



## OPEN

New single-stranded DNA virus with a unique genomic structure that infects marine diatom *Chaetoceros setoensis*Yuji Tomaru<sup>1</sup>, Kensuke Toyoda<sup>2</sup>, Hidekazu Suzuki<sup>3</sup>, Tamotsu Nagumo<sup>4</sup>, Kei Kimura<sup>4,5</sup> & Yoshitake Takao<sup>6</sup>

## SUBJECT AREAS:

MICROBIOLOGY

MICROBIAL ECOLOGY

VIRAL PATHOGENESIS

WATER MICROBIOLOGY

Received  
9 May 2013Accepted  
11 November 2013Published  
26 November 2013Correspondence and  
requests for materials  
should be addressed to  
Y.T. (tomaruy@affrc.  
go.jp)

<sup>1</sup>National Research Institute of Fisheries and Environment of Inland Sea, Fisheries Research Agency, 2-17-5 Maruishi, Hatsukaichi, Hiroshima 739-0452, Japan, <sup>2</sup>Research and Education Center for Natural Sciences, Keio University, 4-1-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8521, Japan, <sup>3</sup>Graduate School of Marine Science and Technology, Tokyo University of Marine Science and Technology, 4-5-7 Konan, Minato-ku, Tokyo 108-8477, Japan, <sup>4</sup>Department of Biology, Nippon Dental University, 1-9-20 Fujimi, Chiyoda-ku, Tokyo 102-9159, Japan, <sup>5</sup>Research Fellow of the Japan Society for the Promotion of Science, <sup>6</sup>Department of Marine Bioscience, Fukui Prefectural University, 1-1 Gakuen-cho, Obama, Fukui 917-0003, Japan.

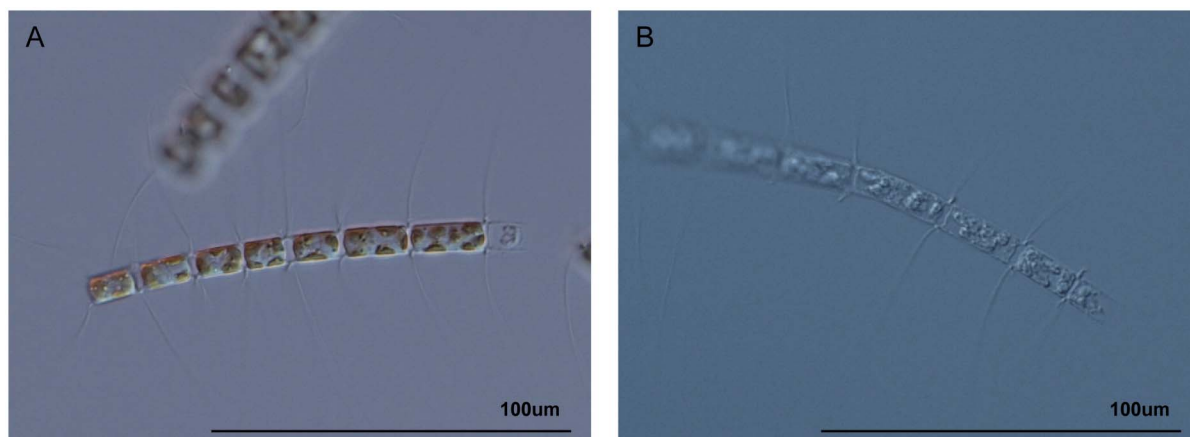
Diatoms are among the most abundant organisms in nature; however, their relationships with single-stranded DNA (ssDNA) viruses have not yet been defined in detail. We report the isolation and characterisation of a virus (CsetDNAV) that lytically infects the bloom-forming diatom *Chaetoceros setoensis*. The virion is 33 nm in diameter and accumulates in the nucleus of its host. CsetDNAV harbours a covalently closed-circular ssDNA genome comprising 5836 nucleotides and eight different short-complementary fragments (67–145 nucleotides), which have not been reported in other diatom viruses. Phylogenetic analysis based on the putative replicase-related protein showed that CsetDNAV was not included in the monophyly of the recently established genus *Bacilladnavirus*. This discovery of CsetDNAV, which harbours a genome with a structure that is unique among known viruses that infect diatoms, suggests that other such undiscovered viruses possess diverse genomic architectures.

Since the late 1980s, the study of marine viruses has revealed an unexpected abundance of virus particles, ranging from  $10^6$  to  $10^9$  particles per millilitre of seawater. Furthermore, these viruses have been shown to play an important role in marine ecosystems and biogeochemical cycles<sup>1</sup>. More recently, metagenomic studies of marine viruses<sup>1–3</sup> have revealed large diversities of virus-like sequences. This information has made a significant contribution to our knowledge of single-stranded DNA (ssDNA) viruses<sup>4,5</sup>. Furthermore, these studies have revealed that these ssDNA viruses are widely distributed across diverse aquatic environments; however, their host organisms remain unknown.

Diatoms are unicellular, photosynthetic, eukaryotic algae that are found worldwide in oceans and bodies of freshwater. Diatoms (Bacillariophyta) account for a large part of the marine biomass, comprising as much as 35% and 75% of the biomass of oligotrophic oceans and nutrient-rich systems, respectively<sup>6</sup>. Diatoms form the foundation of short, energy-efficient food webs that support large-scale coastal fisheries. Therefore, understanding the life cycles of diatoms in the sea is of primary importance for marine ecology. Although research indicates that the population dynamics of diatoms are principally affected by physical and chemical factors, recent studies have also revealed the potential importance of viruses, including ssRNA and ssDNA viruses, in these processes<sup>7</sup>.

