



**HAL**  
open science

# Mutant swarms of a totivirus-like entities are present in the red macroalga *Chondrus crispus* and have been partially transferred to the nuclear genome

Sylvie Rousvoal, Betty Bouyer, Camilo López-Cristoffanini, Catherine Boyen, Jonas Collén

## ► To cite this version:

Sylvie Rousvoal, Betty Bouyer, Camilo López-Cristoffanini, Catherine Boyen, Jonas Collén. Mutant swarms of a totivirus-like entities are present in the red macroalga *Chondrus crispus* and have been partially transferred to the nuclear genome. *Journal of Phycology*, 2016, 52 (4), pp.493-504. 10.1111/jpy.12427 . hal-01320013

**HAL Id: hal-01320013**

**<https://hal.sorbonne-universite.fr/hal-01320013>**

Submitted on 23 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 MUTANT SWARMS OF A TOTIVIRUS-LIKE ENTITY ARE PRESENT IN THE  
2 RED MACROALGA *CHONDRUS CRISPUS* AND HAVE BEEN PARTIALLY  
3 TRANSFERRED TO THE NUCLEAR GENOME

4  
5 Sylvie Rousvoal<sup>1,2</sup>, Betty Bouyer<sup>1,2</sup>, Camilo López-Cristoffanini<sup>1,2,3</sup>, Catherine  
6 Boyen<sup>1,2</sup> & Jonas Collén<sup>1,2</sup>

7 <sup>1</sup>CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de  
8 Roscoff, CS 90074, 29688 Roscoff cedex, France

9 <sup>2</sup>UMR 8227, Integrative Biology of Marine Models , Station Biologique de Roscoff,  
10 Sorbonne Universités, UPMC Univ Paris 06, CS 90074, 29688 Roscoff cedex,  
11 France

12 <sup>3</sup>Present address Departament de Biologia Vegetal, Universitat de Barcelona, 08028  
13 Barcelona, Spain

14

15 **ABSTRACT**

16 *Chondrus crispus* Stackhouse (Gigartinales) is a red seaweed found on North  
17 Atlantic rocky shores. Electrophoresis of RNA extracts showed a prominent band  
18 with a size of around 6,000 bp. Sequencing of the band revealed several sequences  
19 with similarity to totiviruses, double-stranded RNA viruses that normally infect  
20 fungi. This virus-like entity was named CcV. It should probably be regarded as an  
21 extreme viral quasispecies or a mutant swarm since low identity (<65%) was found  
22 between sequences. Totiviruses typically code for two genes: one capsid gene (*gag*)  
23 and one RNA-dependent RNA polymerase gene (*pol*) with a pseudoknot structure

24 between the genes. Both the genes and the intergenic structures were found in the  
25 CcV sequences. A non-identical *gag* gene was also found in the nuclear genome of  
26 *C. crispus*, with associated EST and upstream regulatory features. The gene was  
27 presumably horizontally transferred from the virus to the alga. Similar dsRNA bands  
28 were seen in extracts from different life cycle stages of *C. crispus* and from all  
29 geographical locations tested. In addition, similar bands were also observed in RNA  
30 extractions from other red algae; however, the significance of this apparently  
31 widespread phenomenon is unknown. No phenotype caused by the infection nor any  
32 virus particles, or capsid proteins were identified; thus, the presence of viral particles  
33 has not been validated. These findings increase the known host range of totiviruses to  
34 include marine red algae.

35

36 **Key words:** *Chondrus*, dsRNA, mutant swarm, Rhodophyta, totivirus, virus

37

## 38 INTRODUCTION

39 Viruses play a key role in the functioning of marine ecosystems (Fuhrman 1999,  
40 Suttle 2005, 2007). Their prevalence and diversity contribute to the control of the  
41 abundance, the diversity, and the productivity of many marine organisms and  
42 influence global nutrient and energy cycles. For example, it is estimated that viruses  
43 are responsible for the mortality of 20% of the oceans' biomass every day and that  
44 their abundance is 15 times higher than that of prokaryotes, even if the biomass is  
45 less. Because of the increasing understanding of the relevance of marine viruses a  
46 substantial research effort has been made to study their ecology, physiology, and  
47 molecular biology. Marine viruses have been shown to infect a multitude of various

48 organisms from cyanobacteria to whales. Viruses have also been suggested to  
49 represent the most genetically diverse biological entities in the ocean.

50 There are various types of viruses that differ in terms of their genetic material:  
51 single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), single-stranded  
52 RNA (ssRNA), and double-stranded RNA (dsRNA). The sizes of the viral genomes  
53 vary between a couple of thousand to several million base pairs with RNA viruses  
54 normally having smaller genomes (Philippe et al. 2013). Plant viruses can be divided  
55 into four types based on their lifestyle: persistent, chronic, endogenous and acute.  
56 Infections with persistent plant viruses typically do not induce any phenotype and are  
57 transmitted via gametes. In plants, this type of lifestyle has hitherto been known  
58 exclusively from dsRNA viruses. Acute viruses can normally be transported between  
59 plants and typically produce a phenotype (Roossinck 2010).

60 One area where effects of viruses on marine organisms have been more intensely  
61 studied is the interaction between viruses and phytoplankton (e.g. Etten et al. 1991,  
62 Reisser 1993, Brussaard et al. 2004, Lang et al. 2009). It has been observed (Suttle  
63 2007) that viruses have a key role in controlling the progress, and sometimes the  
64 collapse, of many phytoplankton blooms (e.g. *Emiliana huxleyi*, *Phaeocystis*  
65 *globosa* and *Heterosigma akashiwo*). Another system that has been studied in more  
66 detail is the ecological consequences of the interaction between the green unicellular  
67 alga *Ostreococcus* and its DNA virus (e.g. Derelle et al. 2008, 2015, Moreau et al.  
68 2010). It has therefore been suggested that viral dynamics are one of the explanations  
69 to the paradox for the plankton (Fuhrman 1999); successful plankton species, due to  
70 their abundance, are more likely to succumb to virus-induced collapses of the  
71 populations than are rare species. This preserves the diversity of the phytoplankton.

72 We have, thus, some information on the interaction between unicellular algae and  
73 viruses as well as its importance.

74 Seaweeds are an important component of the coastal ecosystem; however, much less  
75 is known about viruses that infect seaweeds or macroalgae than their unicellular  
76 counterparts (Schroeder 2011). One exception is a virus that was found in the brown  
77 alga *Ectocarpus siliculosus*. This dsDNA virus, EsV-1, belonging to the  
78 Phaeoviruses with a genome of 335,593 bp, becomes associated to the host's genome  
79 (Müller et al. 1990, Klein et al. 1993, Delaroque et al. 2001, Cock et al. 2010). The  
80 virus particles are most expressed during reproduction and spread through infection  
81 of spores and gametes. It is manifested by malformed gametangia and lysis of  
82 swollen hyaline cells (Müller et al. 1990, 2000). Similar phaeoviruses have also been  
83 found in other brown algae, including, *Feldmannia irregularis*, *F. simplex*, (Müller  
84 and Frenzer 1993, Friess-Klebl et al. 1994), *Hincksia hincksiae* (Kapp et al. 1997),  
85 *Botrytella micromora* (as *Sorocarpus uvaeformis*) (Oliveira and Bisalputra 1978), *E.*  
86 *fasciculatus* (Sengco et al. 1996), *Streblonema* sp. (La Claire II and West 1977), and  
87 *Pylaiella littoralis* (Maier et al. 1998). Virus infections have also been observed in  
88 the green coenocytic macroalga *Bryopsis cinicola* (Koga et al. 1998, 2003). Here one  
89 4.5 kbp sequence belonging to the dsRNA virus family Partitiviridae was found  
90 associated with mitochondria and for four others of about 2 kbp with chloroplasts  
91 (Ishihara et al. 1992).

92 Virus infections in red algae are much less studied than in brown algae; however,  
93 virus-like particles have been found, by electron microscopy-studies, in the  
94 cytoplasm of the freshwater red alga *Sirodotia tenuissima* (Lee 1971) and in the  
95 marine species *Acrochaetium (Audouinella) saviana* (Pueschel 1995), *Gracilaria*  
96 *epihippisora* (Apt and Gibor 1991), and *G. verrucosa* (Tripodi and Beth 1976). In

97 addition, gall-like structures have been found in several species of *Bostrychia*  
98 including *B. kelanensis*, *B. radicata*, *B. simpliciuscula*, and *B. tenella*, and in some  
99 cases the galls have been associated with virus-like particles (West et al. 2013). For  
100 all of these possible red algal viruses the type of virus is unknown, no sequences are  
101 known, and the only reported phenotype is the formation of galls. Thus, the  
102 prevalence and importance of viruses in red algal biology is virtually unknown.

103 *Chondrus crispus*, or Irish moss, is a common, often intertidal red macroalga that has  
104 been promoted as a model species for florideophyte red algae (Collén et al. 2014). It  
105 has a triphasic life cycle with morphologically similar tetrasporophytes (diploid),  
106 male, and female gametophytes (haploid) and morphologically different  
107 carposporophytes (diploid) on the female gametophyte. The 105-Mbp genome was  
108 recently sequenced and codes for 9,606 genes (Collén et al. 2013). During the  
109 purification of RNA for the *Chondrus* genome project a prominent band of unusual  
110 size was seen on agarose gels; sequencing of this band showed sequence similarities  
111 to totiviruses, dsRNA viruses that normally infect fungi.

112

## 113 **MATERIALS AND METHODS**

114 *Algal material and culture.* Unless explicitly stated otherwise studies were carried  
115 out on a strain of the red alga *Chondrus crispus* Stackhouse (Gigartinales) grown in  
116 unialgal culture (Collén et al. 2013) at the *Station Biologique de Roscoff*, from a male  
117 gametophyte collected at Peggy's Cove, Nova Scotia, Canada (44°29'N, 63°55'W)  
118 in 1985 by Juan Correa. For the other specified *C. crispus* life cycle stages, samples  
119 were collected near Roscoff or in other specified locations.

