Viral assemblage variation in an Arctic shelf seafloor

Running Title: Viromes of Barents Sea sediment

Tan Thi Nguyen[1,2], Espen Mikal Robertsen[3], Bjarne Landfald[2,*]

[1]Centre for Research-based Innovation on Marine Bioactives and Drug Discovery (MabCent-SFI), UiT

The Arctic University of Norway, N-9037 Tromsø

[2]Norwegian College of Fishery Science, Faculty of Biosciences, Fisheries and Economics, UiT The Arctic

University of Norway, Breivika, N-9037 Tromsø, Norway

[3]NorStruct, Department of Chemistry, Faculty of Science and Technology, UiT The Arctic University of

Norway, Breivika, N-9037 Tromsø, Norway

*Corresponding author: bjarne.landfald@uit.no

ASTRACT: Spatial differences in microbial communities are observable even in habitats with

moderate environmental variation, such as within the pelagic zone or seafloor of

geographically finite regions of the oceans. Here we explore if biogeographical variations are

manifested at this level also in the structure of viral assemblages by comparing DNA viromes

from the Barents Sea upper seafloor, collected at five geographically separated locations.

Twenty-seven to forty-four percent of the open reading frames showed significant similarity

to genes of viral genomes in the Refseq database. The majority of the identified open reading

frames, i.e. 86 to 95%, were affiliated with sequences of single-stranded DNA (ssDNA)

viruses, but the ssDNA virus genetic material was likely strongly overrepresented due to the

use of phi29 DNA polymerase for amplifying viral DNA. The majority of ssDNA virus

sequences originated from the *Microviridae* family of phages and the eukaryotic Circular

Rep-encoding ssDNA (CRESS-DNA) viruses. The sediment virus assemblages showed

higher overall similarity to counterparts from deep sea sediment of the Pacific Ocean than to

27   e.g. Arctic Ocean pelagic viromes, supporting the presence of common compositional features

28   in sediment viral assemblages across continental-scale geographical separations. The Barents

29   Sea viromes grouped biogeographically in accordance with the south-north environmental

30   division of this Arctic sea by the oceanic polar front, thereby mirroring a corresponding 16S

31   rRNA gene based biogeographical division of the bacterial communities. However,

32   compositional differences in the eukaryotic virus assemblages rather than the bacteriophages

33   appeared as the primary basis for this spatial separation.

34

35

36
38

INTRODUCTION

Viruses and virus-like particles are the most abundant biological entities in marine ecosystems and they have an impact on processes like horizontal gene transfer, microbial community structuring and biogeochemical cycling (Bergh et al. 1989, Fuhrman 1999, Weinbauer 2004, Weinbauer & Rassoulzadegan 2004, Suttle 2005, Bouvier & del Giorgio 2007, Yang et al. 2010). Their genomes may have double-stranded (ds) or single-stranded (ss) DNA or ds or ss RNA. Traditionally the dsDNA group has been considered the most abundant in marine environments (Wommack & Colwell 2000, Breitbart et al. 2002, Breitbart et al. 2004, Steward & Preston 2011, Fancello et al. 2013). In recent years, high abundancies of ssDNA viruses have also been reported in various aquatic environments, including marine waters (Angly et al. 2006, Tucker et al. 2011, Labonté & Suttle 2013a, b), fresh water (Lopez-Bueno et al. 2009, Roux et al. 2012), deep seafloor sediments (Yoshida et al. 2013) and methane seep sediments (Bryson et al. 2015). However, it has become increasingly clear that the high prevalence of ssDNA viruses to a large extent is a consequence of amplification bias caused by the use of phi29 DNA polymerase based multiple-displacement DNA amplification (MDA) to obtain sufficient metagenomic material for high-throughput sequencing (Kim & Bae 2011). While the majority of marine dsDNA viral genomes range from 25 to 70 kilobases (kb) in size (Steward et al. 2000), circular ssDNA viruses are small both with respect to physical and genomic size, as the latter largely vary in the range from 1 to 9 kb (Fauquet et al. 2005). Subsequently, the ssDNA viruses are not easily enumerated by epifluorescent microscopy or flow cytometry viral counts because of weak fluorescence signal (Tomaru & Nagasaki 2007, Holmfeldt et al. 2012).

Temporal and geographical variations in aquatic viral assemblages are well-documented (Bergh et al. 1989, Wommack & Colwell 2000, Short & Suttle 2003, Clasen et al. 2008, Danovaro et al. 2008, Yang et al. 2010, Helton et al. 2012, Parsons et al. 2012,

64 Hurwitz et al. 2014) and these variations are at some level presupposed to be related to

65 concomitant variations in the host communities although these relationships may be complex.

66 Uncertainty factors include e.g. lytic versus temperate strategies of the viruses and r versus K

67 type strategies of the hosts (Suttle 2007). Conceptual models like "kill the winner", implying

68 that the fastest growing host populations in a community are the ones most susceptible to viral

69 infections (Thingstad 2000), have been fruitful in explaining the viral contribution to

70 maintenance of high host diversities in natural environments.

71   While there has been a vast increase in the available genomic information from marine

72 microbiotas in the last decade owing to the implementation of next-generation sequencing

73 technologies, the number of isolated and characterized virus-host systems is still limited. Due

74 to their ecological importance marine primary producers, including the ubiquitous

75 *Prochlorococcus* and *Synechococcus* among the prokaryotes and diatoms and dinoflagellates

76 among the eukaryotes, have been principal targets of such studies (Sullivan et al. 2003,

77 Nagasaki 2008, Avrani et al. 2011, Kimura & Tomaru 2015). Among heterotrophic bacteria,

78 phages infecting the *Bacteroidetes* species *Cellulophaga baltica* has been thoroughly

79 characterized by Holmfeldt et al. (2007) and later infection studies of the ubiquitous, but hard-

80 to-cultivate pelagic groups SAR11 and SAR116 have been added to the list (Kang et al. 2013,

81 Zhao et al. 2013). To evade the limitations caused by the consistent lack of cultivability of the

82 large majority of marine prokaryotes, single-cell genomics based approaches have been

83 introduced to identify and characterize individual bacterioplankton-virus systems (Roux et al.

84 2014b, Labonté et al. 2015). To our knowledge, no isolated virus-host system from marine

85 sediments has been characterized.

86   Although a large fraction of the individual or contiguous gene sequences resulting

87 from metagenomic sequencing projects on aquatic viromes have been classified as having

88 unknown function and taxonomic affiliation (Breitbart et al. 2004, Edwards & Rohwer 2005,

4

89 Angly et al. 2006, Desnues et al. 2008, Lopez-Bueno et al. 2009, Roux et al. 2012, Fancello et

90 al. 2013, Yoshida et al. 2013), such projects have still given valuable insights into

91 compositional and biogeographical variations. The pioneering shotgun cloning based study by

92 Breitbart et al. (2004) indicated a higher viral diversity in Californian coastal sediment than in

93 the adjacent pelagic water masses. The two assemblage types were also phylogenetically

94 distinct. Yoshida et al. (2013) have compared pyrosequence data from three deep-sea

95 sediment viromes and conclude they show marked similarities that distinguish them from e.g.

96 sea water viromes, despite spatial separations by more than 1000 km and marked host

97 community and environmental differences.

