolated cyanophages of the RIM groups (see above), the *g20* marker-based division in five RIM strains was corroborated (15, 90). This was an unexpected finding, as the genes in the MLST analysis were scattered throughout the genome. The coexistence of this set of genes suggests that recombination between groups is, at best, a rare event and that the cyanophage genome is extremely stable over long time periods. The same observation was made with marine T7-like cyanophages infecting *Synechococcus* and *Prochlorococcus* species, where the phylogenetic trees of the T7-like DNA polymerase and the T7-like major capsid protein resulted in the same two clades, unrelated to host range (91). In this study, *psbA* was also investigated, but it proved less well suited for species diversity studies, since it was found in phages belonging to only one of the two clades. Using two signature genes, *g20* and *psbA*, a collection of over 900 *Synechococcus* phages, collected at three different coastal sites over a 15-month period, was investigated (92). No spatial variations were found, but distinct seasonal communities were observed, with higher abundances in summer, consistent with previous studies described above (15–18, 77).

In a study of the freshwater Bourget Lake and Annecy Lake (France), pulsed-field gel electrophoresis profiling was used with a range of primers to determine the genome size ranges for the different groups of viruses: *mcp* and *polB* for phycodnaviruses, *g23* for myoviruses, and *g20* and *psbA* for cyanophages (93). The *Phycodnaviridae* were the largest, ranging from 79 kb to 486 kb; the myoviruses ranged from 41 kb to 317 kb, and the cyanophages were between 65 kb and 317 kb in size. When the amplicons of the phycodnavirus markers *polB* and *mcp* in these two lakes were sequenced, diversity values were found to depend on the marker used, probably due to differences in the specificities of the primers (94). A comparison of the phycodnavirus community structure with these two marker genes resulted in the differentiation of three distinct environmental types: exclusively marine, exclusively freshwater, and freshwater or glacier water invaded by seawater. This result indicated that water salinity is important in shaping these communities (94). A multimarker PCR-DGGE-based study (using *mcp*, *polB*, *g20*, *g23*, and *psbA*) of Lake Annecy and Lake Bourget over a 1-year period revealed a temporal variability which varied between the marker genes and the two lakes, with the community structure changes linked to several abiotic factors or shifts in host community abundance or structure (95). In nearby Lake Geneva, a similar PCR-DGGE study of multiple signature genes (*mcp*, *polB*, *g20*, *g23*, *psbA*, and *psbD*) revealed seasonal changes in diversity, with a higher richness in the summer months and a correlation between the diversity of *psbD* and picocyanobacterial abundance (96).

SIGNATURE GENES USED IN METAGENOMIC DATA SETS

Nucleotide sequences of some of the marker genes described above have been successfully used in metagenomics analyses to chart viral richness, such as for the very large GOS data set (*g23*, *polB*, *psbA*, and *psbD* [28, 97, 98]) and the Tara Oceans microbial metagenome (99). In the latter study, 16 conserved genes from NCLDVs were chosen from a comparative study to be used as markers in the metagenomics data set (99, 100). Mining the GOS data set for PolB sequences resulted in a set of almost 2,000 sequences ranging in length from 25 to 562 amino acid residues (98). With these sequences, a phylogenetic analysis identified eight viral groups: mimivirus-related algal viruses, chloroviruses, herpesviruses, baculoviruses, poxviruses, irodoviruses/ascoviruses, phaeoviruses, and a phage group. The latter group contained the majority of the GOS PolB sequences, consistent with the observation that phages are the dominant entity in marine environments.

Other viral markers are currently being used in metagenomics, such as the terminase large subunit gene *terL*, present in all members of the *Caudovirales* and for which the phylogeny can indicate the packaging mechanism used by the phages (101–103), and VP1, the capsid protein of ssDNA microphages belonging to the *Microviridae* family (101, 104, 105).

Currently, one metagenomics pipeline, MetaVir, offers the possibility of automated computing of phylogenetic trees based on the signature genes discussed in this minireview (101, 106). As an example, we have generated a *g23* tree that combines several viral metagenomics data sets from freshwater environments (Lake Pavin, France [102] and El Berbera Saharan pond [107]) and ma-