120 *RNA extraction and purification. Chondrus crispus* RNA was extracted based on a  
121 protocol adapted from Apt et al. (1995). Briefly, frozen tissue (~1 g) was ground  
122 with a small amount of sand and liquid N<sub>2</sub> in a mortar and pestle. The powder was  
123 transferred to a 50 ml Teflon tube, and extracted with 15 ml extraction buffer (100  
124 mM Tris-HCl pH 7.5, 1.5 M NaCl, 50 mM EDTA, 50 mM DTT, 2 % CTAB) for 60  
125 min, with continuous shaking at 200 rpm at room temperature. Cellular debris was  
126 removed after centrifugation with one volume of chloroform-isoamyl alcohol  
127 (24:1) at 10,000 g and 4 °C for 20 min. The aqueous phase was transferred to a  
128 Teflon tube and polysaccharides were precipitated with 0.25 volumes of 100 %  
129 ethanol and centrifuged with one volume of chloroform-isoamyl alcohol (24:1) at  
130 10,000 g and 4 °C for 20 min. Total RNA was precipitated overnight with 3 M LiCl  
131 and 1 % of β-mercaptoethanol; after centrifugation, the pellet was dissolved in TE  
132 buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). RNA quality was assessed on 1.5 %  
133 agarose gel stained with 0.1 µg/ml ethidium bromide (EtBr) and quantified using a  
134 NanoDrop ND-2000 spectrophotometer (NanoDrop products, Wilmington, DE,  
135 USA). CsCl separation of total nucleic acids was used to purify dsRNA in 2 ml  
136 polyallomer tubes with 0.95 g/ml CsCl in TE and 50 µl of 10 mg/ml EtBr solution.  
137 The solution was centrifuged for 24 h at 90,000 rpm. The dsRNA band was  
138 recovered and the EtBr was extracted with TE-saturated butanol. The remaining  
139 aqueous phase containing the dsRNA was diluted three times and isopropanol-  
140 precipitated over-night at -20 °C. After centrifugation for 45 min at 18,000 g and 4  
141 °C, the pellet was washed with 200 µL 80 % ethanol and recentrifuged for 30 min at  
142 18,000 g and 4 °C; the pellet was finally dissolved in 100 µL of TE and kept at -20  
143 °C until use.

144 *Cloning and sequencing.* Cloning of RNAs was adapted from Djikeng et al. (2008)  
145 using random-tagged primers and RNA purified from agarose gels. Briefly, 100 ng  
146 of RNA and 2 pmol of random tagged primers were denaturated for 5 min at 70 °C.  
147 The first cDNA strand was synthesized using the Improm-RT polymerase Kit  
148 (Promega), after annealing 5 min at 25 °C, elongation 1 h at 38 °C, and inactivation  
149 15 min at 70 °C the RNA strand was degraded with 2 units of RNase H, 20 min at 37  
150 °C. The second strand was generated using 2 pmol of random tagged primer and 2.5  
151 units of Klenow. After purification cDNAs were amplified by PCR and cloned in  
152 TOPO-TA (Invitrogen). Race PCR was performed to obtain longer sequences using  
153 75 ng of gel-purified viral RNA with 1.2 µM specific primers, drawn from a first  
154 sequence alignment in the polymerase region, and 1.2 µM universal primers., Reverse  
155 transcription was carried out at 42 °C for 90 min, using SmartScribe reverse  
156 transcriptase Protocol (Clontech). After a 1/12.5 dilution, 1 µl of cDNA was  
157 amplified in a total volume of 25 µl, with 400 nM of each specific and universal  
158 primers, 0.2 mM dNTP, Pfu buffer, using 1.25 units Pfu DNA polymerase  
159 (Invitrogen). After a denaturation step for 4 min at 94 °C, samples were cycled in a  
160 ABI2700 thermocycler as following: 30 sec at 94 °C, 30 sec at 60 °C, 6 min at 72 °C  
161 for 5 cycles followed by 5 cycles of 30 sec at 94 °C, 30 sec at 56 °C, 6 min at 72 °C  
162 and 25 cycles of 30 sec at 94 °C, 30 sec at 52 or 56 °C, depending on the considered  
163 specific primer, 6 min at 72 °C and a final step of 10 min at 72 °C. PCR products (2  
164 µL) were directly cloned with TOPO-TA cloning (Invitrogen) and transformants  
165 were screened for size by PCR and sequenced using universal primers M13Forward  
166 and M13Reverse. Tail-PCR experiments were performed using combined random  
167 and specific primers; 200 ng RNA were reverse transcribed using 200 nM GSP1 and  
168 2 µM random primer RP1, using one step Access quick RT-PCR 2-fold master Mix

169 (Promega) in a 20  $\mu$ l total volume, according to the manufacturer specifications. The  
170 second step included 2.5  $\mu$ l of the first 50-fold diluted amplified sample, 2  $\mu$ M of  
171 RP1 and 200 nM internal GSP2, 3 mM MgCl<sub>2</sub>, 0.2 mM dNTP, Go-Taq buffer and 1  
172 unit Go-Taq polymerase (Promega). The third step is consisted in the same protocol  
173 as the second step, including a 50-fold dilution of the sample and 2  $\mu$ M of RP1 and  
174 200 nM of internal GSP3.

175 The cloning and sequencing effort generated only small sequence fragments in the  
176 capsid part and in the polymerase sequence region; in addition, race or tail-PCR  
177 never permitted amplification of the intergenic part, which can be described as a  
178 complex pseudoknot structured sequence. Because we demonstrated that at least  
179 seven different sequences could be cloned, with regard to the partial sequences motif  
180 obtained, we decided to try to extend some of the contigs obtained; using specific  
181 primers and one-step RT-PCR cross amplification between ends of the capsid part  
182 and the RdRp part to test whether extremities were from the same virus fragment or  
183 not. We started from 30 ng of CsCl purified dsRNA, denaturated for 4 min at 99 °C,  
184 reverse transcribed for 45 min at 45 °C and the PCR cycling parameters were  
185 programmed as follows: 4 min at 95 °C, and 32 cycles of 30 sec at 95 °C, 30 sec at  
186 54 °C, 3 min at 72 °C in a 2700 ABI thermocycler. Sequencing was performed using  
187 capillary electrophoresis on an ABI prism 3100 sequencer (Applied Biosystems,  
188 Foster City, CA, USA) with 50-cm capillaries.

189 *DNase and RNase essays.* To discriminate between DNA and RNA, we compared  
190 DNase and RNase digestion on total nucleic acid extracts. We subjected 1.5  $\mu$ g of  
191 total nucleic acid to 2 units of Turbo DNaseI (Ambion) or 1 unit of RNaseA  
192 (Macherey Nagel) for 1h at 40 °C. After precipitation, the samples were loaded on 1  
193 % agarose gel in TAE stained with EtBr.

194

## 195 **RESULT AND DISCUSSION**

196 Native agarose gel electrophoresis of RNA extracts from *Chondrus crispus* showed,  
197 in addition to rRNA and mRNA (and some DNA contamination), one prominent  
198 band with a size of approximately 6,000 bp. This band was not affected by DNase  
199 treatment but disappeared after treatment with RNase (Figure 1A). The bands were  
200 much less visible after denaturation of the RNA extract on a denatured gel (Figure  
201 1B) than on a native gel. This sensitivity to RNase indicated that the bands were  
202 made of RNA and likely double-stranded since the intensity was greatly diminished  
203 on a denatured gel. This is caused by the stronger affinity of the ethidium bromide  
204 stain to double-stranded RNA than single-stranded RNA (Sambrook and Russell  
205 2006).

206 The probably double-stranded RNA band was purified, reverse transcribed using  
207 random priming, and partially sequenced. Several sequences were attained which  
208 showed very high sequence diversity and could, thus, not be assembled in a  
209 continuous sequence. Several fragments were found to be similar to  
210 *Xanthophyllomyces dendrorhous* virus (XdV) (Figure 2) and this sequence was used  
211 to align the sequences from *C. crispus*. XdV is a double-stranded RNA virus,  
212 belonging to the totivirus (Baeza et al. 2012). The genome of totiviruses, typically  
213 with a genome size 4.6-6.3 kbp, normally codes for two proteins, one capsid (*gag*)  
214 gene and one RNA-dependent RNA polymerase (*pol*) gene with partially over-  
215 lapping open reading frames (ORFs). The sequences from the tentative *C. crispus*  
216 virus (CcV), contained partial sequences from both genes (Figure 3 and 4).  
217 Considerable efforts were made to try to obtain a full sequence, but each approach,

218 including SISPA method, race- or tail PCR, with degenerated, specific, or random  
219 PCR primers gave sequences that corresponded to new viral sequences instead of one  
220 specific sequence. This indicated a considerable sequence diversity within the CcV.  
221 It should be noted that the host strain of *C. crispus* used was artificially forced  
222 through a bottleneck of reduced size to reduce the possibility of the strain being a  
223 chimera before the genome sequencing (Collén et al. 2013) and thereafter grew  
224 vegetatively. The origin of the strain represents milligrams of tissue so that no  
225 nuclear sequence diversity was detected during genome sequencing.