98   The objective of the present study was to complement the hitherto fragmentary

99 knowledge on diversity variations in seafloor viromes by adding data from an oceanic area

100 that deviate both geographically and environmentally from the ones addressed in preceding

101 studies in the field. Sampling at different sites, separated by up to 640 km, in the Barents Sea

102 shelf seafloor opened for elucidation of viral biogeographical variations within a coherent and

103 environmentally moderately variable shelf sediment area. In a recent study based on the same

104 sediment material, 16S rRNA gene based prokaryotic community variation was linked to the

105 environmental and geographic separation of the Barents Sea by the oceanic polar front

106 (Nguyen & Landfald 2015). Hence, the present analysis opened the possibility of elucidating

107 if coincident biogeographical patterns could be observed for bacterial and bacteriophage

108 assemblages originating from the same material.

109

110

111   **MATERIALS AND METHODS**

112

113   **Sampling**

114   Sediment samples were collected by Van Veen grab within three days (20 - 23 of May 2009)

115   from five locations separated by up to 640 km (Fig. 1) in the western part of the Barents Sea.

116   The three southern sampling stations (1 - 3) were south of the polar front, hence dominated by

117   relatively warm and saline Atlantic water (temperature 2.5 - 2.7ºC at seafloor at time of

118   sampling), while the northern stations 4 and 5 were dominated by colder Arctic water

119   (temperature 0.9 - 1.2ºC). Furthermore, ocean depth was largest at the southern stations (442 -

120   474 m versus 230 - 290 m at the northern ones), while other recorded physicochemical variables

121   (salinity, grain size distribution, organic carbon content) showed minimal or inconsistent

122   differences between sampling sites. Additional details of sampling, locations and sediment

123   characteristics are given in Nguyen & Landfald (2015). The present stations 1 to 5 correspond

124   to stations 1, 3, 4, 7 and 10 in that study.

125

126   **Purification and isolation of viral DNA**

127   Viral purification was performed with modifications of the procedure of Thurber et al. (2009).

128   Three mL sediment was suspended in 3 mL of 2% (v/w) tetrasodium pyrophosphate and

129   incubated for 15 min in the dark at 4ºC. To ensure release of the viral particles from the

130   sediment samples, the suspension was sonicated on ice for 3 x 45 sec with a Branson 3210

131   sonicator (Triad Scientific, Inc. Manasquan, USA), and centrifuged at 1000 x g for 5 min.

132   Thereafter, the sediment pellets were twice re-suspended in 3 mL of 2% (v/w) NaCl and

133   centrifuged at 2500 x g for 5 min. The supernatants were pooled and successively passed

134   slowly through 0.45 μm and 0.2 μm Whatman filters to remove residual cells. The filtrate was

135   concentrated by using a 30 kDa NMWCO centrifugal ultrafiltration device (Ultracel PL-30

6

136 membrane, Millipore). The concentrate was treated with DNase I (2.5 U ml^{-1}) and incubated

137 for 1 hour at 37ºC to remove free DNA. To check for purity of the viral preparation, each

138 sample was stained with SYBRGold (Invitrogen) and inspected under an epifluorescence

139 microscope (Leica DM6000 B). If no contamination by prokaryotic cells was revealed, the

140 viral nucleic acids were extracted using the formamide and phenol/chloroform/isoamyl

141 alcohol procedure (Thurber et al. 2009). After the extraction, the absence of detectable

142 quantities of bacterial DNA was confirmed by subjecting the DNA preparation to PCR

143 amplification with universal bacterial 16S rRNA gene primers 27F (5′-

144 AGAGTTTGATCMTGGCTCAG-3′) and 1492R (5′-TACCTTGTTACGACTT-3′). The PCR

145 master mix contained 2.5 µl of 10X PCR buffer (Invitrogen, Waltham, MA, USA), 2.0 µl of

146 0.2 mM dNTPs (Invitrogen), 2.5 µl of 0.5 µM of each primer (Eurofins MWG, Ebersberg,

147 Germany), 1.25 U of Taq polymerase (Invitrogen), 10 ng of genomic DNA template, and

148 Milli-Q water to a total volume of 25 µl. The thermocycler (Applied Biosystems) conditions

149 were initial denaturation step at 95ºC for 5 min; 30 cycles at 95ºC for 30 s, 55ºC for 30 s, and

150 72ºC for 1 min; a final extension at 72ºC for 5 min. Only preparations where no bacterial cells

151 were observed and no bacterial PCR products were generated from the DNA extracts were

152 used for subsequent viral amplification.

153

154 **Amplification of viral DNA and sequencing**

155 Viral genomic DNA (10 ng) was amplified using the Genomiphi V2 DNA Amplication kit

156 (GE Healthcare Life Sciences, Freiburg, Germany) according to the manufacturer's

157 instructions. After amplification, amplicon DNA was purified on silica columns (Qiagen

158 Dnase Kit), precipitated with ethanol and resuspended in pure water. The concentration and

159 quality of amplified DNA were determined by a NanoDrop ND-1000 spectrophotometer

160 (Wilmington, DE, USA). The amplifications were done in duplicate for each sample to reduce

161 bias. The samples were subjected to shotgun pyrosequencing (GS FLX Titanium; 454 Life

162 Sciences, Branford, CT, installed at The Norwegian High-Throughput Sequencing Centre,

163 University of Oslo, Norway; http://www.sequencing.uio.no).

164

Sequence assembly

166 Flowgram sequence data in sff format from the GS FLX Titanium were assembled using

167 Newbler version 2.5.3 for all five stations. Several runs with varying parameters were tested

168 to produce optimal assemblies based on N50 and total contig yield (bp). Default settings for

169 *de novo* assembly and "autotrimming" were considered the most successful. A minimum

170 contig size threshold of 500 bp was applied to discard the smallest contigs.

171

Taxonomic classifications

173 The metagenomic sequence and contig data were analyzed using the Metavir (Roux et al.

174 2011) and Metavir2 (Roux et al. 2014a) pipeline (http://metavir-med.univbpclermont.fr/) with

175 taxonomic affiliations based on BLASTx queries in the NCBI Refseq complete viral genomes

176 protein sequence database (release of 2016-01-19). BLAST hits with E-values $<10^{-3}$ were

177 considered significant. Sequences affiliated with individual *Circoviridae*, *Geminiviridae* and

178 *Nanoviridae* genomes or metagenomic genomes categorized as Circular Rep-encoding

179 ssDNA (CRESS-DNA) viruses were classified as CRESS-DNA virus sequences in

180 accordance with Rosario et al. (2015). Taxonomic assignments were also generated via MG-

181 RAST (Meta Genome Rapid Annotation using Subsystem Technology, v3.1) (Meyer et al.

182 2008) based on BLASTx searches against the M5NR non-redundant database. An E-value

183 cutoff of 10^{-3}, a minimum identity cutoff of 60% and minimum alignment length of 15 bp

184 were used as parameters for this analysis.

185

186

Phylogenetic analyses of marker genes

187

188 The protein sequences of the major capsid protein F (Vp1) of the *Microviridae* family

189 (Desnues et al. 2008) and the replication-associated protein (Rep) of the

190 *Circoviridae/Nanoviridae/Geminiviridae* families (Rosario et al. 2009a) were used to infer

191 phylogenetic trees. Representative sequences covering the whole or major parts of the

192 complete gene products from the Barents Sea sediment viromes were aligned with

193 homologous sequences obtained from translated genomic and metagenomic sequences of

194 public data and the Metavir server, by using the ClustalW algorithm in MEGA 6 program

195 package (Tamura et al. 2013). Phylogenetic trees were constructed using the maximum

196 likelihood method, with Jones-Taylor-Thornton substitution model, gamma distribution

197 among sites, partial deletion of missing data and bootstrapping with 100 resamplings, as

198 implemented in the MEGA package.

