Photosynthetic genes in viral populations with a large genomic size range from Norwegian coastal waters.

Ruth-Anne Sandaa[1] , Martha Clokie[2], Nicholas H. Mann[3]

[1] University of Bergen, P.O. Box 7800, Department of Biology, N-5020 Bergen, Norway.
[2]Department of Infection, Immunity and Inflammation, Maurice Shock Medical Sciences Building, University of Leicester, Leicester LE1 9HN, United Kingdom
[3]Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom
[1] Corresponding author
e-mail: ruth.sandaa@bio.uib.no

Journal: FEMS Microbial Ecology

Pages: 20

Figures: 2

Tables: 2

Running Title: Photosynthetic genes in uncultured viral populations

Keywords: Cyanophage, *Synechococcus*, photosynthetic genes, viral genome size, Pulse field gel electrophoresis (PFGE).

Abstract

Here we report the diversity of uncultured environmental viruses harbouring photosynthetic genes (*psbA* and *psbD*) in samples from cold sea water (latitude above 60^{o}). The viral community in coastal Norwegian waters was separated according to genome size using pulse field gel electrophoresis. Viral populations within a wide genome size range (31 to 380 kb) were investigated for the presence of the *psbA* and *psbD* genes using PCR, combined with cloning and sequencing. The results show the presence of photosynthetic genes in viral populations from all size ranges. Thus we are able to obtain valuable information about the size class to which viral particles that encode photosynthesis genes belong. The wide genomic size range detected implies that we have observed a different cyanophage profile than has previously been reported. Thus, the method of phage gene detection applied here may represent a truer picture of phage diversity in general or that there is a larger range of size profile for viruses with *psbA* and *psbD* in higher latitudes than for the better studied lower latitudes. Alternatively we may be observing a picture of diversity based on a different set of biases than that from either isolation-based research, or from conventional metagenomic approaches.

1. Introduction

The main prokaryotic component of the picophytoplankton in the photic zone of the world´s oceans consist of marine unicellular cyanobacteria from the genera *Synechococcus* and *Prochlorococcus.* Together they contribute to up to 89 % of the primary production in the oligotrophic regions of the oceans (Li, 1995, Liu*, et al.*, 1997, Partensky*, et al.*, 1999). Cyanophages are viruses that infect cyanobacteria and are, like their hosts, ubiquitous in marine environments. They are found in concentrations up to 10^6 particles/ml in coastal waters during the summer period and are considered a significant factor in determining the dynamics of cyanobacterial populations (Suttle & Chan, 1993, Waterbury & Valois, 1993, Suttle & Chan, 1994).

Genes involved in photosynthesis have recently been detected in cyanophages (Mann, 2003). These genes (*psbA* and *psbD*) codes for two proteins, D1 and D2, that form the reaction centre dimer of photosystem II (PSII). DI is common to all oxygenic phototrophs and has a high turnover rate as a result of photo-damage (Aro*, et al.*, 1993) . Both *psbA* and *psbD* genes have been reported in cyanophages infecting *Synechococcus* (Mann, 2003, Millard*, et al.*, 2004, Mann*, et al.*, 2005, Sullivan*, et al.*, 2006), *Prochlorococcus* (Lindell*, et al.*, 2005, Sullivan*, et al.*, 2005, Sullivan*, et al.*, 2006) and identified in BAC clones and amplicons from environmental samples (Zeidner*, et al.*, 2005). Studies have shown that these phage-encoded photosynthetic genes are expressed in the host after infection (Lindell*, et al.*, 2005, Clokie*, et al.*, 2006a). The expression of these genes may increase the fitness of the phage by ensuring the provision of energy for extended viral replication.

Prochlorococcus is essentially ubiquitous between 40°N and 40°S, but is not found in water where the temperature is less than 10°C (Olson*, et al.*, 1990). This contrasts to *Synechococcus*, which is somewhat broader distributed in that it tolerates a wider range of temperature. *Prochlorococcus* is normally about 10-fold more abundant than *Synechococcus* in the oligotrophic regions of the open oceans (Vaulot*, et al.*, 1995), while *Synechococcus* populations are most abundant both in coastal and colder waters (Reviewed in Partensky*, et al.*, 1999). Most *Synechococcus* phage research has been carried out in low-latitude temperate and tropical waters and there is a lack of knowledge about the diversity of these phages in waters of latitudes above 60°.

The fact that most viral hosts have not been cultured has severely limited studies of viral diversity. One characteristic of viruses, which varies over a wide range and is readily

determined, is the genome size. Reported viral genomes range from a few to 1200 kb (Raoult, *et al.*, 2004, Cann, 2005). Pulsed field gel electrophoresis (PFGE) is a method that provides a separation over the full range of intact viral genome sizes. Lately, this approach has been used in several studies to explore the dynamics in the communities of dsDNA viruses in the marine environment (Wommack, *et al.*, 1999, Steward, *et al.*, 2000, Castberg, *et al.*, 2001, Larsen, *et al.*, 2001, Riemann & Middelboe, 2002, Jiang, *et al.*, 2003, Ovreas, *et al.*, 2003, Larsen, *et al.*, 2004, Sandaa & Larsen, 2006). These studies have shown that the viral assemblage in the marine environment is distributed in a genome size range from approximately 20 to 560 kb. The most dominant populations have genome sizes between 20 and 100 kb (Wommack, *et al.*, 1999, Steward, *et al.*, 2000, Ovreas, *et al.*, 2003, Larsen, *et al.*, 2004, Sandaa & Larsen, 2006), which is also the size range of most isolated marine bacteriophages with dsDNA genomes (Ackermann & DuBow, 1987, Jiang, *et al.*, 2003). Infections by cyanophages were first reported in 1990 (Proctor & Fuhrman, 1990, Suttle, *et al.*, 1990) and isolates of these cyanoviruses have recently been characterised and sequenced (Suttle & Chan, 1993, Waterbury & Valois, 1993, Wilson, *et al.*, 1993, Chen & Lu, 2002, Mann, *et al.*, 2005). Most of the *Synechococcus* phages are tailed phages with dsDNA genomes mostly with genomes in the range 100-200 kb belonging to the family *Myoviridae* (Mann, 2003).

The objective of this study was to investigate the diversity of photosynthetic genes in uncultured cyanophages from Norwegian costal waters. All previous studies of photosynthetic viral genes have been either on isolated viruses (Lindell, *et al.*, 2004, Millard, *et al.*, 2004, Lindell, *et al.*, 2005, Sullivan, *et al.*, 2005, Sullivan, *et al.*, 2006) or using whole viral fractions (Zeidner, *et al.*, 2005, Sullivan, *et al.*, 2006). The approach described here is unique as it allows us to investigate the presence of photosynthesis genes in non-cultured viruses, but where the genome size of the virus is known. This type of information can not be obtained from standard metagenomic data sets. To do this, viral PFGE bands within a wide genomic size range were investigated for the presence of the photosynthetic genes, *psbA* and *psbD*. PCR products with amplicons of these two genes were cloned, sequenced, and phylogenetic analyses were performed.