226 The high error rate of viral replication with an RNA-dependent RNA polymerase  
227 compared to a DNA polymerase generates sequence diversity within viral  
228 populations that is higher than in cellular organisms. This causes the formation of  
229 quasispecies or mutant swarms (Andino & Domingo 2015). The CcV seems to be an  
230 extreme example of this with very high sequence diversity, represented by the 20-  
231 65% amino acid identity found between sequences (Table 1). Compared with other  
232 dsRNA viruses, such as the polio virus with two nucleotide mutations per genome or  
233 less than 0.03% difference in nucleotidic sequences (Vignuzzi et al. 2006) and a  
234 similarly low diversity in the St. Louis encephalitis virus (Ciota et al. 2011). The  
235 genetic diversity of CcV is thus several orders of magnitude higher. One explanation  
236 for this higher sequence diversity could be the absence (or rarity) of horizontal  
237 spreading of totiviruses compared to the polio and encephalitis viruses. To our  
238 knowledge, the sequence diversity of totiviruses within one host has only been  
239 reported as different viruses e.g. the *X. dendrorhous* virus XdVL1 and XdVL2  
240 (Baeza et al. 2012) which are less divergent than the CcV sequences. It should be  
241 noted that despite the large sequence variability only one distinct band was seen,  
242 implying sequence diversity rather than size diversity and also that size is conserved

243 even though the sequence is not. In addition, the acquired sequences typically coded  
244 for open reading frames. The CcV system is therefore a potentially interesting system  
245 to study the biology of quasispecies and a model to study viral evolution with  
246 previously unknown features.

247 Compared to totiviruses from, for example, *X. dendrorhous*, *Saccharomyces*  
248 *cerevisiae* and *Tuber aestivum*, the major conserved amino acid residues described in  
249 the consensus sequence of the totivirus *pol* gene (Baeza et al. 2012) are present in the  
250 CcV sequence (Figure 3). This is also true for domains 5 and 6, which have been  
251 demonstrated to be essential for viral function (Ribas 1992, Routhier 1998).  
252 Together, this strongly indicates that functional dsRNA from a totivirus-like entity is  
253 present in *C. crispus*.

254 The totiviruses belong to the mycoviruses which are widespread in yeasts and  
255 filamentous fungi but have not been reported from algae. Typically, they do not  
256 possess extracellular infections modes, are not associated with obvious disease  
257 symptoms, and do not exhibit harmful effects (Ghabrial and Suzuki, 2009). Only a  
258 few reports of phenotype modifications like hypovirulence or antifungal activity  
259 have been described (Pearson et al, 2009, Magliani 1997). Moreover, natural vectors  
260 are so far unknown (Ghabrial and Suzuki, 2009). It has been suggested that the  
261 dominant mode of diversification in most totiviruses is codivergence, parallel  
262 diversification between host and parasite (Göker et al, 2011). The large sequence  
263 diversity found in CcV indicates that other phenomena are probably occurring in the  
264 interaction between *C. crispus* and CcV.

265 Viruses from the totiviridae family are mainly found in fungi and protozoa. Penaeid  
266 shrimp infectious myonecrosis virus (IMNV), isolated from the Pacific white shrimp,

267 was the first totivirus described infecting a host other than a fungus or a protozoan  
268 (Nibert, 2007). Another virus, DTV, genetically related to IMNV was found to infect  
269 a *Drosophila* cell line (Wu et al., 2010). Totiviruses have also been isolated from  
270 mosquitoes (*Armigeres subalbatus*) (Zhai et al., 2010, Isawa et al., 2011). The  
271 viruses isolated from mosquitoes are genetically similar to IMNV and DTV (Isawa et  
272 al., 2011). A totivirus, related to the Giardiavirus, has also been found in salmon.  
273 (Haugland et al., 2011). Together with our data presented here, this shows that the  
274 host range for totivirus is relatively large within the eukaryotes.

275 The totivirus genome organization is characterized by a non-segmented dsRNA, with  
276 two genes, *gag* and *pol*, encoding two overlapping proteins, a capsid protein (CP)  
277 and an RNA-dependent RNA polymerase (RdRp). In most totiviruses, the (+) strand  
278 viral transcript is flanked by a 5' and a 3'-UTR and directs the translation of the CP  
279 protein and a smaller CP-RdRp fusion protein via a -1 ribosomal frameshift. The  
280 ribosomal -1 frameshift site consists of a heptamer with the consensus motif  
281 XXXYYYYZ, (X represents any nucleotide, Y represents A or U, and Z represents A,  
282 C or U) preceded by a GC dinucleotide (Jacks et al., 1988; Bekaert et al., 2003). The  
283 heptamer is followed by a short region preceding a stable RNA secondary structure,  
284 such as a pseudoknot or hairpin. The translational efficiency has been demonstrated  
285 to be directly related to the stability of these structures (Bidou et al. 1997). Slippery  
286 sites can promote a low level of frameshifting and downstream pseudoknot structures  
287 stimulate this process (Giedroc et al. 2000). In the 166 bp overlapping part of the  
288 sequence of CcV2 a putative slippery site, GGGUUUU, is located 18 nucleotides  
289 upstream a pseudoknot type structure. The structure has an estimated free energy at -  
290 18.85 kcal/mol (Figure 4). Thus, the CcV genome is organized as described in other  
291 totiviruses with two overlapping ORF including a 5'-ORF coding a capsid protein

292 and a 3'-ORF coding an RNA-dependent RNA polymerase. This involves an ORF  
293 disruption with an internal stop codon and -1 ribosomal frameshift.

294 Viral sequences from various families are sometimes found integrated into the host  
295 organism's genome (Holmes 2011). A search using the CcV sequence in the  
296 sequenced *C. crispus* nuclear genome showed that one non-identical copy of the *gag*  
297 gene was present (Fig 4B). RNAseq data showed expression of this gene in *C.*  
298 *crispus* (not shown). Sequence analysis suggests that the gene also contains  
299 regulatory sequences (see below). This indicates that a transfer has occurred from the  
300 virus to the *C. crispus* genome. This type of transfer between endogenous viral  
301 elements from non-retroviral RNA viruses is known from insects (Chrochu et al.  
302 2004), other animals (Katzourakis and Gifford 2010), and plants (Chiba et al. 2011).  
303 For the viral genes found in plant genomes they represented both single- and double-  
304 stranded viruses of various families. There is evidence that totiviral and partitiviral  
305 genes are transferred frequently into the nuclear genome of eukaryotes, and that  
306 some transferred genes have functions in the recipient genomes since totivirus-like  
307 sequences have been identified in, for example, *Medicago truncatula*, *Lotus*  
308 *japonica*, and *Populus trichocarpa*, where they have been matched to the *gag* gene of  
309 the *Vicia* cryptic virus M (Liu et al. 2010). ESTs from totiviruses have also been  
310 found in a cDNA library from the diatom *Phaeodactylum tricornutum* (Liu et al.  
311 2012), but it is unknown if this was due to incorporation in the genome or caused by  
312 the presence of viral entities. Normally, after integration of viral gene sequences in a  
313 genome the genes start to degenerate (Zwart et al. 2014). Considering that the *gag*  
314 gene in the nuclear genome is full length and has EST support, this suggests that  
315 either the gene provides an advantage to *C. crispus* or that the integration is recent. It  
316 could be noted that transgenic maize expressing the totiviral protein KP4 is more

317 resistant to fungal infection (Allen et al. 2011) and thus, incorporation of viral genes  
318 may have positive effects. No ESTs or presence in the genome were found for the  
319 *pol* gene in *C. crispus*.

320 A close analysis of nuclear genome regions around the *gag* gene gave unexpected  
321 results (Figure 6). Consensus sequences of prokaryote-type promoters were found in  
322 the upstream part of the gene. This includes a sigma70 type promoter located -11 of  
323 the TATAAA-box and -38 of the TTGAGA-box preceded by a CAAT-box located -  
324 67 before a putative start codon. Using yeast mitochondrial codon table instead of the  
325 eukaryote nuclear one results in a unique reading frame for the *cap* gene from the  
326 methionine codon found downstream the TATA-box. Interestingly, 35 aa preceding  
327 the putative *gag* start codon show similarities with a bacterial protein from *Simkania*  
328 *nevegensis*. Furthermore, in the same reading frame 24 aa sequence homologous to  
329 the cytochrome P450 (CYP) CYP808A2 (XP\_005710778.1) from *C. crispus*.  
330 Preceding the putative CYP gene are several putative regulatory boxes, three  
331 successive adjacent E-Boxes, followed by a G-Box, and a GCC-Box.

332 An E-Box is a cis-acting element, a conserved DNA sequence involved in regulation  
333 of gene expression linked to environmental and physiological signaling. These  
334 elements serve as binding sites for transcriptional activators, and have been  
335 demonstrated to be important for circadian gene expression (Giuliano et al., 1988,  
336 Staiger et al, 1989, Williams et al., 1992). For example, in the green alga  
337 *Chlamydomonas reinhardtii*, upregulation of the C3 subunit of the CHLAMY1  
338 RNA-binding protein is mediated by an E-box element in its promoter; causing  
339 temperature-dependent up-regulation of C3, its circadian expression, and its co-  
340 regulation by C1 (Seitz et al, 2010). The G-Box is a conserved DNA sequence in  
341 plants that has been described as a putative sugar sensor involved in starch

342 catabolism (Reinhold et al., 2011). A close proximity of the G-Box and the GCC-box  
343 is often found in tobacco and linked to the regulation of gene expression during plant  
344 defense responses (Buttner, 1997). GCC-box sequences in plants are promoter  
345 elements involved in the regulation of expression of plant defense genes against  
346 biotic or abiotic stressors and play an important role in relaying pathogen-initiated  
347 signals and the activation of plant defense responses (Solano et al, 1998, Brown et al,  
348 2003; Shenk et al, 2000; Reymond et al, 1998). GCC boxes have also been  
349 characterized in tomato endochitinase, glucanase and osmotin genes (Tournier et al,  
350 2003). This gene structure rich in sensor motifs and a viral-type sequence indicates  
351 that gene expression is modulated by changes in environmental or physiological  
352 conditions.

