J. Mar. Biol. Ass. U.K. (2006), **86**, 475–483
Printed in the United Kingdom

Genomic and phylogenetic analysis of a single-stranded RNA virus infecting *Rhizosolenia setigera* (Stramenopiles: Bacillariophyceae)

Y. Shirai*[‡], Y. Takao[†], H. Mizumoto*, Y. Tomaru*, D. Honda[†], K. Nagasaki*[‡ʃ]

*National Research Institute of Fisheries and Environment of Inland Sea, Fisheries Research Agency, 2-17-5 Maruishi, Hatsukaichi, Hiroshima 739-0452, Japan. [†]Department of Biology, Faculty of Science and Engineering, Konan University, 8-9-1 Okamoto, Higashinada, Kobe 658-8501, Japan. [‡]Y.S. and K.N. equally contributed to this study.
[ʃ]Corresponding author, e-mail: nagasaki@affrc.go.jp

We report the first complete genome sequence of the marine diatom-infecting, positive-sense single-stranded RNA (ssRNA) virus, *Rhizosolenia setigera* RNA virus (RsRNAV). The genome is 8877 nucleotides (nt), polyadenylated, lacking a cap structure, and has two large open reading frames (ORFs): ORF-1 (4818 nt), a polyprotein gene coding for replicases, e.g. RNA helicase, RNA-dependent RNA polymerase (RdRp); and ORF-2 (2883 nt), a polyprotein gene coding for structural proteins. The ORFs are separated by a 323 nt intergenic region (IGR), flanked by a 624 nt 5′-untranslated region (UTR) and a 229 nt 3′-UTR. The deduced amino acid sequences for ORF-1 and ORF-2 respectively show considerable similarities to the non-structural and structural proteins of a marine raphidophyte-infecting virus HaRNAV (*Heterosigma akashiwo* RNA virus). Phylogenetic analyses of concatenated amino acid sequences of RNA helicase and RdRp domains supported the monophyly of RsRNAV, HaRNAV and a marine protist-infecting virus SssRNAV (*Schizochytrium* single-stranded RNA virus) with moderate bootstrap values of 79–83%, but not at the family level, whilst their monophyly was only weakly supported (50–55%) in the phylogenetic tree based on RdRp whole domain. As a result, comparison of the genome organization and sequence suggests RsRNAV is not a member of any currently defined virus family. In the RdRp tree, the positive-sense ssRNA viruses infecting Stramenopiles (RsRNAV, HaRNAV and SssRNAV) and Alveolata (HcRNAV (*Heterocapsa circularisquama* RNA virus)) were categorized into phylogenetically distant clades, which suggests a host/virus coevolution. Our study supports the hypothesis that a diverse array of ssRNA viruses exists in marine environments.

INTRODUCTION

Until recently, the viral impact on diatoms was little understood due to the lack of cultivable viruses that infect diatoms. The first isolation of a diatom-infecting virus RsRNAV was reported by Nagasaki et al. (2004). RsRNAV is an icosahedral, 32 nm virus that harbours a ssRNA genome estimated to be 11.2 kb long using denaturing gel electrophoresis. This virus accumulates in the cytoplasm of *Rhizosolenia setigera* but not in the nucleus. The second diatom-infecting virus reported harboured a DNA genome that infects *Chaetoceros salsugineum* (CsNIV: C. *salsugineum* nuclear inclusion virus) (Nagasaki et al., 2005b). Its accumulation occurs in the nucleus. CsNIV has a unique genome structure consisting of a single molecule of covalently closed circular single-stranded DNA (ssDNA; ~6 kb) as well as a segment of linear ssDNA (~1 kb); the linear segment is complementary to a portion of the closed circle creating a partially double-stranded genome. The third virus isolated and preliminarily characterized is a small icosahedral virus, CspNIV (*Chaetoceros* nuclear inclusion virus), infectious to *C.* cf. *gracilis* (Bettarel et al., 2005). CspNIV accumulates in the host's nucleus often forming paracrystalline arrays. No data concerning its genome have been reported.

The study of RNA viruses infecting marine eukaryotic micro-organisms is in its infancy; there are only five viruses successfully cultured. Four are positive-sense ssRNA viruses: HaRNAV (Tai et al., 2003; Lang et al., 2004) infecting the bloom-forming raphidoflagellate *Heterosigma akashiwo* (Stramenopiles: Raphidophyceae) (classification of the host organism basically follows Cavalier-Smith (1998)); HcRNAV (Tomaru et al., 2004; Nagasaki et al., 2005a) infecting the bloom-forming dinoflagellate *Heterocapsa circularisquama* (Alveolata: Dinophyceae); RsRNAV (Nagasaki et al., 2004) infecting the bloom-forming diatom *R. setigera* (Stramenopiles: Bacillariophyceae); and SssRNAV (Takao et al., 2005, 2006) infecting the marine fungoid protist *Schizochytrium* sp. (Stramenopiles: Labyrinthulea). The other virus is a double-stranded RNA virus, MpRNAV (Brussaard et al., 2004), infecting the cosmopolitan picoplankton *Micromonas pusilla* (Chlorophyta: Prasinophyceae). Among the four ssRNA viruses, full genomic sequences of HaRNAV (Lang et al., 2004), HcRNAV (Nagasaki et al., 2005a) and SssRNAV (Takao et al., 2006) were determined; however, genomic data of RsRNAV have not been reported (Nagasaki et al., 2004).

By comparing the genome features of these viruses and using a phylogenetic analysis of the conserved regions in

the viral replicases (e.g. RdRp, RNA helicase), further information regarding the evolutionary history of marine viruses is expected (Koonin & Dolja, 1993; Zanotto et al., 1996). Recently, Culley et al. (2003) constructed a phylogenetic tree to show the existence of a diverse array of picorna-like viruses by analysing the RdRp sequences amplified from marine virus communities using a reverse-transcription polymerase chain reaction (RT-PCR) technique. In the tree, the amplified sequences were categorized into four phylogenetically distant groups (Groups A–D, Culley et al. (2003)); Group A includes a corresponding sequence for HaRNAV.

Here we: (i) determine the full genome sequence of RsRNAV; (ii) predict the genome structure; (iii) genetically and phylogenetically analyse the conserved domains in the RsRNAV genome; and (iv) conclude by discussing further studies of positive-sense ssRNA viruses infecting marine eukaryotic micro-organisms. This is the first report describing the genomic features of diatom-infecting ssRNA viruses and comparing the evolutionary relatedness among positive-sense ssRNA viruses infectious to Stramenopiles (RsRNAV, HaRNAV, and SssRNAV) and to Alveolata (HcRNAV) (Patterson, 1989; Cavalier-Smith, 1991).