199

Comparative metagenomics of viral communities

200

201 The viral metagenomics data from our study were compared with public virome projects from

202 other environments, available in the MetaVir web server (Roux et al. 2011). The imported

203 data only included studies in which multiple displacement amplification (MDA) had been

204 used to increase DNA quantities and comprised data from deep-sea surface sediment

205 (Yoshida et al. 2013), sea water (Angly et al. 2006, Cassman et al. 2012) and fresh water

206 (Roux et al. 2012, Fancello et al. 2013). In the MetaVir workflow, the similarity of the virus

207 metagenomes were compared and scored as tBLASTx sequence similarity scores. Only

208 viromes that contained more than 50 000 sequences with length exceeding 100 bp on average

209 were used for the comparision. Similarity scores for two by two viromes were computed as

210 the sum of the best BLAST hit scores of a sequence component in one virome community

9

211 against its counterpart in the other virome. The score sums for the virome pairs were stored in

212 a matrix format and used to hierarchically cluster the communities in the Metavir web server.

213   Separate hierarchical clustering trees for the *Microviridae* and the eukaryotic

214 subgroups of virus sequences of the Barents Sea sediment data were inferred by the *pvclust*

215 package with default parameters (Suzuki & Shimodaira 2006) in the R software

216 (http://www.r-project.org). Bootstrap values (1000 resamplings) were calculated.

217

218 **Bacterial community comparisons.** The total DNA was extracted using the PowerSoil[TM]

219 DNA Isolation kit (Mo Bio Labs, Inc., Carlsbad, CA, USA) according to the manufacturer's

220 instruction. The bacterial 16S rRNA genes were amplified by use the primers 27F (5′-

221 AGAGTTTGATCMTGGCTCAG-3′) and 338R (5′-TGCTGCCTCCCGTAGGAGT-3′).

222 Subsequently, the amplicons were sequenced by multiplex pyrosequencing on a 454/Roche

223 GS-FLX Titanium instrument, and the sequences were binned into operational taxonomic

224 units (OTUs) at 97% identity level, as described in Nguyen and Landfald (2015). Sequence-

225 based hierarchical clustering of the bacterial communities was done by average linkage in the

226 R package *pvclust*.

227

228 **Virus/bacteria community relationships.** Correlations between variations in bacterial

229 communities and viral assemblages (complete or subsections) were quantified as Spearman

230 rank correlation coefficients calculated from Bray-Curtis distance matrices. Bacterial

231 distances were based on OTU tables, while viral distances were based on tables of significant

232 BLAST hits ($E < 10^{-3}$). The statistical significance of the correlations were tested by the

233 Mantel test with 10^4 Monte Carlo permutations in the *vegan* package in R.

234

235   **Virome accession numbers.** Viral metagenomes from the five sampling stations are

236   accessible in the Metavir server under the name Barents Sea sediment. In the MG-RAST

237   annotation server, they are available under accession numbers: 4564380.3; 4564381.3;

238   4564382.3; 4564383.3; 4565015.3. The 14 completed circular genomes are deposited in

239   GenBank under accession numbers KX828610 to KX829623.

240

241

242

243

244   **RESULTS AND DISCUSSION**

245

246   **The Barents Sea sediment virome**

247   The shotgun pyrosequencing produced a total of 674 115 sequence reads with an average

248   length of 507 bp from the five Barents Sea sediment samples (Table 1). Between 81% and

249   87% of reads in the different samples were assembled into 17805 contiguous sequences

250   (minimum 500 bp) by the Newbler assembler. The median contig length varied in the range

251   1131 to 1228 bp, the longest being 5745 bp. A total of 39756 complete or partial open reading

252   frames (ORFs) were identified in the contigs and 37 to 51% of the predicted ORFs in the

253   separate samples showed significant homology to sequences present in the NCBI Refseq

254   database. Fourteen contigs, ranging from 819 to 4217 bp in size, were categorized as circular

255   genomes by employing the Metavir2 pipeline, but only four of them produced significant hits

256   (E-value $<10^{-3}$) to established viral genomes. The largest identified genome (4217 bp) was

257   similar to the one of the complete marine *Gokushovirinae* GOM (KC131021; Labonte &

258   Suttle 2013a) while two genomes of size 2163 bp and 1938 bp, respectively, showed highest

259   similarity to the eukaryotic circular Rep-encoding ssDNA (CRESS-DNA) viruses Avon-

11

260   Heathcote Estuary associated circular virus 5 and 6 (KM874301 and KM874301) (Dayaram et

261   al. 2015). The former was categorized as circo-like by encoding both the conserved Rep and a

262   putative capsid (Cap) protein (Rosario et al. 2015) while the latter encoded just the Rep

263   protein, as did the fourth, smallest genome (1265 bp), which was similar to *Circoviridae* 3

264   LDMD-2013 (KF133810; McDaniel et al. 2014).

265   Analyses of the individual sequences identified 27 to 44% as having significant

266   similarity (BLASTx, E-values $< 10^{-3}$) to genes within the complete viral genomes of the

267   Refseq database (Table 1). Hence, both the contiguous and individual sequence data

268   confirmed the consistent pattern from preceding environmental virome studies that the

269   majority of sequences remain unassigned. However, the proportions of virus-affiliated

270   sequences compared favorably with preceding metagenomic studies employing the Refseq

271   database, as 1 - 35% viral genes have been reported at $< 10^{-3}$ significance level (Breitbart et al.

272   2002, Breitbart et al. 2004, Edwards & Rohwer 2005, Angly et al. 2006, Desnues et al. 2008,

273   Lopez-Bueno et al. 2009, Roux et al. 2012, Fancello et al. 2013, Yoshida et al. 2013).

274   Successful enrichment of viral metagenomic material was also indicated by low proportions

275   of prokaryotic (6.7%) and eukaryotic (0.6%) sequences by search in the M5NR database of

276   MG-RAST (Meyer et al. 2008). The rather straightforward method employed in the present

277   study for isolation of virus particles from sediment therefore appeared efficient for enrichment

278   of viral genomic material.

279   Further phylogenetic classification showed 86.1 to 94.8% of reads affiliated to viruses

280   from the five sampling stations to be confidently associated with ssDNA virus genes of the

281   Refseq database (E-value $< 10^{-3}$), while reads of dsDNA virus origin constituted 3.0 to 11.6%

282   (Table 2). In accordance with the pattern for the ssDNA viral groups (see below), both a

283   bacteriophage taxon (*Caudovirales*) and a eukaryote-infecting taxon (*Phycodnaviridae*) were

284   represented among of the most predominant dsDNA groups. Several preceding metagenomic

285 studies have indicated a strong presence of ssDNA viruses in marine environments (Desnues

286 et al. 2008, Kim & Bae 2011, Rosario & Breitbart 2011, Tucker et al. 2011, Yoshida et al.