At least 13 different diatom viruses have been reported to date<sup>8,9</sup>. There are six ssDNA viruses (see Table 2): *Chaetoceros debilis* DNA virus (CdebDNAV)<sup>10</sup>, *Chaetoceros lorenzianus* DNA virus (ClorDNAV)<sup>11</sup>, *Chaetoceros salsugineum* DNA virus (CsalDNAV)<sup>12</sup>, *Chaetoceros* sp. strain TG07-C28 DNA virus (Csp05DNAV)<sup>13</sup>, *Chaetoceros tenuissimus* DNA virus (CtenDNAV)<sup>14</sup>, and *Thalassionema nitzschioides* DNA virus (TnitDNAV)<sup>8</sup>. Their virions, which accumulate in the host nucleus during virus replication, are 32–38 nm in diameter. The genomes of these viruses, except for CdebDNAV, comprise a covalently closed circular ssDNA and a segment of linear ssDNA (0.6–1 kb)<sup>9</sup>. Although CdebDNAV also harbours a circular ssDNA genome, linear fragments have not been identified<sup>10</sup>.

Here, we report the discovery, isolation, and characterisation of an ssDNA diatom virus, CsetDNAV, which infects the bloom-forming diatom *Chaetoceros setoensis* Ikari in the Seto Inland Sea of Japan. We also conducted phylogenetic studies of several ssDNA diatom viruses, including viruses that infect both centric and pennate species.



**Figure 1** | *Chaetoceros setoensis*. (A) Micrograph of intact cells. (B) Micrograph of CsetDNAV-infected cells at 72 hpi. Bars indicate 100  $\mu\text{m}$ .

## Results

**Virus isolation.** The isolated virus retained its lytic activity after filtration through a 0.2  $\mu\text{m}$ -pore-size polycarbonate membrane filter (Whatman). The lytic activity was serially transferable to exponentially growing *C. setoensis* cultures. The cytoplasm and photosynthetic pigments in the virus-infected cells were degraded compared to those of uninfected cells (Fig. 1).

**Host range.** The host range of CsetDNAV was tested using 28 phytoplankton strains, including 14 diatom strains. CsetDNAV lysed its host strain, *C. setoensis* IT07-C11, but not any of the other microalgal strains tested (Table 1).

**Morphological analysis of virus-infected cells and virions.** Thin sections of uninfected *C. setoensis* showed cytoplasmic organisation and frustules that were diagnostic of diatoms (Fig. 2A). In

contrast, *C. setoensis* examined 48 h post-infection (hpi) showed the presence of randomly assembled virus-like particles (VLPs) in the nucleus, which were  $32 \pm 3$  nm ( $N = 30$ ) in diameter (Figs. 2C, D and E). VLPs were not detected in the healthy control cultures (Fig. 2B). The VLPs observed in the culture lysates using negative staining were hexagonal in outline, suggesting icosahedral symmetry, were  $33 \pm 2$  nm ( $N = 32$ ) in diameter, lacked a tail and outer membrane, and appeared similar in size to those of the VLPs observed in the host nucleus (Fig. 2F).

Besides the random assembly of VLPs within the host cell, other distinctive morphological features potentially associated with infection were observed in the host nucleus. We observed fibrous structures that were  $17 \pm 2$  nm wide ( $N = 30$ ) and 0.20–0.65  $\mu\text{m}$  long (Fig. 2E), as well as aggregates of electron-dense particles  $17 \pm 1$  nm in diameter ( $N = 30$ ) (Fig. 2E). Considering their similarity in size, the electron-dense particles might represent cross-sections of the

**Table 1** | Infection specificities of CsetDNAV against 28 strains of marine phytoplankton

Famiy	Species	Strain code	Temperature ( $^{\circ}\text{C}$ )	Strains lysed by CsetDNAV
Bacillariophyta	<i>Chaetoceros cf. affinis</i>	Chae5	15	–
	<i>Chaetoceros debilis</i>	Ch48	15	–
	<i>Chaetoceros lorenzianus</i>	ItDia-51	15	–
	<i>Chaetoceros cf. pseudocurvisetus</i>	IT07-C37	15	–
	<i>Chaetoceros salsugineum</i>	Ch42	15	–
	<i>Chaetoceros setoensis</i>	IT07-C11	15	+
	<i>Chaetoceros socialis f. radians</i>	L-4	15	–
	<i>Chaetoceros tenuissimus</i>	020710MA01	15	–
	<i>Detonula pumila</i>	ItDia-14	15	–
	<i>Ditylum brightwellii</i>	ItDia-21	15	–
	<i>Eucampia zodiacus</i>	EzB	15	–
	<i>Rhizosolenia setigera</i>	S2	15	–
	<i>Skeletonema</i> sp.	ItDia-SK4	15	–
	<i>Stephanopyxis</i> sp.	ItDia-Stp8	15	–
Eustigmatophyceae	<i>Nannochloropsis</i> sp.	SFBB	20	–
Cryptophyceae	<i>Teleaulax amphioxeia</i>	Hama3	20	–
Dinophyceae	<i>Alexandrium catenella</i>	ACNG	20	–
	<i>Gymnodinium catenatum</i>	GC27-1	20	–
	<i>Heterocapsa circularisquama</i>	HU9433-P	20	–
	<i>Heterocapsa triquetra</i>	Ht	20	–
	<i>Kalenia mikimotoi</i>	GmlH	20	–
	<i>Prorocentrum micans</i>	M	20	–
	<i>Scrippsiella</i> sp.	SCKR	20	–
	<i>Chattonella antiqua</i>	CaAR	20	–
	<i>Chattonella marina</i>	CMKG-1	20	–
	<i>Chattonella ovata</i>	CoV	20	–
Raphidophyceae	<i>Fibrocapsa japonica</i>	F96	20	–
	<i>Heterosigma akashiwo</i>	H93616	20	–

–, not lysed; +, lysed.