2. Material and Methods

2.1. Sample collection

Coastal water samples were collected at a station in Raunefjorden (60°16.2'N, 5°12.5'E), south of Bergen, Norway at nine different time points between 13th of April, to 2nd of November 2004 (Table 2). A total volume of 30 l was collected from the 2-meter depth. The samples were collected using a hand pump connected to a flask. Temperature was measured using a STD SAIV a/s SD 204 with a Sea Point fluorometer (SAIV A/S, Environmental Sensors & Systems, Bergen, Norway).

2.2. Concentration of viral communities.

Natural viral communities were concentrated from 25 l of seawater by ultrafiltration. The samples were filtrated through a 0.45 µm pore-size low-protein-binding Durapore membrane filters of 142 mm in diameter (Millipore) to remove zooplankton, phytoplankton and some bacteria. The filtered samples were then concentrated down to a final volume of around 150 to 250 ml, using a 30 000 MW cuttoff spiral-wound Millipore ultrafiltration cartidge (Regenerated Cellulose, PLTK Prep/scale TFF 1 ft2, Millipore). One hundred and forty ml of this concentrate was concentrated further by ultracentifugation (Beckman L8-M with SW-28 rotor, Beckman GmbH, Germany) for 2 h at 25 000 rpm at 10 ^{o}C. The viral pellet was dissolved in 400 µl of SM buffer (0.1 M NaCl, 8 mM $MgSO_4 \cdot 7H_2O$, 50 mM Tris-HCl (pH 8.0), 0.005% (w/v) glycerine). 200 µl was stored at –20ºC for quantitative Real-Time PCR analysis, while 200 µl was used for PFGE analysis

2.3. Pulse field gel electrophoresis (PFGE)

Four virioplankton agarose plugs were made from the 200 µl concentrate. The samples were separated on a 1%w/v SeaKem GTG agarose (FMC, Rockland, Maine) gel in 1 X TBE gel buffer using a Bio-Rad DR-II CHEF Cell (Bio-Rad, Richmond Ca, USA) electrophoresis unit (Wommack, *et al.*, 1999). From each sample point three of the plugs were used, each at a different pulse ramp condition in order to separate the large range of viral genome sizes: i): 1-5 s switch time with 20 h run time for separation of small genome sizes (0-130 kb); ii) 8-30 s switch time with 20 h run time for separation of medium genome sizes (130-300 kb); iii) 20-40 s switch time with 22 h run time for separation of large genome sizes (300-600 kb). A

molecular size standard (lambda ladder and 5 kb ladder (Bio-Rad, Richmond, California)) was run on each side of the gel. Further details of the procedure are found in (Larsen, *et al.*, 2001). The gels were visualized and saved as computer files using the Fujifilm imaging system, LAS-3000.

2.4. PCRs, cloning and DNA sequencing.

PFGE bands to be investigated for the presence of photosynthetic genes were excised from the gel and frozen at –20C^{o}. The DNA was extracted from the gel by use of the GeneClean Turbo kit (BIO101) for extraction of large DNA fragments from agarose gel, following the manufacturer's instructions yielding approximately 10 ng/µl of DNA (total 30 µl). The genomic DNA required for sequencing and PCR were produced by the GenomiPhi DNA amplification kit (Amersham Biosciences) according to the manufacturer's instructions yielding approximately 1µg/µl of DNA. This DNA was used in a PCR with both the primer sets targeting a section of the photosynthetic genes *psbA* and *psbD* (Table 1). PCRs were carried out in a total volume of 50 µl containing: sterile distilled water, PCR buffer (10 x PCR buffer B, Promega, Madison, WI), dNTPs (each 200 nM), primers (each 0.5 µM), 1.5 mM MgCl, 2.5 U *Taq* DNA polymerase (Promega) and template amplicon (1-2 ng). Amplification conditions using the *psbA* primers were as follows: 94°C for 5 min, 10 cycles of 94°C for 30 sec, 64°C (-1°C per cycle) for 30 sec, and 72°C for 1 min. There was then an extension of 2 min at 72°C, followed by 25 cycles of 94°C for 30 sec, 56.5°C for 30 sec, and 72°C for 1 min. The final extension was at 72°C for 10 min. Furthermore, the amplification conditions for the *psbD* primers were: 94°C for 5 min, 35 cycles of 94°C for 1 min, 50°C 1 min, and 72°C for 1 min, and a final extension at 72°C for 10 min. The PCR products were cloned with the TOPO PCR cloning kit (Invitrogen, Paisley, UK) following the manufacturer's description. The resulting reactions were used to transform competent *Escherichia coli* TOP10 (Invitrogen). Fifteen positive clones (white colonies) from each library were picked randomly and transferred by streaking onto agar plates. Positive clones were confirmed by PCR using the M13 primers according to the protocol (Invitrogen). Positive PCR products were purified using the DNA Clean & Concentrator-5 kit (Genetix Limited, New Milton, UK). Five positive PCR products from each cloning reaction were sequenced by cycle sequencing according to the protocol from Perkin Elmer (Foster City, USA) using the cloning primer M13f (Invitrogen) as sequencing primer. Sequences were obtained on the ABI PRISM 3700 sequence analyser (Perkin-Elmer Applied Biosystem, USA).

2.5. Phylogenetic analysis.

Between two and five clones resulted in sequences of good quality and were used in the phylogenetic analysis. Analysis of DNA sequences was carried out by alignment to the closest relative in the GenBank database using BLASTX (Altschul, *et al.*, 1990). Alignments were performed using CLUSTALX (Thompson, *et al.*, 1997). Sequences were initially aligned based on protein sequences. The protein alignment was then used to align the corresponding DNA sequences. Maximum parsimony and neighbor-joining (NJ) analysis were conducted on nucleotide dataset by using the test version of PAUP* 4.0 beta10 (Swofford, 2000). Supports for clades was estimated by means of bootstrap analysis, as implemented in PAUP* using 1000 replicates. The trees were viewed using the TreeView program and rooted with either the *psbA* or *psbD* gene of *Synechocystis* PCC 6803. The nucleotide sequences reported in this paper has been submitted to GenBank and assigned the accession numbers: psbD, DQ787206-DQ787235 and psbA, DQ787236-DQ787256.

