AEM
Journals.ASM.org

Unveiling of the Diversity of Prasinoviruses (*Phycodnaviridae*) in Marine Samples by Using High-Throughput Sequencing Analyses of PCR-Amplified DNA Polymerase and Major Capsid Protein Genes

Camille Clerissi,[a,b] Nigel Grimsley,[a,b] Hiroyuki Ogata,[c,d] Pascal Hingamp,[d] Julie Poulain,[e] Yves Desdevises[a,b]

Sorbonne Universités, UPMC Université Paris 06, Observatoire Océanologique, Banyuls-sur-Mer, France[a]; CNRS, UMR 7232, Observatoire Océanologique, Integrative Biology of Marine Organisms, Banyuls-sur-Mer, France[b]; Education Academy of Computational Life Sciences, Tokyo Institute of Technology, Tokyo, Japan[c]; CNRS, Université Aix-Marseille, IGS UMR7256, Mediterranean Institute of Microbiology (FR 3479), Marseille, France[d]; CEA, Institut de Génomique, Génoscope, Evry, France[e]

Viruses strongly influence the ecology and evolution of their eukaryotic hosts in the marine environment, but little is known about their diversity and distribution. Prasinoviruses infect an abundant and widespread class of phytoplankton, the Mamiellophyceae, and thereby exert a specific and important role in microbial ecosystems. However, molecular tools to specifically identify this viral genus in environmental samples are still lacking. We developed two primer sets, designed for use with polymerase chain reactions and 454 pyrosequencing technologies, to target two conserved genes, encoding the DNA polymerase (PolB gene) and the major capsid protein (MCP gene). While only one copy of the PolB gene is present in *Prasinovirus* genomes, there are at least seven paralogs for MCP, the copy we named number 6 being shared with other eukaryotic alga-infecting viruses. Primer sets for PolB and MCP6 were thus designed and tested on 6 samples from the Tara Oceans project. The results suggest that the MCP6 amplicons show greater richness but that PolB gave a wider coverage of *Prasinovirus* diversity. As a consequence, we recommend use of the PolB primer set, which will certainly reveal exciting new insights about the diversity and distribution of prasinoviruses at the community scale.

Members of the *Phycodnaviridae* family are classified in five genera according to the species of eukaryotic algae that they are known to infect. Indeed, *Chlorovirus*, *Raphidovirus*, *Phaeovirus*, *Coccolithovirus*, and *Prasinovirus*, respectively, infect *Chlorella*, raphidophytes, phaeophytes, coccolithophores, and prasinophytes (known hosts of prasinoviruses now belong to the new class Mamiellophyceae [1], which remains in the subphylum Prasinophytina [2]) (see also references 3 to 6 for further information about these viral groups). They are part of the nucleocytoplasmic large DNA viruses (NCLDV) and share nine conserved core genes with other NCLDV (7). Among them, the DNA polymerase gene and the major capsid protein gene (herein called the PolB and MCP genes, respectively) have been to date the molecular markers most commonly used to investigate the diversity of phycodnaviruses (e.g., see references 8 to 13).

Prasinovirus was the first described phycodnaviral genus (14), and a metagenomic survey recently proposed that these viruses are the most abundant members of the *Phycodnaviridae* in marine ecosystems (15). *Ostreococcus* virus (OV), *Micromonas* virus (MpV), and *Bathycoccus* virus (BpV) represent three major groups of prasinoviruses and are known to infect the three most widespread and abundant genera of the Mamiellophyceae (1, 16–18). The PolB gene is the only gene used so far to examine *Prasinovirus* diversity, and previous analyses suggested that *Ostreococcus lucimarinus* viruses show no biogeographic pattern, i.e., genetic and geographic distances are not correlated (19), that there is a link between genetic distances and host range (9), and that trophic conditions influence *Prasinovirus* abundances (20).

However, all PolB analyses performed so far were culture dependent, and no studies took advantage of deep-sequencing amplicons using high-throughput sequencing, such as 454 technology. Current primers to amplify PolB and MCP genes of phycodnaviruses (11, 21) are not well suited to study of *Prasinovirus* diversity using 454 pyrosequencing. Indeed, even though the present technology allows up to 700 bp to be sequenced, in practice the sequencing of multiplexed amplicons is limited to 350 to 400 bp. This limitation is particularly important in the case of the PolB sequence, since ~900-bp selfish genetic elements called inteins were found within this gene for the *Phycodnaviridae*, as in other eukaryotic viruses (3, 10, 11, 22–25). These elements are inserted at the conserved amino acid motif YGDTDS between the AVS primers commonly used to study the *Phycodnaviridae* family (8, 9, 12, 21, 26, 27). Moreover, current MCP primers do not seem optimal to investigate *Prasinovirus* diversity, as observed by Larsen et al. (11), who failed to amplify MpV, while complete genome sequencing revealed the existence of MCP paralogs in *Chlorovirus* and *Prasinovirus* (28–32).

As a consequence, we decided to design specific primers for prasinoviruses to solve these issues and to target suitably sized parts of their PolB or MCP genes. This study aimed at the following: (i) to analyze the phylogenetic diversity of MCP paralogs, toward identifying orthologous copies shared between *Prasinovirus* and other *Phycodnaviridae*, in order to design a specific primer set for prasinoviruses and (ii) to compare the diversity revealed by the *Prasinovirus*-specific PolB and MCP primers. Our results suggest that the capsid-like gene copy 6 (here named MCP6 because it encodes the 6th gene copy along the linear *Ostreococcus tauri* virus

Received 22 January 2014 **Accepted** 5 March 2014
Published ahead of print 14 March 2014
Editor: K. E. Wommack
Address correspondence to Nigel Grimsley, nigel.grimsley@obs-banyuls.fr.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.00123-14.

doi:10.1128/AEM.00123-14

OtV5 genome, showing similarity to the MCP gene [28]) is shared with other eukaryotic alga-infecting viruses, including *Paramecium bursaria* chlorella virus (PBCV), where it is has been shown experimentally to function as the major capsid protein (33). While the MCP6 primers revealed a higher richness, the use of PolB gave a wider coverage of the actual *Prasinovirus* diversity.

MATERIALS AND METHODS

Sampling and DNA extraction. Seawater samples were collected at 2 depths from 3 stations in the Indian Ocean: at the surface (about 5 meters deep) and at the deep chlorophyll maximum (DCM) (from 30 to 72 meters deep, depending on the station). Samples were collected by pumping seawater up a tube immersed to the appropriate depth at the sampling location, using a large peristaltic pump (A40; Tech-Pompes, Sens, France). Twenty liters of seawater was first passed through 200-μm and 20-μm mesh filters to remove larger plankton and then in series through 1.6-, 0.22-, and 0.1-μm filters (142-mm GF/A glass microfiber prefilter, no. 1825-142 [Whatman]; 142-mm, 0.22-μm-pore-size Express Plus membrane, no. GPWP14250 [Millipore]; 142-mm, 0.1-μm-pore-size MF membrane, no. VCWP14250 [Millipore] [respectively]) using a peristaltic pump (Masterflex; EW-77410-10). The filters were kept for 1 month at −20°C on board and at −80°C in the laboratory until the DNA extraction. DNA extractions were done only on the 0.1-μm filters (0.1- to 0.2-μm fraction), following a modified CTAB protocol (34): (i) the filters were crushed in liquid nitrogen and then (ii) incubated at 60°C for 1 h in a CTAB buffer (2% CTAB [hexadecyltrimethylammonium bromide], 100 mM Tris-HCl [pH = 8], 20 mM EDTA, 1.4 M NaCl, 0.2% β-mercaptoethanol, 0.1 mg/ml proteinase K, and 10 mM dithiothreitol [DTT]), (iii) DNAs were purified using an equal volume of chloroform-isoamylalcohol (24:1) and a 1-h-long RNase digestion step, and (iv) DNAs were precipitated with a 2/3 volume of isopropanol and washed with 1 ml of a solution containing 76% (vol/vol) ethyl alcohol (EtOH) and 10 mM ammonium acetate solution. Finally, the extracted DNA samples were dissolved in 100 μl of laboratory-grade deionized water and stored at −20°C until the sequencing steps. An approximate yield of 65 ng/μl was obtained on average for each sample.