Table 2 | Basic characters for single-stranded DNA diatom viruses ever isolated

Virus	Host	Particle		Rod shape particles	Major proteins (kDa)	Latent period	Burst size (infectious units per cell)	Genome structure	Genome length (nt)	Complementary fragments (nt)	Data	Reference
		size (nm)	assemblage site and aggregation patterns									
CdebDNAV	<i>Chaetoceros debilis</i>	32	nucleus, random	nd	37.5, 41	12–24	55	nd	ca. 7 kb, not fully sequenced	nd	AB504376	10
ClorDNAV	<i>Chaetoceros lorenzianus</i>	34	nucleus, random, ring formation	Yes	225	<48	$2.2 \times 10^4$	covalently closed circular	5,813	979	AB553581	11
CsalDNAV	<i>Chaetoceros salsugineum</i>	38	nucleus, random	nd	43.5, 46	12–24	325	covalently closed circular	6,000	997	AB193315	12
CsetDNAV	<i>Chaetoceros setoensis</i>	33	nucleus, random	Yes	33, 39	48	$2.0 \times 10^4$	covalently closed circular	5,836	67, 70, 72, 76, 90, 107, 109, 145	AB781089	This study
CtenDNAV	<i>Chaetoceros tenuissimus</i>	37	nucleus, random	Yes	38.5	96	320	covalently closed circular	5,639	875	AB597949	14
Csp05DNAV	<i>Chaetoceros</i> sp. strain TG07-C28	33	nucleus, random	Yes	40, 75	<24	nd	covalently closed circular	5,785*	890*	AB647334	13
TnitDNAV	<i>Thalassiosira nitzschoides</i>	35	nucleus, para-crystalline array	nd	nd	nd	nd	covalently closed circular	5,573*	ca. 600, not sequenced	AB781284	8

nd, not determined.  
\*unpublished data.

fibrous structures. In contrast, the fibrous structures were not detected in the lysates by the negative staining.

**Thermal stability of virions.** CsetDNAV suspensions containing  $5.1 \times 10^8$  infectious units/mL were stored under different temperature conditions. The titers of the virus suspensions after 82 days in the dark at 20°C, 10°C, 4°C, -20°C, -80°C, and -196°C were 14%, 27%, 37%, 36%, 41%, and 184%, respectively.

**Analysis of the viral genome.** Agarose gel electrophoresis of viral DNA revealed the presence of two major bands at approximately 3.8 kb and 3.2 kb (Fig. 3, lane 1); neither of which were sensitive to heat treatment at 100°C for 2 min (Fig. 3, lane 5). Both of the bands were sensitive to S-1 nuclease and DNase I, but not to RNase A (Fig. 3, lanes 2–4), indicating an ssDNA genome. Polymerase chain reaction (PCR) experiments using primer sets designed to confirm the structure of the CsetDNAV genome resulted in the amplification of bands with the expected sizes (see Methods), which demonstrated that the genome is a closed-circular form (Fig. 4). Full sequencing of the CsetDNAV genome revealed that the circular ssDNA comprises 5836 nucleotides (Fig. 5), which was deposited in the DNA Data Bank of Japan (DDBJ; accession number, AB781089).

We also detected a very small amount of S1 nuclease resistant fragments at ~100 bp after S1 nuclease treatment, which were difficult to visualise using normal brightness and contrast levels (Supplementary Fig. S1), suggesting that the genome contains short complementary ssDNA or ssRNA fragments in addition to the covalently closed circular ssDNA. The types of nucleic acid, DNA or RNA, of these complementary fragments were not determined in the present study. The S1 nuclease-resistant fragments were successfully sequenced. Their lengths varied (67, 70, 72, 76, 90, 107, 109, and 145 bp (AB781089)), and their sequences were identical to different regions of the closed-circular genome (Fig. 5).

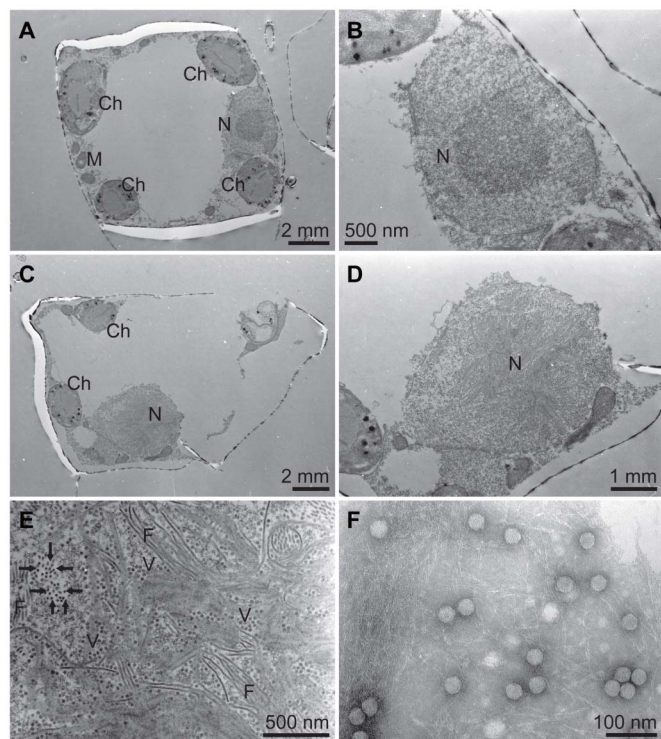
Three major open reading frames (ORFs) comprising 385 (VP1), 396 (VP2), and 489 (VP3) amino acid residues were identified. The sequence of the largest ORF was found to be highly similar to that of putative replication-associated proteins of some ssDNA diatom viruses, including CdebDNAV (E-value:  $4e^{-63}$ ; % identities/positives, 47/59), ClorDNAV ( $2e^{-93}$ ; 39/54), CsalDNAV ( $5e^{-98}$ ; 42/56), Csp05DNAV ( $5e^{-108}$ ; 41/57), and CtenDNAV ( $3e^{-95}$ ; 37/57), and was less similar to the replication protein of other ssDNA viruses, including the rodent stool-associated circular genome virus ( $2e^{-04}$ ; 24/39)<sup>15</sup> and the beak and feather disease virus ( $1e^{-03}$ ; 30/42)<sup>16</sup>. The VP2 was found to be similar to putative structural proteins of ClorDNAV ( $7e^{-31}$ ; 29/46), CsalDNAV ( $1e^{-35}$ ; 29/46) and CtenDNAV ( $4e^{-44}$ ; 30/48). The VP1 showed relatively lower similarities to predicted proteins of ClorDNAV ( $1e^{-17}$ ; 30/48), CsalDNAV ( $8e^{-07}$ ; 34/47), and CtenDNAV ( $4e^{-13}$ ; 35/48).