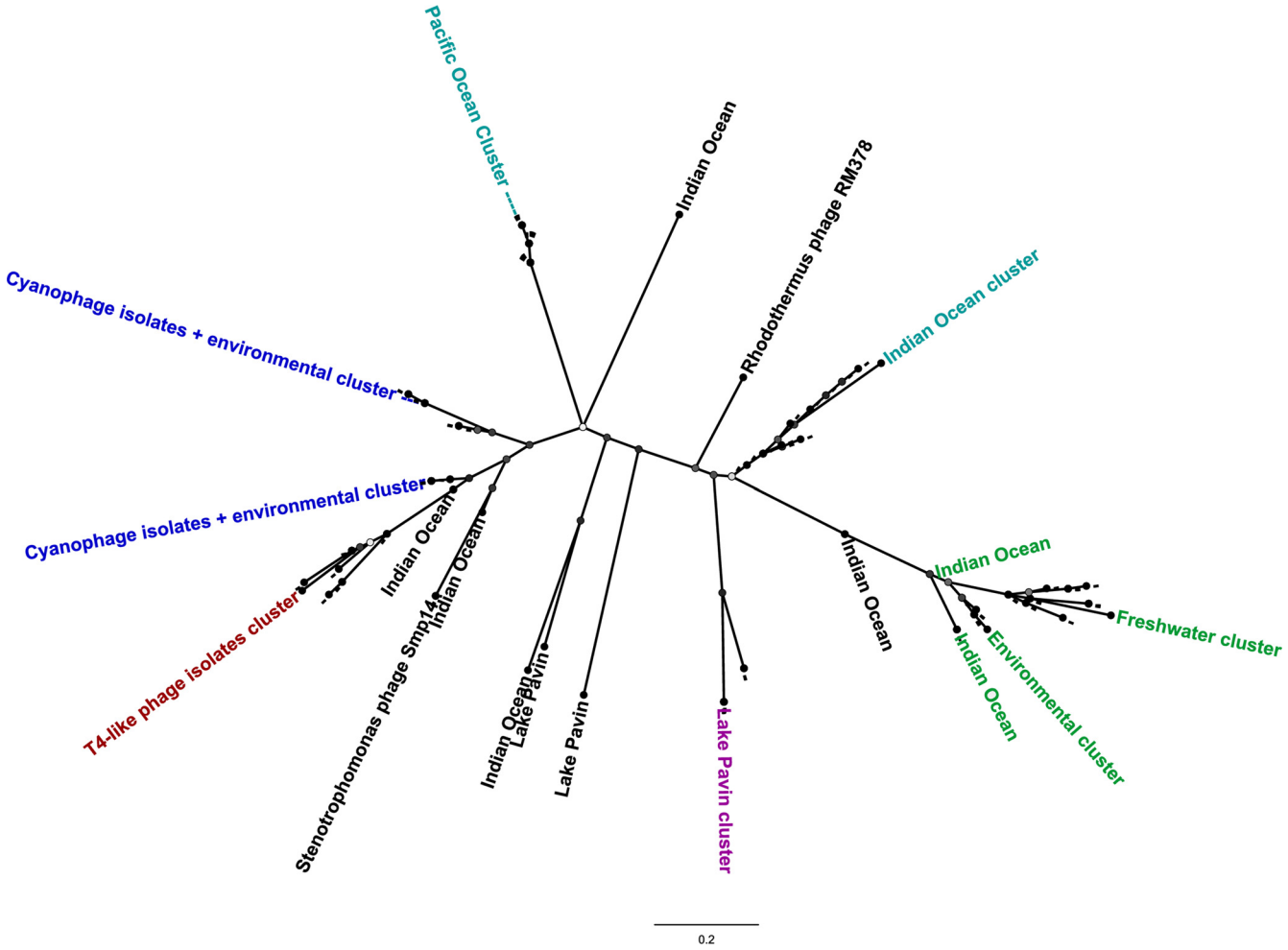

FIG 1 Phylogenetic tree (PhyML) of *g23* marker sequences generated by MetaVir. Metagenomic sequence reads (Lake Pavin, El Berbera, Indian Ocean, and Pacific Ocean) were aligned with reference marker sequences from selected isolated T4-like phages by using MUSCLE (at least 35 bp; 98% identity). The tree was visualized using FigTree v1.4.0 and rooted at *Rhodothermus* phage RM378. Bootstrap support (on 100 bootstraps) of the nodes is indicated by the circles on each node, from white (low support) to black (high support). The scale bar represents the number of substitutions per site. Distinct clades are colored according to their origin: green, environmental sequences from all investigated environments; light blue, marine; purple, freshwater; dark blue, sequences clustering with isolated cyanophages; brown, isolated non-cyano-T4-like phage signatures.

rine environments (M6O1K, Indian Ocean [108] and LJ12S and LJ26S, Pacific Ocean [109]) with isolated T4-like phages (Fig. 1). In accordance with the studies described above, the environmental sequences were much more diverse than the sequences that belonged to isolated phages. The cyanophage isolates clustered with several environmental sequences from freshwater and marine data sets, while the non-cyanophage T4-like isolates clustered separately. We could also identify several clusters according to geographical origin: two marine clusters (Indian Ocean cluster and Pacific Ocean cluster) and one freshwater cluster (Lake Pavin cluster). Many of the metagenomics sequences formed a separate, diverse environmental clade comprising sequences from all environments (Fig. 1, shown in green).

DISCUSSION

To date, general ecological observations can only be made for the well-studied viral group *Myoviridae*, while for other viral groups diversity explorations are still in the charting phase. For marine myoviruses, irrespective of the marker used, temporal changes in community structure have been found at each of the sites investigated and are most likely linked to host density fluctuations. A similar trend has been observed in freshwater habitats. Geographical patterns, on the other hand, are more difficult to generalize due to the presence of both universal and geographically distinct myovirus groups. As more metagenomic data sets become available, it is possible that distinct biogeographical patterns will emerge.

Considering the observation that a community of related viruses is apparently spread globally in marine environments and taking into account all marker genes which have been investigated on a global scale, we propose to amend the hypothesis of "everything is everywhere" to "every virus group is everywhere," which to date appears to be true for the myoviruses, podoviruses, gokushoviruses, phycodnaviruses, and picornaviruses. We can speculate that for any given host, all virus groups are present as a predator, in constant competition with each other and using different infection strategies.

None of the signature gene primer sets described have suc-

ceeded in capturing the full diversity of viruses in a single environment. Of the classified viral taxa (ICTV), only two orders have been investigated in any detail: the *Picornavirales* and *Caudovirales*, comprising eight families, and three unassigned families, the *Microviridae* family and the *Phycodnaviridae* and *Mimiviridae*, leaving the vast majority of viruses largely uninvestigated. The use of a combination of existing markers can only marginally solve this problem, by providing better coverage of different virus families or by including more families in one study, but this will still fail to fully assess within-family richness. Moreover, metagenomics studies have shown that the majority of viral sequences found in any habitat have no database counterpart, suggesting that PCR-based studies will continue to provide inadequate coverage of true viral diversity.