Primer design, PCR, sequencing, and data processing. Based on the amino acid sequences of the PolB and MCP6 genes of OtV1 (35), OtV5 (28), *Ostreococcus lucimarinus* virus 1 (OlV1), MpV1, BpV1, BpV2 (32), PBCV1, PBCV_NY2A, PBCV_AR158 (29, 33, 36), *Acanthocystis turfacea* chlorella virus 1 (ATCV1) (30), PBCV_MT325, and PBCV_FR483 (31), *Prasinovirus*-specific primer sets consisting of VpolAS4 (5′-GAR GGI GCI ACI GTI YTN GA-3′)-VpolAAS1 (5′-CCI GTR AAI CCR TAI ACI SWR TTC AT-3′) and VmcpAS3 (5′-GGI GGI CAR MGI RTI GAY AA-3′)-VmcpAAS1 (5′-TGI ACY TGY TCD ATI ARR TAY TCR TG-3′) were inferred and designed to amplify and sequence ~320- and ~350-bp-long fragments of the genes, respectively. The reverse primer of the PolB gene is located to the 5′ side of the YGDTDS intein insertion site to avoid any bias (see Fig. S1 in the supplemental material) (10, 25), and the MCP6 primer corresponds to the conserved region amplified by Larsen et al. (see Fig. S2 in the supplemental material) (11). In order to multiplex different samples for sequencing (454 pyrosequencing), the adapter A (5′-CCA TCT CAT CCC TGC GTG TCT CCG AC-3′), a key of four letters (5′-TCAG-3′), and a "MID" tag were added before the PCR step at the 5′ end of the forward primer, and the adapter B plus the four-letter key (5′-CCT ATC CCC TGT GTG CCT TGG CAG TC TCAG-3′) was added at the 5′ end of the reverse primer. Different MIDs were used for each sample and were selected from the Roche list. PCRs for the PolB and MCP6 fragments were done in duplicate and set up as follows: 2 μl of environmental DNA (100 ng) was added to a 48-μl reaction mixture which contained 0.2 mM (each) deoxyribonucleoside triphosphates, 30 pmol of each primer, 1/10 of Advantage 2 PCR buffer, and 1 U of Advantage 2 polymerase mix (catalog no. 639201; Clontech). The PCR was carried out in a Mastercycler instrument (Eppendorf) with an initial step at 95°C (3 min) followed by 38 rounds at 95°C (30 s), 50°C (60 s) for PolB or 48°C (60 s) for MCP6, and 72°C (90 s) and a final extension at 72°C (4 min). PCR products were electrophoresed in 0.5% Tris-acetate-EDTA (TAE) buffer in a 0.8% agarose gel, and the expected band was excised by UV visualization after ethidium bromide staining. PCR bands were purified directly using a gel extraction kit (QIAquick, catalog no. 28704; Qiagen). The DNA concentrations were estimated using a Nanodrop spectrophotometer (Nanodrop 2000; Thermo Scientific), and amplicons were pooled stoichiometrically before being sequenced on a GS-FLX Titanium plate. Because 454 pyrosequencing produces error-containing sequences, we used the software suite AmpliconNoise v.1.25 (37) to reduce the number of erroneous sequences (errors introduced in PCR and pyrosequencing steps, as well as chimeric sequences), and we added a step to remove stop codon-containing sequences. The AmpliconNoise analysis was performed using the program Qiime v.1.3.0 (38). Reads lacking a correct primer, having less than 360 successful pyrosequencing flows, or both were removed. Then, after the removal of primer sequences, the following steps and default parameters were used: PyroDist, PyroNoise (-s 60.0 -i 0.01), SeqDist; SeqNoise (-s 25.0 -i 0.08), and Perseus. Denoised sequences of the PolB and MCP6 genes were trimmed to ~300 bp after the protein motifs GKQLAYK and RRRFAQKG, respectively. Finally, blastp searches (39) (using the PolB and MCP6 genes of OtV5 as a query with 35.4 and 48.48% identity thresholds, respectively) and phylogenetic analyses were performed to retain only potential *Prasinovirus* sequences, i.e., those positioned phylogenetically within known prasinoviruses or between *Chlorovirus* and *Prasinovirus*.

Alignment and phylogenetic reconstruction. All alignments were performed according to codons or amino acids, using the software program Muscle v3.8.31 (40) implemented in Seaview (41, 42). Phylogenetic reconstructions were carried out first with protein alignments for the capsid-like genes, the complete MCP6 and complete PolB genes belonging to available reference prasinoviruses and chloroviruses: OtV1 and OtV2 (43), OtV5 and OtV6 (J. Kegel, R. Thomas, E. Derelle, and H. Moreau, unpublished data), OlV1, MpV1, BpV1, BpV2, PBCV1, PBCV_NY2A, ATCV1, PBCV_AR158, and PBCV_CVK2 (44), and PBCV_CVG-1 (45). Poorly aligned and very variable regions of the alignments were automatically removed using the software program GBlocks 0.91b (46) (http://molevol.cmima.csic.es/castresana/Gblocks_server.html). Different models of protein evolution and associated parameters were investigated using the program ProtTest 2.4 (47), and the best parameters were identified based on the Akaike information criterion (AIC). In addition, Bayesian inference (BI) was used with a mixed amino acid model (48). Phylogenetic trees based on nucleotide sequences were obtained via BI and maximum-likelihood analysis (ML). In ML, different models of DNA evolution and associated parameters were estimated using the program jModeltest 2 (49), and the best parameters were identified based on the AIC. In BI, coding DNA sequences were considered using an evolutionary model that takes into account the genetic code ("codon model" [50, 51]; see also reference 52). The ML estimate for the nucleotide and protein alignments was performed using the program PhyML 3.0 (53) using the appropriate model and validated via a bootstrap procedure with 100 replicates. Bayesian analyses were performed using MrBayes 3.2 software (48), with 4 chains of 7×10^6 generations, trees sampled every 700 generations, and burn-in value set to 20% of the sampled trees (4 chains of 2×10^6 generations and trees sampled every 200 generations were used for the MCP-like tree). We checked that standard deviation of the split frequencies fell below 0.01 to ensure convergence in tree search.

***Prasinovirus* annotation.** In order to describe the environmental *Prasinovirus* diversity, 23 PolB gene reference sequences, representing 475 *Prasinovirus* and *Chlorovirus* sequences, were selected for a nucleotide distance of 10% (see below; GenBank accession numbers are indicated in parentheses or at the end of this section). These are 1 *Bathycoccus* virus, BpV87 (FJ267515); 14 *Micromonas* viruses (viral names indicate which host strain or host clade was used for isolation; for example, Mi1109V14 is virus number 14 isolated using the RCC1109 strain, and MicCV32 is virus number 32 isolated with a clade C *Micromonas* strain), Mi1109V14, MicCV32, Mi497V14, MicAV11, Mi829V1, MicAV17, MicAV8,

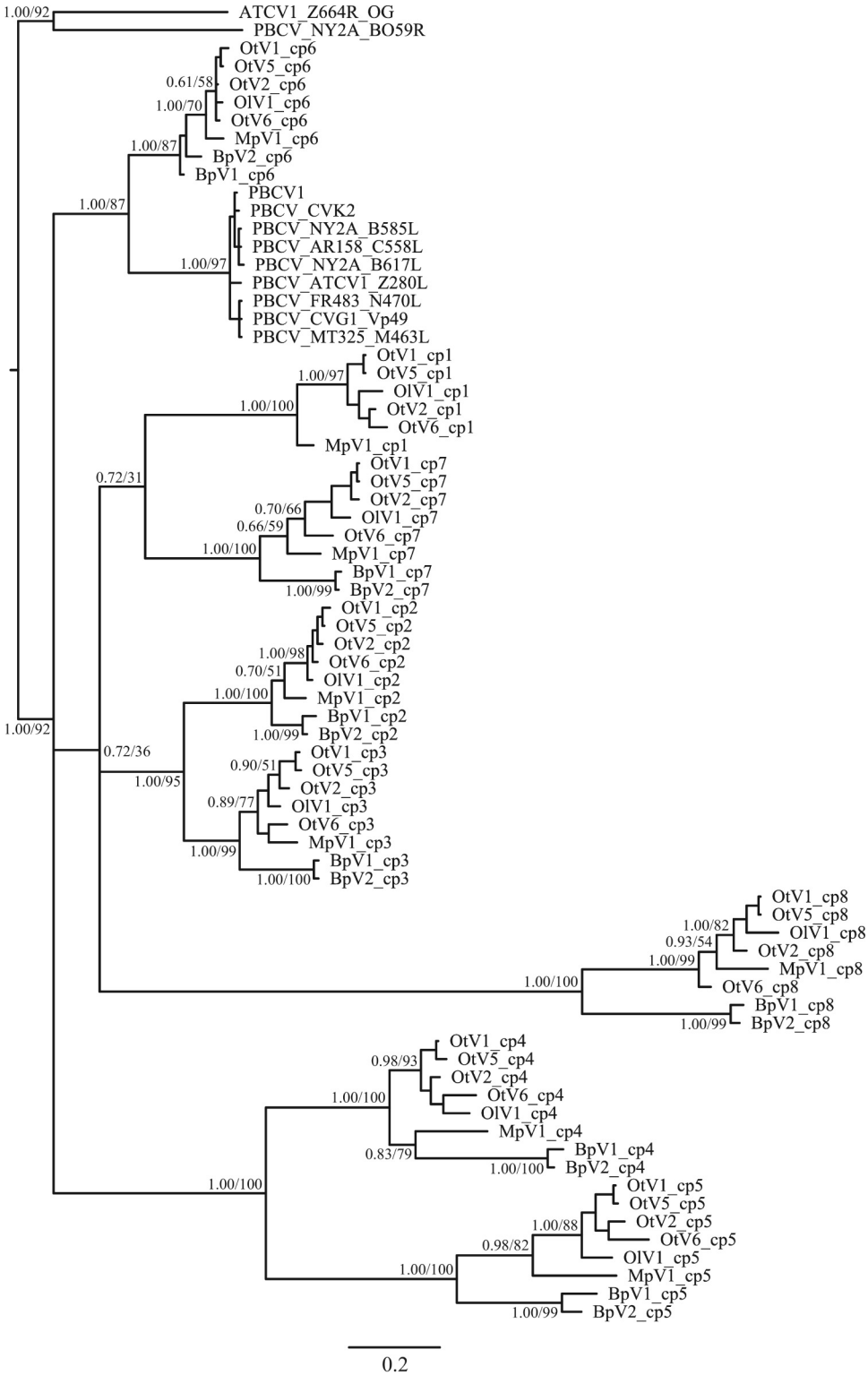

FIG 1 Phylogenetic tree of MCP-like sequences belonging to *Prasinovirus* and *Chlorovirus*. The phylogenetic tree was built using Bayesian inference (BI) and maximum likelihood (ML). Numbers are posterior probabilities (BI) and bootstrap proportions (ML) reflecting clade support. The tree was rooted using an unshared *Chlorovirus* MCP-like copy.