MATERIALS AND METHODS

Host algal culture

The host for the virus strain tested (RsRNAV06 (RsRNAV strain 06)) was *Rhizosolenia setigera* strain S2. The host diatom culture was grown in modified SWM3 medium enriched with 2 nm Na_2SeO_3 and incubated using a 12L:12D h cycle of $\sim 110\,\mu$mol photons m^{-2} s^{-1} provided by cool white fluorescent illumination (FL40S D EDL D65, Toshiba) at 15°C (Nagasaki et al., 2004).

Purification of virus

Four hundred and fifty millilitres of exponentially growing culture of *R. setigera* strain S2 was inoculated with 22.5 ml of RsRNAV06 suspension ($\sim 10^7$ infectious particles ml^{-1}) and incubated. Virus particles were purified using the method described by Tomaru et al. (2004).

Purification and sequencing of the viral genome

Viral RNAs were isolated and purified using the RNeasy Mini Kit (Qiagen). The purified RNA was reverse-transcribed to construct cDNAs with the Time Saver cDNA Synthesis Kit (Amersham Biosciences Corp.) using random primers according to the manufacturer's recommendation. A NotI/EcoRI adaptor (Amersham Biosciences Corp.) was ligated to the resultant dsDNA fragments and the 5′-ends were phosphorylated using T4 polynucleotide kinase (Amersham Biosciences Corp.). The resulting fragments were ligated to the EcoRI-cleaved dephosphorylated pBlueScript SK(+) plasmid vector (Stratagene) using a Ligation High Kit (Toyobo Co., Ltd) according to the manufacturers' recommendations. The ligated dsDNA fragments were transformed into *Escherichia coli* DH5*a* competent cells (Toyobo Co., Ltd) and sequenced by the dideoxy method using an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems). The fragment sequences were assembled using DNASIS-Mac software (Hitachi Software Engineering) and Genetyx version 7.03 for Windows software (Genetyx Corp.).

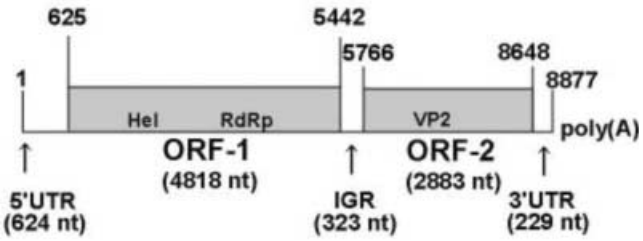

Figure 1. Schematic genome structure of RsRNAV06. Numbers indicate base positions from the 5′ terminus in the nucleotide sequence. Hel, RNA helicase domain; RdRp, RNA-dependent RNA polymerase domain.

A 5′-RACE (rapid amplification of cDNA ends) analysis was performed to determine the 5′-end sequence of the RsRNAV genomic RNA using the 5′-RACE System for Rapid Amplification of cDNA Ends, version 2.0 (Invitrogen). The 5′-terminal nucleotide sequence was determined by comparing the sequences of seven clones (data not shown). To reduce secondary structure in the isolated RNA, the first-strand cDNA synthesis was performed in the presence of 5% dimethyl sulphoxide (Wako Pure Chemical Industries Ltd). A 3′-RACE analysis was performed to determine the 3′-end sequence of the viral genome by using oligo-dT primer. The 3′-terminal nucleotide sequence was determined by comparing the sequences of five clones (data not shown).

The RsRNAV genome RNA was fully sequenced. Potential open reading frames were identified by using the ORF finder system (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Automated comparisons were conducted on the RsRNAV sequence using the BLAST (Basic Local Alignment Research Tool) algorithm (http://www.ncbi.nlm.nih.gov/blast/).

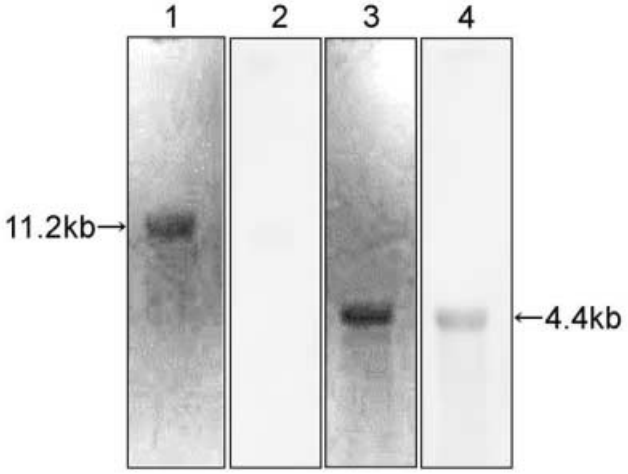

Figure 2. Electrophoresis of RsRNAV genomic RNA (lanes 1 and 2) and a 5′-capped HcRNAV transcript (lanes 3 and 4) stained with methylene blue (lanes 1 and 3) or immunostained using a monoclonal antibody specific against 2,2,7-trimethyl guanosine containing the cap structure (lanes 2 and 4).

https://doi.org/10.1017/S0025315406013373 Published online by Cambridge University Press