287 2013). However, a common feature of both the present and the referred studies is the use of

288 MDA with phi29 polymerase for amplification of viral metagenomic DNA. MDA biases the

289 genome type distribution towards ssDNA viruses and, in particular, towards small circular

290 ssDNA viruses. For seawater samples this has been demonstrated by comparing viral DNA

291 amplified by the linker amplified shotgun library (LASL) method of Breitbart et al. (2002)

292 with the same samples amplified by MDA (Kim & Bae 2011), while McDaniel et al. (2014)

293 find large differences in abundances of ssDNA virus genes between MDA and unamplified

294 DNA preparations from seawater. Chromatographic separation of viral groups prior to LASL

295 amplification, as devised by Andrews-Pfannkoch et al. (2010), represents an improvement,

296 but it still remains unclarified to what extent the diversities of the resulting metagenomic data

297 are truly representative of the source material. In our study, the proportion of dsDNA viruses

298 may have been lowered also by the filtration step used for removal of cellular and abiotic

299 particles. Electron micrographs show some tailed or filamentous environmental virions to

300 have physical dimensions that limit their passage through the 0.2 μm pore size filters used in

301 the present study (Bergh et al. 1989, Zhao et al. 2013), while circular ssDNA viruses are less

302 likely hindered due to their overall smaller size. To secure the inclusion of large-size virions

303 in the preparations, some recent virome studies have preferred the use of 0.45 μm pore size

304 filters or no filtration step at all (Lopez-Bueno et al. 2009, Fancello et al. 2013, Martínez

305 Martínez et al. 2014). On the other hand, environmental viral diversity studies have also

306 demonstrated variable proportions of ssDNA virus genes and even dominance of dsDNA

307 virus genes in 0.2 μm filtrates amplified by MDA (Rosario et al. 2009b, Cassman et al. 2012,

308 McDaniel et al. 2014), indicating a relationship between abundances of ssDNA viruses in the

309 source material and ssDNA virus genes in the amplified metagenome. Hence, the apparent

310 predominance by ssDNA viral material in our amplified virome DNA likely reflects a lower

311 but substantial proportion of this virus group in the source material.

312   The International Committee on Taxonomy of Viruses has classified the ssDNA

313 viruses into 11 families and one genus not assigned to a family (Krupovic & Forterre 2015,

314 Krupovic et al. 2016). The BLASTx search affiliated the open reading frames of the Barents

315 Sea sediment material with six of these families, i.e. *Microviridae*, *Circoviridae*, *Nanoviridae*,

316 *Germiniviridae*, *Parvoviridae* and *Inoviridae*, among which the *Microviridae* phage family

317 constituted 8 - 44% whereas a larger fraction, 46 - 81%, were assigned to the three eukaryotic

318 families of circular Rep-encoding ssDNA viruses (*Circoviridae*, *Nanoviridae*,

319 *Germiniviridae*), collectively named CRESS-DNA viruses (Rosario et al. 2015). More than

320 half of the CRESS DNA virus sequences were affiliated with the *Circoviridae* family by

321 showing highest similarity to genes in complete genomes carrying both Rep and putative

322 capsid encoding genes.

323   There are still rather few described examples of ssDNA viruses infecting marine

324 microorganisms, the first one being a novel virus that infects bloom-forming diatoms

325 (Nagasaki et al. 2005). More recently, putative *Microviridae* infecting marine *Bacteroidetes*

326 phylotypes have been reported (Holmfeldt et al. 2012). Identification of the virus-host

327 associations by *in silico* analysis generally requires complete or near-complete genomic

328 information of both virus and host (Roux et al. 2015). This is a challenging task by

329 metagenomic approaches even in rather well-characterized microbiotas like the human gut

330 (Pérez-Brocal et al. 2013), and even more so in highly diverse habitat types like marine

331 environments, which are dominated by hitherto uncultivated and genomically uncharacterized

332 phylogenetic groups. However, Mizuno et al. (2013) were able to assign putative hosts to

333 more than 500 viral conigs in a sea water fosmid library by combining several *in silico*

334 analytical approaches. Potential bacterial and archaeal hosts of viral assemblages in

14

335   hydrothermal vent environments have been identified by clustered regularly interspaced short

336   palindromic repeat (CRISPR) analysis (Anderson et al. 2011), while Cassman et al. (2012)

337   were unable to identify virus-host pairs by CRISPR analysis in Pacific sea water material.

338   Recently single-cell genome amplification has been demonstrated as a viable new approach to

339   identify virus-bacteria associations even among the uncultivated majority of bacteria in a

340   pelagic marine system (Roux et al. 2014b).

341

342

343

344   **Marker gene phylogenies**

345   Phylogenetic relationships of the major capsid protein F (Vp1) as marker gene for the

346   *Microviridae* and the replication-associated protein (Rep) as marker gene for the CRESS-

347   DNA virus were inferred by comparing sequences covering conserved regions with

348   corresponding sequences from relevant metagenome and genome studies (Fig. 2A, B).

349   *Gokushovirinae* originating from intracellular parasitic bacteria (*Chlamydiae, Bdellovibrio*

350   *and Spiroplasma* spp.) (Brentlinger et al. 2002), and *Microvirus* from enterobacterial species

351   (including φX174) were the sources of the most closely related Vp1 genes of established

352   genomic origin. The actual host groups of the Vp1-carrying phages within the Barents Sea

353   sediment communities were not established. A preceding study of bacterial community

354   structure in the same material identified *Gammaproteobacteria* and *Deltaproteobacteria* as

355   predominant groups by jointly constituting nearly two-thirds of the 16S rRNA gene pool

356   (Nguyen & Landfald 2015). Outside the marine realm, the *Microviridae* have been shown to

357   infect a diverse range of hosts, including *Proteobacteria*, *Spiroplasma* and *Chlamydiales*

358   (Read et al. 2000, Brentlinger et al. 2002, Garner et al. 2004) and a recent, comprehensive

359   search for viral genomes in prokaryotic genomic data sets (Roux et al. 2015) also ties the

15

360   *Microviridae* to *Alpha-*, *Gamma-* and *Deltaproteobacteria* as important hosts classes. It

361   therefore appears likely that the mentioned classes of *Proteobacteria* constitutes substantial

362   parts of the *Microviridae* host range.

363   Different Barents Sea Rep-encoding sequences showed affiliation to both animal-

364   infecting (*Circoviridae*) and plant-infecting (*Nanoviridae*, ssDNA Alphasatellites,

365   *Geminiviridae*) genomic virus sequences and metagenomic sequences from sea water, fresh

366   water and deep-sea sediments were also distributed throughout the Rep gene diversity. No

367   distinct grouping of the picked Vp1 or Rep metagenomic sequences in accordance with the

368   south-north climatic division of the Barents Sea was observed, indicating that important

369   groups of host organisms were abundant throughout the sampling range.

370

371   **Community comparisons**

372   A hierarchical clustering tree, including both the five Barents Sea sampling sites and

373   additional aquatic and sediment viromes (Angly et al. 2006, Cassman et al. 2012, Roux et al.

374   2012, Fancello et al. 2013, Yoshida et al. 2013), was inferred from BLAST-based

375   comparisons (Fig. 3) by tools integrated in the MetaVir server (Roux et al. 2011). The three

376   Pacific Ocean marine sediment viromes (Yoshida et al. 2013) clustered with the Barents Sea

377   samples and the environmentally most comparable location, i.e. pelagic sediment (1181 m

378   water depth) off the Japanese coast, also showed the highest assemblage similarity. Even the

379   viromes from two extremely deep Pacific trenches were more similar to the Barents Sea

380   material than any sea water or freshwater sample, including the ones collected from Arctic

381   regions. These common phylogenetic features in seafloor sediment viromes across large

382   geographical separations and extensive variations in water depth likely reflect corresponding

383   similarities in the prokaryotic and eukaryotic host communities, but the specific basis for the

384   observed commonness remains vague as community data is lacking for the microorganisms.