MicBV10, MicAV16, MicAV29, MicBV39, MicBV30, MicBV26, and MpV1 (HM004429); 6 *Ostreococcus* viruses, O356V303, OlV537 (GQ412097), OtV63 (FJ267501), OtV09_559, OlV464 (GQ412092), and OtV6 (JN225873); and 2 chloroviruses, PBCV_CVK2 (AB011500) and PBCV_NY2A (DQ491002). In addition, 12 MCP6 reference sequences were selected from genomic data sets (except for PBCV_CVK2) of 2 *Bathycoccus* viruses, BpV1 (NC_014765) and BpV2 (HM004430); 1 *Micromonas* virus, MpV1 (HM004429); 5 *Ostreococcus* viruses, OlV1

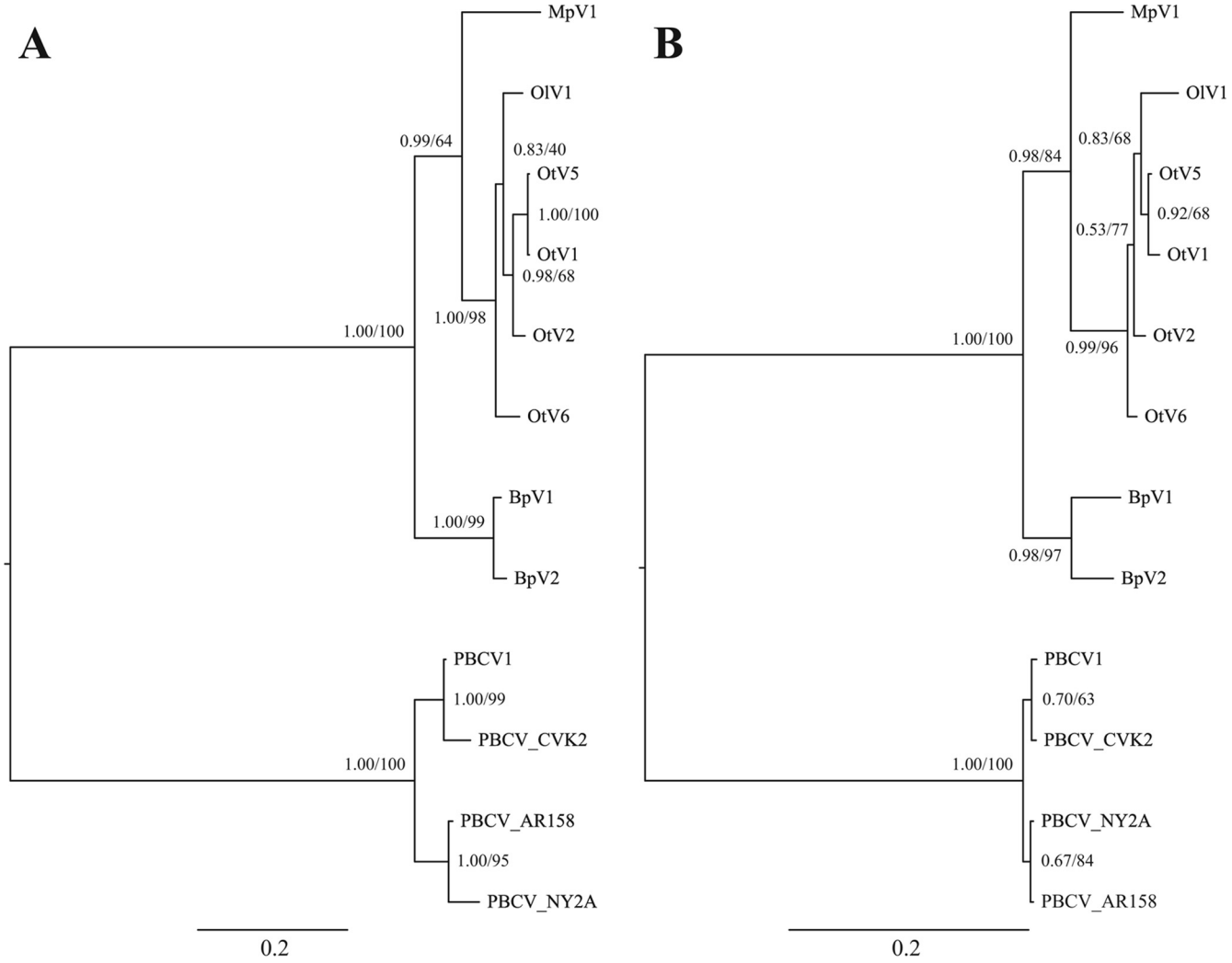

FIG 2 Phylogenetic analysis of the complete PolB (A) and MCP6 (B) genes of *Prasinovirus* and *Chlorovirus*. The phylogenetic trees were built using Bayesian inference (BI) and maximum likelihood (ML). Numbers are posterior probabilities (BI) and bootstrap proportions (ML) reflecting clade support. Trees were rooted using *Chlorovirus* sequences.

(NC_014766), OtV1 (NC_013288), OtV2 (NC_014789), OtV5 (NC_010191), and OtV6 (JN225873); and 4 chloroviruses, PBCV1 (JF411744), PBCV_AR158 (DQ491003), PBCV_NY2A (DQ491002), and PBCV_CVK2 (AB011506). All reference sequences were aligned with the corresponding environmental sequences and clustered at different nucleotide distances (from 30% to 10%). The environmental PolB gene sequences were annotated for a distance of 26%, i.e., when all the reference sequences of OV, MpV, and BpV clustered into three different groups. The annotation was done using a representative sequence from each operational taxonomic unit (OTU), as a query for the Basic Local Alignment Search Tool (BLAST) (39) search against the NCBI nucleotide collection (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Because this database contains fewer reference sequences for the MCP6 gene than for the PolB gene, MCP6 was also annotated for a distance of 26%. As a consequence, a cutoff of 26% was used for both markers to describe the major groups of *Prasinovirus*, such as OV, BpV, and MpV. Twenty-six percent is a conservative regrouping based on known host genera, but each genus regroups more than one host algal species and may also contain species that are not yet represented in cultures. We retained this conservative estimation in order to be certain that the new groups of prasinoviruses that we predict must really represent novel groups of viruses.

Sequence analysis. The nucleotide diversity π (54) of the reference alignment sequences was obtained using the software program DnaSP version 5 (55). All other sequence analyses were performed using the program mothur (56). Because both the PolB and MCP6 genes were amplified with degenerate primers, we did not consider abundances of identical sequences since PCR and sequencing steps generate bias on sequence proportions. The representatives of identical sequences are henceforth named genotypes. Sequences were aligned and clustered at different nucleotide distances from 30% to 0%. Rarefaction curves were produced at different thresholds and allowed us to define the OTU when the curves began to level off, i.e., at 10% for both the PolB and MCP6 genes. Richness estimation was assessed using the bootstrap calculator for an OTU definition of 10%. The environmental samples were clustered according to their composition in *Prasinovirus* and the Bray-Curtis index using the function tree.shared. Finally, an OTU definition of 26% was used to analyze the major groups of *Prasinovirus* and particularly to compare the genetic backgrounds of groups obtained with PolB and MCP6 primers. Indeed, the distances between the environmental OTU and 12 reference sequences (partial PolB and MCP6 from genome sequences of BpV1, BpV2, OlV1, OtV5, OtV1, OtV2, OtV6, MpV1, PBCV1, PBCV_AR158, PBCV_CVK2, and PBCV_NY2A) were assessed for both markers with an