3. Results and Discussion.

Here we report the first study to investigate the diversity of uncultured environmental viruses harbouring photosynthetic genes (*psbA* and *psbD*) in samples from cold sea water. The study was performed by investigating nineteen PFGE bands in a broad size range (31-380 kb) from samples taken at different time periods (Fig. 1) based on information on the viral diversity and cyanophage seasonal dynamics in Raunefjorden (Sandaa & Larsen, 2006). The bands which were chosen represented the brightest bands in each of the different size classes. The viral community in Raunefjorden have shown a pronounced seasonal dynamic that correlates with changes in the abundance of possible hosts (Sandaa & Larsen, 2006). Most of the bands examined in the present study were from samples collected between late September and early November. This is at the same time there is a bloom in the *Synechococcus* population in the fjord (Sandaa & Larsen, 2006). The bloom is followed by a peak in the cyanophage numbers, and is accompanied by a major change in the viral community structure (Sandaa & Larsen, 2006). The water temperature in Raunefjorden, at the sampling times was between 5,3 and 13,3°C (Table 2). Over a 10 months period, the water temperature at the sampling depth in Raunefjorden may vary between 5,2 to 15,4°C (Sandaa & Larsen, 2006). As *Prochlorococcus* has not been reported in water with temperatures below 10°C (Olson, *et al.*, 1990) it is reasonable to conclude that the putative uncultured cyanophages presented in this study infect *Synechococcus* strains.

Eleven out of 19 investigated PFGE bands in this study contained detectable photosynthetic genes. The genes were detected in PFGE bands with genomic sizes from 31 to 380 kb (Table 2). Although the size range of genomes containing *psbA* and *psbD* is larger than has been described before, it may be even larger than we report, as bands smaller or larger than this size range were not investigated. Eight of the PFGE bands contained both the *psbA* and *psbD* genes, while three of the PFGE bands had only one of the genes (Table 2). The *psbA* and *psbD* genes are highly conserved (Lindell, *et al.*, 2004, Millard, *et al.*, 2004, Sullivan, *et al.*, 2006), suggesting that they encode functional proteins that may be involved in maintaining host photosynthesis during infections (Clokie, *et al.*, 2006a). Most cultured cyanophages carry both the *psbA* and *psbD* genes (Lindell, *et al.*, 2004, Millard, *et al.*, 2004, Sullivan, *et al.*, 2006), however, some only contain *psbA* (Millard, *et al.*, 2004, Sullivan, *et al.*, 2006). Two of the PFGE bands in this study contained only the *psbA* gene. However, in contrast to earlier published findings, we amplified only the *psbD* gene from one of the PFGE bands (Table 2). This is the first observation of cyanophages carrying the *psbD* gene, only. These observations, along with the fact that *psbD* and *psbA* are very far apart from each other

in *Synechococcus* genomes, implies that the genes have been acquired independently from their hosts.

Our present knowledge on the occurrence and diversity of the photosynthetic genes in cyanophages is mainly based on isolated phages, belonging to two viral families, *Myoviridae* and *Podoviridae* (Mann, 2003, Lindell*, et al.*, 2004, Millard*, et al.*, 2004, Sullivan*, et al.*, 2005, Sullivan*, et al.*, 2006) and from their abundance in metagenomic data sets (Zeidner*, et al.*, 2005, Sullivan*, et al.*, 2006). Metagenomic data sets have been very valuable in raising our awareness of the widespread nature of phage encoded photosynthesis genes but do not provide further information about the phage biology. Although whole and partial genome sequencing of cultured cyanophages has revealed *psbA* and *psbD* genes in *Synechococcus* myoviruses, none have been shown to be in any *Synechococcus* podoviruses (Sullivan*, et al.*, 2006), however it should be noted that very few *Podoviridae* infecting *Synechococcus* have been isolated or sequenced. The genome size in cultured cyanophages varies from approximately 48-200 kb and the trend is for myoviral genomes (normally in the size range 100-200 kb) to be larger than podoviral genomes which normally range from 38-48 kb (Wichels*, et al.*, 1998, Mann, 2003). Forty-five percent of the observed PFGE bands with photosynthetic genes, had genome sizes in this range (95-200 kb), however, 55 % fell outside this normal size range. Three of the populations (S9, S12 and S13) were in the genomic size range 31 to 67 kb, which according to size suggests they might belong to the *Podo-* or *Siphoviridae*. So far, there is no report of *Synechococcus* phages from S*iphoviridae*, and only few isolates reported belonging to *Podoviridae* (Waterbury & Valois, 1993, Fuller*, et al.*, 1998, Chen & Lu, 2002, Sullivan*, et al.*, 2005, Sullivan*, et al.*, 2006). On the other hand, if these viral populations belong to *Myoviridae*, this is also interesting as it suggests that the size range of this taxonomic group must be much broader that earlier reported. The other viral populations (3b, 5b, and S1) were in the genomic size range of 240 to 380 kb. Although it is most likely that these are from *Myoviridae*, they are much bigger than any reported thus far. Another plausible explanation is that these photosynthetic genes are from viral populations infecting photosynthetic picoeukariotes, as viruses infecting algae do have genome sizes in this size range (Sandaa*, et al.*, 2001, Castberg*, et al.*, 2002,Van Etten & Meints, 1999). These *psbA* and *D* genes, however, exhibited highest similarity to other cyanophage and *Synechococcus* sequences in GenBank, not to photosynthetic genes in picoeukariotes. We therefore believe that these genes are present in cyanophages infecting *Synechococcus*. It should though be stressed that these observations are based on genome size alone, and that

physiological properties of the phage, e.g. viral morphology, is essential for correct definition of phage taxonomy.

Phylogenetic analysis of the *psbA* and *psbD* was carried out in PAUP* using both maximum parsimony and a neighbour-joining analysis (NJ). The two methods gave congruent results and those presented here are from the NJ analysis (Fig 2a and b). Most of the clusters of the environmental sequences from this study lack sequences from cultured cyanophage or hosts, which suggests that these sequences may belong to phages that infects as-yet uncultured *Synechococcus* hosts. This is a common observation when using culture-independent approaches for investigation of environmental viruses. The *psbA* sequences fell into several distinct groups. From the top of the tree and downwards, one clade of 7 *psbA* sequences has 87% bootstrap support and consists only of environmental sequences determined in this study. There is then an interesting group (with 76% bootstrap support) which contains both a new *psbA* sequences from this study, an environmental clone from the Monterey Bay, California (Zeidner*, et al.*, 2003) and a known *Synechococcus* myophage S-RSM2. Two smaller clades again contain 2 and 3 new environmental *psbA* sequences, respectively. The two *Synechococcus* strains included in this analysis form a clade with an environmental *psbA* sequence from the Monterey Bay, California (Zeidner*, et al.*, 2003). There is one final large group which contains the remainder of the environmental *psbA* sequenced from Raunefjorden and which at its higher level contains 2 environmental sequences, one from the Red Sea (BAC9D04) and the other from the Mediterranean Sea (V141) (Zeidner*, et al.*, 2005). The sequence BAC9DO4 is from a putative podophage based on the fact that genes upstream of it are podophage like (Zeidner & Béjà, 2004), but the phage has not been isolated so the actual morphology has not been confirmed. Thirty eight percent of the *psbA* sequences from this study clustered in this large clade. If these viral populations are podophages, infecting *Synechococcus*, this is in contrasts to what has been shown using different methods in other environments. Therefore it may be that the method of phage gene detection employed here may represent a truer picture of phage diversity in general, or that phage diversity in high latitudes has a different profile to that from better studied lower latitudes. It may alternatively be that the method used in this study simply has a different set of biases than isolation-based or strict metagenomic approaches. Although the phage *psbA* sequences isolated here cluster with the sequence in BAC9D4, it should be emphasised that there is significant variation at the nucleotide level, as there were a sequence differences of approximately 25 %.