353 We used the *gag* and *pol* sequences from CcV and the nuclear *gag* sequence to  
354 understand the phylogeny of the CcV. The sequences for the *gag* and *pol* genes were  
355 compared with other sequences of totiviruses found in GenBank in December 2015.  
356 The *pol* genes from CcV cluster within the totivirus and is most closely related to the  
357 XdV gene (Figure 7). It is to be noted is that all of these sequences are from terrestrial  
358 organisms and most from Fungi. It is, however, unclear if the sequences obtained  
359 from terrestrial plants are due to fungal infections or if they are using the plant cells  
360 as hosts (Roossinck 2010). Similarly, the CcV *gag* genes cluster with the XdV gene.  
361 However, the genomic *gag* gene did not cluster with the CcV sequences, but rather  
362 with the *Saccharomyces* and *Scheffersomyces* virus gene sequences. The *gag* gene it  
363 is sometimes reported to be of viral origin and sometimes to be of nuclear origin  
364 (Figure 7), but no clear phylogenetic pattern emerged.

365 Even though the nuclear *gag* gene was expressed and dsRNA coding for the *gag* and  
366 *pol* genes were present in *C. crispus*, inspections of TEM images showed no clearly

367 identifiable virus-like particles (not shown) nor were any attempts to purify viral  
368 particles successful. Therefore, we are not presently convinced that viral particles are  
369 produced in *C. crispus* during normal conditions.

370 After finding the CcV in our gametophytic laboratory strain we wanted to know if  
371 this was specific for this strain and this life cycle stage. We therefore collected *C.*  
372 *crispus* from various locations and in different life cycle stages. The dsRNA could be  
373 visualized in all life stages tested, including tetrasporophytes and male and female  
374 gametophytes of *C. crispus* from Roscoff and in each part of the thallus tested. A  
375 band could also be seen in *C. crispus* samples from every location tested; in Europe  
376 from the northernmost location, Tjärnö in Sweden, via Plymouth in the UK, to the  
377 southernmost location, Porto in Portugal (Table 2). The band was also present with  
378 similar intensity in *C. crispus* with visible epiphytes and lesions. This indicates that  
379 the presence of dsRNA is wide-spread within *C. crispus*, at least along the European  
380 coast. For the laboratory strain with an origin in Canada, we cannot exclude an  
381 “infection” after the arrival of the strain in the laboratory in 1986.

382 We also wanted to verify if the presence of dsRNA was a phenomenon specific for  
383 *C. crispus* or if it was a more general occurrence. We therefore collected samples of  
384 other red algae around Roscoff, France (Table 3). A visible band was found in most  
385 of the red algae, except *Palmaria palmata* and *Furcellaria lumbricalis*, including one  
386 member of the Bangiales, a *Porphyra/Pyropia* species. This indicates that the  
387 presence of dsRNA identities of probable viral origin is wide-spread in red algae.

388 When the first sequence signatures showed similarities with totiviruses, which are  
389 mainly found in fungi or protozoa, we first suspected a contamination of *C. crispus*  
390 by epi- or endophytes in the culture medium. To exclude this, we decided to extract

391 nucleic acids from protoplasts. In protoplasts, the cell wall is enzymatically digested  
392 with cellulase and carrageenase (Le Gall et al, 1990). The total nucleic acid  
393 extraction from protoplast was then obtained and a band of similar intensity was seen  
394 (Figure 7).

395 We consider CcV to be associated with *C. crispus* and not other eukaryotic  
396 organisms since it was found in extractions of RNA performed simultaneously with  
397 the DNA extractions for the genome project and no DNA from other eukaryotes were  
398 found during the genome analysis. The cultures of *C. crispus* are not axenic, so it is  
399 possible that the dsRNA virus is associated with bacteria. However, the virus was  
400 also present in extractions of RNA from protoplasts (Figure 7), where bacteria  
401 associated with the cell wall were drastically reduced. Totiviruses are also unknown  
402 from bacteria (Ghabrial and Suzuki 2009) and considering the amounts of dsRNA  
403 found on the gels we find it is highly likely that CcV is associated with *C. crispus*  
404 and not an associated organism. This is, at least, true for the Peggy's Cove strain. For  
405 the other samples microscopic analysis to localize endophytic fungi gave negative  
406 results (data not shown), but we cannot formally exclude that endophytic fungi  
407 associated with algae might be the host of totiviruses for these strains and species.

408

## 409 **CONCLUSIONS**

410 A totivirus-like entity closely related to viral strains that mostly infect fungal species  
411 exists in the red macroalga *Chondrus crispus*. Double-stranded viral RNA was  
412 found in relatively high quantities, but no viral particles were found, nor has any  
413 phenotype been identified. It is present in all tested life cycle stages and ubiquitous in  
414 *C. crispus* isolates from different geographic areas. In a limited sample very large

415 sequence diversity was seen. Similar bands of dsRNA are present in other red algae.  
416 To our knowledge, this is the first time a virus-related sequence has been reported  
417 from a red alga and the first time a totivirus is reported from algae. The CcV genome  
418 is organized as described in other totiviruses with two overlapping ORFs including a  
419 5'-ORF coding a capsid protein (CP) and a 3'-ORF coding an RNA-dependent RNA  
420 polymerase. This involves an ORF disruption with an internal stop codon and a -1  
421 ribosomal frameshift. A copy of the gene encoding the capsid was identified in the  
422 nuclear genome of *C. crispus*. Despite the fact that the presence of dsRNA seems to  
423 be widespread in red algae nothing is known about the physiological and ecological  
424 consequences of the association with totivirus-like entities.

425

## 426 **ACKNOWLEDGMENTS**

427 Work at the *Station Biologique de Roscoff* was supported by the *Centre National de*  
428 *Recherche Scientifique*, the University Pierre and Marie Curie, and IDEALG Grants  
429 ANR-10-BTBR-04-02 and 04-04 “*Investissements d’avenir, Biotechnologies-*  
430 *Bioressources*”. We also want to thank Marion Azidrou and Léa Cabioch for work  
431 performed early in the project, Anne-Claire Baudoux for expertise on viral particle  
432 precipitation, Gabriel Markov for CYP analysis, Simon Dittami for helpful  
433 comments, and Gunilla Toth and Isabel Sousa Pinto for collecting *Chondrus crispus*  
434 in Sweden and Portugal.

435

436 **REFERENCES**

- 437 Allen, A., Islamovic, E., Kaur, J., Gold, S., Shah, D. & Smith, T. J. 2011. Transgenic  
438 maize plants expressing the Totivirus antifungal protein, KP4, are highly resistant to  
439 corn smut. *Plant Biotech. J.* 9:857-64.
- 440 Apt, K. E. & Gibor, A. 1991. The ultrastructure of galls on the red alga *Gracilaria*  
441 *epihippisora*. *J. Phycol.* 27:409-13.
- 442 Apt, K., Clendennen, S., Powers, D. & Grossman, A. 1995. The gene family  
443 encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis*  
444 *pyrifera*. *Mol. Gen. Genet.* 246:455-64.
- 445 Baeza, M., Bravo, N., Sanhueza, M., Flores, O., Villarreal, P. & Cifuentes, V. 2012.  
446 Molecular characterization of totiviruses in *Xanthophyllomyces dendrorhous*. *Virolog.*  
447 *J.* 9:140.
- 448 Bekaert, M., Bidou, L., Denise, A., Duchateau-Nguyen, G., Forest, J. P., Froidevaux,  
449 C., Hatin, I., Rousset, J. P. & Termier, M. (2003). Towards a computational model  
450 for- 1 eukaryotic frameshifting sites. *Bioinformatics* 19:327-35.
- 451 Bidou, L., Stahl, G., Grima, H., Liu, H., Cassan, M. & Rousset, J. P. 1997. In vivo  
452 HIV -1 frameshifting efficiency is directly related to the stability of the stem-loop  
453 stimulatory signal. *RNA* 3:1153-8.
- 454 Brown, R. L., Kazan, K., McGrath, K. C., Maclean, D. J. & Manners, J. M. 2003. A  
455 role for the GCC-box in jasmonate-mediated activation of the PDF1. 2 gene of  
456 *Arabidopsis*. *Plant Physiol.* 132:1020-32.
- 457 Bruenn, J. A. 1993. A closely related group of RNA-dependent RNA polymerases  
458 from double-stranded RNA viruses. *Nucleic Acids Res.* 21:5667-9.

459 Brussaard, C. P., Noordeloos, A. A., Sandaa, R. -A. A., Heldal, M. & Bratbak, G.  
460 2004. Discovery of a dsRNA virus infecting the marine photosynthetic protist  
461 *Micromonas pusilla*. *Virology* 319:280-91.

462 Buttner, M. & Singh, K. B. 1997. *Arabidopsis thaliana* ethylene-responsive element  
463 binding protein (AtEBP), an ethylene-inducible, GCC-box DNA binding protein  
464 interacts with an ocs element binding protein. *Proc. Natl. Acad. Sci. USA* 94:5961-6.

465 Byun, Y. & Han, K. 2009. PseudoViewer3: generating planar drawings of large-scale  
466 RNA structures with pseudoknots. *Bioinformatics* 25:1435-7.

467 Cao, S. & Chen, S. J. 2006. Predicting RNA pseudoknot folding thermodynamics.  
468 *Nucleic Acid Res.* 34:2634-52.

469 Chen, C. & Montelaro, R. C. 2003. Characterization of RNA elements that regulate  
470 Gag-Pol ribosomal frameshift in equine infectious anemia virus. *J. Virol.* 77:10280-  
471 7.

472 Chen, S. J. & Cao, S. 2009. Predicting structures and stabilities for H-type  
473 pseudoknots with interhelix loops. *RNA* 15:696-706.

474 Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S. & Suzuki,  
475 N. 2011. Widespread endogenization of genome sequences of non-retroviral RNA  
476 viruses into plant genomes. *PLoS pathogens* 7:e1002146.

477 Cock, J. M., Sterck, L., Rouzé, P., Scornet, D., Allen, A. E., Amoutzias, G.,  
478 Anthouard, V. et al. 2010. The Ectocarpus genome and the independent evolution of  
479 multicellularity in the brown algae. *Nature* 465:617-21.

480 Collén, J., Cornish, M. L., Craigie, J., Ficko-Blean, E., Hervé, C., Krueger-Hadfield,  
481 S. A., Leblanc, C. et al. 2014. *Chondrus crispus*- a present and historical model  
482 organism for red seaweeds. *Adv. Bot. Res.* 71:53-90.