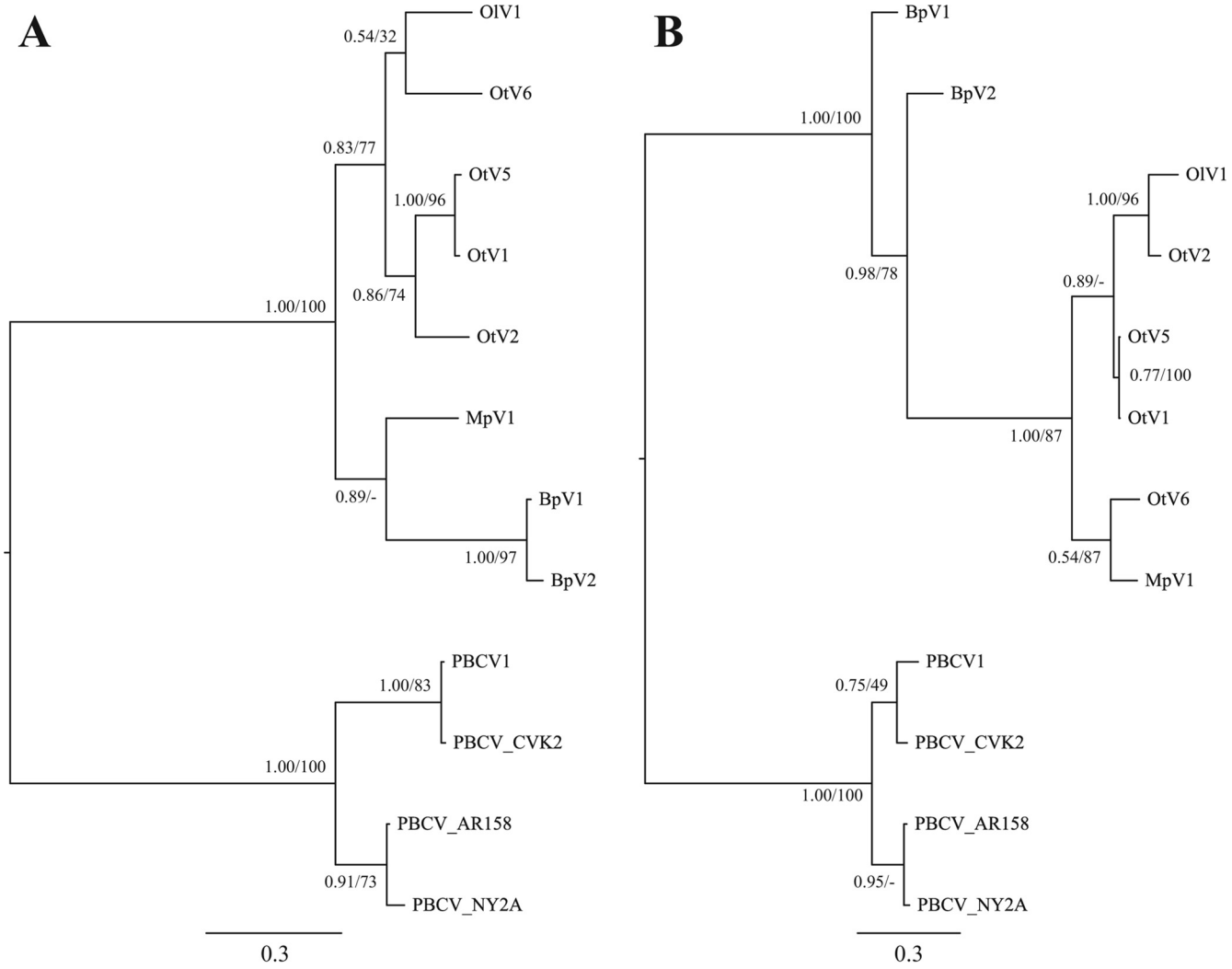

FIG 3 Phylogenetic analysis of the partial PolB (A) and MCP6 (B) genes of *Prasinovirus* and *Chlorovirus*. The phylogenetic trees were built using Bayesian inference (BI) and maximum likelihood (ML). Numbers are posterior probabilities (BI) and bootstrap proportions (ML) reflecting clade support. Trees were rooted using *Chlorovirus* sequences.

OTU definition of 26%, using the mothur dist.seq function. A hierarchical clustering analysis (Euclidean metric; unweighted-pair group method using average linkages [UPGMA]) of these distances was then computed with the R "stats" software package using the hclust function in order to visualize the genetic proximities of the major groups sampled by both primer sets.

Statistical analysis. All statistical analyses were performed using R v2.15.0 (R: a language and environment for statistical computing, 2008; R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria [http://www.R-project.org]). The congruence of both the PolB and MCP6 trees was assessed from patristic distances and Mantel tests using the R "vegan" software package and the mantel function. In order to estimate the link between sequencing effort and OTU richness, we used the rank-based Spearman ρ correlation coefficient (instead of Pearson, since there were only 6 samples) within the R "stats" package and the cor.test function. A richness comparison was done for a distance of 10% between the PolB and MCP6 genes with the Wilcoxon test, using the R "stats" package and the wilcox.test function. Finally, the shapes of rank-abundance plots that illustrate the major groups of *Prasinovirus* were described using the equitability indices of the PolB and MCP6 genes for a distance of 26 to 25% and 26 to 24% (respectively) and were computed using the R "pgirmess" package and the "shannon" function.

Nucleotide sequence accession numbers. The 454 sequence data sets were submitted to the Sequence Read Archive of the European Nucleotide Archive (accession number PRJEB4356). Sequences for the following *Prasinovirus* isolates were submitted to GenBank (accession numbers are indicated in parentheses): Mi1109V14 (KF378564), MicCV32 (KF378579), Mi497V14 (KF378565), MicAV11 (KF378568), Mi829V1 (KF378567), MicAV17 (KF378570), MicAV8 (KF378574), MicBV10 (KF378575), MicAV16 (KF378569), MicAV29 (KF378572), MicBV39 (KF378578), MicBV30 (KF378577), MicBV26 (KF378576), O356V303 (KF378581), and OtV09_559 (KF378582).

RESULTS

General features of the MCP and PolB genes. Because BI and ML trees were very similar, we show only BI trees with posterior probabilities and bootstrap support values from ML analyses. The capsid-like-protein tree was rooted using distantly related *Chlorovirus* MCP-like sequences as an outgroup (Fig. 1). All of the analyzed *Prasinovirus* genomes contain eight copies of the MCP gene (Fig. 1), except for two BpVs lacking copy number 1. This tree suggests that all the MCP copies in prasinoviruses were acquired from their common ancestor, since their phylogenetic relationships were similar

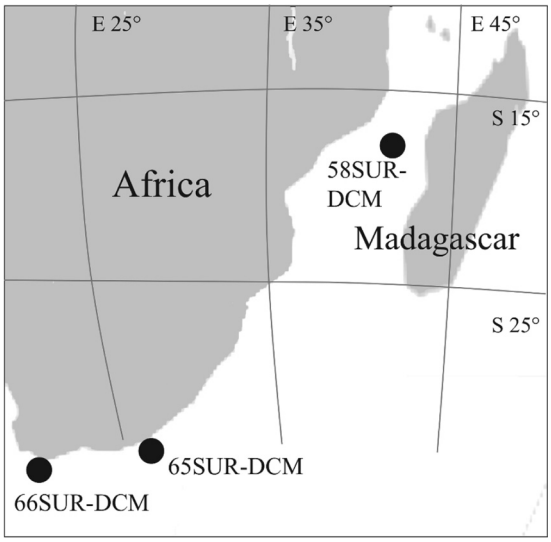

FIG 4 Locations of sampling sites. Numbers in station names are in chronological order. SUR, surface; DCM, deep chlorophyll maximum. (Adapted from a map on Wikipedia [http://en.wikipedia.org/wiki/File:Africa_%28orthographic_projection%29.svg#file], published under a Creative Commons license.)

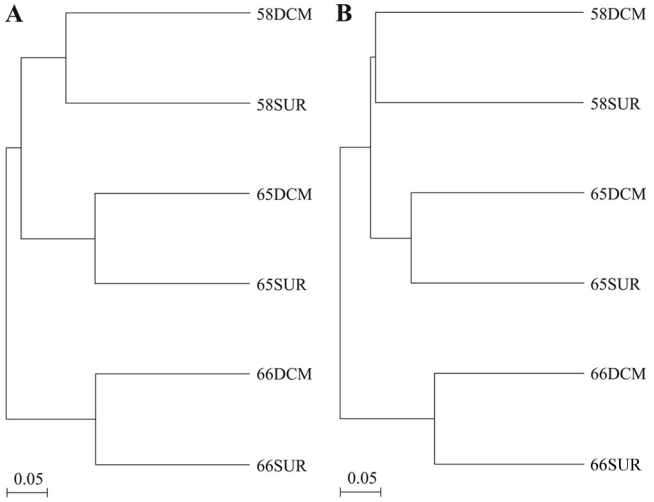

FIG 5 Clustering of the 6 samples based on *Prasinovirus* assemblages. (A) PolB; (B) MCP6. Bray-Curtis dissimilarities and UPGMA were used. Numbers in station names are in chronological order. SUR, surface; DCM, deep chlorophyll maximum. OTU are defined for a nucleotide distance of 10%.

across the different copies. The lack of copy number 1 in BpV is thus likely due to gene loss in this lineage. Notably, OtV6 displays different histories for the MCP copies 3 and 8. This analysis also revealed that copy number 6 has close homologs in *Chlorovirus* (Fig. 1). Phylogenetic trees of complete PolB and complete MCP6 genes were congruent (Fig. 2) ($P = 0.001$), but the partial MCP6 gene sequence amplified by PCR gave a phylogeny slightly different from the complete alignment, since OtV6 clustered with MpV (Fig. 3). This is probably due to reduced phylogenetic signals; however, both trees built with the region amplified for PolB and MCP6 are generally congruent with each other and with the respective complete gene alignment ($P = 0.001$). Finally, the nucleotide diversity π of the partial MCP6 alignment is comparable to that for partial PolB: 0.31303 ± 0.000763 and 0.30475 ± 0.000527, respectively.