The environmental *psbD* sequences obtained in this study fell into several distinct groups (Fig 2b), showing greatest similarity to three *Synechococcus* myophages, syn1, syn10 and S-RSM2 (Millard, *et al.*, 2004, Sullivan, *et al.*, 2006). It is worth mentioning that the *psbD* gene has only been detected in cyanophage isolates belonging to *Myoviridae* (Sullivan, *et al.*, 2006). The isolate, S-RSM2 was isolated from the Gulf of Aquaba (Millard, *et al.*, 2004). *PsbD* appears to be a much better phylogenetic marker than *psbA* as all of these new *psbD* sequences were distinguishable from their host *psbD* genes, in contrast to the phage-encoded *psbA* genes that form a clade together with their host genes (Fig 2a and b). Thus, compared to the phage encoded *psbA* genes, these new *psbD* genes might have had a longer purifying selection time resulting in a clear divergence of the viral and host *psbD* genes. It may alternatively be the case that there is less evolutionary constraint on *psbD* than *psbA* genes and thus the phage-encoded versions can evolve more freely than phage-encoded *psbA*.

Interestingly, the *psbD* genes from the different clones partly cluster according to their original PFGE band. In contrast, the *psbA* genes did not show any such relationship. Thus, it seems that the level of diversity observed in the *psbD* gene is less than that observed in *psbA* where multiple very closely related genes were recovered from a single band. This might imply that the genes are moving or evolving (or both) independently, with different selection pressures that also would result in phylogenetic differences. Indeed, it may be that a specific phage 'population' may have multiple versions of psbA. Alternatively, differences between the trees may also be attributed to the fact that each PFGE band may consist of more than one viral population, with a similar genome sizes and so the clone library of one viral band can thus contain photosynthetic genes from different, but genetically similar cyanophages. Although, some of the photosynthetic genes in this study might originate from host genes, the fact that all the *psbD* sequences from Raunefjorden displayed greatest similarity to isolated cyanophages, and not to their hosts, is inconsistent with this assumption, confirming the reliability of both the viral concentration step and the PFGE analysis. In summary the *psbD* analysis shows a clear diversity of phage encoded photosynthetic gene, whilst the *psbA* analysis appears to suffer from homoplasy, despite this, it is still useful as it has hinted at the family affiliation to which 8 of the newly sequenced phage populations may belong.

Using PFGE for studies of the viral community, it is possible to detect between 10^5 - 10^6 viral particles per ml (Wommack, *et al.*, 1999, Steward, 2001). One important issue is the size of the viral genome that will influence the sensitivity of the method. For viruses with larger genome sizes, the detection limit will be lower, e.g. a viral genome of 300 kb might be detected down to 10^5 particles per ml (Steward, 2001). In our analysis we collected 25 l of

sample that was concentrated by a factor of approximately 100. If we assume a detection limit of 10^5, we need to have 10^3 particles per ml of one population to result in a signal on the PFGE gel.

In order to produce sufficient DNA from the different PFGE bands, we applied a technique based on whole genome amplification, using the enzyme φ 29 DNA polymerase. The φ 29 DNA polymerase is an enzyme which is widely used for rolling-circle amplification of plasmids and circular DNA templates (Dean, *et al.*, 2002, Detter, *et al.*, 2002). The strand-displacing enzyme has proofreading activity, is extremely sensitive, and has been shown to amplify DNA of up to 70 kb (Blanco, *et al.*, 1989). This polymerase has also been used for environmental whole-genome amplification of bacterial and viral communities(Abulencia, *et al.*, 2006, Angly, 2006) and for a variety of genetical applications including sequencing and microsatellite marker and single nucleotide polymorphism (SNP) analysis of single individuals (Jiang, *et al.*, 2005). These results have demonstrated that the introduction of bias may be due to the sizes of DNA templates (possibly sheared during extraction and mixing), random primer availability, and stochastic effects of amplifying from very low concentrations of template (Jiang, *et al.*, 2005, Abulencia, *et al.*, 2006). Another bias may be the differential cloning efficiency of amplified DNA compared to un-amplified DNA (Abulencia, *et al.*, 2006). Most of these biases will not be valid in our study as we used the multiple displacement amplification (MDA) approach to produce DNA that were used in a second PCR with specific primers. However, as our technique does apply several amplification steps, we should be aware of errors in the PCR product that might place clones from the same viral population at different positions in the tree.

Our objective with this study was to describe the distribution and diversity of photosynthetic genes in different genomic size ranges of viral populations, using a culture-independent approach such as PFGE in combination with cloning and sequencing. With PFGE it is possible to gain information about the dominant and presumably the most active viral populations in the samples, without the need for a cultureable host. So far, most knowledge on photosynthetic genes in phages is based on cultured cyanophages isolated using two cyanobacterial strains; *Synechococcus* WH7803 and *Prochlorococcus* Med4 as hosts. Recent estimates have shown that more than 99 % of marine bacteria (potential phage hosts) are unculturable (Rappé & Giovannoni, 2003), supporting the importance of a culture- independent approach for studies of genetically and functional viral diversity in environmental samples. Compared to other culture-independent approaches, e.g. metagenome cloning (Zeidner, *et al.*, 2005), PFGE also provides us information about the genome size of the viruses. Such

information can be used to suggest the likely family that these viruses belong to. Using this approach we discovered a much higher diversity in viral populations with photosynthetic genes than previously reported. Our results document the naturally occurring genetic diversity among uncultured environmental viruses in samples from cold sea water. Furthermore, the interesting findings that viral populations of different sizes harbour photosynthetic genes that were phylogenetically similar supports the idea of promiscuous horizontal gene transfer of gene modules within a common cyanophage gene pool. Thus, it might be hypothesized that marine cyanophages could have played a vital role in the evolution of photosynthetic gene diversity by providing an accessible pool of portable genes and facilitating the reshuffling, acquisition and exchange of such genetic material.

Acknowledgement

We are grateful to Thomas Sørlie for collecting the sea water samples and Mette Hordnes (UiB) for providing the water temperature data. This study was performed with financial support from Research Council of Norway project: Biodiversity patterns: blooms versus stable coexistence in the lower part of marine food webs and EU project: Bacterial single-cell approaches to the relationship between diversity and function in the sea (BASIC), Contract number: EVK3-CT-2002-00078 and

Table 1.