The phylogenetic tree of the ssDNA diatom viruses based on the putative replication-associated protein revealed that the monophyly of ClorDNAV, CsalDNAV, CtenDNAV, and Csp05DNAV, which are all assumed to be members of the genus *Bacilladnavirus*, was supported by 100% and 98% bootstrap values in the neighbour-joining (NJ) and maximum likelihood (ML) analyses, respectively (Fig. 6). The genome of CsetDNAV also forms a closed circle; however, it was grouped into another clade that included CdebDNAV and TnitDNAV (Fig. 6 & Supplementary Fig. S2).

**Viral proteins.** Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis revealed two major protein bands migrating at positions corresponding to 39.0 kDa and 33.0 kDa (Fig. 7). These results are similar to those reported for other diatom DNA viruses, CsalDNAV, CdebDNAV, and CtenDNAV (Table 2).

**Virus replication.** The viral infectious units increased as of two days post-infection (dpi), and a decrease in host cell number was observed after three dpi (Fig. 8). In this study, the burst size was estimated to be





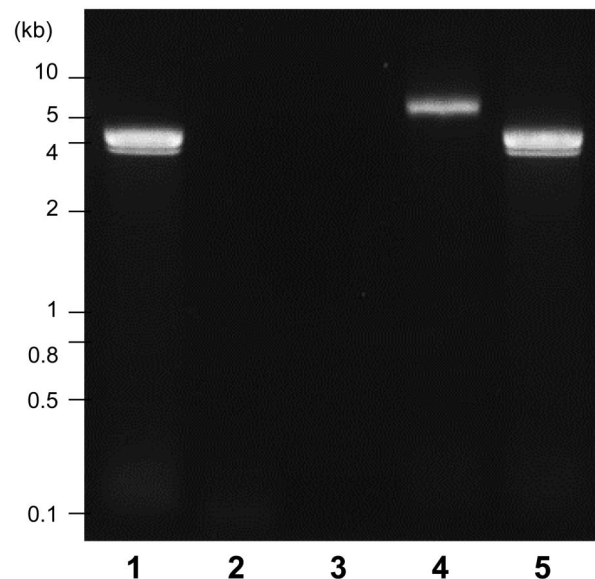
**Figure 2** | TEM micrographs of ultra-thin sections of *C. setoensis* and negatively stained CsetDNAV particles. (A) Uninfected cell. (B) Nucleus of the uninfected cell in panel A. (C) CsetDNAV infected cell at 48 hpi. (D) Nucleus in panel C. (E) Higher magnification of the infected host cell nucleus in panel D. (F) Negatively stained CsetDNAV particles in the culture lysate. Ch: chloroplast; M: mitochondrion; N: nucleus; F: fibrous structure; V: virus-like particles. Arrows indicate electron dense particles, which are smaller than the mature virus-like particles.

$2.0 \times 10^4$  infectious units per cell, based on the ratio of cells to viruses of 4–5 dpi (Fig. 8). However, it is important to note that this value was quite variable, and was  $4.7 \times 10^3$  infectious units per cell in a second experiment.

## Discussion

The virion assembly site and particle diameter of the isolated virus were similar to those of known ssDNA diatom viruses<sup>11</sup> (Table 2). The algicidal pathogen induced lysis when used to infect a fresh algal culture, VLPs were observed in the lysed culture, and VLPs were not found in healthy cultures. Therefore, we concluded that the 33 nm particles observed within the infected cells and in the algal lysates represented a pathogen of *C. setoensis*. This new virus was termed *C. setoensis* DNA virus (CsetDNAV) after its host species and genomic structure.

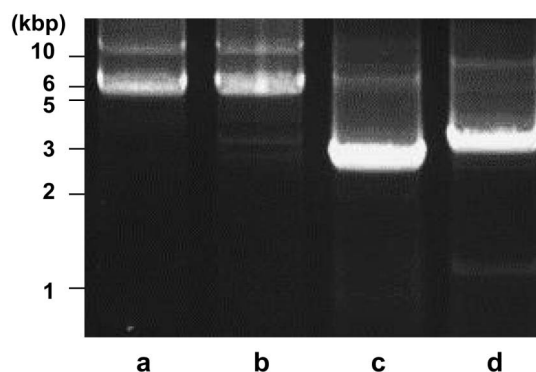
Fibrous or rod-shaped structures in virus-infected host cells have been reported previously. In the case of *Plutella xylostella* granulovirus (PlxyGV) (Baculoviridae), which harbours a dsDNA genome and infects the diamondback moth, an array of rod-shaped particles are apparent in the host cytoplasm during the early steps of the viral maturation process<sup>17</sup>. The rod-shaped particles of PlxGV are nucleocapsids, and ultimately take on an enveloped form. The tobacco mosaic virus (TMV) in the genus *Tobamovirus*, which harbours an ssRNA genome and infects common plants, shows various formations of rod-shaped particles in the infected host cytoplasm, and the mature virions are composed of rod-shaped particles of diverse lengths<sup>18</sup>. In these viruses, however, icosahedral virions are not produced, which differs from the present results. The observations similar to those of this study have been reported in an RNA plant virus, *Solanum nodiflorum* mottle virus, of the genus *Sobemovirus*, which



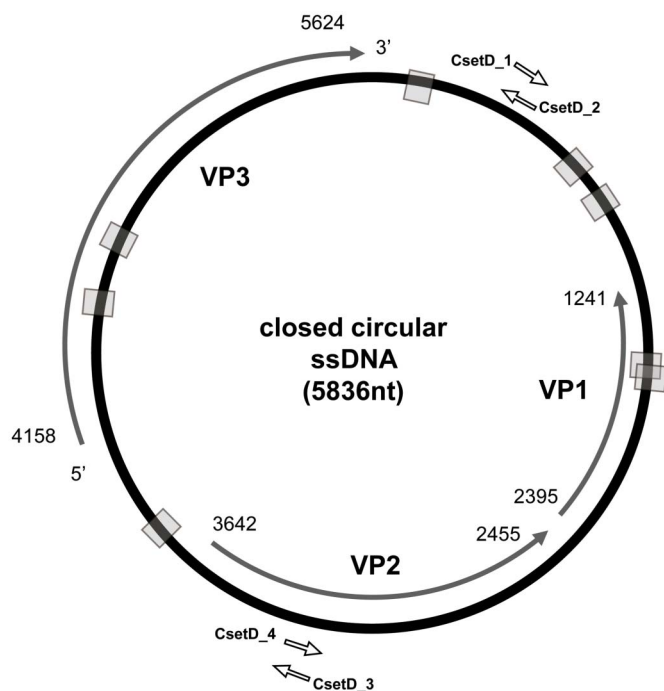
**Figure 3** | Nucleic acids of CsetDNAV untreated (lane 1) or treated with S-1 nuclease (lane 2), DNase I (lane 3), RNase A (lane 4), and heated at 100 °C for 2 min (lane 5).