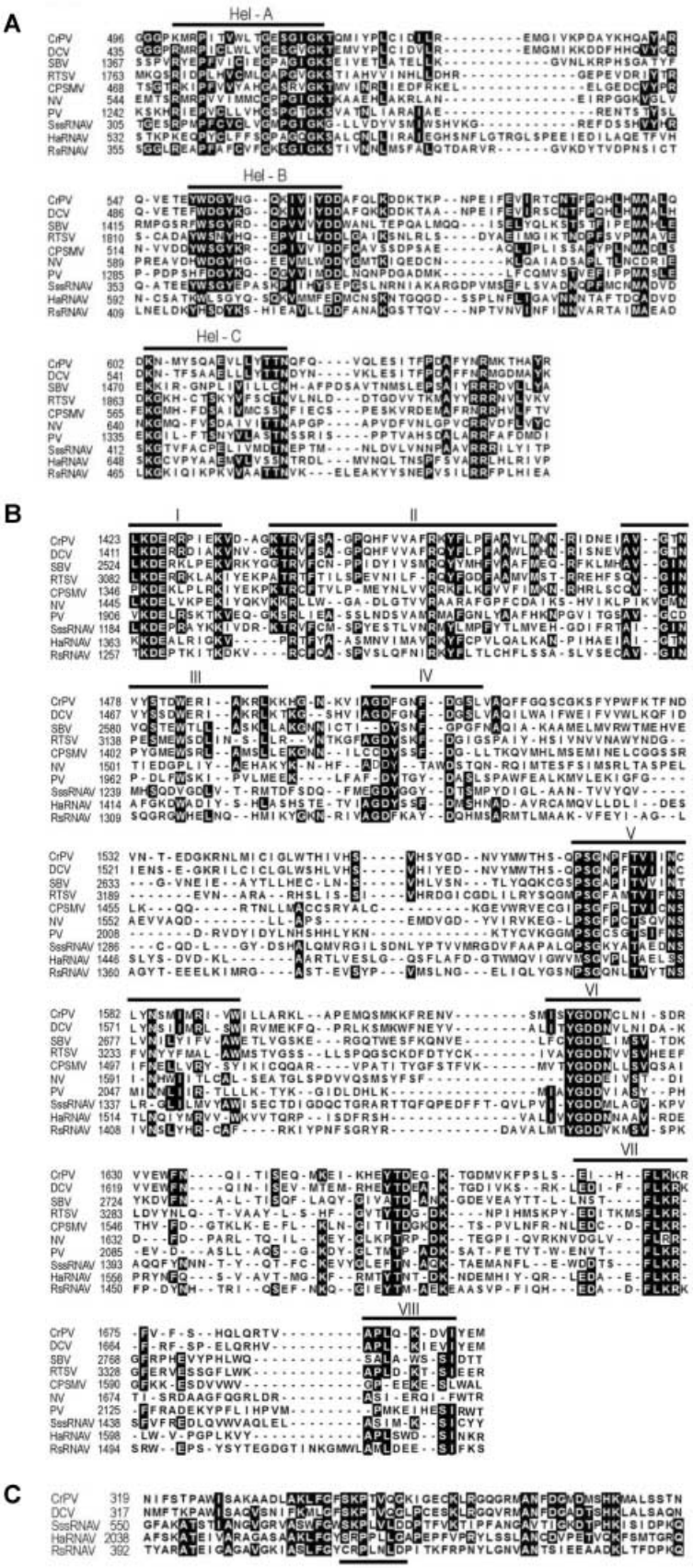

Figure 3. Alignment of the conserved regions in: (A) the putative RNA helicase; (B) RNA-dependent RNA polymerase (RdRp); and (C) structural protein(s) of RsRNAV and picorna-like viruses. The number to the left of each sequence indicates the amino acid position relative to the corresponding protein. Residues identical in more than 5 and 3 viruses are shaded in (A, B) and (C), respectively. The conserved regions of the RNA helicase and RdRp (Koonin & Dolja, 1993) are marked as Hel-A, B, and C in (A), and I–VIII in (B), respectively. The amino acid position that corresponds to the N-terminus region of VP2 is underlined in (C).

https://doi.org/10.1017/S0025315406013373 Published online by Cambridge University Press

Table 1. *Results of BLAST similarity search of RNA helicase domain (A) and RNA-dependent RNA polymerase domain (B) in RsRNAV ORF-1.*

BLASTP hit	Score	Expect value	Accession number (NCBI)
(A) Triatoma virus [non-structural protein precursor]	50.8 bits (120)	4E-05	NP_620562
Heterosigma akashiwo RNA virus SOG263 [polyprotein]	50.4 bits (119)	6E-05	AAP97137
Honeybee virus—Israel [non-structural polyprotein]	47.8 bits (112)	4E-04	AAV64179
Broad bean wilt virus 1 [polyprotein]	47.8 bits (112)	4E-04	NP_945134
Broad bean wilt virus 1 [NTP-binding protein]	47.8 bits (112)	4E-04	NP_951027
Plautia stali intestine virus [non-structural polyprotein]	47.8 bits (112)	4E-04	NP_620555
Schizochytrium single-stranded RNA virus [polyprotein]	47.4 bits (111)	5E-04	BAE47143
Kashmir bee virus [non-structrual polyprotein]	47.0 bits (110)	6E-04	NP_851403
Broad bean wilt virus 1 [polyprotein 1]	46.4 bits (109)	8E-04	AAX12375
(B) Unidentified picorna-like virus JP700-2 (Group D*) [RNA-dependent RNA polymerase]	132 bits (332)	2E-29	NP_620560
Broad bean wilt virus 1 [RNA-dependent RNA polymerase]	114 bits (286)	5E-24	NP_945134
Strawberry mottle virus [RNA-dependent RNA polymerase]	113 bits (283)	1E-23	NP_599086
Broad bean wilt virus 1 [polyprotein 1]	109 bits (272)	2E-22	AAX12375
Unidentified picorna-like virus FRP896-1 (Group B*) [RNA-dependent RNA polymerase]	108 bits (270)	4E-22	AAQ21147
Unidentified picorna-like virus JP800-6 (Group B*) [RNA-dependent RNA polymerase]	107 bits (268)	7E-22	AAQ21162
Unidentified chinese clam virus 16-50 [RNA-dependent RNA polymerase]	105 bits (262)	3E-21	AAL57344
Unidentified picorna-like virus JP700-1 (Group B*) [RNA-dependent RNA polymerase]	105 bits (261)	4E-21	AAQ21155
Heterosigma akashiwo RNA virus SOG263 [polyprotein]	103 bits (256)	2E-20	AAP97137
Homo sapiens [unnamed protein product]	103 bits (256)	2E-20	BAC87499
Unidentified picorna-like virus JP800-7 (Group B*) [RNA-dependent RNA polymerase]	102 bits (254)	3E-20	AAQ21163
Schizochytrium single-stranded RNA virus [polyprotein]	102 bits (254)	3E-20	BAE47143
Unidentified picorna-like virus JP800-10 (Group C*) [RNA-dependent RNA polymerase]	101 bits (252)	5E-20	AAQ21166
Unidentified picorna-like virus JP800-5 (Group C*) [RNA-dependent RNA polymerase]	100 bits (250)	8E-20	AAQ21161
Unidentified picorna-like virus JP800-3 (Group C*) [RNA-dependent RNA polymerase]	100 bits (250)	8E-20	AAQ21159
Unidentified picorna-like virus JP800-2 (Group C*) [RNA-dependent RNA polymerase]	100 bits (250)	8E-20	AAQ21158
Unidentified picorna-like virus JP800-9 (Group C*) [RNA-dependent RNA polymerase]	97.4 bits (241)	9E-19	AAQ21165
Drosophila C virus [replicase polyprotein]	97.4 bits (241)	9E-19	NP_044945
Deformed wing virus [polyprotein]	96.7 bits (239)	2E-18	AAP49008
Varroa destructor virus 1 [polyprotein]	95.9 bits (237)	3E-18	YP_145791
Cricket paralysis virus [non-structural polyprotein]	95.5 bits (236)	3E-18	NP_647481

BLASTP, Basic Local Alignment Research Tool Protein; NCBI, National Center for Biotechnology Information. *, Based on the phylogenetic tree constructed by Culley et al. (2003).