385   In the present study, some level of coincidence in bacterial and viral biogeography

386   patterns among the Barents Sea sediment sampling stations was indicated by the hierarchical

387   clustering patterns. A previous comprehensive study including 10 sampling stations,

388   established significant differences among bacterial communities in accordance with the main

389   environmental division of this ocean area into a southern and a northern region by the oceanic

390   polar front (Nguyen & Landfald 2015). Here we reconfirmed this bacterial pattern in a study

391   comprising just the five stations included in the virome analysis. The two northernmost

392   stations 4 and 5 constituted a distinct cluster based on 16S rRNA gene sequence data (Fig.

393   4A). BLAST based similarity scores indicated the same partition of the viral assemblages if

394   the analysis was based on the complete virome sequence data (Fig. 3). However, if restricting

395   the analysis to the dominant group of bacteria-associated virus in the assemblages, i.e. the

396   *Microviridae* family of phages, the viromes showed no clustering pattern reflecting the one

397   exerted by the bacterial communities (Fig. 4B). On the other hand, the northern sampling

398   stations 4 and 5 showed a distinct grouping of the eukaryotic CRESS-DNA virus assemblages

399   (Fig. 4C), suggesting that these taxa were inclined to stronger host variations along the south-

400   north axis than the bacteria-infecting viruses. The absence of relationship between bacteria

401   and bacteriophage community variations was also confirmed by a Mantel test of correlation

402   between the respective community distance matrices (P = 0.29). One factor that may have

403   contributed to this lack of coherence is the aforementioned expected bias towards ssDNA

404   virus in MDA-amplified viromes. This would lead to an overestimation of the importance of

405   the *Microviridae* as infective agents in the bacterial communities. Secondly, the bacterial

406   community profiling was based on ribosomal RNA gene frequency comparisons. This implies

407   that the phylogenetic distribution of actively proliferating bacterial phylotypes, which

408   expectedly are the more influential ones on phage assemblage structuring, may have deviated

409   from the observed total 16S rRNA gene based biogeographical variation. The elevated

17

410 relative abundance of eukaryotic virus types in the north (Table 2) coincided with higher

411 levels of chlorophyll *a* and chloroplast 16S rRNA gene frequencies at the same locations, as

412 reported elsewhere (Nguyen & Landfald 2015). Hence, increased seafloor deposition of fresh

413 algal material in the northern region, due to the recent ice margin spring bloom, may have

414 contributed to the observed differences among the eukaryotic virus assemblages.

415   In conclusion, the study confirmed that virome biogeographical variation within a

416 coherent and environmentally moderately variable shelf seafloor area was small as compared

417 with differences among more distinctly separated ecosystems, such as geographically distant

418 lakes or oceans. It furthermore indicated some degree of commonness in composition of

419 seafloor viral assemblages at a global scale. The observed Barents Sea virome variation was

420 principally associated with changes in relative abundance and composition of eukaryotic

421 ssDNA virus genes, while the established bacterial community variation within the same

422 ocean range was not mirrored in the bacteriophage metagenomic material. Our work

423 underlined the need to complement this kind of community-wide approaches with

424 identification and characterization of representative virus-microorganism associations to get

425 comprehensive pictures of the interactions between the groups in their natural environment.

426

431

432 **LITERATURE CITED**

433 Anderson RE, Brazelton WJ, Baross JA (2011) Is the genetic landscape of the deep

434   subsurface biosphere affected by viruses? Front Microbiol 2:219

435 Andrews-Pfannkoch C, Fadrosh DW, Thorpe J, Williamson SJ (2010) Hydroxyapatite-

436   mediated separation of double-stranded DNA, single-stranded DNA, and RNA

437   genomes from natural viral assemblages. Appl Environ Microbiol 76:5039-5045

438 Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C, Chan AM, Haynes M,

439   Kelley S, Liu H, Mahaffy JM, Mueller JE, Nulton J, Olson R, Parsons R, Rayhawk S,

440   Suttle CA, Rohwer F (2006) The marine viromes of four oceanic regions. PLoS Biol

441   4:e368

442 Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011) Genomic island variability

443   facilitates *Prochlorococcus*–virus coexistence. Nature 474: 604–608.

444 Bergh O, Børsheim KY, Bratbak G, Heldal M (1989) High abundance of viruses found in

445   aquatic environments. Nature 340:467-468

446 Bouvier T, del Giorgio PA (2007) Key role of selective viral-induced mortality in determining

447   marine bacterial community composition. Environ Microbiol 9:287-297

448 Breitbart M, Felts B, Kelley S, Mahaffy JM, Nulton J, Salamon P, Rohwer F (2004) Diversity

449   and population structure of a near-shore marine-sediment viral community. Proc Biol

450   Sci 271:565-574

451 Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F

452   (2002) Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci

453   U S A 99:14250-14255

454 Brentlinger KL, Hafenstein S, Novak CR, Fane BA, Borgon R, McKenna R, Agbandje-

455   McKenna M (2002) Microviridae, a family divided: Isolation, characterization, and

456   genome sequence of phi MH2K, a bacteriophage of the obligate intracellular parasitic

457   bacterium *Bdellovibrio bacteriovorus*. J Bacteriol 184:1089-1094

458 Bryson SJ, Thurber AR, Correa AM, Orphan VJ, Vega Thurber R (2015) A novel sister clade

459   to the enterobacteria microviruses (family Microviridae) identified in methane seep

460   sediments. Environ Microbiol 17:3708-3721

461 Cassman N, Prieto-Davo A, Walsh K, Silva GG, Angly F, Akhter S, Barott K, Busch J,

462   McDole T, Haggerty JM, Willner D, Alarcon G, Ulloa O, DeLong EF, Dutilh BE,

463   Rohwer F, Dinsdale EA (2012) Oxygen minimum zones harbour novel viral

464   communities with low diversity. Environ Microbiol 14:3043-3065

465 Clasen JL, Bigden SM, Payet JP, Suttle CA (2008) Evidence that viral abundance across

466   oceans and lakes is driven by different biological factors. Freshw Biol 53:1090-1100.

467 Danovaro R, Corinaldesi C, Filippini, M, Fischer UR, Gessner MO, Jacquet S, Magangnini

468   M, Velimirov B (2008) Viriobenthos in freshwater and marine sediments: a review.

469   Freshw Biol 53:1186-1213.

470 Dayaram A, Goldstien S, Argüello-Astorga GR, Zawar-Reza P, Gomez C, Harding JS,

471   Varsani A (2015) Diverse small circular DNA viruses circulating amongst estuarine

472   molluscs. Infect Genet Evol 31:284-295

473 Desnues C, Rodriguez-Brito B, Rayhawk S, Kelley S, Tran T, Haynes M, Liu H, Furlan M,

474   Wegley L, Chau B, Ruan Y, Hall D, Angly FE, Edwards RA, Li L, Thurber RV, Reid

475   RP, Siefert J, Souza V, Valentine DL, Swan BK, Breitbart M, Rohwer F (2008)

476   Biodiversity and biogeography of phages in modern stromatolites and thrombolites.

477   Nature 452:340-343

478 Edwards RA, Rohwer F (2005) Viral metagenomics. Nat Rev Microbiol 3:504-510

479   Fancello L, Trape S, Robert C, Boyer M, Popgeorgiev N, Raoult D, Desnues C (2013)

480   Viruses in the desert: a metagenomic survey of viral communities in four perennial

481   ponds of the Mauritanian Sahara. ISME J 7:359-369

482   Fauquet C, Mayo M, Maniloff J, Desselberger U, Ball L (2005) Virus Taxonomy. Eighth

483   Report of the International Committee on Taxonomy of Viruses. San Diego, CA,

484   USA: Academic Press.

485   Fuhrman JA (1999) Marine viruses and their biogeochemical and ecological effects. Nature

486   399:541-548

487   Garner SA, Everson JS, Lambden PR, Fane BA, Clarke IN (2004) Isolation, molecular

488   characterisation and genome sequence of a bacteriophage (Chp3) from *Chlamydophila*

489   *pecorum*. Virus Genes 28:207-214

490   Helton RR, Wang K, Kan J, Powell DH, Wommack KE (2012) Interannual dynamics of

491   viriobenthos abundance and morphological diversity in Chesapeake Bay sediments.