harbours an ssRNA genome and is 23 nm in diameter<sup>19</sup>. Virions are scattered throughout the nucleus and cytoplasm in the virus-infected leaf cells, sometimes forming crystalline arrays. In addition to these particles, linear or tubular structures are present in the virus-infected cells, but are not found in healthy cells<sup>19</sup>. Similar observations have been reported for known *Chaetoceros* viruses, CwNIV, ClorDNAV, CtenDNAV, and Csp05DNAV<sup>11,13,14,20</sup>. These studies showed accumulations of rod-like VLPs in virus-infected *Chaetoceros* cell nuclei, which were not found in the lysate, and it was suggested that they might represent a morphological maturation stage of the virus. This morphology may be a common feature of this *Chaetoceros* virus group; however, we could not exclude the hypothesis that the fibrous structures are another virus infecting *C. setoensis* that co-infect with CsetDNAV. To better elucidate the role of the fibrous structures, further observations using immunological methods might be necessary.

No significant loss of the infectious titers was observed under any of the temperature tested. Instead, the titers increased when the virus suspensions were preserved at  $-196^{\circ}\text{C}$ . This phenomenon has also been reported in other microalgal viruses<sup>21</sup>. The virus particles of



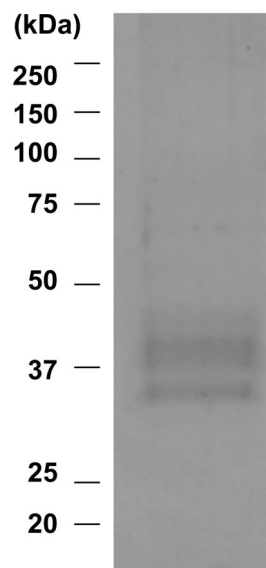
**Figure 4** | PCR products amplified with primer sets CsetD\_1 and CsetD\_2 (a), CsetD\_3 and CsetD\_4 (b), CsetD\_1 and CsetD\_4 (c), and CsetD\_2 and CsetD\_3 (d). The expected sizes of the products in lanes a–d are 5857 bp, 5856 bp, 2724 bp, and 3153 bp, respectively. The most intense bands correspond to target amplicons. The faint bands were likely generated due to overlapping nucleotide sequences.



**Figure 5 | Structure of the CsetDNAV genome.** Open arrows indicate PCR primers (see Fig. 4). Grey arrows indicate the positions of the VP1, VP2, and VP3 ORFs. Grey boxes indicate the positions of the small, complementary ssDNA fragments.

CsetDNAV are aggregated in the host cytoplasm, and a portion of them may form small aggregations in the lysate. The increase in this viral titer following cryopreservation is presumably caused by the diffusion of aggregated viruses. The low sensitivities of algal viruses to temperature are considered to be one of the most important characteristic for their survival in nature, as well as their low light sensitivity<sup>22</sup> and crystalline arrays of virions<sup>23</sup>. The diatom viruses CtenRNAV<sup>24</sup> and CtenDNAV<sup>14</sup>, which both infect *Chaetoceros tenuissimus* and have high thermal stability, are assumed to propagate their populations during a host bloom period, and survive in sediments during other seasons in the year in Hiroshima Bay, Japan, where the water temperature ranges from 7 to 28 °C<sup>7</sup>. The high thermal stability of CsetDNAV might be one of its main advantages for surviving in natural environments.

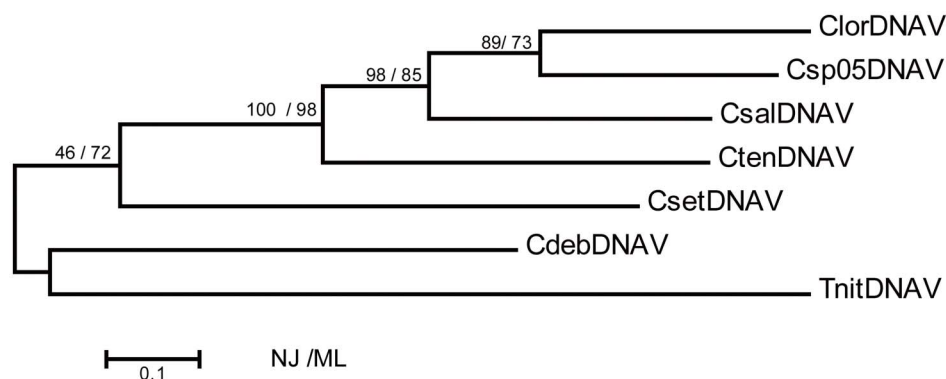
Two major bands were detected in the agarose gel after electrophoresis of viral DNA. The larger and smaller bands are considered



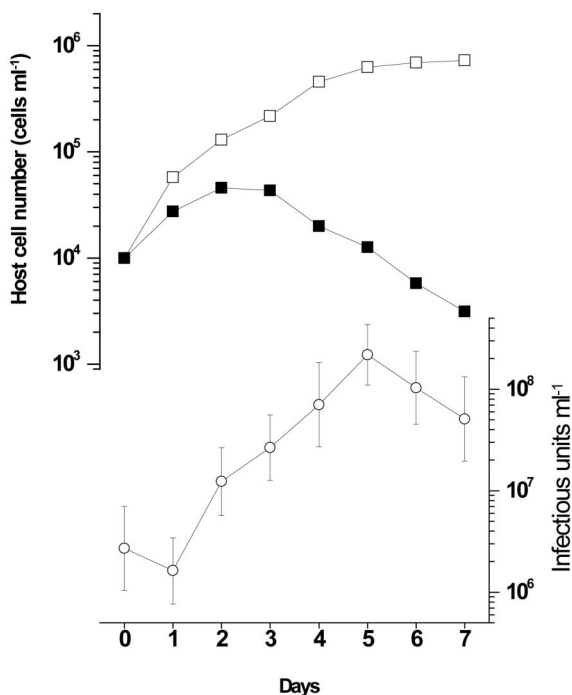
**Figure 7 | Structural proteins of CsetDNAV.**

to be the covalently closed circular form and the linear form of the same molecule, but with a slower electrophoretic mobility, respectively. Similar results were reported for other closed-circular ssDNA virus genomes, e.g., CsalDNAV<sup>12</sup>.