Protein sequencing

Purified RsRNAV virions were electrophoresed using a denaturing sodium dodecyl sulphate (SDS) polyacrylamide gel, blotted onto a PVDF (polyvinylidine difluoride) membrane (Millipore Immobion-P), and stained with a 1% Amidoblack solution (Katayama Chemical Industries Co., Ltd). Each stained protein band was excised and sequenced using an Edman degradation.

Detection of the 5′ cap structure

Specific detection of capped RNA was performed according to the method described by Mizumoto et al. (2003) with a minor modification. An RNA transcript was synthesized *in vitro* from the HcRNAV-cDNA clone (Mizumoto et al., unpublished data) using T7 RNA polymerase in the presence of the cap structure analogue (m^7GpppG; New England Biolabs) according to the manufacturer's recommendation; then, the resultant RNA was purified using a MEGAclear (Ambion, Inc.) according to the manufacturer's recommendation. This served as a positive control.

RsRNAV genomic RNA and HcRNAV 5′-capped RNA were fractionated on a 1.5% agarose gel, transferred to a Hybond-N+ membrane (Amersham Pharmacia Biotech.), and cross-linked onto the blots by incubation at 80°C for 2 h. After blotting, the membrane was immersed

https://doi.org/10.1017/S0025315406013373 Published online by Cambridge University Press

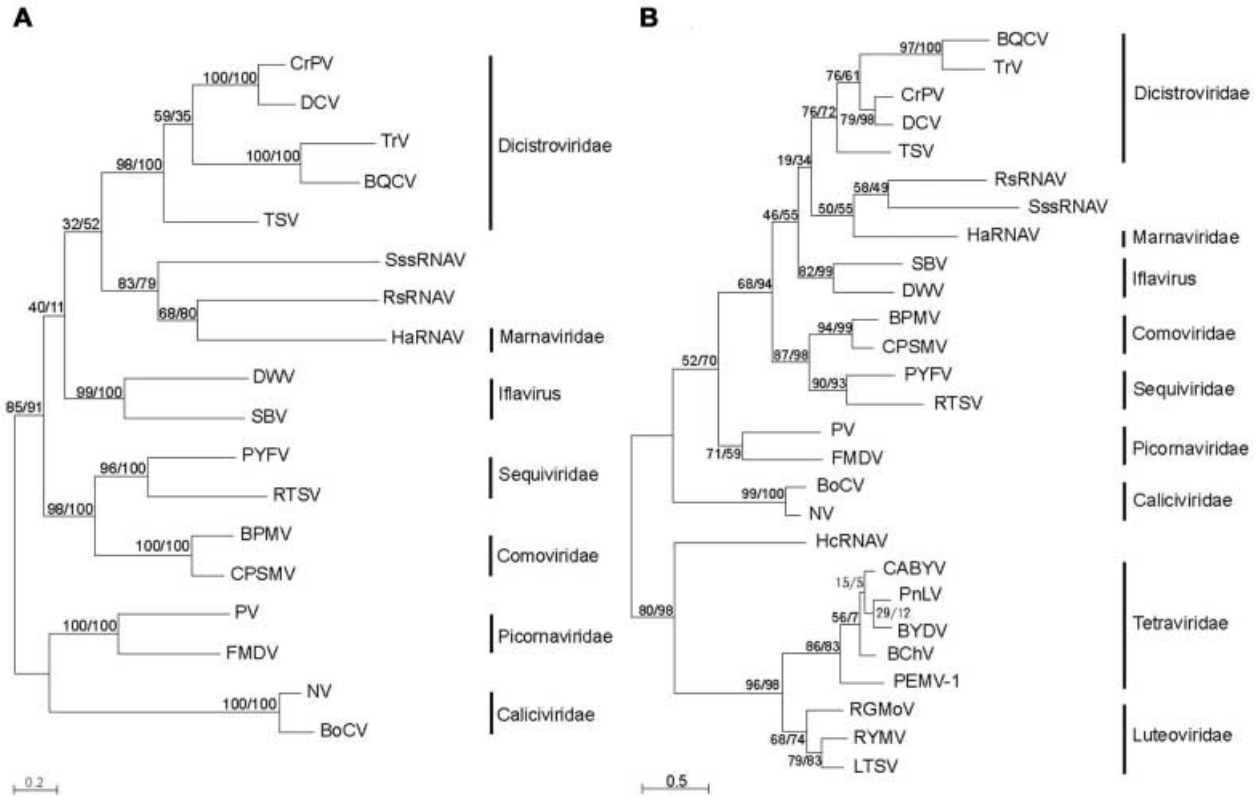

Figure 4. Maximum likelihood (ML) trees calculated from confidently aligned regions of concatenated amino acid sequences of RNA helicase domain and RNA-dependent RNA polymerase (RdRp) domain (A), and amino acid sequences of RdRp whole domain (B). ML bootstrap values (%) from 100 samples are shown at the nodes followed by bootstrap values based on neighbour-joining analysis (%) from 100 samples. The ML distance scale bars are shown.

in a 5% blocking solution (Amersham Biosciences Corp.) for 60 min and incubated with a 1/5000-dilluted monoclonal antibody reactive against the 2,2,7-trimethyl guanosine containing cap structure (Synaptic Systems GmbH) for 60 min. After washing with PBS–Tween buffer (80 mM Na_2HPO_4, 20 mM NaH_2PO_4, 100 mM NaCl, 0.05% Tween20, pH 7.5), the membrane was incubated with a 1/25000-diluted anti-mouse IgG, HRP-linked whole antibody (Amersham Biosciences Corp.), and visualized with ECL Western blotting detection reagents (Amersham Biosciences Corp.) according to the manufacturers' recommendations. Chemiluminescence was detected using a luminescence image analyser (LAS 3000 mini, Fuji Photo Film). Total RNA blotted onto the membrane was detected using methylene blue staining (Molecular Research Center, Inc.) according to the manufacturer's recommendation.