The most obviously underinvestigated group in signature gene analyses is the *Siphoviridae* family, a taxon that dominates both cultured isolates and metagenomic studies. However, newly available metagenomic data sets can be used to develop new degenerate primers for existing signature genes to increase coverage, or to develop primer sets for new signature genes for siphoviruses or other less-studied viral groups. For the latter, the program PhiSiGns was specifically developed to find signature genes in environmental phage data sets and design the appropriate PCR primers (110), and it has been implemented for the gokushovirus *mcp* primer set (59).

The use of viral signature genes is a relatively easy and cheap method to assess viral diversity (compared with viral shotgun metagenomics), especially when there are large numbers of samples to be processed. All these "diagnostic" techniques rely on amplification with specific viral primers, after which the amplicon fingerprints are compared using appropriate methods (DGGE or [T]RFLP) or are sequenced (clone library sequencing or amplicon sequencing). The former is less time-consuming and cheaper but only generates comparative intergroup information, while the latter is more time-consuming for clone library sequencing and more expensive but provides qualitative and taxonomic data, leading to greater insights. Currently, PCR-based studies remain the best option when dealing with multiple samples, locations, or time points, and the choice of signature gene(s) will continue to depend on the focus of the study. The use of viral whole-genome shotgun metagenomics is potentially a much more effective method for assessing the full diversity (and ecology) of viruses in any environment, and we project that, in a not-so-distant future in which next-generation sequencing costs continue to fall, this will become the gold standard for viral diversity studies.

ACKNOWLEDGMENTS

We thank the Genomics Research Institute of the University of Pretoria for financial support. E.M.A. currently holds a Vice-Chancellor's Postdoctoral Fellowship of the University of Pretoria.

REFERENCES

1. **Breitbart M, Rohwer F.** 2005. Here a virus, there a virus, everywhere the same virus? Trends Microbiol. **13:**278–284. http://dx.doi.org/10.1016/j.tim.2005.04.003.
2. **Weinbauer MG.** 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. **28:**127–181. http://dx.doi.org/10.1016/j.femsre.2003.08.001.
3. **Hambly E, Suttle CA.** 2005. The viriosphere, diversity, and genetic exchange within phage communities. Curr. Opin. Microbiol. **8:**444–450. http://dx.doi.org/10.1016/j.mib.2005.06.005.
4. **Wilhelm SW, Suttle CA.** 1999. Viruses and nutrient cycles in the sea. Bioscience **49:**781–788. http://dx.doi.org/10.2307/1313569.
5. **Thurber RV.** 2009. Current insights into phage biodiversity and biogeography. Curr. Opin. Microbiol. **12:**582–587. http://dx.doi.org/10.1016/j.mib.2009.08.008.
6. **Ackermann H-W.** 2007. 5500 phages examined in the electron microscope. Arch. Virol. **152:**227–243. http://dx.doi.org/10.1007/s00705-006-0849-1.
7. **Koonin EV, Yutin N.** 2010. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology **53:**284–292. http://dx.doi.org/10.1159/000312913.
8. **Rohwer F, Edwards R.** 2002. The Phage Proteomic Tree: a genome-based taxonomy for phage. J. Bacteriol. **184:**4529–4535. http://dx.doi.org/10.1128/JB.184.16.4529-4535.2002.
9. **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.
10. **Wilson WH, Fuller NJ, Joint IR, Mann NH.** 1999. Analysis of cyanophage diversity and population structure in a south-north transect of the Atlantic ocean. Bull. Inst. Océanogr. **19:**209–216.
11. **Wilson W, Fuller N, Joint IR, Mann NH.** 2000. Analysis of cyanophage diversity in the marine environment using denaturing gradient gel electrophoresis, p 565–570. *In* Bell C, Brylinsky M, Johnson-Green P (ed), Proceedings of the 8th International Symposium on Microbial Ecology, Halifax, Canada. International Symposium on Microbial Ecology, Malvern, Victoria, Australia.
12. **Frederickson CM, Short SM, Suttle CA.** 2003. The physical environment affects cyanophage communities in British Columbia inlets. Microb. Ecol. **46:**348–357. http://dx.doi.org/10.1007/s00248-003-1010-2.
13. **Jameson E, Mann NH, Joint I, Sambles C, Mühling M.** 2011. The diversity of cyanomyovirus populations along a North-South Atlantic Ocean transect. ISME J. **5:**1713–1721. http://dx.doi.org/10.1038/ismej.2011.54.
14. **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. http://dx.doi.org/10.1128/AEM.68.4.1576-1584.2002.
15. **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. http://dx.doi.org/10.1128/AEM.69.8.4639-4647.2003.
16. **Wang K, Chen F.** 2004. Genetic diversity and population dynamics of cyanophage communities in the Chesapeake Bay. Aquat. Microb. Ecol. **34:**105–116. http://dx.doi.org/10.3354/ame034105.
17. **Sandaa R-A, 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. http://dx.doi.org/10.1128/AEM.00168-06.
18. **Mühling M, Fuller NJ, Millard A, Somerfield PJ, Marie D, Wilson WH, Scanlan DJ, Post AF, Joint I, Mann NH.** 2005. Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: evidence for viral control of phytoplankton. Environ. Microbiol. **7:**499–508. http://dx.doi.org/10.1111/j.1462-2920.2005.00713.x.
19. **Dorigo U, Jacquet S, Humbert J-F.** 2004. Cyanophage diversity, inferred from g20 gene analyses, in the largest natural lake in France, Lake Bourget. Appl. Environ. Microbiol. **70:**1017–1022. http://dx.doi.org/10.1128/AEM.70.2.1017-1022.2004.
20. **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. http://dx.doi.org/10.1128/AEM.00349-06.
21. **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. http://dx.doi.org/10.1128/AEM.05879-11.
22. **Wang G, Murase J, Asakawa S, Kimura M.** 2010. Unique viral capsid assembly protein gene (g20) of cyanophages in the floodwater of a Japanese paddy field. Biol. Fertil. Soils **46:**93–102. http://dx.doi.org/10.1007/s00374-009-0410-y.
23. **Wang G, Asakawa S, Kimura M.** 2011. Spatial and temporal changes of cyanophage communities in paddy field soils as revealed by the capsid