Comparison of environmental diversity. Six samples were collected between 29 June and 15 July 2010 from the subsurface and DCM at three stations located in the southwestern region of the Indian Ocean (Fig. 4). Rarefaction curves were produced for the whole sample at different thresholds and led us to define OTU using a threshold of 10% where the curve levels off (data not shown). Thus, we assumed that each sample was representative of the populations at this OTU cutoff. Moreover, since the abundance of environmental sequences may be biased by PCR, we analyzed richness by focusing on genotypes. An average of 81.5 and 122.3 genotypes per sample, corresponding to an average OTU richness of 35.1 and 60.5 (nucleotide distance of 10%), was found for PolB and MCP6, respectively (Table 1). These values were not normalized, since we did not find a link between sequencing efficiency and OTU richness (Spearman's rank correlation rho = 0.0420; $P = 0.9037$). Finally, the OTU richness for MCP6 was significantly higher than that for PolB (Wilcoxon test, $P = 0.001$).

OTU comparisons. The six samples were analyzed regarding their community compositions, based on genotypes and on an OTU cutoff of 10% (Bray-Curtis dissimilarities; UPGMA). Both PolB and MCP6 displayed the same pattern, where samples from the same station clustered together (Fig. 5).

In order to describe the major groups of *Prasinovirus*, an OTU cutoff of 26% was chosen, since all PolB reference sequences of OV, MpV, and BpV clustered into three distinct OTU at this threshold (Fig. 6; Table 2). In the case of both PolB and MCP6, OTU containing MpV (OTU7_PolB and OTU4_MCP6) and BpV (OTU2_PolB and OTU3_MCP6) sequences represented the richest clusters, and a cluster containing all or most of known OV sequences (OTU1_PolB and OUT9_MCP6) was ranked at the fifth position. There were two environmental OTU (OTU3/

TABLE 1 Total numbers of sequences and diversity of *Prasinovirus*[a]

Sample	No. of PolB gene sequences	No. of PolB genotypes	PolB gene OTU richness (10%)	No. of MCP6 gene sequences	No. of MCP6 genotypes	MCP6 gene OTU richness (10%)
58DCM	22,967	64	31.4	45,692	115	56.4
58SUR	26,114	100	31.9	7,314	87	51.6
65DCM	150,216	103	28.3	4,553	99	50.8
65SUR	32,484	55	22.9	32,176	113	55.6
66DCM	15,568	86	48.9	32,424	158	71.8
66SUR	33,741	81	47.1	46,802	162	77.1
Mean	46,848	81.5	35.1	34,989	122.3	60.5

[a] The estimation of richness was assessed using the bootstrap calculator for an OTU definition of 10%.

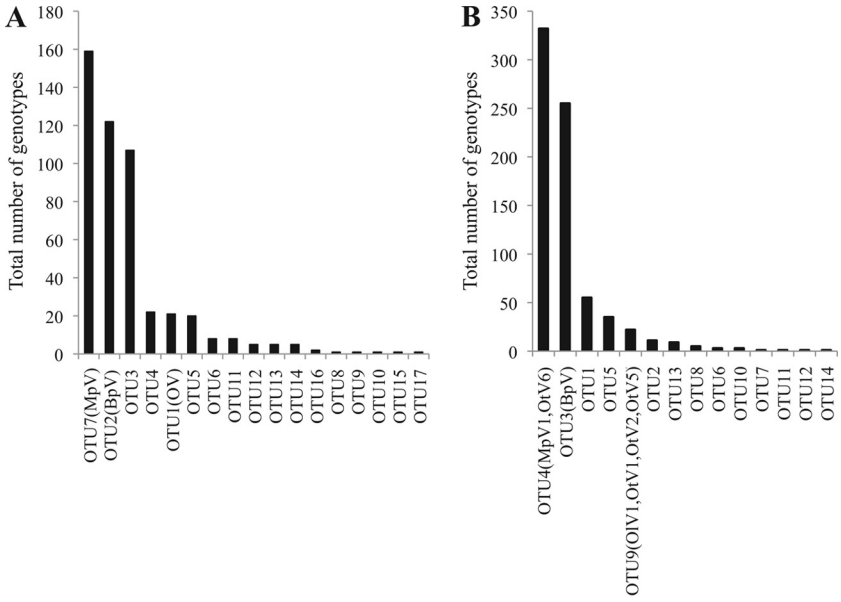

FIG 6 Rank-abundance of *Prasinovirus* genotypes in the 6 samples. (A) PolB; (B) MCP6. OTU are defined for a nucleotide distance of 26%, i.e., when all the reference sequences of OV, MpV, and BpV clustered into three different groups for PolB.

4_PolB and OTU1/5_MCP6) composed solely of environmental sequences (Table 2) between the OV and BpV clusters. Notably, as highlighted by the phylogenetic analysis of the partial MCP6 sequence (Fig. 3), OtV6 clustered with MpV1 and not with the other OV, whatever the threshold used (from 30 to 20%; data not shown).

Finally, to compare the genetic backgrounds of the major environmental groups obtained for partial PolB and MCP6 (particularly the unknown OTU3/4_PolB and OTU1/5_MCP6), we assessed their proximities using a clustering analysis (Fig. 7). Because partial PolB and MCP6 trees were congruent, we expected to find a single *Prasinovirus* group at the same position in both trees, i.e., genetic distances between one special group and the other taxa should be similar or proportionally similar when comparing both partial genes. Using this assumption, we assessed the nucleotide distances between environmental OTU and 12 reference sequences of viral cultures and then performed a clustering analysis (Euclidean metric; UPGMA), which suggests that these two partial genes do not give the same picture of *Prasinovirus* diversity (Fig. 7). Indeed, while the OTU containing OV, BpV, and MpV clustered as expected for the two markers (OTU1_PolB and OTU9_MCP6, indicated by solid-line black boxes; OTU2_PolB and OTU3_MCP6, indicated by dashed black boxes; and OTU7_PolB and OTU4_MCP6, indicated by gray boxes, respectively), the four most diverse OTU (indicated by an asterisk) did not (OTU3/4_PolB were not similar to OTU1/5_MCP6), suggesting different genetic backgrounds for the origin of these PolB and MCP sequences.

Equitability indices and diversity distributions. The shapes of the rank-abundance plots computed using genotypes and an OTU cutoff of 26% (Fig. 6) showed that PolB mostly sampled genotypes of three major groups of *Prasinovirus*, while MCP6 targeted only two of them. This difference in shapes was also highlighted by their equitability indices (Table 3), based on the analysis done for a distance of 26% or for the same number of OTU (nucleotide distances of 25 and 24% for PolB and MCP6, respectively). Indeed, the lower the equitability index is, the less different OTU are represented by equal numbers of genotypes. Furthermore, the division of the clustering analysis (Fig. 7) into four groups with a similar number of sequences (group 1, from OTU11_PolB to OTU3_PolB; group 2, from OTU10_MCP6 to OTU12_MCP6; group 3, from OTU4_MCP6 to OTU9_MCP6; group 4, from OTU15_PolB to OTU11_MCP6) produced a regular distribution of the PolB OTU (mean greater than variance) compared to a clumped distribution of the MCP6 OTU (mean less than variance) within the dendrogram (Table 3).

DISCUSSION

In order to describe *Prasinovirus* communities using high-throughput sequencing technology (Titanium 454 pyrosequencing), we developed two primer sets targeting the genes most frequently chosen for studying diversity in aquatic eukaryote-infecting viruses, the PolB and MCP genes. They produce ~350-bp-long fragments and are thus well adapted to current sequencing technologies. Our analysis suggests that prasinoviruses have at least seven capsid-like genes and that copy 6 shows the most similarity to other *Phycodnaviridae* MCP genes. The MCP6 primers designed in this study revealed a richer set of environmental *Prasinovirus* sequences than those designed for PolB, but overall the MCP6 primers used target a smaller taxonomic diversity.