Phylogenetic analysis

Conserved domains of RNA helicase, RdRp and protease were identified in RsRNAV ORF-1, and the deduced amino acid sequences of the corresponding regions were compared to the other viruses. They were automatically aligned using ClustalW and manually refined. Phylogenetic trees were constructed by neighbour-joining (NJ) and maximum-likelihood (ML) methods by using the PHYLIP 3.65 Program with the Jones–Taylor–Thornton matrix (JTT model): they are based on the deduced amino acid sequences of the concatenated sequences of RNA helicase and RdRp domains, or the RdRp whole domain. Amino acid sequences used for comparison in the analyses are as follows with the organisms' scientific names, with abbreviations in parentheses if necessary, and the database accession numbers (referring to the US National Center for Biotechnology Information (NCBI) unless otherwise stated): beet chlorosis virus (BChV), AAK49964; bovine enteric calicivirus (BoCV), AJ011099; bean pod mottle virus (BPMV), NC_003496; black queen cell virus (BQCV), NC_003784; barley yellow dwarf virus (BYDV), BAA01054; cucurbit aphid-borne yellows virus (CABYV), CAA54251; cowpea severe mosaic virus (CPSMV), M83830; cricket paralysis virus (CrPV), NC_003924; *Drosophila* C virus (DCV), NC_001834; deformed wing virus (DWV), NC_004830; foot-and-mouth disease virus (FMDV), P03306; *Heterosigma akashiwo* RNA virus (HaRNAV), NC_005281; *Heterocapsa circularisquama* RNA virus 109 (HcRNAV), AB218609; Lucerne transient streak virus (LTSV), NP_736596; Norwalk virus (NV), M87661; pea enation mosaic virus 1 (PEMV-1), AAA72297; *Poinsettia* latent virus (PnLV), CAI34771; human poliovirus 1 Mahoney (PV), V01149; parsnip yellow fleck virus (PYFV), D14066; ryegrass mottle virus (RGMoV), NP_736587; 1KHVA; *Rhizosolenia setigera* RNA virus (RsRNAV), DDBJ Accession number AB243297; rice turgo spherical virus (RTSV), AAA66056; rice yellow mottle virus (RYMV), CAE81345; sacbrood virus (SBV), NC_002066; *Schizochytrium* single-stranded RNA virus (SssRNAV), BAE47143; triatoma virus (TrV), NC_003783; Taura syndrome virus (TSV), NC_003005.

https://doi.org/10.1017/S0025315406013373 Published online by Cambridge University Press

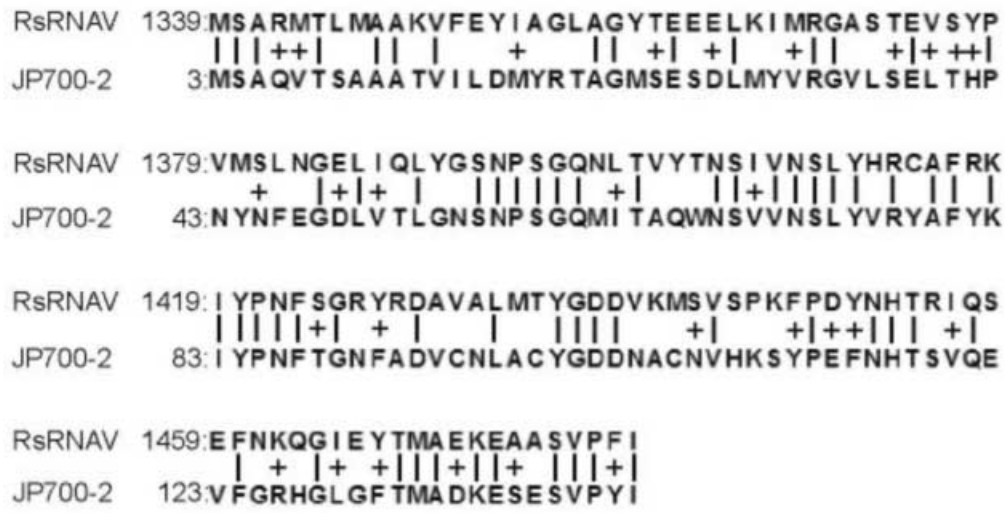

Figure 5. Comparison of the partial amino acid sequence of RNA-dependent RNA polymerase of RsRNAV and an unidentified picorna-like virus sequence JP700-2 (Culley et al., 2003). Numbers correspond to the amino acid position. |, identical; +, positive.

Multiple alignment

A fragmental amino acid sequence of RsRNAV RdRp was automatically aligned with an environmental virus sequence JP700-2 (NCBI Accession number, NP_620560) reported by Culley et al. (2003) by using ClustalW.

RESULTS AND DISCUSSION

Features of the RsRNAV genome

The RsRNAV genome was shown to have a 3′ poly(A) tail because the RsRNAV genome was successfully reverse-transcribed using oligo-dT primer (data not shown). By integrating the results of cDNA analysis, 5′-RACE and 3′-RACE, the entire RsRNAV genome was fully sequenced and estimated to be 8877 nt excluding the poly(A) tail (Figure 1). The RsRNAV genome did not react with the monoclonal antibody against 2,2,7-trimethyl guanosine containing cap structure (Figure 2); therefore, we conclude the RsRNAV genome is uncapped at its 5′-termini. Here we found the viral genome was smaller than the previously estimated size (11.2 kb) using denaturing gel electrophoresis (Nagasaki et al., 2004). This may reflect the additional length of the 3′ poly(A) tail or the addition of a viral genome-linked protein (VPg); however, we do not entirely understand this. The base composition is A, 31.2%; U, 32.5%; G, 19.5%; C, 16.7%; and Y (U or C; among 12 sequencing trials, 6 and 6 respectively resulted in U and C for nt 7409), 0.01%; hence, the AU ratio is high at 63.7%. The AU-richness of the RsRNAV genome is comparable to those of dicistroviruses and the other insect-infecting picorna-like viruses, e.g. deformed wing virus, Varroa destructor virus 1, Kakugo virus. Compared to RsRNAV, the AU ratios of HaRNAV, SssRNAV, and HcRNAV are much lower at 53.1, 50.2, and 44.9%, respectively. A computer-assisted analysis shows the genome contains two large ORFs coded in different reading frames: ORF-1 (nt 625–5442) 4818 nt and ORF-2 (nt 5766–8648) 2883 nt (Figure 1).