483 Collén, J., Porcel, B., Carre, W., Ball, S. G., Chaparro, C., Tonon, T., Barbeyron, T.  
484 et al. 2013. Genome structure and metabolic features in the red seaweed *Chondrus*  
485 *crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA*  
486 110:5247–52.

487 Crochu, S., Cook, S., Attoui, H., Charrel, R. N., De Chesse, R., Belhouchet, M,  
488 Lemasson, J. J. et al. 2004. Sequences of flavivirus-related RNA viruses persist in  
489 DNA form integrated in the genome of *Aedes* spp. mosquitoes. *J. Gen. Virol.*  
490 85:1971-80.

491 Cui, J. & Holmes, E. C. 2012. Endogenous RNA viruses of plants in insect genomes.  
492 *Virology* 427:77-9.

493 Delaroque, N., Boland, W., Müller, D. G. & Knippers, R. 2003. Comparisons of two  
494 large phaeoviral genomes and evolutionary implications. *J. Mol. Evol.* 57:613-2.

495 Delaroque, N., Müller, D. G., Bothe, G., Pohl, T., Knippers, R. & Boland, W. 2001.  
496 The complete DNA sequence of the *Ectocarpus siliculosus* virus EsV-1 genome.  
497 *Virology* 287:112-32.

498 Derelle, E., Ferraz, C., Escande, M. L., Eychenié, S., Cooke, R., Piganeau, G.,  
499 Desdevises, Y., Bellec, L., Moreau, H. & Grimsley, N. 2008. Life-cycle and genome  
500 of OtV5, a large DNA virus of the pelagic marine unicellular green alga  
501 *Ostreococcus tauri*. PLoS One 3, e2250.

502 Derelle, E., Monier, A., Cooke, R., Worden, A. Z., Grimsley, N. H. & Moreau, H.  
503 2015. Diversity of viruses infecting the green microalga *Ostreococcus lucimarinus*.  
504 *J. Virol.* 89:5812-21.

505 Djikeng, A., Halpin, R., Kuzmickas, R., DePasse, J., Feldblum, J., Sengamalay, N.,  
506 Afonso, C. et al. 2008. Viral genome sequencing by random priming methods. *BMC*  
507 *Genomics* 9:5.

508 Etten, J., Lane, L. & Meints, R. 1991. Viruses and virus like particles of eukaryotic  
509 algae. *Microbiol. Rev.* 55:586-620.

510 Friess-Klebl, A. K., Knippers, R. & Müller, D.G. 1994. Isolation and characterization  
511 of a DNA virus infecting *Feldmannia simplex* (Phaeophyceae). *J. Phycol.* 30:653–8.

512 Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological  
513 effects. *Nature* 399:541–8.

514 Ghabrial, S. A. & Nibert, M. L. 2009. Victorivirus, a new genus of fungal viruses in  
515 the family Totiviridae. *Arch. Virol.* 154:373-9.

516 Ghabrial, S. A. & Suzuki, N. 2009. Virus of plant pathogenic fungi. *Annu. Rev.*  
517 *Phytopathol.* 47:353-84.

518 Giedroc, D. P., Theimer, C. A. & Nixon, P. L. 2000. Structure, stability and function  
519 of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J. Mol. Biol.*  
520 298:167-85.

521 Giuliano, G., Pichersky, E., Malik, V. S., Timko, M.P., Scolnik, P. A. & Cashmore,  
522 A. R. 1988. An evolutionary conserved protein binding sequence upstream of a plant  
523 light-regulated gene. *Proc. Natl. Acad. Sci. USA* 85:7089-93.

524 Göker, M., Scheuner, C., Klenk, H. P., Stielow, J. B. & Menzel, W. 2011.  
525 Codivergence of mycoviruses with their hosts. *PLoS One* 6:e22252.

526 Holmes, E. C. 2011. The evolution of endogenous viral elements. *Cell Host Microbe*  
527 10:368-77.

528 Ishihara, J., Pak, J. Y., Fukuhara, T. & Nitta, T. 1992. Association of particles that  
529 contain double-stranded RNAs with algal chloroplasts and mitochondria. *Planta*  
530 187:475-82.

531 Ivey, R. G., Henry, E. C., Lee, A. M., Klepper, L., Krueger, S. K. & Meints, R. H.  
532 (1996) A *Feldmannia* algal virus has two genome size-classes. *Virology* 220:267-73.

533 Jacks, T., Madhani, H. D., Masiarz, F. R. & Varmus, H. E. 1988. Signals for  
534 ribosomal frameshifting in the Rous sarcoma virus gag-pol region. *Cell* 55:447-58.

535 Kapp, M., Knippers, R. & Müller, D. G. 1997. New members of a group of DNA  
536 viruses infecting brown algae. *Phycol. Res.* 45:85-90.

537 Katzourakis, A. & Gifford, R. J. 2010. Endogenous viral elements in animal  
538 genomes. *PLoS Genet.* 6:e1001191.

539 Klein, S. T. L. M., Ramsperger, U., Müller, D. G. & Knippers, R. 1993. Genome  
540 structure of a virus infecting the marine brown alga *Ectocarpus siliculosus*. *Virology*  
541 193:802-11.

542 Koga, R., Fukuhara, T. & Nitta, T. 1998. Molecular characterization of a single  
543 mitochondria-associated double-stranded RNA in the green alga *Bryopsis*. *Plant Mol.*  
544 *Biol.* 36:717-24.

545 Koga, R., Horiuchi, H. & Fukuhara, T. 2003. Double-stranded RNA replicons  
546 associated with chloroplasts of a green alga, *Bryopsis cinicola*. *Plant Mol. Biol.*  
547 51:991-9.

548 Koonin E. V. 1991. Genome replication/expression strategies of positive-strand RNA  
549 viruses: a simple version of a combinatorial classification and prediction of new  
550 strategies. *Virus Genes* 5:273-81.

551 La Claire II, J. W. & West, J. A. 1977. Virus-like particles in the brown alga  
552 *Streblonema*. *Protoplasma* 93:127-30.

553 Lang, A., Rise, M., Culley, A. & Steward, G. 2009. RNA viruses in the sea. *FEMS*  
554 *Microbiol. Rev.* 33:295-323.

555 Le Gall, Y., Braud, J. P. & Kloareg, B. 1990. Protoplasts production in *Chondrus*  
556 *crispus* gametophytes (Gigartinales, Rhodophyta). *Plant Cell Rep.* 8:582-5.

557 Lee, R. E. 1971. Systemic viral material in the cells of the freshwater red alga  
558 *Sirodotia tenuissima* (Holden) Skuja. *J. Cell Sci.* 8:623-31.

559 Liu, H., Fu, Y., Jiang, D., Li, G., Xie, J., Cheng, J., Peng, Y., Ghabrial, S. A. & Yi,  
560 X. 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to  
561 eukaryotic nuclear genomes. *J. Virol.* 84:11876-87.

562 Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Peng, Y., Yi, X. & Jiang,  
563 D. 2012. Evolutionary genomics of mycovirus-related ds RNA viruses reveals cross-  
564 family horizontal gene transfer and evolution of diverse viral lineages. *BMC Evol.*  
565 *Biol.* 12:91

566 Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Yi, X. & Jiang, D. 2012.  
567 Discovery of novel dsRNA viral sequences by in silico cloning and implications for  
568 viral diversity, host range and evolution. *PLoS one* 7(7).

569 Magliani W., Conti S., Gerloni M., Bertolotti D. & Polonelli L. 1997. Yeast killer  
570 systems. *Clin. Microbiol. Rev.* 10:369–400.

571 Maier, I., Wolf, S., Delaroque, N., Müller, D. G. & Kawai, H. 1998. A DNA virus  
572 infecting the marine brown alga *Pilayella littoralis* (Ectocarpales, Phaeophyceae) in  
573 culture. *Europ. J. Phycol.* 33:213-20.

574 Minoia, S., Navarro, B., Covelli, L., Barone, M., García-Becedas, M. T., Ragozzino,  
575 A., Alioto, D., Flores, R. & Di Serio, F. 2014. Viroid-like RNAs from cherry trees  
576 affected by leaf scorch disease: further data supporting their association with  
577 mycoviral double-stranded RNAs. *Arch. Virol.* 159:589-93.

578 Monier, A., Pagarete, A., de Vargas, C., Allen, M. J., Read, B., Claverie, J. -M. M. &  
579 Ogata, H. 2009. Horizontal gene transfer of an entire metabolic pathway between a  
580 eukaryotic alga and its DNA virus. *Genome Res.* 19:1441-9.

581 Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E. & Grimsley, N.  
582 2010. Marine prasinovirus genomes show low evolutionary divergence and  
583 acquisition of protein metabolism genes by horizontal gene transfer. *J. Virol.*  
584 84:12555-63.

585 Müller, D. G. & Frenzer, K. 1993. Virus infections in three marine brown algae:  
586 *Feldmannia irregularis*, *F. simplex*, and *Ectocarpus siliculosus*. In Chapman, A. R.  
587 O., Brown, M. T. & Lahaye, M. [Eds.] *Fourteenth International Seaweed*  
588 *Symposium*. Springer Netherlands, pp. 37-44.

589 Müller, D. G., Kawai, H., Stache, B. & Lanka, S. 1990. A virus infection in the  
590 marine brown alga *Ectocarpus siliculosus* (Phaeophyceae). *Bot. Acta* 103:72-82.

591 Müller, D. G., Westermeier, R., Morales, J., Reina, G. G., Del Campo, E., Correa, J.  
592 A. & Rometscha, E. 2000. Massive prevalence of viral DNA in *Ectocarpus*  
593 (Phaeophyceae, Ectocarpales) from two habitats in the North Atlantic and South  
594 Pacific. *Bot. Mar.* 43:157-9.