Primers targeting the photosynthetic genes *psb*A and *psb*D.

Primer name	**Sequence 5'-3'**	**Approximate product size (bp)**	**References**
psbAf	GTNGAYATHGAYGGNATHMGNGARCC	800	(Millard, *et al.*, 2004)
psbAr	GGRAARTTRTGNGC		
psbDf	GGNTTYATGCTNMGNCARTT	380	(Clokie, *et al.*, 2006b)
psbDr	CKRTTNGCNGTVAYCAT		

Table 2. Information about size range and sampling dates of investigated PFGE bands for the presence of the photosynthetic genes *psb*A and *psb*D, water temperature and number of cyanophages in the samples. The number of cyanophages was determined by real-time PCR, respectively (Sandaa & Larsen, 2006).

Viral bands	Date of sampling	Genome size (Kb)	*psbA* gene	*psbD* gene	Water temperature (C^o)	Cyanophages Particles (10^3/ml)[a]
2b	29.jun	300	-	nd	7,8	2
3b	29.jun	255	+	-	7,8	2
4b	29.jun	192	+	+	7,8	2
5b	19.aug	380	+	+	9,2	17
6b	10.aug	140	+	+	8,0	nd
7b	11.may	140	+	-	6,1	11
8b	11.may	200	-	-	6,1	11
1C2	13.apr	75	-	nd	5,3	3
S1	02.nov	240	+	+	13,3	21
S2	02.nov	200	+	+	13,3	21
S5	02.nov	165	-	-	13,2	21
S6	02.nov	145	-	nd	13,2	21
S7	02.nov	139	-	nd	13,2	21
S8	02.nov	95	+	+	13,2	21
S9	02.nov	67	+	+	13,2	21
S10	06.oct	64	-	nd	11,3	30
S11	06.oct	46	-	nd	11,3	30
S12	12.oct	44	+	+	11,6	72
S13	28.sep	31	-	+	11,5	21

nd= not determined

[a] Data published in (Sandaa & Larsen, 2006).

Figure legends:

Fig.1

Schematic outline of viral populations determined by PFGE. Viral populations are defined by genome size, and the outline is based on three different electrophoresis runs for each viral concentrate. Numbers indicate bands that have sliced out and tested for the presence of *psbA* and *psbD* genes by PCR. Red dots: presence of both the *psbA* and *psbD* genes, Green dots: presence of the *psbA* gene, Blue dots: presence of the *psbD* gene.

Fig. 2.

Phylogenetic affiliation of the *psbA* (a) and *psbD* (b) sequences from the clones isolated from PFGE bands (the first number indicates the PFGE band of origin (Table 2) and the second number indicates the clone number), and representative *psbA* and *psbD* sequences from six marine cyanomyoviruses; P-SSM4, S-RSM2, S-RSM28, S-WHM1, S-PM2, S-RM88, (GenBank accession no: NC006884, AJ628768, AJ629221, AJ628769, AJ630128, AJ629075), and the *psbA* gene from a putative podovirus BAC9D04 (a) (GenBank accession no: AY456121). *PsbA* genes from three clones; eBAC65-3, RED-132-6-6 and V141 were also included (a) (GenBank accession no: AY176623, AY176632, AY713429), and *psbD* genes from three marine cyanomyoviruses; syn1, syn10 and syn30 (DQ473702, DQ473703, DQ473700). Both *psbA* and *psbD* genes from the cyanobacteria *Prochlorococcus* MED4, *Synechococcus* sp. WH7803 and *Synechococcus* sp. WH8103 were included in the analysis (GenBank accession no: NC 005072 and DQ473718, AF156980, DQ473687, DQ473716, respectively). Trees were rooted with either the *psbA* or *psbD* gene from the freshwater cyanobacterium *Synechocystis* sp. PCC 6803 (GenBank accession no: X58825, NC 000911, respectively). Text in bold and colour indicates the clones from this study retrieved from PFGE bands in the size range; Blue: < 100 kb, Black: 100-199 kb, Green: 200-299 kb, Red: >300 kb. Bootstrap values were generated with 1000 replicates; values < 75 is not shown. Scale bar represents 0.1 substitutions per site.

References

Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie E L, Hazen TC & Keller, M. (2006) Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl Environ Microbiol* **72**: 3291-3301.

Ackermann H-W & DuBow MS (1987) *Viruses of prokaryotes*. CRC Press Inc., Boca Raton, FL.

Altschul S, Gish W, Miller W, Myers E & Lipman D (1990) Basic local aligment search tool. *J. Mol. Biol.* **215**: 403-410.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M, Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S, Suttle, CA & Rohwer F (2006) The Marine Viromes of Four Oceanic Regions. *PLoS Biol* **4**: e368.

Aro E-M, Virgin I & Andersson B (1993) Photoinhibition of photosystem-2-inactivation, protein damage and turnover. *Biochim Biophys Acta.* **1143**: 113-134.

Blanco L, Bernad A, LÃ¡zaro JM, MartÃ-n G, Garmendia C & Salas M (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem* **264**: 8935-8940.

Cann A (2005) *Principles of Molecular Virology.* Elsevier Academic Press,, Burlington, MA, USA.

Castberg T, Larsen A, Sandaa RA, Brussaard CPD, Egge JK, Heldal M, Thyrhaug R, van Hannen EJ & Bratbak G (2001) Microbial population dynamics and diversity during a bloom of the marine coccolithophorid *Emiliania huxleyi* (Haptophyta). *Marine Ecology-Progress Series* **221**: 39-46.

Castberg T, Thyrhaug R, Larsen A, Sandaa RA, Heldal M, Van Etten JL & Bratbak G (2002) Isolation and characterization of a virus that infects *Emiliania huxleyi* (Haptophyta). *Journal of Phycology* **38**: 767-774.

Chen F & Lu J (2002) Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol* **68**: 2589-2594.

Clokie MRJ, Shan J, Bailey S, Jia Y, Krisch HM, West S & Mann NH (2006a) Transcription of a "photosynthetic" T4-type phage that infects marine cyanobacteria. *Environmental Microbiology* **8**: 827-835.

Clokie MRJ, Millard AD, Mehta JY & Mann NH (2006b) Virus isolation studies suggest short-term variations in abundance in natural cyanophage populations of the Indian Ocean. *J Mar Biol Ass UK* **86**: 499-505.

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M & Lasken RS (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A* **99**: 5261-5266.

Detter JC, Jett JM, Lucas SM, Dalin E, Arellano AR, Wang M, Nelson JR, Chapman J, Lou Y, Rokhsar D, Hawkins TL & Richardson PM (2002) Isothermal strand-displacement amplification applications for high-throughput genomics. *Genomics* **80**: 691-698.

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. *Applied and Environmental Microbiology* **64**: 2051-2060.