492   FEMS Microbiol Ecol 79:474-486

493   Holmfeldt K, Middelboe M, Nybroe O, Riemann L (2007) Large variabilities in host strain

494   susceptibility and phage host range govern interactions between lytic marine phages

495   and their *Flavobacterium* hosts. Appl Environ Microbiol 73:6730-6739

496   Holmfeldt K, Odic D, Sullivan MB, Middelboe M, Riemann L (2012) Cultivated single-

497   stranded DNA phages that infect marine Bacteroidetes prove difficult to detect with

498   DNA-binding stains. Appl Environ Microbiol 78:892-894

499   Hurwitz BL, Westveld AH, Brum JR, Sullivan MB (2014) Modeling ecological drivers in

500   marine viral communities using comparative metagenomics and network analyses.

501   Proc Natl Acad Sci U S A 111:10714-10719

502   Kang I, Oh HM, Kang D, Cho JC (2013) Genome of a SAR116 bacteriophage shows the

503   prevalence of this phage type in the oceans. Proc Natl Acad Sci U S A 110:12343-

504   12348

505   Kim KH, Bae JW (2011) Amplification methods bias metagenomic libraries of uncultured

506   single-stranded and double-stranded DNA viruses. Appl Environ Microbiol 77:7663-

507   7668

508   Kimura K, Tomaru Y (2015) Discovery of two novel viruses expands the diversity of single-

509   stranded DNA and single-stranded RNA viruses infecting a cosmopolitan marine

510   diatom. Appl Environ Microbiol 81:1120–1131

511   Krupovic M, Forterre P (2015) Single-stranded DNA viruses employ a variety of

512   mechanisms for integration into host genomes. Ann NY Acad Sci 1341: 41-53

513   Krupovic M, Ghabrial SA, Jiang D, Varsani A (2016) Genomoviridae: a new family of

514   widespread single-stranded DNA viruses. Arch Virol, 161 (9):2633-44

515   Labonté JM, Suttle CA (2013a) Metagenomic and whole-genome analysis reveals new

516   lineages of gokushoviruses and biogeographic separation in the sea. Front Microbiol

517   4:404

518   Labonté JM, Suttle CA (2013b) Previously unknown and highly divergent ssDNA viruses

519   populate the oceans. ISME J 7:2169-2177

520   Labonté JM, Swan BK, Poulos B, Luo H, Koren S, Hallam SJ, Sullivan MB, Woyke T,

521   Wommack KE, Stepanauskas R (2015) Single-cell genomics-based analysis of virus-

522   host interactions in marine surface bacterioplankton. ISME J 9:2386-2399

523   Lopez-Bueno A, Tamames J, Velazquez D, Moya A, Quesada A, Alcami A (2009) High

524   diversity of the viral community from an Antarctic lake. Science 326:858-861

525   Martínez Martínez J, Swan BK, Wilson WH (2014) Marine viruses, a genetic reservoir

526   revealed by targeted viromics. ISME J 8:1079-1088

527 McDaniel LD, Rosario K, Breitbart M, Paul JH (2014) Comparative metagenomics: natural

528   populations of induced prophages demonstrate highly unique, lower diversity viral

529   sequences. Environ Microbiol 16:570-585

530 Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A,

531   Stevens R, Wilke A, Wilkening J, Edwards RA (2008) The metagenomics RAST

532   server - a public resource for the automatic phylogenetic and functional analysis of

533   metagenomes. BMC Bioinformatics 9:386

534 Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R (2013) Expanding the marine

535   virosphere using metagenomics. PLoS Genet. 9

536 Nagasaki K, Tomaru Y, Takao Y, Nishida K, Shirai Y, Suzuki H, Nagumo T (2005)

537   Previously unknown virus infects marine diatom. Appl Environ Microbiol 71:3528-

538   3535

539 Nagasaki K (2008) Dinoflagellates, diatoms, and their viruses. J Microbiol: 46:235-43

540 Nguyen TT, Landfald B (2015) Polar front associated variation in prokaryotic community

541   structure in Arctic shelf seafloor. Front Microbiol 6:17

542 Parsons RJ, Breitbart M, Lomas MW, Carlson CA (2012) Ocean time-series reveals recurring

543   seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. ISME J

544   6:273-284

545 Pérez-Brocal V, Garcia-Lopez R, Vazquez-Castellanos JF, Nos P, Beltran B, Latorre A, Moya

546   A (2013) Study of the viral and microbial communities associated with Crohn's

547   disease: a metagenomic approach. Clin Transl Gastroenterol 4:e36

548 Read TD, Brunham RC, Shen C, Gill SR, Heidelberg JF, White O, Hickey EK, Peterson J,

549   Utterback T, Berry K, Bass S, Linher K, Weidman J, Khouri H, Craven B, Bowman C,

550   Dodson R, Gwinn M, Nelson W, DeBoy R, Kolonay J, McClarty G, Salzberg SL,

551   Eisen J, Fraser CM (2000) Genome sequences of *Chlamydia trachomatis* MoPn and

552   *Chlamydia pneumoniae* AR39. Nucl Acids Res 28:1397-1406

553 Rosario K, Breitbart M (2011) Exploring the viral world through metagenomics. Curr Opin

554   Virol 1:289-297

555 Rosario K, Duffy S, Breitbart M (2009a) Diverse circovirus-like genome architectures

556   revealed by environmental metagenomics. J Gen Virol 90:2418-2424

557 Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M (2009b) Metagenomic analysis of

558   viruses in reclaimed water. Environ Microbiol 11:2806-2820

559 Rosario K, Schenck RO, Harbeitner RC, Lawler SN, Breitbart M (2015) Novel circular

560   single-stranded DNA viruses identified in marine invertebrates reveal high sequence

561   diversity and consistent predicted intrinsic disorder patterns within putative structural

562   proteins. Front Microbiol 6:696

563 Roux S, Faubladier M, Mahul A, Paulhe N, Bernard A, Debroas D, Enault F (2011) Metavir:

564   a web server dedicated to virome analysis. Bioinformatics 27:3074-3075

565 Roux S, Enault F, Robin A, Ravet V, Personnic S, Theil S, Colombet J, Sime-Ngando T,

566   Debroas D (2012) Assessing the diversity and specificity of two freshwater viral

567   communities through metagenomics. PLoS One 7:e33641

568 Roux S, Tournayre J, Mahul A, Debroas D, Enault F (2014a) Metavir 2: new tools for viral

569   metagenome comparison and assembled virome analysis. BMC Bioinformatics 15:76

570 Roux S, Hawley AK, Beltran MT, Scofield M, Schwientek P, Stepanauskas R, Woyke T,

571   Hallam SJ, Sullivan MB (2014b) Ecology and evolution of viruses infecting

572   ucultivated SUP05 bacteria as revealed by single-cell and meta-genomics. eLife