595 Oliveira, L. & Bisalputra, T. 1978. A virus infection in the brown alga *Sorocarpus*  
596 *uvaeformis* (Lyngbye) Pringsheim (Phaeophyta, Ectocarpales). *Ann. Bot.* 42:439-45.

597 Pearson, M. N., Beever, R. E., Boine, B. & Arthur, K. 2009. Mycoviruses of  
598 filamentous fungi and their relevance to plant pathology. *Mol. Plant Pathol.* 10:115–  
599 28.

600 Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D.  
601 et al. 2013. Pandoraviruses: Amoeba viruses with genomes up to 2.5 Mb reaching  
602 that of parasitic eukaryotes. *Science* 341:281–6.

603 Pueschel, C. M. 1995. Rod-shaped virus-like particles in the endoplasmic reticulum  
604 of *Audouinella saviana* (Acrochaetiales, Rhodophyta). *Can. J. Bot.* 73:1974-80.

605 Reinhold, H., Soyk, S., Šimková, K., Hostettler, C., Marafino, J., Mainiero, S.,  
606 Vaughan, C. K., Monroe, J. D. & Zeeman, S. C. 2011.  $\beta$ -Amylase-like proteins  
607 function as transcription factors in Arabidopsis, controlling shoot growth and  
608 development. *Plant Cell* 23:1391-1403.

609 Reisser, W. 1993. Viruses and virus-like particles of freshwater and marine  
610 eukaryotic algae—a review. *Arch. Protistenkunde* 143:257-65.

611 Ribas, J. C. & Wickner, R. B. 1998. The Gag domain of the Gag-Pol fusion protein  
612 directs incorporation into the LA double-stranded RNA viral particles in  
613 *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273:9306-11.

614 Roossinck, M. J. 2010. Lifestyles of plant viruses. *Phil. Trans. R. Soc. B* 365:1899-  
615 1905.

616 Routhier, E. & Bruenn, J. A. 1998. Functions of conserved motifs in the RNA-  
617 dependent RNA polymerase of a yeast double-stranded RNA virus. *J. Virol.*  
618 72:4427-9.

619 Sambrook, J. & Russell, D. W. 2006. Detection of DNA in agarose gels. *CSH*  
620 *Protocols* doi: 10.1101/pdb.prot4022.

621 Schenk, P. M., Elliott, A. R. & Manners, J. M. 1998. Assessment of transient gene  
622 expression in plant tissues using the green fluorescent protein as a reference. *Plant*  
623 *Mol. Biol. Rep.* 16:313–22.

624 Schroeder, D. C. 2011. Viruses of seaweeds. In Hurst, J. C. [Ed.] *Studies in viral*  
625 *ecology: Microbial and botanical host systems 1*. John Wiley & Sons, Inc., pp. 205-  
626 15.

627 Seitz, S. B., Voytsekh, O., Mohan, K. M. & Mittag, M. 2010. The role of an E-box  
628 element: multiple functions and interacting partners. *Plant Signal. Behav.* 5:1077-80.

629 Seitz, S. B., Weisheit, W. & Mittag, M. 2010. Multiple roles and interaction factors  
630 of an E-box element in *Chlamydomonas reinhardtii*. *Plant Physiol.* 152:2243-57.

631 Sengco, M., Bräutigam, M., Kapp, M. & Müller, D. 1996. Detection of virus DNA in  
632 *Ectocarpus siliculosus* and *E. fasciculatus* (Phaeophyceae) from various geographic  
633 areas. *Europ. J. Phycol.* 31:73-8.

634 Solano, R., Stepanova, A., Chao, Q. & Ecker, J. R. 1998. Nuclear events in ethylene  
635 signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and  
636 ETHYLENE-RESPONSE-FACTOR1. *Genes Dev.* 12: 3703-14.

637 Soyk, S., Šimková, K., Zürcher, E., Luginbühl, L., Brand, L. H., Vaughan, C. K.,  
638 Wanke, D. & Zeeman, S. C 2014. The enzyme-like domain of *Arabidopsis* nuclear  $\beta$ -  
639 amylases is critical for DNA sequence recognition and transcriptional activation.  
640 *Plant Cell* 26:1746-63.

641 Staiger, D., Kaulen, H. & Schell, J. 1989. A CACGTG motif of the *Antirrhinum*  
642 *majus* chalcone synthase promoter is recognized by an evolutionary conserved  
643 nuclear protein. *Proc. Natl. Acad. Sci. USA* 86:6930-4.

644 Stevens, K., Weynberg, K., Bellas, C., Brown, S., Brownlee, C., Brown, M. T. &  
645 Schroeder, D. C. 2014. A novel evolutionary strategy revealed in the phaeoviruses.  
646 *PLoS ONE* 9:e86040.

647 Suttle, C. A. 2005. Viruses in the sea, *Nature* 437:356-61.

648 Suttle, C. A. 2007. Marine viruses-major players in the global ecosystem. *Nature*  
649 *Rev. Microbiol.* 5:801-12.

650 Tournier, B., Sanchez-Ballesta, M. T., Jones, B., Pesquet, E., Regad, F., Latché, A.,  
651 Pech, J. C. & Bouzayen, M. 2003. New members of the tomato ERF family show  
652 specific expression pattern and diverse DNA-binding capacity to the GCC box  
653 element. *FEBS letters* 550:149-54.

654 Tripodi, G. & Beth, K. 1976. Unusual cell structures in tumor-like formations of  
655 *Gracilaria* (Rhodophyta). *Arch. Microbiol.* 108:167-74.

656 Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. 2006.  
657 Quasispecies diversity determines pathogenesis through cooperative interactions in a  
658 viral population. *Nature* 439:344-8.

659 West, J. A., Pueschel, C. M., Klochkova, T. A., Kim, G. H., de Goër, S. &  
660 Zuccarello, G. C. 2013. Gall structure and specificity in *Bostrychia* culture isolates  
661 (Rhodomelaceae, Rhodophyta). *Algae* 28:83-92.

662 Williams, M. E., Foster, R. & Chua, N. H. 1992. Sequences flanking the hexameric  
663 G-box core CACGTG affect the specificity of protein binding. *Plant Cell* 4:485-96.

664 Zhai, Y., Attoui, H., Jaafar, F. M., Wang, H. Q., Cao, Y.X., Fan, S. P., Sun, Y. X. et  
665 al. 2010. Isolation and full-length sequence analysis of *Armigeres subalbatus*  
666 totivirus, the first totivirus isolate from mosquitoes representing a proposed novel  
667 genus (Artivirus) of the family Totiviridae. *J. Gen. Virol.* 91:2836-45.

668 Zwart, M. P., Willemsen, A., Daròs, J. A. & Elena, S. F. 2014. Experimental  
669 evolution of pseudogenization and gene loss in a plant RNA virus. *Mol. Biol. Evol.*  
670 31:121-34.

671

672 Table 1. Comparisons of % identity between amino acid sequences obtained by  
 673 sequencing a dsRNA band from *Chondrus crispus* and comparing multiple  
 674 sequences for the *pol* gene. Sequences were translated to amino acids, aligned and  
 675 compared pairwise using Clustal Omega ([www.ebi.ac.uk](http://www.ebi.ac.uk)).

	<b>CcV1</b>	<b>CcV2</b>	<b>CcvP5</b>	<b>CcvP6</b>	<b>CcvP7</b>	<b>CcvP8</b>	<b>CcvP9</b>
<b>CcV1</b>	100						
<b>CcV2</b>	37	100					
<b>CcvP5</b>	44	52	100				
<b>CcvP6</b>	29	24	28	100			
<b>CcvP7</b>	50	57	55	29	100		
<b>CcvP8</b>	23	21	26	19	26	100	
<b>CcvP9</b>	50	61	52	32	65	27	100

676

677 Table 2. The presence of a detectable dsRNA band around 6 kbp in RNA  
 678 extractions from *Chondrus crispus*.

Life cycle stage	Location		Virus detected
Gametophyte male	Roscoff, France	(48° 43' N, 3° 58' W)	Yes
Gametophyte female	Roscoff		Yes
Tetrasporophyte	Roscoff		Yes
Gametophytes	Roscoff		Yes
stressed*			
Gametophyte	Tjämnö, Sweden	(58° 86' N, 11° 13' E)	Yes
Gametophyte	Plymouth, UK	(50° 39' N, 4° 9' W)	Yes
Gametophyte	Helgoland, Germany	(54° 11' N, 7° 53' E)	Yes
Gametophyte	Chaucre, France	(45° 59' N, 1° 22' W)	Yes
Gametophyte	Lorient, France	(47° 42' N, 3° 23' W)	Yes
Gametophyte	Viana do Castelo, Portugal	(41° 42' N, 8° 51' W)	Yes

679 \*Algae with visible epiphytes and lesions.

680

681 Table 3. The presence of a detectable dsRNA band with a size around 6  
 682 kbp in different species of red macroalgae collected around Roscoff,  
 683 France.

Species	Class	Band detectable
<i>Dumontia contorta</i> (S.G. Gmelin) Ruprecht	Florideophyceae	Yes
<i>Grateloupia turuturu</i> Yamada	Florideophyceae	Yes
<i>Chondracanthus acicularis</i> (Roth) Fredericq	Florideophyceae	Yes
<i>Porphyra/Pyropia</i> sp	Bangiophyceae	Yes
<i>Laurencia pinnatifida</i> (Hudson) J.V. Lamouroux	Florideophyceae	Yes
<i>Mastocarpus stellatus</i> (Stackhouse) Guiry	Florideophyceae	Yes
<i>Polysiphonia elongata</i> (Hudson) Sprengel	Florideophyceae	Yes
<i>Palmaria palmata</i> (L.) Weber & Mohr	Florideophyceae	No
<i>Corallina elongata</i> J. Ellis & Solander	Florideophyceae	Yes
<i>Furcellaria lumbricalis</i> (Hudson) J.V. Lamouroux	Florideophyceae	No

684

685 FIGURE LEGENDS

686

687 Figure 1. Agarose gel electrophoresis of nucleic acid extract from *Chondrus crispus*:

688 A. native control (1), native control incubated 1h at 40°C (2), DNase1 treated (3),  
689 RNase treated (4); L Smart Ladder (Eurogentec). B. Native control with 10%  
690 DMSO (1), after denaturing 5 min at 95 °C with 10% DMSO.