assembly protein gene g20. FEMS Microbiol. Ecol. **76:**352–359. http://dx.doi.org/10.1111/j.1574-6941.2011.01052.x.
24. **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. http://dx.doi.org/10.1128/AEM.71.1.480-486.2005.
25. **McDaniel L, DelaRosa M, Paul JH.** 2006. Temperate and lytic cyanophages from the Gulf of Mexico. J. Mar Biol. Assoc. UK **86:**517–527. http://dx.doi.org/10.1017/S0025315406013427.
26. **Tétart F, Desplats C, Kutateladze M.** 2001. Phylogeny of the major head and tail genes of the wide-ranging T4-type bacteriophages. J. Bacteriol. **183:**358–366. http://dx.doi.org/10.1128/JB.183.1.358-366.2001.
27. **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. U. S. A. **102:**12471–12476. http://dx.doi.org/10.1073/pnas.0503404102.
28. **Comeau AM, Krisch HM.** 2008. The capsid of the T4 phage superfamily: the evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. Mol. Biol. Evol. **25:**1321–1332. http://dx.doi.org/10.1093/molbev/msn080.
29. **Needham DM, Chow C-ET, 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. http://dx.doi.org/10.1038/ismej.2013.19.
30. **Chow C-ET, Fuhrman JA.** 2012. Seasonality and monthly dynamics of marine myovirus communities. Environ. Microbiol. **14:**2171–2183. http://dx.doi.org/10.1111/j.1462-2920.2012.02744.x.
31. **Thingstad T, Lignell R.** 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. Aquat. Microb. Ecol. **13:**19–27. http://dx.doi.org/10.3354/ame013019.
32. **Breitbart M.** 2012. Marine viruses: truth or dare. Annu. Rev. Mar. Sci. **4:**425–448. http://dx.doi.org/10.1146/annurev-marine-120709-142805.
33. **Chow C-ET, Kim DY, Sachdeva R, Caron DA, Fuhrman JA.** 2013. Top-down controls on bacterial community structure: microbial network analysis of bacteria, T4-like viruses and protists. ISME J. http://dx.doi.org/10.1038/ismej.2013.199.
34. **Pagarete A, Chow C-ET, Johannessen T, Fuhrman JA, Thingstad TF, Sandaa RA.** 2013. Strong seasonality and interannual recurrence in marine myovirus communities. Appl. Environ. Microbiol. **79:**6253–6259. http://dx.doi.org/10.1128/AEM.01075-13.
35. **Jamindar S, Polson SW, Srinivasiah S, Waidner L, Wommack KE.** 2012. Evaluation of two approaches for assessing the genetic similarity of virioplankton populations as defined by genome size. Appl. Environ. Microbiol. **78:**8773–8783. http://dx.doi.org/10.1128/AEM.02432-12.
36. **Jia Z, Ishihara R, Nakajima Y, Asakawa S, Kimura M.** 2007. Molecular characterization of T4-type bacteriophages in a rice field. Environ. Microbiol. **9:**1091–1096. http://dx.doi.org/10.1111/j.1462-2920.2006.01207.x.
37. **Fujii T, Nakayama N, Nishida M, Sekiya H, Kato N, Asakawa S, Kimura M.** 2008. Novel capsid genes (g23) of T4-type bacteriophages in a Japanese paddy field. Soil Biol. Biochem. **40:**1049–1058. http://dx.doi.org/10.1016/j.soilbio.2007.11.025.
38. **Nakayama N, Tsuge T, Asakawa S, Kimura M.** 2009. Morphology, host range and phylogenetic diversity of *Sphingomonas* phages in the floodwater of a Japanese paddy field. Soil Sci. Plant. Nutr. **55:**53–64. http://dx.doi.org/10.1111/j.1747-0765.2008.00332.x.
39. **Nakayama N, Asakawa S, Kimura M.** 2009. Comparison of g23 gene sequence diversity between *Novosphingobium* and *Sphingomonas* phages and phage communities in the floodwater of a Japanese paddy field. Soil Biol. Biochem. **41:**179–185. http://dx.doi.org/10.1016/j.soilbio.2008.06.008.
40. **Wang G, Murase J, Taki K, Ohashi Y, Yoshikawa N, Asakawa S, Kimura M.** 2009. Changes in major capsid genes (g23) of T4-type bacteriophages with soil depth in two Japanese rice fields. Biol. Fertil. Soils **45:**521–529. http://dx.doi.org/10.1007/s00374-009-0362-2.
41. **Cahyani VR, Murase J, Ishibashi E, Asakawa S, Kimura M.** 2009. T4-type bacteriophage communities estimated from the major capsid genes (g23) in manganese nodules in Japanese paddy fields. Soil Sci. Plant Nutr. **55:**264–270. http://dx.doi.org/10.1111/j.1747-0765.2009.00363.x.
42. **Wang G, Hayashi M, Saito M, Tsuchiya K, Asakawa S, Kimura M.** 2009. Survey of major capsid genes (g23) of T4-type bacteriophages in Japanese paddy field soils. Soil Biol. Biochem. **41:**13–20. http://dx.doi.org/10.1016/j.soilbio.2008.07.008.
43. **Fujihara S, Murase J, Tun CC, Matsuyama T, Ikenaga M, Asakawa S, Kimura M.** 2010. Low diversity of T4-type bacteriophages in applied rice straw, plant residues and rice roots in Japanese rice soils: Estimation from major capsid gene (g23) composition. Soil Sci. Plant Nutr. **56:**800–812. http://dx.doi.org/10.1111/j.1747-0765.2010.00513.x.