573   3:e03125

574 Roux S, Hallam SJ, Woyke T, Sullivan MB (2015) Viral dark matter and virus–host

575   interactions resolved from publicly available microbial genomes. eLife. 4: e08490

576 Short SM, Suttle CA (2003) Temporal dynamics of natural communities of marine algal

577   viruses and eukaryotes. Aquat Microb Ecol 32:107-119

578 Steward GF, Montiel JL, Azam F (2000) Genome size distributions indicate variability and

579   similarities among marine viral assemblages from diverse environments. Limnol

580   Oceanogr 45:1697-1706

581 Steward GF, Preston CM (2011) Analysis of a viral metagenomic library from 200 m depth in

582   Monterey Bay, California constructed by direct shotgun cloning. Virol J 8:287

583 Sullivan MB, Waterbury JB, Chisholm SW (2003) Cyanophages infecting the oceanic

584   cyanobacterium *Prochlorococcus*. Nature 424:1047-1051

585 Suttle CA (2005) Viruses in the sea. Nature 437:356-361

586 Suttle CA (2007) Marine viruses--major players in the global ecosystem. Nat Rev Microbiol

587   5:801-812

588 Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in

589   hierarchical clustering. Bioinformatics 22:1540-1542

590 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular

591   Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-2729

592 Thingstad TF (2000) Elements of a theory for the mechanisms controlling abundance,

593   diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. Limnol

594   Oceanogr 45:1320-1328

595 Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F (2009) Laboratory procedures to

596   generate viral metagenomes. Nat Protoc 4:470-483

597 Tomaru Y, Nagasaki K (2007) Flow cytometric detection and enumeration of DNA and RNA

598   viruses infecting marine eukaryotic microalgae. J Oceanogr 63:215-221

599 Tucker KP, Parsons R, Symonds EM, Breitbart M (2011) Diversity and distribution of single-

600   stranded DNA phages in the North Atlantic Ocean. ISME J 5:822-830

601   Weinbauer MG (2004) Ecology of prokaryotic viruses. FEMS Microbiol Rev 28:127-181

602   Weinbauer MG, Rassoulzadegan F (2004) Are viruses driving microbial diversification and

603   diversity? Environ Microbiol 6:1-11

604   Wommack KE, Colwell RR (2000) Virioplankton viruses in aquatic ecosystems. Microbiol

605   Mol Biol Rev 64:69-114

606   Yang Y, Motegi C, Yokokawa T, Nagata T (2010) Large-scale distribution patterns of

607   virioplankton in the upper ocean. Aquat Microb Ecol 60:233-246

608   Yoshida M, Takaki Y, Eitoku M, Nunoura T, Takai K (2013) Metagenomic analysis of viral

609   communities in (hado)pelagic sediments. PLoS One 8:e57271

610   Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, Ellisman M,

611   Deerinck T, Sullivan MB, Giovannoni SJ (2013) Abundant SAR11 viruses in the

612   ocean. Nature 494:357-360

613

614

615   Table 1. Outcome of the virome sequence analyses at the individual sampling sites

Sampling site	Number of reads	Average length (bp)	Assembled reads (% of reads)	Number of contigs (>500bp)	Viral hits[1] (% of reads)
ST.1	118 313	497	81.3	3642	40.5
ST.2	158 119	511	80.1	5110	27.3
ST.3	154 403	516	78.5	4526	40.3
ST.4	111 206	502	79.9	2308	32.2
ST.5	132 074	509	86.6	2219	43.8

616   [1]Significant similarity ($E < 10^{-3}$) to the NCBI Refseq database
617
618
619

620   Table 2. Phylogenetic affiliations of viral sequences from the five sampling stations

Group	Taxonomic composition	ST.1 (%) (n=47948)	ST.2 (%) (n=43193)	ST.3 (%) (n=62148)	ST.4 (%) (n=36110)	ST.5 (%) (n=57784)
ssDNA	Microviridae	36.0	15.7	43.9	7.9	12.9
	CRESS-DNA virus	57.8	69.7	45.7	81.2	81.4
	Parvoviridae	0.1	0.1	0.1	0.0	0.0
	Inoviridae	0.0	0.1	0.1	0.0	0.0
	Unclassified ssDNA viruses	0.4	0.5	0.5	0.5	0.5
dsDNA	Phycodnaviridae	2.4	6.3	1.9	2.1	1.1
	Caudovirales	0.5	3.4	4.9	4.4	1.2
	Mimiviridae	0.2	0.5	0.2	0.3	0.2
	Poxviridae	0.2	0.5	0.2	0.2	0.1
	Herpesvirales	0.1	0.2	0.2	0.1	0.1
	Iridoviridae	0.1	0.1	0.1	0.0	0.0
	Unclassified dsDNA	0.1	0.6	0.7	0.8	0.3
Satellites		1.7	2.1	1.1	2.1	2.0
ssRNA viruses		0.1	0.0	0.1	0.0	0.0
Unclassified viruses		0.3	0.2	0.3	0.4	0.2
*Chloroplast 16S rRNA genes (%)**		*0.1*	*0.0*	*0.2*	*2.5*	*1.8*

621   *(Nguyen & Landfald 2015); CRESS DNA, Circular Rep-encoding ssDNA

622
623
624

625 **Figure legends**

626

627 Fig. 1. The western Barents Sea with geographical positions of sampling stations.
628

629 Fig. 2. Unrooted maximum likelihood phylogenetic trees computed from the major capsid

630 protein Vp1 of *Microviridae* (A) and the replication-associated protein (Rep) of CRESS-DNA

631 viruses (B). Marker coding: filled circles, sequences of the Barents Sea virome from southern

632 stations 1-3 (red) and northern stations 4-5 (blue); open circles, sequences from other

633 metagenomic projects, including fresh water (FW; Roux et al. 2012), sea water (SW; Labonté

634 & Suttle 2013a, Labonté & Suttle 2013b) and deep-sea sediment (DSS; Yoshida et al. 2013);

635 black triangles, sequences from completed genomes.

636

637 Fig. 3. tBLASTx similarity based comparison of environmental virome assemblages. All

638 included sequence data originate from multiple-displacement amplified viral DNA. Sources:

639 sea water (Angly et al. 2006, Cassman et al. 2012), deep-sea surface sediment (Yoshida et al.