691

692 Figure 2. Schematic representation of the *Chondrus crispus* viral sequences

693 compared to the sequences of related viral sequences of *Xanthophyllomyces*  
694 *dendrorhous* (XdL1b, XdL2) and *Saccharomyces cerevisiae* (ScVL-A(L1)); note the  
695 presence of the viral capsid sequence integrated into the *C. crispus* genome  
696 (CcVGe). The capsid coding open reading frame (ORF) in blue, RNA dependent  
697 RNA polymerase coding ORF in red, and overlapping ORF in purple.

698

699 Figure 3. Alignment of domains in the RNA-dependent RNA polymerase (*pol*) gene

700 from totiviruses as determined in Bruenn (2003) and Koonin (1991) with the  
701 *Chondrus crispus* dsRNA sequences. The consensus conserved domains described in  
702 Baeza et al. (2012) are highlighted.

703

704 Figure 4. Pseudoknot structure prediction of the CcV2 sequence using DotKnot

705 method (based on Cao and Chen 2006, 2009) and drawn using PseudoViewer3  
706 (Byun et al. 2009).

707

708 Figure 5. Genome region of *Chondrus crispus* containing the *gag* gene homolog.  
709 CYP: fragment of a cytochrome P-450 gene similar to a *C. crispus* nuclear gene (in  
710 blue); E-Box, G-Box and GCC-Box (in grey boxes): cis-regulatory elements located  
711 upstream the *gag* gene CcVGe (in red); open reading frame related to an unknown  
712 protein found in *Simkania nevegensis* (in green); putative sigma-type promoters  
713 (underlined in black); universal genetic code/yeast mitochondrial genetic code (in  
714 italics).

715

716 Figure 6. A. An unrooted neighbor-joining tree of the capsid sequence (n, reported as  
717 nuclear gene; v, reported as viral gene). B. An unrooted neighbor-joining tree of  
718 RNA dependent RNA polymerase sequences. The alignment was based on entire  
719 sequences using Muscle and neighbor-joining analysis was carried out in MEGA.

720

721 Figure 7. Gel electrophoresis of nucleic acid extracts from *C. crispus* protoplasts on a  
722 1% agarose gel stained with EtBr. L: smart Ladder Eurogentec

723

Figure 1

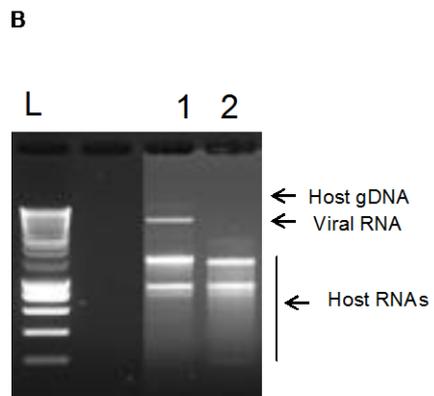
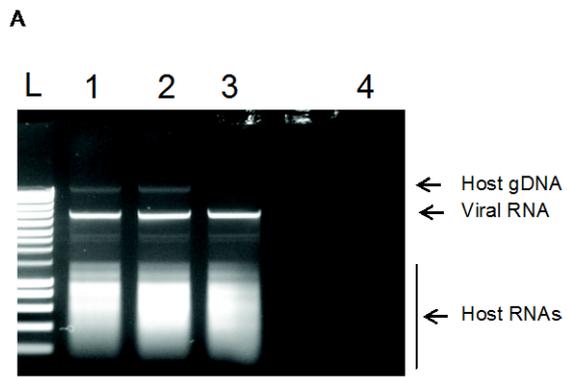
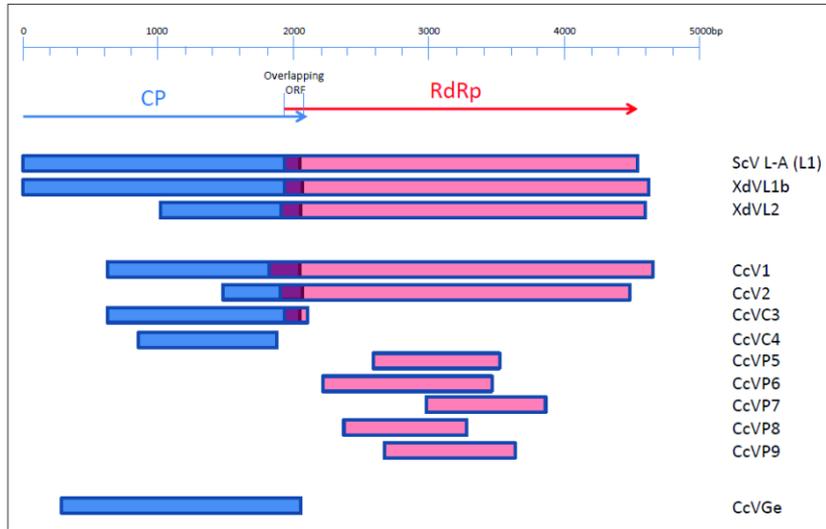
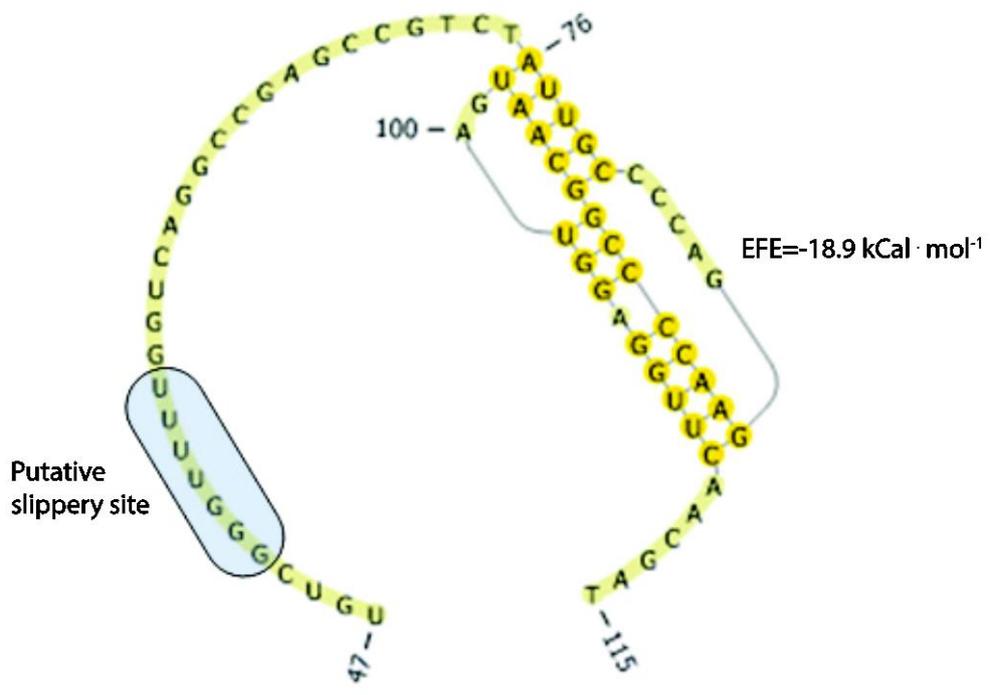


Figure 2



	Domain1	Domain2	Domain3
	XEXXXLXGRXXXXXXEXXXRX X	RXXWXXBXX X	KXEXGKRXI XXDXXXYXXXXXXXXXXE
CcV1	FEMNLI NRQGVVDVDAEYKRS	SRVDFPTGSV	IKYEWGRRAIYGDLTSLTNFAMVCE
CcV2	FEIDVLSNRVGGVDVWGEKENRL	ARVDSATGS	IKYEWGRRAIYGDLTSLTHFAFYNCE
CcVP9	FEAEVLNRSI GQVDVWGEKENRL	ARVDSAGGS	VKYEWGRRAIYGDLTSLMASHFAFYNCE
CcVP7	GVDVWGEKNRM	ARVDSAGGS	VKYEWGRRAIYGDLTSLMASHFAFYNCE
CcVP5	FEIDVLI NRYLEPIXESEKRNRC	ARVDSATGS	IKYEWGRRAIYGTI TSVLTFNFAFYNCE
X. dendrorhous V- L1b	FEIDT LI NRD GEVDWAEERDHRV	SRVDSAGGS	IKYEWGRRAIYGDLTSLWLSNFAFYNCE
X. dendrorhous V- L1a	FELEVLNRD GVDVWGEKENRC	NRVDSAGGS	TKYEWGRRAIYGDLTSLWLSNFAFYNCE
Black raspb erry F	FELQVLNRG GVDVWNERKIRG	SRVEVPTGSV	VKYEWGRRAIYGDLTSSVITNYAMFCE
T. aestivum V1	FELTVLLNRG GVDVWGEKENRV	SRVEVPTGSV	TKYEWGRRAIYGDLTSLITN FAMFCE
S. cerevisiae L- A	FELAVLNRG GVDVWAEKDFRL	NRVEVPTGSV	TKYEWGRRAIYGDLTSLITN FAMFCE
CcVP8	FELQVLNRG GVDVWAEKDFRL	QRSILMPGSV	V
S. segobiensis V- L Y	ELNVLNRDSDVWKEIDHRT	GRVAMPNSV	TKYEWGRRALYGDVTSFLHSDFGMTNCE
S. cerevisiae L- BC	FELNVLNGVDVWGEKENRS	GRVAMPNSV	TKYEWGRRALYGDVTSHTMABGLLQCE
CcVP6	FETEVLNRG GVDVWAEKDFRL	TRVLHVPTGSV	EKYEWGRRAIYGDVTEGFLCDLALPSAE