Jiang S, Fu W, Chu W & Fuhrman JA (2003) The vertical distribution and diversity of marine bacteriophage at a station off Southern California. *Microbial Ecology* **45**: 399-410.

Jiang Z, Zhang X, Deka R & Jin L (2005) Genome amplification of single sperm using multiple displacement amplification. *Nucleic Acids Res* **33**: e91.

Larsen A, Flaten GAF, Sandaa RA, Castberg T, Thyrhaug R, Erga SR, Jacquet S &

Bratbak G (2004) Spring phytoplankton bloom dynamics in Norwegian coastal waters: Microbial community succession and diversity. *Limnology and Oceanography* **49**: 180-190.

Larsen A, Castberg T, Sandaa RA, Brussaard CPD, Egge J, Heldal M, Paulino A, Thyrhaug R, van Hannen EJ & Bratbak G (2001) Population dynamics and diversity of phytoplankton, bacteria and viruses in a seawater enclosure. *Marine Ecology-Progress Series* **221**: 47-57.

Li WKW (1995) Composition of ultraphytoplankton in the central North-Atlantic. *Mar Ecol Prog Ser* **122**: 1-8.

Lindell D, Jaffe JD, Johnson ZI, Church GM & Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86.

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F & Chisholm SW (2004) Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 11013-11018.

Liu HB, Nolla HA & Campbell L (1997) Prochlorococcus growth rate and contribution to primary production in the equatorial and subtropical North Pacific Ocean. *Microb Ecol* **12**: 39-47.

Mann NH (2003) Phages of the marine cyanobacterial picophytoplankton. *FEMS Microbiology Reviews* **27**: 17-34.

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

Millard A, Clokie MRJ, Shub DA & Mann NH (2004) Genetic organization of the psbAD region in phages infecting marine Synechococcus strains. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 11007-11012.

Olson RJ, Chisholm SW, Zettler ER, Altabet MA & Dusenberry JA (1990) Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. *Deep-Sea Research Part a-Oceanographic Research Papers* **37**: 1033.

Ovreas L, Bourne D, Sandaa RA, Casamayor EO, Benlloch S, Goddard V, Smerdon G, Heldal M & Thingstad TF (2003) Response of bacterial and viral communities to nutrient manipulations in seawater mesocosms. *Aquatic Microbial Ecology* **31**: 109-121.

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

Proctor LM & Fuhrman JA (1990) Viral Mortality of Marine-Bacteria and Cyanobacteria. *Nature* **343**: 60-62.

Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M & Claverie J (2004) The 1.2-megabase genome sequence of Mimivirus. *Science* **306**: 1344-1350.

Rappé M & Giovannoni S (2003) The uncultured microbial majority. *Annu Rev Microbiol* **57**: 369–394.

Riemann L & Middelboe M (2002) Stability of bacterial and viral community compositions in Danish coastal waters as depicted by DNA fingerprinting techniques. *Aquatic Microbial Ecology* **27**: 219-232.

Sandaa RA & Larsen A (2006) Seasonal variations in viral-host populations in Norwegian coastal waters: Focusing on the cyanophage community infecting marine Synechococcus species. *Appl Environ Microbiol* **72**: 4610-4618.

Sandaa RA, Heldal M, Castberg T, Thyrhaug R & Bratbak G (2001) Isolation and characterization of two viruses with large genome size infecting Chrysochromulina ericina (Prymnesiophyceae) and Pyramimonas orientalis (Prasinophyceae). *Virology* **290**: 272-280.

Steward GF (2001) Fingerprinting viral assemblages by pulsed field gel electrophoresis (PFGE). *Methods in Microbiology, Vol 30,* pp. 85-103.

Steward GF, Montiel JL & Azam F (2000) Genome size distributions indicate variability and similarities among marine viral assemblages from diverse environments. *Limnology and Oceanography* **45**: 1697-1706.

Sullivan MB, Coleman ML, Weigele P, Rohwer F & Chisholm SW (2005) Three Prochlorococcus cyanophage genomes: Signature features and ecological interpretations. *Plos Biology* **3**: 790-806.

Sullivan MB, Coleman M, Weigele P, Rohwer F & Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *Public Library of Science* **3**: e144.

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 Biology* **4**: e234.

Suttle CA & Chan AM (1993) Marine Cyanophages Infecting Oceanic and Coastal Strains of Synechococcus - Abundance, Morphology, Cross-Infectivity and Growth-Characteristics. *Marine Ecology-Progress Series* **92**: 99-109.

Suttle CA & Chan AM (1994) Dynamics and Distribution of Cyanophages and Their Effect on Marine Synechococcus Spp. *Applied and Environmental Microbiology* **60**: 3167-3174.

Suttle CA, Chan AM & Cottrell MT (1990) Infection of Phytoplankton by Viruses and Reduction of Primary Productivity. *Nature* **347**: 467-469.

Swofford DL (2000) *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods).* Sinauer Associates, Sunderland MA.

Thompson JD, Gibson T, Plewniak F, Jeanmougin F & Higgins DG (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.

Van Etten JL & Meints RH (1999) Giant viruses infecting algae. *Annual Review of Microbiology* **53**: 447-494.

Vaulot D, Marie D, Olson RJ & Chisholm SW (1995) Growth of Prochlorococcus, a photosynthetic prokaryote, in the equatorial Pacific ocean. *Science* **268**: 1480-1482.

Waterbury JB & Valois FW (1993) Resistance to Cooccurring Phages Enables Marine Synechococcus Communities to Coexist with Cyanophages Abundant in Seawater. *Applied and Environmental Microbiology* **59**: 3393-3399.

Wichels A, Biel SS, Gelderblom HR, Brinkhoff T, Muyzer G & Schütt C (1998) Bacteriophage diversity in the North Sea. *Appl Environ Microbiol* **64**: 4128-4133.

Wilson WH, Joint IR, Carr NG & Mann NH (1993) Isolation and Molecular Characterization of 5 Marine Cyanophages Propagated on Synechococcus Sp Strain Wh7803. *Applied and Environmental Microbiology* **59**: 3736-3743.

Wommack KE, Ravel J, Hill RT, Chun JS & Colwell RR (1999) Population dynamics of Chesapeake bay virioplankton: Total-community analysis by pulsed-field gel electrophoresis. *Applied and Environmental Microbiology* **65**: 231-240.

Zeidner G & Béjà O (2004) The use of DGGE analyses to explore eastern Mediterranean and Red Sea marine picophytoplankton assemblages. *Environ Microbiol* **6**: 528–534.

Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G & Beja O (2005) Potential photosynthesis gene recombination between Prochlorococcus and Synechococcus via viral intermediates. *Environmental Microbiology* **7**: 1505-1513.

Zeidner G, Preston C, Delong EF, Massana R, Post AF, Scanlan DJ & Béjà O (2003) Molecular diversity among marine picophytoplankton as revealed by psbA analyses. *Environ Microbiol* **5**: 212–216.