640 2013), and fresh water (Roux et al. 2012, Fancello et al. 2013).
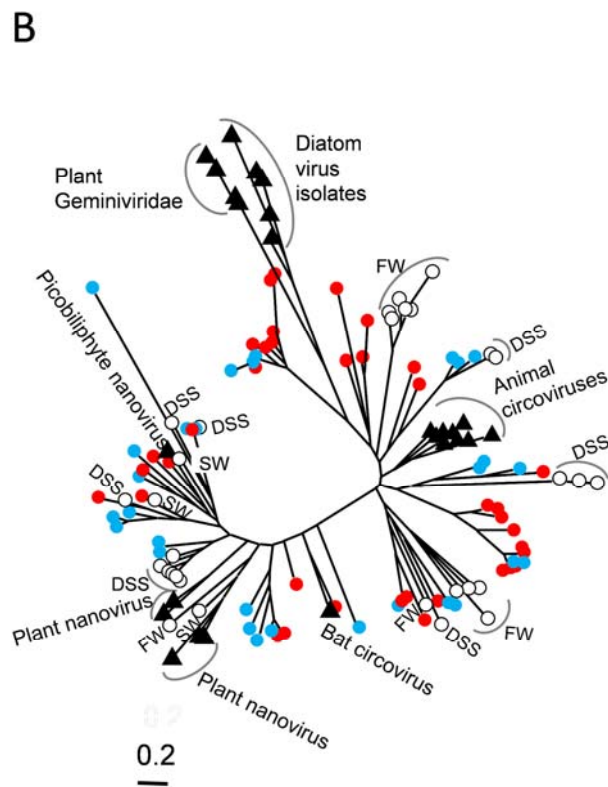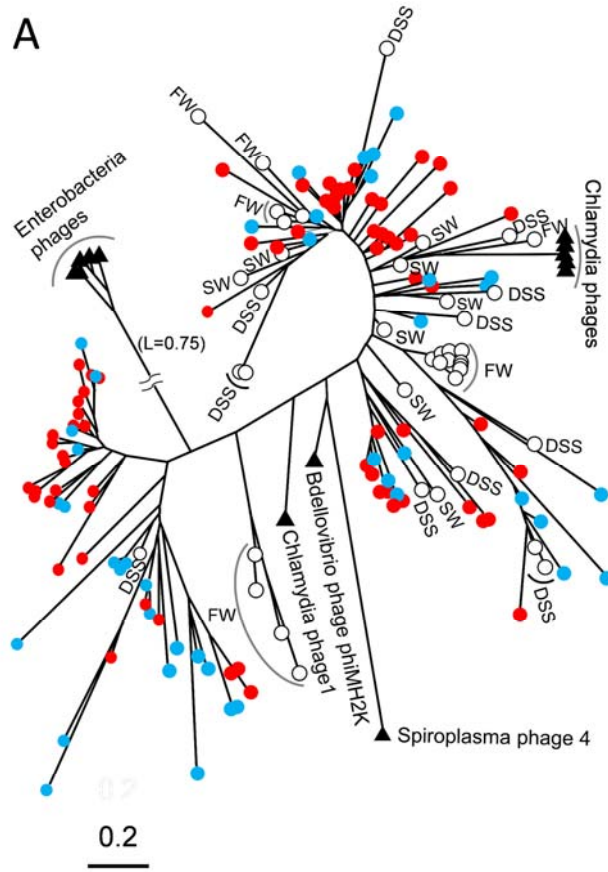
641

642 Fig. 4. Hierachical clustering of (A) bacterial communities, (B) *Microviridae* assemblages and

643 (C) CRESS-DNA virus assemblages of the Barents Sea sampling sites. Bacterial 16S rRNA

644 gene sequence data were obtained from Nguyen & Landfald (2015). Multiscale bootstrap

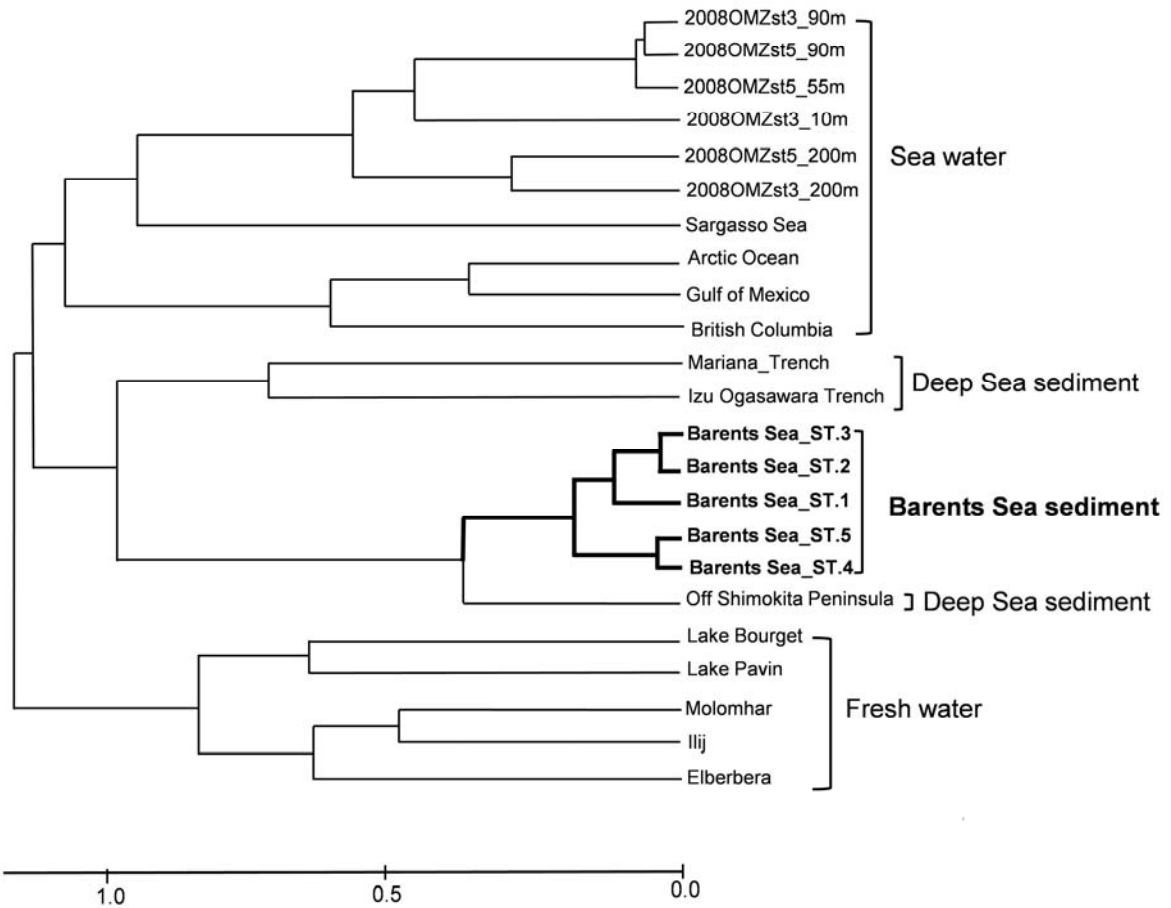645 values (1000 re-samplings) are indicated at the nodes.

646

647
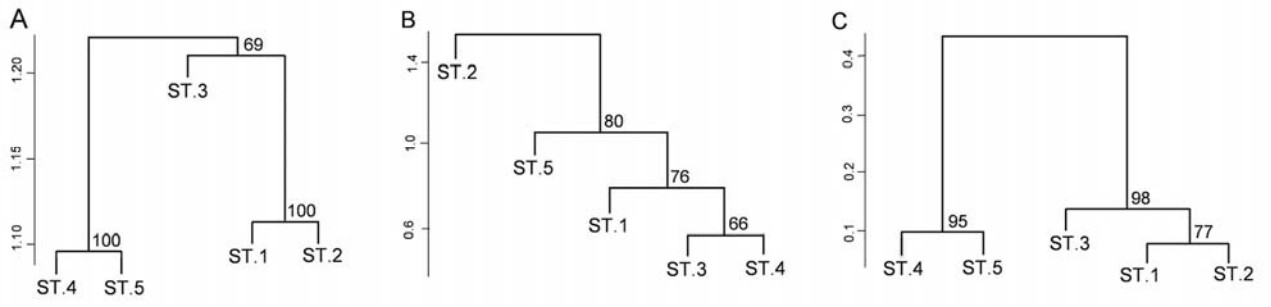648
649
650
651
652   Fig.1
653
654
655
656

657

658
659
660
661   Fig. 2
662

31

663
664
665
666
667   Fig.3
668

32

669
670
671
672   Fig.4
673