General features of MCP and PolB. MCP and PolB were proposed to be good candidates for studying evolution and diversity in NCLDV, either because they contain conserved domains surrounding divergent regions (13, 21) or because a correlation between genetic distances of partial fragments and hybridization of total genomic DNA was found (8). However, no primer sets were specifically tailored to describe *Prasinovirus* diversity in the environment using amplicon sequencing with high-throughput technologies, even though this approach allows diversity at the community level to be described and avoids the biases of culture-dependent methods. The 454 pyrosequencing technology is lim-

TABLE 2 Annotation of *Prasinovirus* OTU for a nucleotide distance of 26%[a]

OTU	No. of genotypes	OTU representative	Best BLAST hit	Accession no.	Max identity (%)	Query coverage (%)
OTU7_PolB	159	PC.27_18416	Clone SO98-2	AF405573	83	100
OTU2_PolB	122	PC.23_8374	Clone KBvp-12	EU889365	85	99
OTU3_PolB	107	PC.21_11386	BpV178	FJ267518	76	100
OTU4_PolB	22	PC.25_79887	Clone KBvp-7	EU889360	90	99
OTU1_PolB	21	PC.31_17112	OlV360	GQ412085	84	100
OTU5_PolB	20	PC.23_6189	Clone CL1a-44	EU336476	75	97
OTU6_PolB	8	PC.21_11235	Clone Lake227_ September2a_2004	EU408237	76	98
OTU11_PolB	8	PC.27_19315	BpV178	FJ267518	79	99
OTU12_PolB	5	PC.23_8155	Clone KBvp-18	EU889371	78	99
OTU13_PolB	5	PC.25_88578	Clone LO.16Jul08.8	HM750210	77	100
OTU14_PolB	5	PC.31_14701	Clone KBvp-14	EU889367	79	98
OTU16_PolB	2	PC.31_15216	Clone LO1b-17	EU336662	74	100
OTU8_PolB	1	PC.21_11300	BpV2	HM004430	72	99
OTU9_PolB	1	PC.21_11437	Unknown	Unknown	Unknown	Unknown
OTU10_PolB	1	PC.23_6198	Clone SOLa09	HQ424379	86	100
OTU15_PolB	1	PC.29_8062	Clone PSC99-1	AF405597	82	100
OTU17_PolB	1	PC.31_17203	Clone KBvp-12	EU889365	83	99
OTU4_MCP6	332	PC.26_22014	Clone OTU/P06-04	EU006619	86	99
OTU3_MCP6	255	PC.28_13239	BpV2	HM004430	80	100
OTU1_MCP6	55	PC.22_12013	OtV1	FN386611	79	100
OTU5_MCP6	35	PC.30_2339	Clone P3	FJ791177	76	99
OTU9_MCP6	22	PC.30_8703	ORtV6	JN225873	82	100
OTU2_MCP6	11	PC.24_1066	OtV2	FN600414	84	100
OTU13_MCP6	9	PC.32_23329	ORtV6	JN225873	94	100
OTU8_MCP6	5	PC.30_13495	OtV1	FN386611	72	99
OTU6_MCP6	3	PC.32_23331	Clone OTU/P06-04	EU006619	73	94
OTU10_MCP6	3	PC.22_18937	MpV1	HM004429	88	99
OTU7_MCP6	1	PC.22_18719	OlV1	HM004431	73	99
OTU11_MCP6	1	PC.22_19267	Unknown	Unknown	Unknown	Unknown
OTU12_MCP6	1	PC.24_1212	MpV1	HM004429	80	97
OTU14_MCP6	1	PC.30_13516	Unknown	Unknown	Unknown	Unknown

[a] OTU were sorted according to their total numbers of sequences. Representative sequences from each of the OTU of the PolB and MCP6 genes were BLASTed against the nucleotide collection of the NCBI database.

ited to 400 to 500 bp, both by the size of sequences that can be produced and by the size of amplicons that can be sequenced in multiplex. In any case, since an ~900-bp-long selfish genetic element called an intein is sometimes inserted in the DNA region amplified by the currently used PolB primer set (10, 24, 25), we needed specific primers for *Prasinovirus* that cover the well-studied partial PolB marker (9, 12, 19, 57) and that excluded the intein insertions, thus producing ~350-bp-long fragments. Although

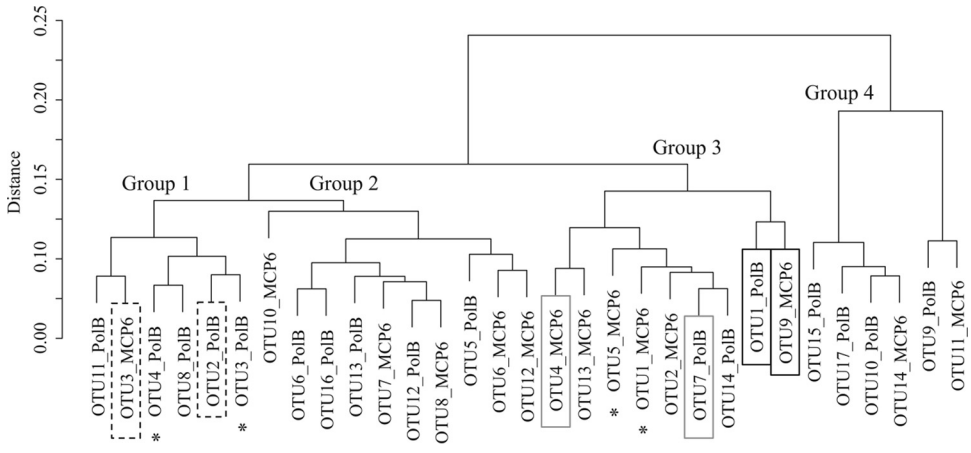

FIG 7 Clustering of the PolB and MCP6 OTU according to their genetic proximity with 12 reference sequences. Euclidean metric and UPGMA were used. OTU are defined for a nucleotide distance of 26%. The reference sequences are BpV1, BpV2, OlV1, OtV5, OtV1, OtV2, OtV6, MpV1, PBCV1, PBCV_AR158, PBCV_CVK2, and PBCV_NY2A. OTU containing OV, MpV, and BpV are outlined in boxes with black, gray, and dashed black lines, respectively. The environmental OTU showing large diversities of sequences for PolB and MCP6 are indicated by an asterisk. OTU were divided into four groups (from group 1 to group 4) to estimate the nature of their distribution, according to the gene markers.

TABLE 3 Diversity of the PolB and MCP6 genes belonging to environmental *Prasinovirus*[a]

Gene	Equitability (26%)	Equitability (20 OTU)	Mean no. of OTU per cluster (26%)	Variance of OTU per cluster (26%)	Distribution (26%)
PolB gene	0.63	0.62	4.3	0.9	Regular
MCP6 gene	0.53	0.55	3.3	6.9	Clumped

[a] Equitability was assessed for a nucleotide distance of 26% (17 and 14 OTU for the PolB and MCP6 genes, respectively) and for an identical number of 20 OTU (nucleotide distances of 25 and 24% for the PolB and MCP6 genes, respectively).

the proportion of intein-containing PolB is not known, this strategy allows both intein-free and intein-containing PolB prasinoviruses to be analyzed, and as a consequence it can be used to describe their diversity more accurately. Furthermore, because NCLDV genomes contain various MCP gene copy numbers, analyses using this gene must be interpreted with caution because of possible paralogs. Although the MCP copies of *Prasinovirus* clearly clustered according to their positions in genomes (from 1 to 8), resulting from the colinearity of common *Prasinovirus* genes (32), *Chlorovirus* copies are not as divergent as those found in prasinoviruses. Indeed, different copies clustered with the other *Phycodnaviridae*, such as B585L and B617L of PBCV_NY2A (Fig. 1), indicating that MCP is not a suitable target for all *Phycodnaviridae*, particularly since paralogs could be amplified. However, our analysis revealed that MCP6 is the orthologous copy shared between prasinoviruses and other *Phycodnaviridae* and that primers could be designed to specifically amplify and study *Prasinovirus* MCP.

In addition, we found that PolB and MCP genes are useful for studying *Prasinovirus* diversity, since their phylogenies are congruent. There may be others genes suitable for such studies. For instance, the ribonucleotide reductase is well conserved and present in three of the four NCLDV families (7). However, horizontal transfers were suspected for this gene, particularly for iridoviruses that could have acquired it from bacteria during a double infection of aquatic vertebrate hosts (58), suggesting that this potential marker should be avoided for studying the diversity within the NCLDV family.

Finally, OtV6 is a virus that is able to infect a strain of *Ostreococcus tauri* resistant to OtV5 (28, 59). Atypical grouping of OtV6 sequences with MpV in our phylogenetic analysis could be the result of recombination events (60), which are often suggested to be causative agents of changes in host range and virulence in viruses (61–64). Future studies should investigate whether the genome of OtV6 and its ability to infect OtV5-resistant cells are linked to a mosaic structure of its genes and genome, resulting from genetic exchanges between distinct prasinoviruses (see reference 10).

Comparison of the two markers. Both the PolB and MCP6 primer sets were designed from amino acid sequences encoded in available *Prasinovirus* and *Chlorovirus* genomes, except for one amino acid per primer set that differed between the two genera and made these primers more specific to *Prasinovirus* (see Fig. S1 and S2 in the supplemental material). Because *Chlorovirus* is a sister genus of *Prasinovirus* (65), this strategy was expected to uncover environmental genotypes lacking cultured representatives. In addition, designing primers based on amino acid sequences within conserved domains already led to efficient amplifications (for example, the AVS primers of reference 21) and allowed the description of environmental clades for podoviruses (66). However, this method requires high primer degeneracies, which means that each unique primer sequence is diluted hundreds or thousands of times. In order to decrease this dilution defect, inosines were added for positions representing the four possible nucleotides. Because inosine-containing primers were shown to amplify rare DNA (67), to amplify a variety of alleles (68), and to be efficient when their number does not exceed four or five substitutions (69), we designed primers that do not contain more than four inosines.