1

Fig 1

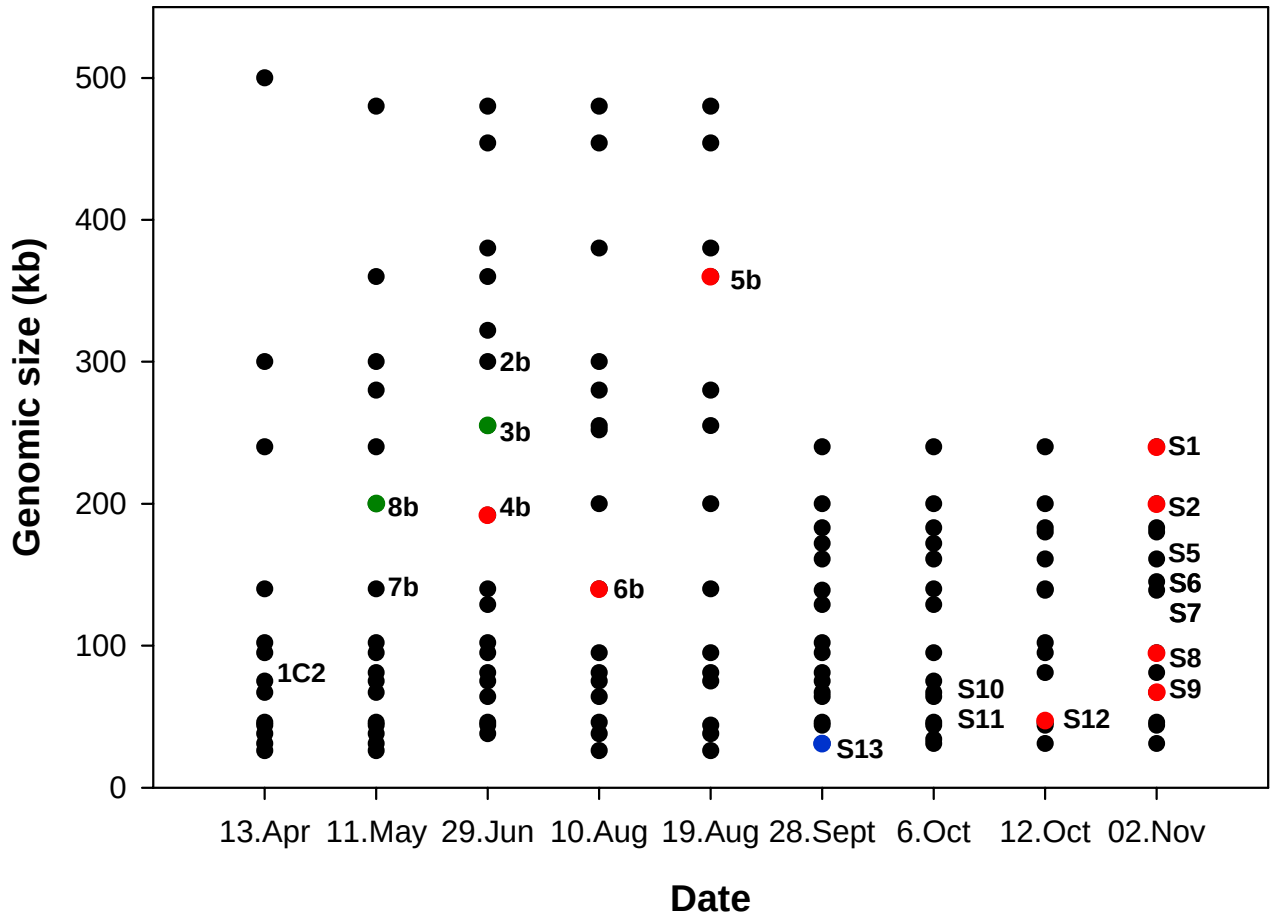

Genomic size (kb)
500
400
300
200
100
0
13.Apr
11.May
29.Jun
10.Aug
19.Aug
28.Sept
6.Oct
12.Oct
02.Nov
Date
1C2
8b
7b
3b
2b
4b
6b
5b
S13
S10
S11
S12
S1
S2
S5
S6
S7
S8
S9