Eight different S1 nuclease resistant regions, i.e., the small linear nucleotide fragments, were recognised in the CsetDNAV genome (AB781089). Considering the resultant sequences, these small fragments, 64–145 nucleotides in length, were not considered to form secondary structures that would not be digested through S-1 nuclease treatments, i.e., the complementary sequences of these fragments were not found in the genome sequence. Previously characterised diatom ssDNA viruses were shown to possess a complementary strand linear DNA, but their sizes were determined to be ~1 kb<sup>11,13,14</sup> (Table 2), which is similar to features of the genomes of mastreviruses, members of Geminiviridae, that infect land plants<sup>25</sup>. The genome of mastreviruses consists of a circular ssDNA and an approximately 80-nucleotide DNA molecule that is complementary to the encapsidated viral strand DNA. The small DNA molecule in the mastrevirus genome is considered to act as a primer at an early stage of viral DNA replication<sup>25</sup>. In the present study, PCR tests using the CsetDNAV genome and DNA polymerase in the absence of added primers also showed that molecules that were larger than the viral genome were amplified (Supplementary Fig. S3). This



**Figure 6 | ML tree based on deduced amino acid sequences of the putative replication protein.** NJ bootstrap values (%) from 1000 samples are shown at the nodes, followed by bootstrap values (%) based on the ML analysis of 1000 samples. The ML distance scale bar is shown. Amino acid sequences and their accession numbers are as follows: *C. debilis* DNA virus (CdebDNAV), AB504376; *C. lorenzianus* DNA virus (ClorDNAV), AB553581; *C. salsgineum* DNA virus (CsalDNAV), AB193315; *C. setoensis* DNA virus (CsetDNAV), AB781089; *C. tenuissimus* DNA virus (CtenDNAV), AB597949; *Chaetoceros* sp. strain TG07-C28 DNA virus (Csp05DNAV), AB647334; *Thalassionema nitzschioides* DNA virus (TnitDNAV), AB781284.



**Figure 8** | Changes in cell numbers of virus-infected (closed squares) or virus-uninfected cultures (open squares) of *C. setoensis* and viral titers (open circles). The error bars indicate 95% confidence intervals.

preliminary experiment, however, did not still fully demonstrate that the small nucleotide molecules act as primers in the virus genome replications. The nature of the small nucleotide fragments and their biological functions should be revealed in future studies.

Considering these findings, we concluded that the viral genome consists of a covalently closed circular ssDNA (5836 nucleotides) and segments of linear single stranded nucleotide (67–145 bp). The linear segment is complementary to a portion of the closed circle, creating a partially double-stranded genome.

The group comprising CsetDNAV, TnitDNAV, and CdebDNAV is considered to be phylogenetically distant from other bacilladnaviruses. These findings suggest that the features of the genomes of diatom viruses might be even more diverse than previously thought. The most significant differences between CsetDNAV and other bacilladnaviruses are in the length and number of the complementary fragments in their genomes. The genome of CdebDNAV is also unlikely to include a long complementary fragment<sup>10</sup>; however, that of TnitDNAV is a closed-circular genome harbouring an approximately 0.6 kb dsDNA region. The sequences of the complementary fragments might have diverged during the evolution of ssDNA diatom viruses.

The latent period of CsetDNAV was estimated to be <48 h based on the growth test. However, the decrease in host cell number and the increase in viral abundance did not simultaneously occur in this host-virus system. Similar results were also observed for previously isolated diatom viruses<sup>10,21,24</sup>. The delay in host cell reduction might have been caused by heterogeneity in the susceptibility of the cells to virus infection.

Metagenomic analyses have revealed the nucleotide sequence diversity of ssDNA viruses. The structures of the genomes of ssDNA viruses and their relationships to host organisms cannot, however, be determined by metagenomics alone. The present study showed that the genome of CsetDNAV harbours small nucleotide fragments that are complementary to the viral ssDNA. The functions of these fragments are unknown, but may play an important role in viral replication processes, and might have diverged during evolution.

Considering the huge biomass and numbers of diatom species in the ocean, they both represent vast ssDNA diatom virus groups and their diversities including their genome architectures.

## Methods

**Algal cultures and growth conditions.** The axenic clonal algal strain used in this study, *C. setoensis* IT07-C11 (Fig. 1A), was isolated from surface water collected at the Itsukaichi Fishing Port (34°21.320'N, 132°21.482'E) in Hiroshima Bay, Japan, on March 16, 2007. This diatom strain was observed using transmission and scanning electron microscopy and was identified as *C. setoensis*<sup>26</sup>. Algal cultures were grown in modified SWM-3 medium enriched with 2 nM Na<sub>2</sub>SeO<sub>3</sub><sup>27</sup> under a 12/12 h light/dark cycle of approximately 110–150 μmol photons/m<sup>2</sup>/s using cool white fluorescent illumination at 15°C.