The analysis of the six samples indicated that the two primer sets clustered them in a concurrent manner according to geography but not to depth. For a nucleotide distance of 10%, MCP6 gave the same community picture, albeit with higher richness values than PolB. This is the result not of sequencing efficiencies but certainly of the slight difference found between the nucleotide diversity π of both partial fragments. Furthermore, the degenerate primers allowed us to find two environmental OTU, which contain a larger number of distinct novel genotypes independent from the known OV, for both the PolB and MCP6 genes (nucleotide distance of 26%). One of them is close to the environmental clone KBvp-7, which clustered within a robust clade in the study of Culley et al. (22).

Nevertheless, our clustering analysis highlighted that dominant unknown *Prasinovirus* groups did not show the same genetic backgrounds for the two markers (Fig. 7). As a consequence, while MCP6 primers seemed biased toward MpV and OV sequences, PolB amplified more BpV sequences. These biases certainly come from the PCR step, particularly from various affinities between primers and the targeted sequences (70).

Equitability indices and diversity distribution. One advantage of using PolB as a genetic marker is the number of available sequences from cultured representatives (9, 10, 19, 20, 26, 28, 32, 35, 43, 59) and environmental clones (22, 71, 72), since such primers were proposed before the use of MCP as a marker (11, 21). Indeed, while 475 sequences from cultured representatives were used to annotate PolB sequences in this study, only 12 could be used for MCP6 (28, 29, 32, 33, 35, 43, 44; Kegel et al., unpublished). Because of this lack of reference sequences for MCP6, we decided to consider the same nucleotide distance of 26% to define OTU that was used for PolB. Our analysis of the nucleotide diversity π suggested that this strategy could have overestimated the equitability and the distribution parameters, since the partial MCP6 sequence is slightly more variable than that of PolB. Hence, the difference between the equitability indices of PolB and MCP6 and the distribution in the clustering analysis (Fig. 7 and Table 3; regular versus clumped, respectively) are minimal estimations and show that PolB sequences better cover the *Prasinovirus* diversity.

To conclude, the two *Prasinovirus*-specific primer sets designed and tested here provided new tools to better understand the diversity and distribution of the most abundant *Phycodnaviridae* genus in marine ecosystems. However, we recommend that the PolB primers are more suitable for general comparisons; since there is only one copy of this gene per genome, they give a wider coverage of *Prasinovirus* diversity, and at the moment, the large

number of reference sequences from cultured viruses provides a broader representation of landmarks.

ACKNOWLEDGMENTS

We thank the Genophy team in Banyuls-sur-Mer, France, for stimulating discussions, especially Rozenn Thomas, Evelyne Derelle, and Hervé Moreau, who allowed us to use the DNA polymerase and major capsid protein sequences of the unpublished *Prasinovirus* OtV6 genome. We are also grateful to Martha Clokie and Ramon Massana for advice and comments and to Nathalie Simon, Elodie Foulon, Roseline Edern, and Laure Bellec, who allowed us to use the DNA polymerase sequences of new *Prasinovirus* isolates.

This work was supported by an Agence Nationale de Recherche grant, TARA-GIRUS, ANR-09-PCS-GENM-218 (coordinator, H. Ogata). Camille Clerissi benefited from a doctoral fellowship from the AXA Research Fund.

We are keen to thank the commitment of the following people and sponsors who made the Tara Oceans expedition possible: CNRS, EMBL, Genoscope/CEA, VIB, Stazione Zoologica Anton Dohrn, UNIMIB, ANR (projects POSEIDON, BIOMARKS, and PROMETHEUS), FWO, BIO5, Biosphere 2, Agnès b., the Veolia Environment Foundation, Region Bretagne, World Courier, Illumina, Cap L'Orient, the EDF Foundation EDF Diversiterre, FRB, the Prince Albert II de Monaco Foundation, Etienne Bourgois, and the *Tara* schooner and its captain and crew. Tara Oceans would not exist without the continuous support of the participating 23 institutes (see http://oceans.taraexpeditions.org).

This is contribution number 0013 of the Tara Oceans Expedition 2009–2012.

REFERENCES

1. **Marin B, Melkonian M.** 2010. Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear-and plastid-encoded rRNA operons. Protist **161:** 304–336. http://dx.doi.org/10.1016/j.protis.2009.10.002.
2. **Round FE.** 1963. The taxonomy of the Chlorophyta. Br. Phycol. Bull. **2:**224–235. http://dx.doi.org/10.1080/00071616300650061.
3. **Fischer MG, Allen MJ, Wilson WH, Suttle CA.** 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc. Natl. Acad. Sci. U. S. A. **107:**19508–19513. http://dx.doi.org/10.1073/pnas.1007615107.
4. **Schroeder DC, Oke J, Malin G, Wilson WH.** 2002. *Coccolithovirus* (*Phycodnaviridae*): characterisation of a new large dsDNA algal virus that infects *Emiliania huxleyi*. Arch. Virol. **147:**1685–1698. http://dx.doi.org/10.1007/s00705-002-0841-3.
5. **Van Etten JL, Graves MV, Müller DG, Boland W, Delaroque N.** 2002. *Phycodnaviridae*—large DNA algal viruses. Arch. Virol. **147:**1479–1516. http://dx.doi.org/10.1007/s00705-002-0822-6.
6. **Wilson WH, Van Etten JL, Allen MJ.** 2009. The *Phycodnaviridae*: the story of how tiny giants rule the world. Curr. Top. Microbiol. Immunol. **328:**1–42. http://dx.doi.org/10.1007/978-3-540-68618-7_1.
7. **Iyer LM, Aravind L, Koonin EV.** 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. **75:**11720–11734. http://dx.doi.org/10.1128/JVI.75.23.11720-11734.2001.
8. **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.
9. **Clerissi C, Desdevises Y, Grimsley N.** 2012. Prasinoviruses of the marine green alga *Ostreococcus tauri* are mainly species specific. J. Virol. **86:**4611–4619. http://dx.doi.org/10.1128/JVI.07221-11.
10. **Clerissi C, Grimsley N, Desdevises Y.** 2013. Genetic exchanges of inteins between prasinoviruses (*Phycodnaviridae*). Evolution **67:**18–33. http://dx.doi.org/10.1111/j.1558-5646.2012.01738.x.
11. **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.
12. **Manrique JM, Calvo AY, Jones LR.** 2012. Phylogenetic analysis of *Ostreococcus* virus sequences from the Patagonian coast. Virus Genes **45:** 316–326. http://dx.doi.org/10.1007/s11262-012-0762-5.
13. **Tidona CA, Schnitzler P, Kehm R, Darai G.** 1998. Is the major capsid protein of iridoviruses a suitable target for the study of viral evolution? Virus Genes **16:**59–66. http://dx.doi.org/10.1023/A:1007949710031.
14. **Mayer JA, Taylor FJR.** 1979. A virus which lyses the marine nanoflagellate *Micromonas pusilla*. Nature **281:**299–301. http://dx.doi.org/10.1038/281299a0.
15. **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.
16. **Not F, Latasa M, Marie D, Cariou T, Vaulot D, Simon N.** 2004. A single species, *Micromonas pusilla* (Prasinophyceae), dominates the eukaryotic picoplankton in the western English Channel. Appl. Environ. Microbiol. **70:**4064–4072. http://dx.doi.org/10.1128/AEM.70.7.4064-4072.2004.
17. **Viprey M, Guillou L, Ferréol M, Vaulot D.** 2008. Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. Environ. Microbiol. **10:** 1804–1822. http://dx.doi.org/10.1111/j.1462-2920.2008.01602.x.
18. **Zhu F, Massana R, Not F, Marie D, Vaulot D.** 2005. Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. FEMS Microbiol. Ecol. **52:**79–92. http://dx.doi.org/10.1016/j.femsec.2004.10.006.
19. **Bellec L, Grimsley N, Desdevises Y.** 2010. Isolation of prasinoviruses of the green unicellular algae *Ostreococcus* spp. on a worldwide geographical scale. Appl. Environ. Microbiol. **76:**96–101. http://dx.doi.org/10.1128/AEM.01799-09.
20. **Bellec L, Grimsley N, Derelle E, Moreau H, Desdevises Y.** 2010. Abundance, spatial distribution and genetic diversity of *Ostreococcus tauri* viruses in two different environments. Environ. Microbiol. Rep. **2:**313–321. http://dx.doi.org/10.1111/j.1758-2229.2010.00138.x.
21. **Chen F, Suttle CA.** 1995. Amplification of DNA polymerase gene fragments from viruses infecting microalgae. Appl. Environ. Microbiol. **61:** 1274–1278.
22. **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.
23. **La Scola B, Campocasso A, N'Dong R, Fournous G, Barrassi L, Flaudrops C, Raoult D.** 2010. Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. Intervirology **53:** 344–353. http://dx.doi.org/10.1159/000312919.
24. **Nagasaki K, Shirai Y, Tomaru Y, Nishida K, Pietrokovski S.** 2005. Algal viruses with distinct intraspecies host specificities include identical intein elements. Appl. Environ. Microbiol. **71:**3599–3607. http://dx.doi.org/10.1128/AEM.71.7.3599-3607.2005.
25. **Ogata H, Raoult D, Claverie J-M.** 2005. A new example of viral intein in Mimivirus. Virol. J. **2:**8. http://dx.doi.org/10.1186/1743-422X-2-8.
26. **Bellec L, Grimsley N, Moreau H, Desdevises Y.** 2009. Phylogenetic analysis of new prasinoviruses (*Phycodnaviridae*) that infect the green unicellular algae *Ostreococcus*, *Bathycoccus* and *Micromonas*. Environ. Microbiol. Rep. **1:**114–123. http://dx.doi.org/10.1111/j.1758-2229.2009.00015.x.
27. **Pietrokovski S.** 1998. Modular organization of inteins and C-terminal autocatalytic domains. Protein Sci. **7:**64–71. http://dx.doi.org/10.1002/pro.5560070106.
28. **Derelle E, Ferraz C, Escande M-L, Eychenié S, Cooke R, Piganeau G, Desdevises Y, Bellec L, Moreau H, Grimsley N.** 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. PLoS One **3:**e2250. http://dx.doi.org/10.1371/journal.pone.0002250.
29. **Fitzgerald LA, Graves MV, Li X, Feldblyum T, Nierman WC, Van Etten JL.** 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A. Virology **358:**472–484. http://dx.doi.org/10.1016/j.virol.2006.08.033.
30. **Fitzgerald LA, Graves MV, Li X, Hartigan J, Pfitzner AJP, Hoffart E, Van Etten JL.** 2007. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic *Chlorella* strain of the heliozoon *Acanthocystis turfacea*. Virology **362:**350–361. http://dx.doi.org/10.1016/j.virol.2006.12.028.
31. **Fitzgerald LA, Graves MV, Li X, Feldblyum T, Hartigan J, Van Etten JL.** 2007. Sequence and annotation of the 314-kb MT325 and the 321-kb