44. **Cahyani VR, Murase J, Asakawa S, Kimura M.** 2009. Change in T4-type bacteriophage communities during the composting process of rice straw: Estimation from the major capsid gene (g23) sequences. Soil Sci. Plant Nutr. **55:**468–477. http://dx.doi.org/10.1111/j.1747-0765.2009.00391.x.
45. **Li Y, Watanabe T, Murase J, Asakawa S, Kimura M.** 2013. Identification of the major capsid gene (g23) of T4-type bacteriophages that assimilate substrates from root cap cells under aerobic and anaerobic soil conditions using a DNA-SIP approach. Soil Biol. Biochem. **63:**97–105. http://dx.doi.org/10.1016/j.soilbio.2013.03.026.
46. **Wang G, Jin J, Asakawa S, Kimura M.** 2009. Survey of major capsid genes (g23) of T4-type bacteriophages in rice fields in northeast China. Soil Biol. Biochem. **41:**423–427. http://dx.doi.org/10.1016/j.soilbio.2008.11.012.
47. **Wang G, Yu Z, Liu J, Jin J, Liu X, Kimura M.** 2011. Molecular analysis of the major capsid genes (g23) of T4-type bacteriophages in an upland black soil in northeast China. Biol. Fertil. Soils **47:**273–282. http://dx.doi.org/10.1007/s00374-010-0533-1.
48. **Liu J, Wang G, Zheng C, Yuan X, Jin J, Liu X.** 2011. Specific assemblages of major capsid genes (g23) of T4-type bacteriophages isolated from upland black soils in northeast China. Soil Biol. Biochem. **43:**1980–1984. http://dx.doi.org/10.1016/j.soilbio.2011.05.005.
49. **Liu J, Wang G, Wang Q, Liu J, Jin J, Liu X.** 2012. Phylogenetic diversity and assemblage of major capsid genes (g23) of T4-type bacteriophages in paddy field soils during rice growth season in northeast China. Soil Sci. Plant Nutr. **58:**435–444. http://dx.doi.org/10.1080/00380768.2012.703610.
50. **Zheng C, Wang G, Liu J, Song C, Gao H, Liu X.** 2013. Characterization of the major capsid genes (g23) of T4-type bacteriophages in the wetlands of northeast China. Microb. Ecol. **65:**616–625. http://dx.doi.org/10.1007/s00248-012-0158-z.
51. **Butina TV, Belykh OI, Maksimenko SY, Belikov SI.** 2010. Phylogenetic diversity of T4-like bacteriophages in Lake Baikal, East Siberia. FEMS Microbiol. Lett. **309:**122–129. http://dx.doi.org/10.1111/j.1574-6968.2010.02025.x.
52. **Yan-Ming L, Xiu-Ping Y, Qi-Ya Z.** 2006. Spatial distribution and morphologic diversity of virioplankton in Lake Donghu, China. Acta Oecol. **29:**328–334. http://dx.doi.org/10.1016/j.actao.2005.12.002.
53. **Butina TV, Belykh OI, Potapov SA, Sorokovikova EG.** 2013. Diversity of the major capsid genes (g23) of T4-like bacteriophages in the eutrophic Lake Kotokel in East Siberia, Russia. Arch. Microbiol. **195:**513–520. http://dx.doi.org/10.1007/s00203-013-0884-8.
54. **López-Bueno A, Tamames J, Velázquez D, Moya A, Quesada A, Alcamí A.** 2009. High diversity of the viral community from an Antarctic lake. Science **326:**858–861. http://dx.doi.org/10.1126/science.1179287.
55. **Bellas CM, Anesio AM.** 2013. High diversity and potential origins of T4-type bacteriophages on the surface of Arctic glaciers. Extremophiles **17:**861–870. http://dx.doi.org/10.1007/s00792-013-0569-x.
56. **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. http://dx.doi.org/10.1128/AEM.00270-06.
57. **Larsen JB, Larsen A, Bratbak G, Sandaa R-A.** 2008. Phylogenetic analysis of members of the *Phycodnaviridae* virus family, using amplified fragments of the major capsid protein gene. Appl. Environ. Microbiol. **74:**3048–3057. http://dx.doi.org/10.1128/AEM.02548-07.
58. **Rowe JM, Fabre M-F, Gobena D, Wilson WH, Wilhelm SW.** 2011. Application of the major capsid protein as a marker of the phylogenetic diversity of *Emiliania huxleyi* viruses. FEMS Microbiol. Ecol. **76:**373–380. http://dx.doi.org/10.1111/j.1574-6941.2011.01055.x.
59. **Hopkins M, Kailasan S, Cohen A, Roux S, Tucker KP, Shevenell A, Agbandje-McKenna M, Breitbart M.** 3 April 2014. Diversity of environmental single-stranded DNA phages revealed by PCR amplification of the partial major capsid protein. ISME J. http://dx.doi.org/10.1038/ismej.2014.43.
60. **Takashima Y, Yoshida T, Yoshida M, Shirai Y, Tomaru Y, Takao Y, Hiroishi S, Nagasaki K.** 2007. Development and application of quantitative detection of cyanophages phylogenetically related to cyanophage Ma-LMM01 infecting *Microcystis aeruginosa* in fresh water. Microbes Environ. **22:**207–213. http://dx.doi.org/10.1264/jsme2.22.207.