Non-structural proteins

The best-fit BLAST comparison for whole amino acid sequence of RsRNAV ORF-1 was to the polyprotein of HaRNAV, particularly to the region coding for the RdRp and RNA helicase (Figure 1) (e-value=5e-32); hence, ORF-1 is considered to be a polyprotein gene coding for enzymes essential for viral replication. Considerable BLAST similarities were also found to a protist-infecting virus SssRNAV (1e-23), an unidentified picorna-like virus RdRp sequence amplified from marine natural viral communities (1e-28) (JP700-2 in Group D; Culley et al. (2003)), and those of plant viruses (e.g. strawberry mottle virus (1e-23), broad bean wilt virus 1 (2e-23)) and insect viruses (e.g. deformed wing virus (1e-19), Kakugo virus (1e-18)). These data suggest there is a significant relationship among the positive-sense ssRNA viruses infecting marine eukaryotic micro-organisms as well as among land and marine positive-sense ssRNA viruses.

The first conserved motifs of the RNA helicase, GXXGXGKS (motif A), was found at amino acid sequence 369–376 assuming that the first (5′-proximal) AUG (nt 625–627) is the initiation codon; the second conserved motif, QX_5DD (motif B), was IX_5DD in the amino acid sequence 423–430; and the third domain, KGX_4SX_5STN (motif C), was found at 466–481 (Figure 3A). The conserved domain of RdRp, KDE (motif I), SGX_3TX_3N (motif V), YGDD (motif VI), and FLKR (motif VII), were found in the sequence of amino acids 1258–1260, 1397–1407, 1438–1441, and 1489–1492, respectively (Figure 3B). Although no significant BLAST similarity for a viral protease domain was detected in RsRNAV ORF-1, an amino acid sequence GMCM resembles the cysteine protease-like motif sequence (GXCG) (Koonin & Dolja, 1993). This suggests the polyprotein is translated and then enzymatically (proteolytically) cleaved into smaller functional proteins including RNA helicase and RdRp. This is the case with other positive-sense ssRNA viruses such as poliovirus. This also shows amino acid sequences of positive-sense ssRNA viruses'

https://doi.org/10.1017/S0025315406013373 Published online by Cambridge University Press

protease domain may be more diverse than was generally suspected: e.g. those of HaRNAV and HcRNAV are GTSG and GPHCG, respectively. Further analysis of undiscovered marine ssRNA viruses' sequences will be required to estimate the diversity.

The deduced amino acid sequence of the RNA helicase domain (amino acids 359–492) and the RdRp domain (amino acids 1177–1597) were analysed using the BLAST algorithm. Although the amino acid sequence of the three RNA helicase motifs (A–C) of RsRNAV were similar to those of the other positive-sense ssRNA viruses (Figure 3A), the helicases showed relatively low similarities to polyproteins of marine ssRNA viruses (HaRNAV (6e-5), SssRNAV (5e-4)) and those infectious to insects or plants (Table 1A). In contrast, the RdRp amino acid sequence showed the highest similarity (1e-28) to an unidentified picorna-like virus sequence, JP700-2, that was categorized into Group D by Culley et al. (2003). It also showed similarity to marine positive-sense ssRNA viruses (HaRNAV (2e-20), SssRNAV (3e-20)), unidentified picorna-like virus sequences (Group B (4e-22∼3e-20), Group C (5e-20∼9e-19); Culley et al. (2003); see below), and those infectious to plants or insects (Table 1B).

There was also a BLAST similarity in the RdRp region (2e-20; amino acid 1269–1730) to an unnamed protein product of *Homo sapiens* (Table 1B). Considering the conserved domains showed an agreement (only one amino acid substitution was detected in motifs I, V, VI, and VII) and the possible function of a previously unidentified RNA polymerase is suggested by Lai (2005), this result suggests the function of the product might be related to RNA-dependent RNA polymerization. It may be that the product functions in the process of RNA interference (Jørgensen, 2003).

Structural proteins

The deduced amino acid sequence of RsRNAV ORF-2 showed a significant similarity to structural proteins of positive-sense ssRNA viruses infecting marine eukaryotic micro-organisms, HaRNAV (2e-86) and SssRNAV (4e-41), and also to those of insect viruses belonging to the family Dicistroviridae, e.g. *Rhopalosiphum padi* virus (5e-33), aphid lethal paralysis virus (8e-29) and cricket paralysis virus (1e-20). Hence, ORF-2 is considered to be a gene coding for a polyprotein for the structural proteins.

Nagasaki et al. (2004) reported that RsRNAV has three major structural proteins (41.5, 41.0 and 29.5 kDa) and two minor proteins (155 and 69 kDa) by sodium dodecylsulphate polyacrylamide gel electrophoresis. However, considering the maximum molecular mass of a protein that can be encoded by ORF-2 is 107 kDa (data not shown), the 155 kDa band is assumed to be an incompletely monomerized protein (e.g. dimer or trimer) or reflecting the contamination of a host protein; the possibility of protein glycosilation is unknown. Then, we designated the other viral proteins as VP1 (69 kDa), VP2 (41.5 kDa), VP3 (41.0 kDa) and VP4 (29.5 kDa). The N-terminal sequence of VP2 (41.5 kDa) was found to be CRPLNLD by the Edman degradation technique. This coincides with the deduced amino acid sequence from ORF-2 (corresponding to nt 7005–7025 in the genome) (Figure 3C). This amino acid sequence is considerably similar to the N-terminal sequence of major capsid proteins of HaRNAV, SssRNAV and three dicistroviruses (Figure 3C). These results support the speculation that ORF-2 encodes a polyprotein comprising proteins involved in the encapsidation of the virus. The N-terminal sequence of the other capsid proteins (VP1, 3 and 4) could not be determined by Edman degradation presumably due to an N-terminal block. Further analysis using mass spectrometry protein sequencing will be required to determine the translation of the structural proteins encoded by ORF-2.