**Virus isolation.** Sediment samples (0–1 cm) were collected using an Ekman-Birge bottom sampler from the Itsukaichi Fishing Port (ca. 5 m in depth) in Hiroshima Bay on October 15, 2007. Twelve grams of the sediment sample was shaken with 12 mL of the SWM-3 medium (400 rpm, room temperature, for 30 min), and was centrifuged at 860 × g at 4°C for 10 min. The supernatants were passed through 0.2 μm Dismic-25cs filters (Advantec TOYO, Tokyo, Japan). Aliquots (0.2 mL) of the filtrates obtained from the sediment samples were inoculated into exponentially growing *C. setoensis* cultures (0.8 mL) followed by incubation at 15°C using the same lighting conditions described above. Algal cultures inoculated with the SWM3 medium served as controls. Consequently, a *C. setoensis* culture inoculated with the filtrate showed inhibition of algal growth (Fig. 1B). We then cloned the causative pathogen from the lysed culture through two extinction dilution cycles<sup>28,29</sup>. The lysate from the highest dilution well of the second assay was filtered through a 0.1 μm polycarbonate membrane filter (Whatman, Kent, UK) to remove bacteria, and was transferred to a fresh, exponentially growing host culture. The virus preparation was considered as clonal stock, and was designated CsetDNAV01.

**Host range.** The host range of CsetDNAV was determined by adding 5% (v/v) aliquots of fresh lysate passed through 0.2 μm filters (Whatman) into duplicate cultures of 28 exponentially growing clonal strains of the algal species listed in Table 1. The strains were cultured under the same conditions described earlier, at either 15°C or at 20°C. Growth of each algal culture after virus inoculation was monitored by visible light microscopy and compared with that of control cultures inoculated with SWM-3. The strains not lysed at 14 dpi were considered unsuitable hosts for CsetDNAV.

**Transmission electron microscopy (TEM).** An exponentially growing culture of *C. setoensis* was inoculated with CsetDNAV (9% v/v, multiplicity of infection = 52). A fresh host culture inoculated into SWM-3 served as a control. An aliquot of cell suspension was sampled at 48 hpi, harvested by centrifugation at 860 × g at 4°C for 10 min, and fixed with 1% glutaraldehyde in SWM-3 for 4 h at 4°C. The cell pellets were then fixed for 3 h using 2% osmic acid in 0.1 M phosphate buffer (pH 7.2–7.4), dehydrated through a graded ethanol series (50 to 100%), and embedded in Quetol 812 resin (Nisshin EM, Tokyo, Japan). Ultrathin sections were stained with 4% uranyl acetate and 3% lead citrate, and were observed at an acceleration voltage of 80 kV using a JEOL JEM-1010 transmission electron microscope. CsetDNAV particles negatively stained with uranyl acetate were also observed using TEM according to the method of Tomaru et al.<sup>11</sup>. Particle diameters were estimated based on the negatively stained images.

**Thermal stability.** An exponentially growing culture of *C. setoensis* was inoculated with CsetDNAV and incubated for seven days. The lysate was passed through a 0.8 μm-pore-size polycarbonate membrane filter (Whatman) to remove cellular debris. The titer of the resultant virus suspension was estimated using the extinction dilution method<sup>29</sup>, and aliquots of the lysate were then stored at 20°C, 10°C, 4°C, –20°C, –80°C, or –196°C (liquid nitrogen) in the dark without the addition of cryoprotectants. Virus stocks were titrated after 82 days of storage.

**Virus purification.** An exponentially growing *C. setoensis* culture (450 mL) was inoculated with 5 mL of a suspension of CsetDNAV and incubated until complete lysis. The lysate was passed through 0.4 μm polycarbonate membrane filters (Nucleopore) to remove cellular debris. Polyethylene glycol 6000 (Wako Pure Chemical Industries) was added to the filtrate to a final concentration of 10% (w/v), and the suspension was stored at 4°C in the dark overnight. After centrifugation at 57,000 × g at 4°C for 1.5 h, the pellet was washed with 10 mM phosphate buffer (pH 7.2) and added to an equal volume of chloroform. After vortexing, the suspension was centrifuged at 2200 × g for 20 min at 4°C to remove the chloroform. The water phase was collected with a pipette and centrifuged at 217,000 × g for 4 h at 4°C to collect the virus particles. The resultant viral pellets were used for viral genome and protein analyses.

**CsetDNAV nucleic acids.** Nucleic acids were extracted from the viral pellet using the DNeasy Mini Kit (Qiagen, Hilden, Germany), and were dissolved in 100 μL EB buffer, which was supplied with the kit. Aliquots (7 μL) of the nucleic acid solution were digested with 0.025 μg/μL RNase A (Nippon Gene, Tokyo, Japan) at 37°C for 1 h, 0.5 U/μL DNase I (Takara Bio, Shiga, Japan) at 37°C for 1 h, or 0.7 U/μL S1 nuclease (Takara Bio, Inc.) at 23°C for 15 min. With or without treatment at 100°C





for 2 min followed by cooling on ice, nucleic acids were separated by electrophoresis through agarose (1.5% w/v) gels (SeaKem® Gold Agarose; Lonza, Tokyo, Japan) at 50 V for 1 h in parallel with the enzyme-treated samples. Nucleic acids were visualised using SYBR-Gold staining (Molecular Probes, OR, US).

**Nucleotide sequence analysis of the viral genome.** For generating cDNAs, purified viral DNA was treated with a cDNA synthesis kit (MMLV version; Takara Bio) using random primers according to manufacturer recommendations. The 5'-ends of the resultant dsDNA fragments were phosphorylated using T4 polynucleotide kinase (Takara Bio). The resultant cDNA fragments were electrophoresed through agarose gel, and fragments of 1.0–1.5 kb were extracted. The fragments were ligated into a *HincII*-cleaved and dephosphorylated pUC118 plasmid vector (Takara Bio). The ligated dsDNA fragments were used to transform competent *Escherichia coli* DH10B (Invitrogen), and were sequenced after isolation and purification using the Sanger dideoxy method and an ABI 3730xl DNA Analyser (Applied Biosystems, US). The sequenced fragments were assembled using PGA (CAP4) ver. 2.6.2 (Paracel).