61. **Yoshida M, Yoshida T, Kashima A, Takashima Y, Hosoda N, Nagasaki K, Hiroishi S.** 2008. Ecological dynamics of the toxic bloom-forming cyanobacterium *Microcystis aeruginosa* and its cyanophages in freshwater. Appl. Environ. Microbiol. **74:**3269–3273. http://dx.doi.org/10.1128/AEM.02240-07.
62. **Mann NH, Cook A, Millard A, Bailey S, Clokie M.** 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature **424:**741. http://dx.doi.org/10.1038/424741a.
63. **Millard A, Clokie MRJ, Shub DA, Mann NH.** 2004. Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. Proc. Natl. Acad. Sci. U. S. A. **101:**11007–11012. http://dx.doi.org/10.1073/pnas.0401478101.
64. **Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G, Béjà O.** 2005. Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. Environ. Microbiol. **7:**1505–1513. http://dx.doi.org/10.1111/j.1462-2920.2005.00833.x.
65. **Chénard C, Suttle CA.** 2008. Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. Appl. Environ. Microbiol. **74:**5317–5324. http://dx.doi.org/10.1128/AEM.02480-07.
66. **Wang G, Murase J, Asakawa S, Kimura M.** 2009. Novel cyanophage photosynthetic gene psbA in the floodwater of a Japanese rice field. FEMS Microbiol. Ecol. **70:**79–86. http://dx.doi.org/10.1111/j.1574-6941.2009.00743.x.
67. **Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW.** 2006. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol. **4:**e234. http://dx.doi.org/10.1371/journal.pbio.0040234.
68. **Clokie MRJ, Millard AD, Mehta JY, Mann NH.** 2006. Virus isolation studies suggest short-term variations in abundance in natural cyanophage populations of the Indian Ocean. J. Mar. Biol. Assoc. UK **86:**499–505. http://dx.doi.org/10.1017/S0025315406013403.
69. **Wang K, Chen F.** 2008. Prevalence of highly host-specific cyanophages in the estuarine environment. Environ. Microbiol. **10:**300–312. http://dx.doi.org/10.1111/j.1462-2920.2007.01452.x.
70. **Goldsmith DB, Crosti G, Dwivedi B, McDaniel LD, Varsani A, Suttle C a, Weinbauer MG, SandaaR-A, Breitbart M.** 2011. Development of phoH as a novel signature gene for assessing marine phage diversity. Appl. Environ. Microbiol. **77:**7730–7739. http://dx.doi.org/10.1128/AEM.05531-11.
71. **Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, Chisholm SW.** 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ. Microbiol. **12:**3035–3056. http://dx.doi.org/10.1111/j.1462-2920.2010.02280.x.
72. **Breitbart M, Miyake JH, Rohwer F.** 2004. Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol. Lett. **236:**249–256. http://dx.doi.org/10.1111/j.1574-6968.2004.tb09654.x.
73. **Lavigne R, Seto D, Mahadevan P, Ackermann H-W, Kropinski AM.** 2008. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. Res. Microbiol. **159:**406–414. http://dx.doi.org/10.1016/j.resmic.2008.03.005.
74. **Labonté JM, Reid KE, Suttle CA.** 2009. Phylogenetic analysis indicates evolutionary diversity and environmental segregation of marine podovirus DNA polymerase gene sequences. Appl. Environ. Microbiol. **75:**3634–3640. http://dx.doi.org/10.1128/AEM.02317-08.
75. **Chen F, Wang K, Huang S, Cai H, Zhao M, Jiao N, Wommack KE.** 2009. Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. Environ. Microbiol. **11:**2884–2892. http://dx.doi.org/10.1111/j.1462-2920.2009.02033.x.
76. **Huang S, Wilhelm SW, Jiao N, Chen F.** 2010. Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. ISME J. **4:**1243–1251. http://dx.doi.org/10.1038/ismej.2010.56.
77. **Marston MF, Taylor S, Sme N, Parsons RJ, Noyes TJE, Martiny JBH.** 2013. Marine cyanophages exhibit local and regional biogeography. Environ. Microbiol. **15:**1452–1463. http://dx.doi.org/10.1111/1462-2920.12062.
78. **Chen F, Suttle C.** 1995. Amplification of DNA polymerase gene fragments from viruses infecting microalgae. Appl. Environ. Microbiol. **61:**1274–1278.
79. **Chen F, Suttle CA.** 1996. Evolutionary relationships among large double-stranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. Virology **219:**170–178. http://dx.doi.org/10.1006/viro.1996.0234.
80. **Chen F, Suttle C, Short S.** 1996. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. **62:**2869–2874.
81. **Short SM, Suttle CA.** 2002. Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. Appl. Environ. Microbiol. **68:**1290–1296. http://dx.doi.org/10.1128/AEM.68.3.1290-1296.2002.
82. **Short SM, Suttle CA.** 2003. Temporal dynamics of natural communities of marine algal viruses and eukaryotes. Aquat. Microb. Ecol. **32:**107–119. http://dx.doi.org/10.3354/ame032107.
83. **Culley AI, Asuncion BF, Steward GF.** 2009. Detection of inteins among diverse DNA polymerase genes of uncultivated members of the *Phycodnaviridae*. ISME J. **3:**409–418. http://dx.doi.org/10.1038/ismej.2008.120.
84. **Short SM, Short CM.** 2008. Diversity of algal viruses in various North American freshwater environments. Aquat. Microb. Ecol. **51:**13–21. http://dx.doi.org/10.3354/ame01183.
85. **Clasen JL, Suttle CA.** 2009. Identification of freshwater *Phycodnaviridae* and their potential phytoplankton hosts, using DNA *pol* sequence fragments and a genetic-distance analysis. Appl. Environ. Microbiol. **75:**991–997. http://dx.doi.org/10.1128/AEM.02024-08.
86. **Gimenes MV, Zanotto PMDA, Suttle CA, da Cunha HB, Mehnert DU.** 2012. Phylodynamics and movement of phycodnaviruses among aquatic environments. ISME J. **6:**237–247. http://dx.doi.org/10.1038/ismej.2011.93.
87. **Culley AI, Lang AS, Suttle CA.** 2003. High diversity of unknown picorna-like viruses in the sea. Nature **424:**1054–1057. http://dx.doi.org/10.1038/nature01886.
88. **Culley AI, Lang AS, Suttle CA.** 2006. Metagenomic analysis of coastal RNA virus communities. Science **312:**1795–1798. http://dx.doi.org/10.1126/science.1127404.
89. **Culley AI, Steward GF.** 2007. New genera of RNA viruses in subtropical seawater, inferred from polymerase gene sequences. Appl. Environ. Microbiol. **73:**5937–5344. http://dx.doi.org/10.1128/AEM.01065-07.
90. **Marston MF, Amrich CG.** 2009. Recombination and microdiversity in coastal marine cyanophages. Environ. Microbiol. **11:**2893–2903. http://dx.doi.org/10.1111/j.1462-2920.2009.02037.x.
91. **Dekel-Bird NP, Avrani S, Sabehi G, Pekarsky I, Marston MF, Kirzner S, Lindell D.** 2013. Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. Environ. Microbiol. **15:**1476–1491. http://dx.doi.org/10.1111/1462-2920.12103.
92. **Clasen JL, Hanson CA, Ibrahim Y, Weihe C, Marston MF, Martiny.** 2013. Diversity and temporal dynamics of Southern California coastal marine cyanophage isolates. Aquat. Microb. Ecol. **69:**17–31. http://dx.doi.org/10.3354/ame01613.
93. **Zhong X, Pradeep Ram AS, Colombet J, Jacquet S.** 2014. Variations in abundance, genome size, morphology, and functional role of the viriopplankton in Lakes Annecy and Bourget over a 1-year period. Microb. Ecol. **67:**66–82. http://dx.doi.org/10.1007/s00248-013-0320-2.
94. **Zhong X, Jacquet S.** 2014. Contrasting diversity of phycodnavirus signature genes in two large and deep western European lakes. Environ. Microbiol. **16:**759–773. http://dx.doi.org/10.1111/1462-2920.12201.
95. **Zhong X, Rimet F, Jacquet S.** 2014. Seasonal variations in PCR-DGGE fingerprinted viruses infecting phytoplankton in large and deep perialpine lakes. Ecol. Res. **29:**271–287. http://dx.doi.org/10.1007/s11284-013-1121-2.
96. **Parvathi A, Zhong X, Jacquet S.** 2012. Dynamics of various viral groups infecting autotrophic plankton in Lake Geneva. Adv. Oceanogr. Limnol. **3:**171–191. http://dx.doi.org/10.1080/19475721.2012.738157.
97. **Williamson SJ, Rusch DB, Yooseph S, Halpern AL, Heidelberg KB, Glass JI, Andrews-Pfannkoch C, Fadrosh D, Miller CS, Sutton G, Frazier M, Venter JC.** 2008. The Sorcerer II Global Ocean Sampling Expedition: metagenomic characterization of viruses within aquatic microbial samples. PLoS One **3:**e1456. http://dx.doi.org/10.1371/journal.pone.0001456.
98. **Monier A, Claverie J-M, Ogata H.** 2008. Taxonomic distribution of large DNA viruses in the sea. Genome Biol. **9:**R106. http://dx.doi.org/10.1186/gb-2008-9-7-r106.