	Domain4	Domain5	Domain6	Domain7	Domain8
	XXDXDFNSQHX X	TLXSGRXTTFXNXLNXXY X	HXGDDX X	KXXXXXEFRLX X	YXXRXXXXXW
CcV1	DFDFDFNSQHS	GTLLSGRLTTFNSVLRNLY	LHNGDBL	TKCVFPGMIEFLRV	QYLTRAVATLV
CcV2	VDFDFNSQHS	GTLLSGRWLTFNSVLRNLY	VHNGDVM	TKGSFGGLAEFLRV	QVTRNATLV
CcVP9	IDFDFNSQHS	GTLLSGRWLTFNSVLRNLY			
CcVP7	IDFDFNSQHS	GTLLSGRWLTFNTLNLY	VHNGDDM		
CcVP5	DFDFDFNSQHS	GTLLSGRWLTFNSVLRNLY			
X. dendrorhous V- L1b	LDFFDFNSQHS	ATLLSGRWLTFNSVLRNLY	LHNGDDV	SKCAFVGGIAEFLRV	QYLTRAIATLM
X. dendrorhous V- L1a	LDFFDFNSQHS	GTLLSGRWLTFNSVLRNLY	LHNGDDV	SKCAFVGGIAEFLRV	QYLSRAATMR
Black raspb erry F	YDFDFNSQHS	GTLLSGRWLTFNTVLRNLY	VHNGDDV	AKQNVFSIGEFLRV	QYLSRAATLV
T. aestivum V1	FDYDFNSQHS	GTLLSGRWLTFNTVLRNLY	VHNGDDM	TKQNVLSISEFLRV	QYLSRSCATLV
S. cerevisiae L- A	FDYDFNSQHS	GTLLSGRWLTFNTVLRNLY	VHNGDDM	AKQNVLSISEFLRV	QYLSRSCATLV
CcVP8	FDYDFNSQHS	GTLLSGRWLTFNTVLRNLY	LHNGDBF	SKTNLGTIGEFLRV	QYLSRSCATLV
S. segobiensis V- L	YDFDFNSQHS	GTLLSGRWLTFNTVLRNLY	LHNGDDV	TKMNIGTIAEFLRV	QYLTRGIATFT
S. cerevisiae L- BC	YDFDFNSQHS	GTLLSGRWLTFNTVLRNLY	LHNGDDV	TKMNIGTIAEFLRV	QYLTRGIATFT
CcVP6	YDFDFNSQHS	GTLLSGRWLTFNTVLRNLY	LHNGDDV	TKMNIGTIAEFLRV	QYLTRGIATFT



727

Figure 5

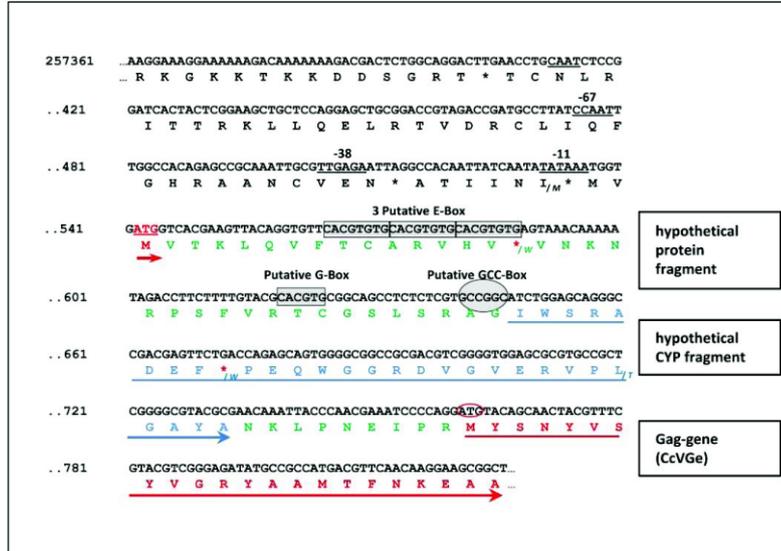


Figure 6

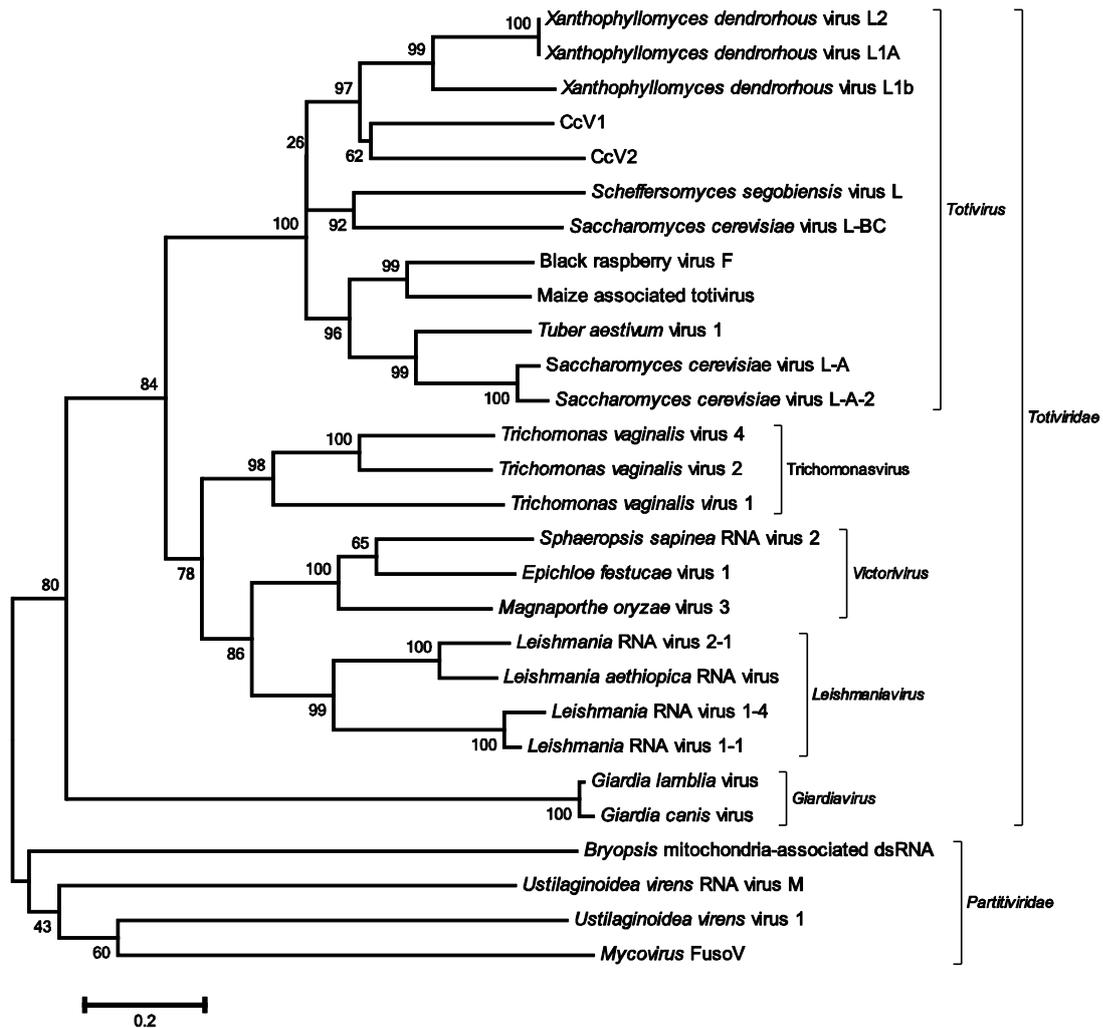
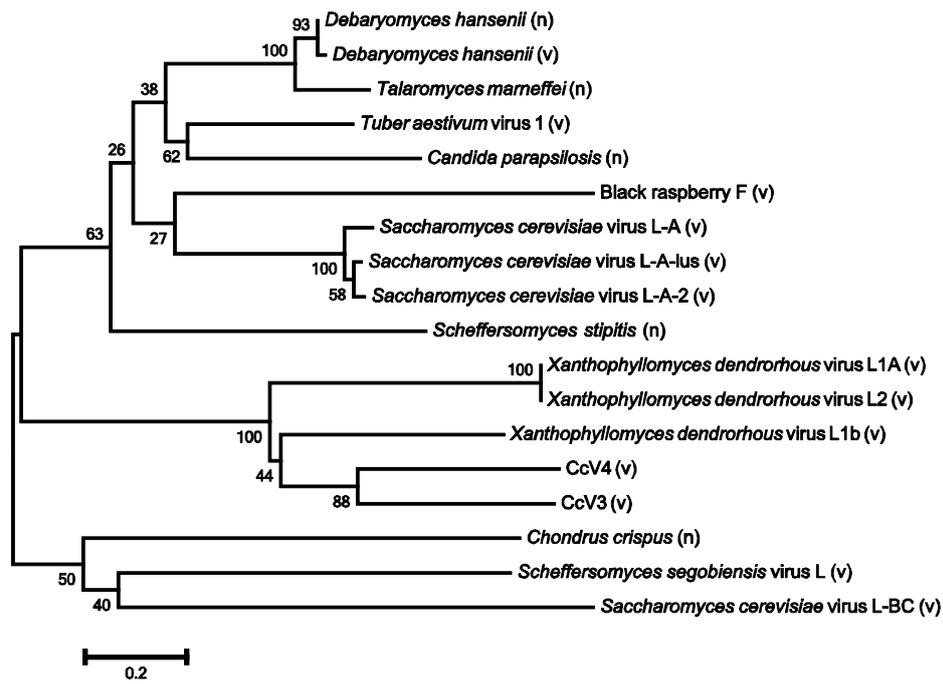


Figure 7