Fig 2a

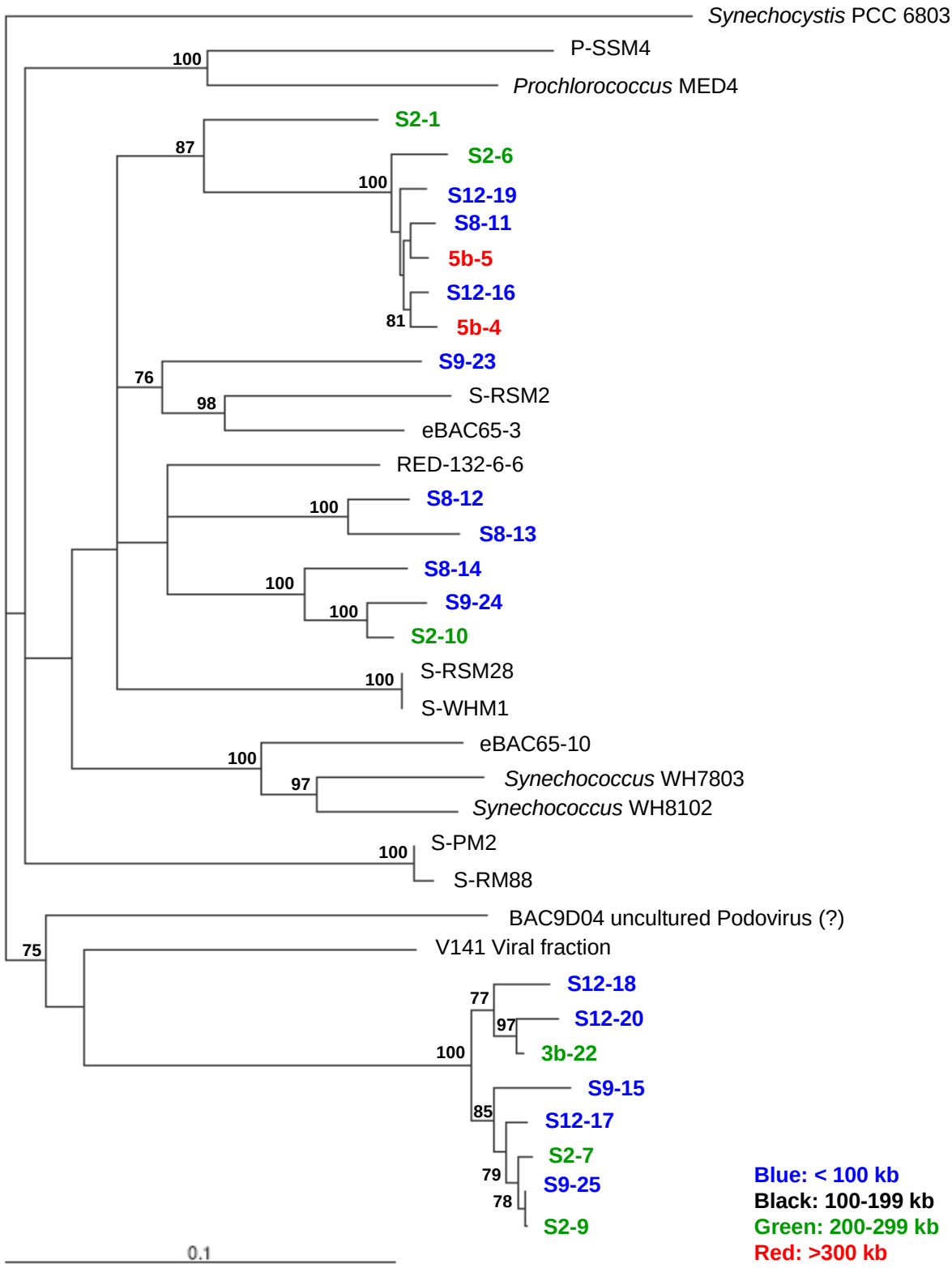
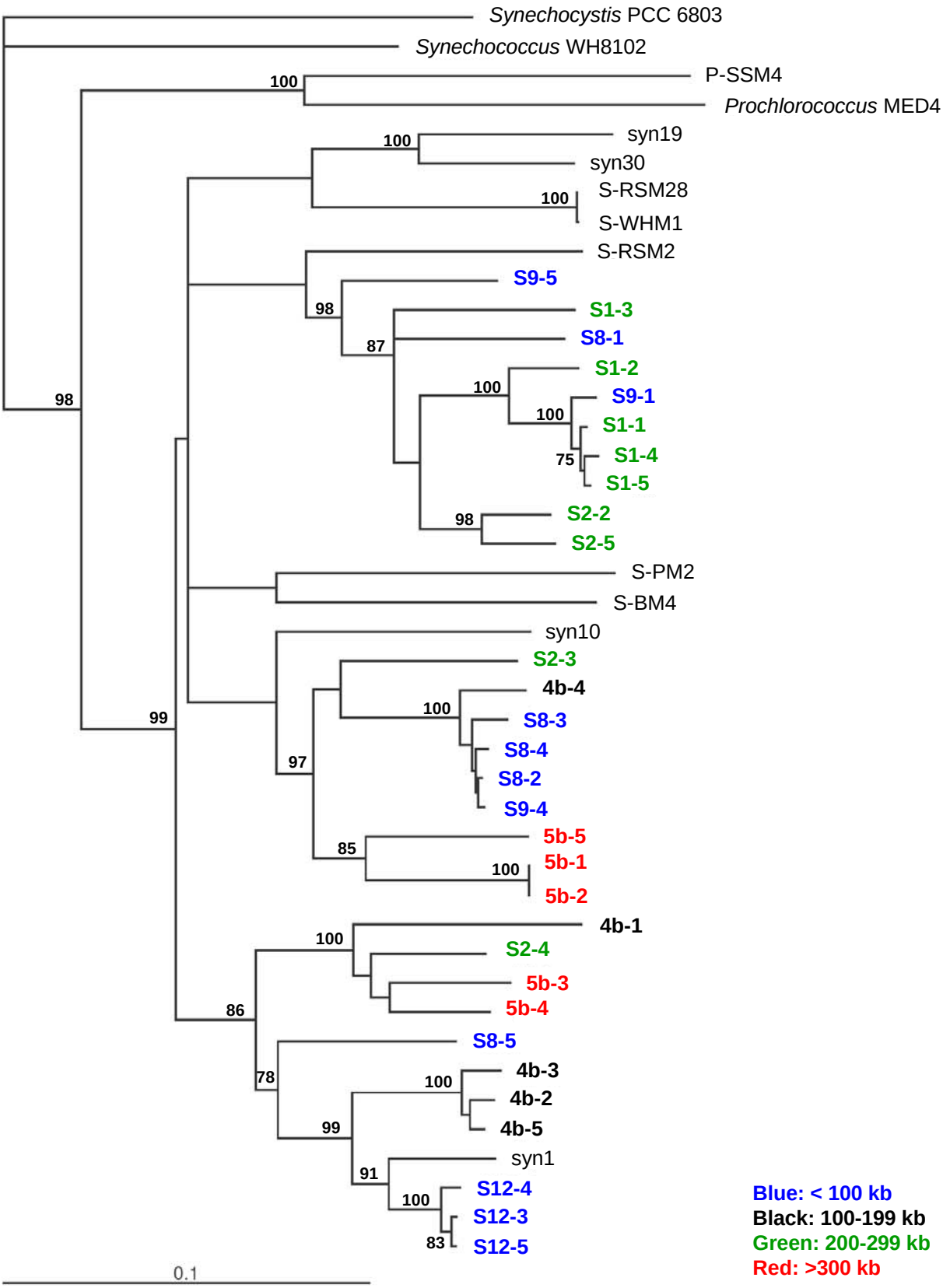

Synechocystis PCC 6803
P-SSM4
100
Prochlorococcus MED4
S2-1
87
S2-6
100
S12-19
S8-11
5b-5
S12-16
81
5b-4
S9-23
76
S-RSM2
98
eBAC65-3
RED-132-6-6
S8-12
100
S8-13
S8-14
100
S9-24
100
S2-10
S-RSM28
100
S-WHM1
eBAC65-10
100
Synechococcus WH7803
97
Synechococcus WH8102
S-PM2
100
S-RM88
BAC9D04 uncultured Podovirus (?)
75
V141 Viral fraction
S12-18
77
S12-20
97
3b-22
100
S9-15
85
S12-17
S2-7
79
S9-25
78
S2-9
0.1
Blue: < 100 kb
Black: 100-199 kb
Green: 200-299 kb
Red: >300 kb

Fig2b


Synechocystis PCC 6803
Synechococcus WH8102
P-SSM4
Prochlorococcus MED4
syn19
syn30
S-RSM28
S-WHM1
S-RSM2
S9-5
S1-3
S8-1
S1-2
S9-1
S1-1
S1-4
S1-5
S2-2
S2-5
S-PM2
S-BM4
syn10
S2-3
4b-4
S8-3
S8-4
S8-2
S9-4
5b-5
5b-1
5b-2
4b-1
S2-4
5b-3
5b-4
S8-5
4b-3
4b-2
4b-5
syn1
S12-4
S12-3
S12-5
100
100
100
98
98
87
100
100
75
98
98
99
97
100
85
100
100
86
78
100
99
91
100
83
0.1
Blue: < 100 kb
Black: 100-199 kb
Green: 200-299 kb
Red: >300 kb