99. **Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L, Poulain J, Ferrera I, Sarmento H, Villar E, Lima-Mendez G, Faust K, Sunagawa S, Claverie J-M, Moreau H, Desdevises Y, Bork P, Raes J, de Vargas C, Karsenti E, Kandels-Lewis S, Jaillon O, Not F, Pesant S, Wincker P, Ogata H.** 2013. Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. ISME J. **7:**1678–1695. http://dx.doi.org/10.1038/ismej.2013.59.
100. **Yutin N, Wolf YI, Raoult D, Koonin EV.** 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. **6:**223. http://dx.doi.org/10.1186/1743-422X-6-223.
101. **Roux S, Faubladier M, Mahul A, Paulhe N, Bernard A, Debroas D, Enault F.** 2011. Metavir: a web server dedicated to virome analysis. Bioinformatics **27:**3074–3075. http://dx.doi.org/10.1093/bioinformatics/btr519.
102. **Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S, Colombet J, Sime-Ngando T, Debroas D.** 2012. Assessing the diversity and specificity of two freshwater viral communities through metagenomics. PLoS One **7:**e33641. http://dx.doi.org/10.1371/journal.pone.0033641.
103. **Sullivan MB, Krastins B, Hughes JL, Kelly L, Chase M, Sarracino D, Chisholm SW.** 2009. The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial "mobilome." Environ. Microbiol. **11:**2935–2951. http://dx.doi.org/10.1111/j.1462-2920.2009.02081.x.
104. **Roux S, Krupovic M, Poulet A, Debroas D, Enault F.** 2012. Evolution and diversity of the *Microviridae* viral family through a collection of 81 new complete genomes assembled from virome reads. PLoS One **7:**e40418. http://dx.doi.org/10.1371/journal.pone.0040418.
105. **Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M, Liu H, Furlan M, Wegley L, Chau B, Ruan Y, Hall D, Angly FE, Edwards RA, Li L, Thurber RV, Reid RP, Siefert J, Souza V, Valentine DL, Swan BK, Breitbart M, Rohwer F.** 2008. Biodiversity and biogeography of phages in modern stromatolites and thrombolites. Nature **452:**340–343. http://dx.doi.org/10.1038/nature06735.
106. **Roux S, Tournayre J, Mahul A, Debroas D, Enault F.** 2014. Metavir 2: new tools for viral metagenome comparison and assembled virome analysis. BMC Bioinformatics **15:**76. http://dx.doi.org/10.1186/1471-2105-15-76.
107. **Fancello L, Trape S, Robert C, Boyer M, Popgeorgiev N, Raoult D, Desnues C.** 2013. Viruses in the desert: a metagenomic survey of viral communities in four perennial ponds of the Mauritanian Sahara. ISME J. **7:**359–369. http://dx.doi.org/10.1038/ismej.2012.101.
108. **Williamson SJ, Allen LZ, Lorenzi HA, Fadrosh DW, Brami D, Thiagarajan M, McCrow JP, Tovchigrechko A, Yooseph S, Venter JC.** 2012. Metagenomic exploration of viruses throughout the Indian Ocean. PLoS One **7:**e42047. http://dx.doi.org/10.1371/journal.pone.0042047.
109. **Hurwitz BL, Sullivan MB.** 2013. The Pacific Ocean Virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS One **8:**e57355. http://dx.doi.org/10.1371/journal.pone.0057355.
110. **Dwivedi B, Schmieder R, Goldsmith DB, Edwards RA, Breitbart M.** 2012. PhiSiGns: an online tool to identify signature genes in phages and design PCR primers for examining phage diversity. BMC Bioinformatics **13:**37. http://dx.doi.org/10.1186/1471-2105-13-37.
111. **Schroeder DC, Oke J, Malin G, Wilson WH.** 2002. Coccolithovirus (*Phycodnaviridae*): characterisation of a new large dsDNA algal virus that infects *Emiliana huxleyi*. Arch. Virol. **147:**1685–1698. http://dx.doi.org/10.1007/s00705-002-0841-3.
112. **Zeidner G, Preston CM, Delong EF, Massana R, Post AF, Scanlan DJ, Beja O.** 2003. Molecular diversity among marine picophytoplankton as revealed by psbA analyses. Environ. Microbiol. **5:**212–216. http://dx.doi.org/10.1046/j.1462-2920.2003.00403.x.