Phylogenetic analyses

In the present study, both the NJ method and ML method were used to assess the phylogenetic relationship among positive-sense ssRNA viruses. Consequently, similar results were obtained by these two methods; thus, only ML phylogenetic trees are shown in Figure 4. In the ML phylogenetic tree constructed based on the concatenated amino acid sequences of RNA helicase and RdRp domains, the monophyly of RsRNAV, HaRNAV and SssRNAV was supported with a moderate bootstrap value of 83% (Figure 4A). In the tree, these three viruses clustered distantly from the cluster of dicistroviruses (including a marine crustacean-infecting virus, TSV), the monophyly of which is supported with a high bootstrap value (100%). While, in the phylogenetic tree based on the RdRp whole domain, monophyly of RsRNAV, HaRNAV and SssRNAV was only very weakly supported (50%) (Figure 4B); in addition, their order of branching differed between the two ML phylogenetic trees (Figure 4A,B). Hence, the phylogenetic relationship among these three viruses still remains obscure. Adding sequences of undiscovered positive-sense ssRNA viruses may have a significant effect on the branching order and/or phylogenetic lineage. At least, it is of great interest that these three positive-sense ssRNA viruses infecting marine micro-organisms form a cluster separate from the other virus families in both the NJ and ML trees.

Because no putative RNA helicase-coding region is found in the genome of the dinoflagellate-infecting virus HcRNAV, the RdRp sequence tree (Figure 4B) was used to look at the relationships among RsRNAV, HaRNAV, SssRNAV and HcRNAV. The tree demonstrates HcRNAV is isolated from the three other viruses and deeply branched from the cluster composed of the families Luteoviridae, Barnaviridae, and Tetraviridae, their monophyly being supported with a moderate bootstrap value (80%); by the NJ method, the corresponding bootstrap value was estimated at 97% (Figure 4B). Comparing the host organism each virus infects, the three viruses (RsRNAV, HaRNAV, and SssRNAV) have a host belonging to Stramenopiles (Cavalier-Smith et al., 1989; Patterson et al., 1998) and form a phylogenetic cluster as mentioned above (Figure 4); whereas, the host for HcRNAV, *Heterocapsa circularisquama*, belongs to Alveolata (Cavalier-Smith et al., 1991; Patterson et al., 1998). In order to further see the evolutionary relationships among them, more positive-sense ssRNA viruses infecting marine micro-organisms need to be isolated, genetically analysed, and phylogenetically compared. This will be of value to understand their coevolution, i.e. an evolutionary history of the host/virus systems.

https://doi.org/10.1017/S0025315406013373 Published online by Cambridge University Press

Comparison with environmental virus sequences

Culley et al. (2003) designed a RT-PCR method for amplifying the picorna-like virus RdRp partial sequence and isolated a number of environmental virus sequences, which were categorized into Groups A–D. Among them, the first-highest similarity of the RdRp corresponding amino acid sequence of RsRNAV was to that of Group D (JP700-2; see above), where the rate of identities (identical residues in the alignment) and positives (conservative substitutions in the alignment) were 45% (65/142) and 65% (93/142), respectively (Figure 5). The second-highest similarity was to Group B, the third to Group C, and a lower similarity was found to Group A including HaRNAV (Table 1B; see above), suggesting that the RT-PCR method also enabled amplification of a picorna-like virus sequence that can be related to the diatom-infecting virus RsRNAV. A great number of unknown ssRNA viruses are most likely hidden in the branch comprising the Stramenopiles-infecting viruses (RsRNAV, HaRNAV, and SssRNAV) and the unidentified viruses that are resources of the environmental RdRp sequences reported by Culley et al. (2003).

Besides, redesign of the degenerate primers referring to the sequences of RsRNAV (this study) and SssRNAV (Takao et al., 2006) may improve the RT-PCR method refining the target virus range.

Conclusions

Based on the results mentioned above, RsRNAV is considered significantly related to the marine raphidophyte-infecting virus HaRNAV (the family Marnaviridae). However, they are different in genome structure; i.e. RsRNAV and HaRNAV respectively have two and one ORFs, as well as in AU-richness (63.7% and 53.1%, respectively). The RsRNAV ORFs showed a considerable similarity also to dicistroviruses (Table 1), and they have a common feature in genome structure, i.e. AU-rich positive-sense ssRNA genome having two ORFs; however, the results of phylogenetic analyses showed RsRNAV is not a member of the family Dicistroviridae the monophly of which was supported with a high ML bootstrap value of 98% (Figure 4A). Hence, we conclude RsRNAV is not a member of any currently-defined virus family.

The genome of picorna-like viruses is a linear positive-sense ssRNA with a protein (VPg) attached to the 5′-end; and is translated into a polyprotein before processing (generally with no overlapping ORFs). Some replicases (e.g. RdRp, RNA helicase) are essential in the replication process. Hence, the essential protein sequences and their nucleotide sequences, arrangement of domains (or genes) coding for them, and the number of ORFs are outstanding points to be compared in considering the relationships among picorna-like viruses as well as virus structures, hosts and epidemiology. HaRNAV is the first ssRNA virus infecting microalgae where the genome is fully sequenced; based on the genome structural features and sequence similarities to other viruses, Lang et al. (2004) proposed a new virus family Marnaviridae, where HaRNAV is the first member. Following this, the genome of the ssRNA viruses infecting microalgae (HcRNAV and RsRNAV) and a fungoid protist (SssRNAV) were sequenced and phylogenetically compared with HaRNAV and the other viruses (Nagasaki et al., 2005a; Takao et al., 2006; this study). These studies show marine positive-sense ssRNA viruses are highly diverse. In the phylogenetic tree based on amino acid sequences of RdRp (Figure 4B), viruses infectious to organisms belonging to Stramenopiles (RsRNAV, HaRNAV, and SssRNAV) did form a monophyly although it was not at the family level. While, HcRNAV (Alveolata-infecting ssRNA virus) showed only a very weak relatedness to the three Stramenopiles-infecting viruses (Figure 4B). In addition, one of the two ORFs in HcRNAV genome codes for the capsid protein and is not a polyprotein gene (Nagasaki et al., 2005a); this is distinctive from picorna-like viruses. Therefore, we conclude the positive-sense ssRNA viruses infecting marine microalgae or eukaryotic micro-organisms cannot be categorized into a single family. Further investigations concerning the classification and nomenclature for these viruses and their relatives will be required. Undoubtedly, RdRp is a key molecule to consider the genetic relationships among positive-sense ssRNA viruses (Koonin & Dolja, 1993; Zanotto et al., 1996) as DNA polymerase is in the field of studies on dsDNA viruses including those infectious to algae (Wilson et al., 2005).