FR483 viruses that infect *Chlorella* Pbi. Virology **358:**459–471. http://dx.doi.org/10.1016/j.virol.2006.08.034.
32. **Moreau H, Piganeau G, Desdevises Y, Cooke R, Derelle E, Grimsley N.** 2010. Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. J. Virol. **84:**12555–12563. http://dx.doi.org/10.1128/JVI.01123-10.
33. **Graves MV, Meints RH.** 1992. Characterization of the major capsid protein and cloning of its gene from algal virus PBCV-1. Virology **188:**198–207. http://dx.doi.org/10.1016/0042-6822(92)90750-J.
34. **Winnepenninckx B, Backeljau T, Wachter RD.** 1993. Extraction of high molecular weight DNA from molluscs. Trends Genet. **9:**407. http://dx.doi.org/10.1016/0168-9525(93)90102-N.
35. **Weynberg KD, Allen MJ, Ashelford K, Scanlan DJ, Wilson WH.** 2009. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. Environ. Microbiol. **11:**2821–2839. http://dx.doi.org/10.1111/j.1462-2920.2009.01991.x.
36. **Grabherr R, Strasser P, Van Etten JL.** 1992. The DNA polymerase gene from chlorella viruses PBCV-1 and NY-2A contains an intron with nuclear splicing sequences. Virology **188:**721–731. http://dx.doi.org/10.1016/0042-6822(92)90527-V.
37. **Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ.** 2011. Removing noise from pyrosequenced amplicons. BMC Bioinformatics **12:**38. http://dx.doi.org/10.1186/1471-2105-12-38.
38. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Gonzalez Peña A, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods **7:**335–336. http://dx.doi.org/10.1038/nmeth.f.303.
39. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
40. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797. http://dx.doi.org/10.1093/nar/gkh340.
41. **Galtier N, Gouy M, Gautier C.** 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12:**543–548.
42. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. **27:**221–224. http://dx.doi.org/10.1093/molbev/msp259.
43. **Weynberg KD, Allen MJ, Gilg IC, Scanlan DJ, Wilson WH.** 2011. Genome sequence of *Ostreococcus tauri* virus OtV-2 throws light on the role of picoeukaryote niche separation in the ocean. J. Virol. **85:**4520–4529. http://dx.doi.org/10.1128/JVI.02131-10.
44. **Nishida K, Kawasaki T, Fujie M, Usami S, Yamada T.** 1999. Aminoacylation of tRNAs encoded by *Chlorella* virus CVK2. Virology **263:**220–229. http://dx.doi.org/10.1006/viro.1999.9949.
45. **Plugge B, Becker B, Wolf AH.** 1999. Several genes in *Chlorella* virus strain CVG-1 encode putative virion components. J. Gen. Virol. **80:**1067–1072.
46. **Castresana J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17:**540–552. http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334.
47. **Abascal F, Zardoya R, Posada D.** 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21:**2104–2105. http://dx.doi.org/10.1093/bioinformatics/bti263.
48. **Ronquist F, Huelsenbeck JP.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572–1574. http://dx.doi.org/10.1093/bioinformatics/btg180.
49. **Darriba D, Taboada GL, Doallo R, Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods **9:**772. http://dx.doi.org/10.1038/nmeth.2109.
50. **Goldman N, Yang Z.** 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11:**725–736.
51. **Muse SV, Gaut BS.** 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. **11:**715–724.
52. **Shapiro B, Rambaut A, Drummond AJ.** 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Mol. Biol. Evol. **23:**7–9. http://dx.doi.org/10.1093/molbev/msj021.
53. **Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. **59:**307–321. http://dx.doi.org/10.1093/sysbio/syq010.
54. **Nei M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York, NY.
55. **Librado P, Rozas J.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25:**1451. http://dx.doi.org/10.1093/bioinformatics/btp187.
56. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.** 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75:**7537–7541. http://dx.doi.org/10.1128/AEM.01541-09.
57. **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.
58. **Lundin D, Gribaldo S, Torrents E, Sjoberg BM, Poole AM.** 2010. Ribonucleotide reduction—horizontal transfer of a required function spans all three domains. BMC Evol. Biol. **10:**383. http://dx.doi.org/10.1186/1471-2148-10-383.
59. **Thomas R, Grimsley N, Escande M-L, Subirana L, Derelle E, Moreau H.** 2011. Acquisition and maintenance of resistance to viruses in eukaryotic phytoplankton populations. Environ. Microbiol. **13:**1412–1420. http://dx.doi.org/10.1111/j.1462-2920.2011.02441.x.
60. **Posada D, Crandall KA, Holmes EC.** 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. **36:**75–97. http://dx.doi.org/10.1146/annurev.genet.36.040202.111115.
61. **Croizier G, Croizier L, Argaud O, Poudevigne D.** 1994. Extension of *Autographa californica* nuclear polyhedrosis virus host range by interspecific replacement of a short DNA sequence in the p143 helicase gene. Proc. Natl. Acad. Sci. U. S. A. **91:**48–52. http://dx.doi.org/10.1073/pnas.91.1.48.
62. **Gibbs MJ, Armstrong JS, Gibbs AJ.** 2001. Recombination in the hemagglutinin gene of the 1918 "Spanish flu." Science **293:**1842–1845. http://dx.doi.org/10.1126/science.1061662.
63. **Javier RT, Sedarati F, Stevens JG.** 1986. Two avirulent herpes simplex viruses generate lethal recombinants in vivo. Science **234:**746–748. http://dx.doi.org/10.1126/science.3022376.
64. **Maeda S, Kamita SG, Kondo A.** 1993. Host range expansion of *Autographa californica* nuclear polyhedrosis virus (NPV) following recombination of a 0.6-kilobase-pair DNA fragment originating from *Bombyx mori* NPV. J. Virol. **67:**6234–6238.
65. **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.
66. **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.
67. **Knoth K, Roberds S, Poteet C, Tamkun M.** 1988. Highly degenerate, inosine-containing primers specifically amplify rare cDNA using the polymerase chain reaction. Nucleic Acids Res. **16:**10932. http://dx.doi.org/10.1093/nar/16.22.10932.
68. **Candrian U, Furrer B, Hofelein C, Luthy J.** 1991. Use of inosine-containing oligonucleotide primers for enzymatic amplification of different alleles of the gene coding for heat-stable toxin type I of enterotoxigenic *Escherichia coli*. Appl. Environ. Microbiol. **57:**955–961.
69. **Zheng L, Gibbs MJ, Rodoni BC.** 2008. Quantitative PCR measurements of the effects of introducing inosines into primers provides guidelines for improved degenerate primer design. J. Virol. Methods **153:**97–103. http://dx.doi.org/10.1016/j.jviromet.2008.07.029.
70. **Dahllöf I.** 2002. Molecular community analysis of microbial diversity. Curr. Opin. Biotechnol. **13:**213–217. http://dx.doi.org/10.1016/S0958-1669(02)00314-2.
71. **Gimenes MV, de Zanotto APM, 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.
72. **Short SM, Rusanova O, Staniewski MA.** 2011. Novel phycodnavirus genes amplified from Canadian freshwater environments. Aquat. Microb. Ecol. **63:**61–67. http://dx.doi.org/10.3354/ame01478.