Evelien Adriaenssens received her B.S. and M.S. from the faculty of Bioscience Engineering of the University of Leuven (KU Leuven), Leuven, Belgium. During her Ph.D. research, she worked on the isolation and characterization of bacteriophages of the potato pathogen *Dickeya solani* for applications in plant protection. This was a collaborative effort between the Laboratory of Gene Technology of Professor Rob Lavigne (KU Leuven), the Plant Production Laboratory of Professor Maurice De Proft (KU Leuven), and the Unit Plant-Crop Protection of the Institute for Agricultural and Fisheries Research with Dr. Martine Maes. After obtaining the degree in 2012, she joined the Centre of Microbial Ecology and Genomics at the University of Pretoria, Pretoria, South Africa, in 2013 to work as a postdoctoral fellow heading the viral metagenomics project.

Don Cowan was educated in New Zealand at the University of Waikato and completed a period of postdoctoral study there before moving to University College London (United Kingdom) as a Lecturer in 1985. After 16 years in London, he accepted the position of Professor of Microbiology in the Department of Biotechnology at the University of the Western Cape, Cape Town, South Africa, where he was a Senior Professor and Director of the Institute for Microbial Biotechnology and Metagenomics. In May 2012 he moved to the University of Pretoria as Director of both the Genomics Research Institute and the Centre for Microbial Ecology and Genomics. His research activities encompass a wide range of projects in the field of ecogenomics: the use of genomic and metagenomic methods to understand the diversity and function of microorganisms in different environments.