Our study supports the hypothesis given by Culley et al. (2003) that a diverse array of ssRNA viruses exists in the marine environment. More analytical studies concerning positive-sense ssRNA viruses infecting eukaryotic marine microalgae and protists are needed. We believe this will allow us to understand the coevolutionary history of hosts and viruses as well as their diversity, evolutionary ecology, and the significance of the ecology of the marine viruses.

This work was supported by funding from Grants-in-Aid for Scientific Research (A) (2) (No. 16208019) from the Ministry of Education, Science and Culture in Japan, the Fisheries Agency of Japan, the Ministry of Agriculture, Forestry and Fisheries of Japan. The authors are grateful to Drs T. Okuno and K. Mise (Kyoto University, Japan) for their helpful advice. Thanks are also due to Dr S. Itakura (Fisheries Research Agency, Japan) for providing a diatom culture and to Dr K. Nishida (Hiroshima Prefectural Institute of Industrial Science and Technology, Japan) for technical assistance.

REFERENCES

Bettarel, Y. et al., 2005. Isolation and preliminary characterisation of a small nuclear inclusion virus infecting the diatom *Chaetoceros* cf. *gracilis*. *Aquatic Microbial Ecology*, **40**, 103–114.

Brussaard, C.P., Noordeloos, A.A., Sandaa, R.A., Heldal, M. & Bratbak, G., 2004. Discovery of a dsRNA virus infecting the marine photosynthetic protist *Micromonas pusilla*. *Virology*, **319**, 280–291.

Cavalier-Smith, T., 1991, Cell diversification in heterotrophic flagellates. In *The biology of free-living heterotrophic flagellates* (ed. D. J. Patterson and J. Larsen), pp. 113–131. Oxford: Clarendon Press.

Cavalier-Smith, T., 1998. A revised six-kingdom system of life. *Biological Reviews*, **73**, 203–266.

Culley, A.I., Lang, A.S. & Suttle, C.A., 2003. High diversity of unknown picorna-like viruses in the sea. *Nature, London*, **424**, 1054–1057.

Jørgensen, R.A., 2003. Sense cosuppression in plants: past, present, and future. In *RNAi: a guide to gene silencing* (ed. G.J. Hannon), pp. 5–22. New York: Cold Spring Harbor Laboratory Press.

Koonin, E.V. & Dolja, V.V., 1993. Evolution and taxonomy of positive-strand RNA viruses: implication of comparative analysis of amino acid sequences. *Critical Reviews in Biochemistry and Molecular Biology*, **28**, 375–430.

Lai, M.M.C., 2005. RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. *Journal of Virology*, **79**, 7951–7958.

Lang, A.S., Culley, A.I. & Suttle, C.A., 2004. Genome sequence and characterization of a virus (HaRNAV) related to picorna-like viruses that infects the marine toxic bloom-forming alga *Heterosigma akashiwo*. *Virology*, **320**, 206–217.

Mizumoto, H., Tatsuta, M., Kaido, M., Mise, K. & Okuno, T., 2003. Cap-independent translational enhancement by the 3′ untranslated region of red clover necrotic mosaic virus RNA1. *Journal of Virology*, **77**, 12113–12121.

Nagasaki, K., Tomaru, Y., Katanozaka, N., Shirai, Y., Nishida, K., Itakura, S. & Yamaguchi, M., 2004. Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Applied and Environmental Microbiology*, **70**, 704–711.

Nagasaki, K., Shirai, Y., Takao, Y., Mizumoto, H., Nishida, K. & Tomaru, Y., 2005a. Comparison of genome sequences of single-stranded RNA viruses infecting the bivalve-killing dinoflagellate *Heterocapsa circularisquama*. *Applied and Environmental Microbiology*, **71**, 8888–8894.

Nagasaki, K., Tomaru, Y., Takao, Y., Nishida, K., Shirai, Y., Suzuki, H. & Nagumo, T., 2005b. Previously unknown virus infects marine diatom. *Applied and Environmental Microbiology*, **71**, 3528–3535.

Patterson, D.J., 1989, Stramenopiles: chromophytes from a protistan perspective. In *The chromophyte algae problems and perspectives* (ed. J.C. Green et al.), pp. 357–379. Oxford: Clarendon Press.

Tai, V., Lawrence, J.E., Lang, A.S., Chan, A.M., Culley, A.I. & Suttle, C.A., 2003. Characterization of HaRNAV, a single-stranded RNA virus causing lysis of *Heterosigma akashiwo* (Raphidophyceae). *Journal of Phycology*, **39**, 343–352.

Takao, Y., Mise, K., Nagasaki, K., Okuno, T. & Honda, D., 2006. Complete nucleotide sequence and genome organization of a single-stranded RNA virus (SssRNAV) infecting the marine fungoid *Schizochytrium* sp. *Journal of General Virology*, **87**, 723–733.

Takao, Y., Nagasaki, K., Mise, K., Okuno, T. & Honda, D., 2005. Isolation and characterization of a novel single-stranded RNA virus infectious to a marine fungoid protist, *Schizochytrium* sp. (Thraustochytriaceae, Labyrinthulea). *Applied and Environmental Microbiology*, **71**, 4516–4522.

Tomaru, Y., Katanozaka, N., Nishida, K., Shirai, Y., Tarutani, K., Yamaguchi, M. & Nagasaki, K., 2004. Isolation and characterization of two distinct types of HcRNAV, a single-stranded RNA virus infecting the bivalve-killing microalga *Heterocapsa circularisquama*. *Aquatic Microbial Ecology*, **34**, 207–218.

Wilson, W.H., Van Etten, J.L., Schroeder, D.S., Nagasaki, K., Brussaard, C., Delaroque, N., Bratbak, G. & Suttle, C., 2005. Phycodnaviridae. In *Virus taxonomy, VIIIth Report of the ICTV* (ed. C.M. Fauquet et al.), pp. 163–175. Chian: Elsevier Academic Press.

Zanotto, P., Gibbs, M., Gould, E. & Holmes, E., 1996. A reevaluation of the higher taxonomy of viruses based on RNA polymerases. *Journal of Virology*, **70**, 6083–6096.

Submitted 16 December 2005. Accepted 17 March 2006.