The S1 nuclease-resistant fragment (~100 bp) was purified using phenol-chloroform extraction and dissolved in ultrapure water. Ends were blunted using T4 DNA polymerase (Takara Bio), and were treated with alkaline phosphatase (Takara Bio). The fragments were electrophoresed through an agarose gel, ~100 bp fragments were extracted using Quantum Prep Freeze N' Squeeze™ DNA Gel Extraction Spin Columns (Bio-Rad Laboratories, Hercules, CA), dA was added, and the fragments were ligated to the TOPO TA cloning vector (Life Technologies, Tokyo, Japan). The addition of dA was conducted using 25  $\mu$ L mixtures containing the fragments, 1  $\times$  ExTaq buffer (Takara Bio), deoxyadenosine triphosphate (200 nM), and 1 U ExTaq DNA polymerase, with a GeneAmp PCR System 9700 (Life Technologies) for 2 h at 72°C. The cloned fragments were sequenced using the dideoxy method on an ABI PRISM 3100 DNA Analyser (Applied Biosystems).

Southern blot analysis was conducted to distinguish the viral (+) and complementary (–) strands of the viral genomic DNA. Based on the predicted sequence, an approximately 1 kb segment of the viral genome was amplified using PCR and ligated to the pCR®4-TOPO® vector (Life Technologies). Digoxigenin-labelled RNA probes specific for the viral or complementary sense strand of the viral genome DNA were transcribed from the plasmids using T7 RNA polymerase or T3 RNA polymerase according to manufacturer protocols (Roche, Basel, Switzerland). The nucleotide sequence of the CsetDNAV covalently closed-circular genome was determined by Southern dot-blot analysis using the probes according to the method of Mizumoto et al.<sup>30</sup>. The signals were detected with a luminescence image analyser (LAS-3000 mini; Fuji Photo Film, Tokyo, Japan).

**PCR experiments.** To confirm whether the viral genome forms covalently closed circular DNA, four primer pairs were designed based on its nucleotide sequence (see Fig. 7): CsetD\_1 (5'-GTG TTT AGC GTA GAA GTA GGG-3'), CsetD\_2 (5'-CCC TAC TTC TAC GCT AAA CAC-3'), CsetD\_3 (5'-CGC CAT TCG TTA ATA TTG TC-3'), and CsetD\_4 (5'-GAC AAT ATT AAC GAA TGG CG-3'). The primers were designed to amplify the entire viral genome (expected size of the amplicon is ~5.8 kb); CsetD\_1 and CsetD\_4, and CsetD\_2 and CsetD\_3 were designed to amplify partial regions of the circular DNA, ~2.7 kb and ~3.2 kb, respectively.

Using the primer sets for amplifying the whole genome, CsetD\_1 and CsetD\_2, and CsetD\_3 and CsetD\_4, PCR amplification was conducted with 20  $\mu$ L mixtures containing ~40 ng of viral template DNA, 1  $\times$  BlendTaq buffer (Toyobo, Osaka, Japan), each deoxynucleoside triphosphate (dNTP) (200 nM), 10 pmol of each primer, and 1 U BlendTaq DNA polymerase, using a GeneAmp PCR System 9700 (Applied Biosystems) according to the following cycle parameters: 10 cycles of denaturation at 94°C (30 s), annealing at 42°C (30 s), and extension at 72°C (6 min); 10 cycles of denaturation at 94°C, annealing at 47°C (30 s), and extension at 72°C (6 min); and 10 cycles of denaturation at 94°C (30 s), annealing at 52°C (30 s), and extension at 72°C (6 min). The conditions were identical to those described above when using the primer sets for amplifying the partial genomes, CsetD\_1 and CsetD\_4 and CsetD\_2 and CsetD\_3, except that the extension time at each step was 3 min. The PCR products were electrophoresed through 1% (w/v) Agarose ME gels (WAKO Pure Chemical Industries), and the nucleic acids were visualised using ethidium bromide staining.

**Phylogenetic analysis.** We identified a conserved nucleotide sequence encoding a putative replication-associated protein using the BLAST algorithm<sup>31</sup>. The deduced amino acid sequence of the corresponding region was compared with that of other viruses. They were automatically aligned using ClustalW<sup>32</sup>, and the alignments were subsequently manually refined. Phylogenetic trees were constructed using the neighbour-joining (NJ) and maximum likelihood (ML) methods with the Jones-Taylor-Thornton matrix (JTT model) packaged in MEGA 5<sup>33</sup>. The amino acid sequences used and their accession numbers (DDBJ database) were as follows: CdebDNAV, AB504376; ClorDNAV, AB553581; CsalDNAV, AB193315; *Chaetoceros setoensis* DNA virus (CsetDNAV), AB781089; CtenDNAV, AB597949; Csp05DNAV, AB647334; TnitDNAV, AB781284.

**Viral proteins.** The virus particles were resuspended in 600  $\mu$ L of ultra-pure water. Aliquots (5  $\mu$ L) of the suspension were mixed with 4 volumes of denaturing sample buffer (62.5 mM Tris-HCl, pH 6.8, 5% (v/v) 2-mercaptoethanol, 2% (w/v) SDS, 20% (v/v) glycerol, and 0.005% (w/v) bromophenol blue) and boiled for 5 min. The proteins were then separated using SDS-PAGE (80  $\times$  40  $\times$  1 mm, 12.5% polyacrylamide (w/v), 150 V) using the XV Pantera System (DRC, Tokyo, Japan).

Proteins were visualised using Coomassie Brilliant Blue. Protein molecular mass standards (BioRad) ranging from 10 to 250 kDa were used for size calibration.

**Growth experiments.** An exponentially growing culture of *C. setoensis* (25 mL) was inoculated with CsetDNAV (1 mL) at a multiplicity of infection of 270. A *C. setoensis* culture inoculated with autoclaved culture medium served as the control. An aliquot of the cell suspension was sampled from each culture at 0, 1, 2, 3, 4, 5, 6, and 7 dpi. The numbers of host cells and viruses were estimated by direct counting and the extinction dilution method<sup>29</sup>, respectively.

- Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
- Culley, A. I., Lang, A. S. & Suttle, C. A. High diversity of unknown picorna-like viruses in the sea. *Nature* **424**, 1054–1057 (2003).
- Tucker, K. P., Parsons, R., Symonds, E. M. & Breitbart, M. Diversity and distribution of single-stranded DNA phages in the North Atlantic Ocean. *ISME J.* **5**, 822–830 (2010).
- Angly, F. E. et al. The marine viromes of four oceanic regions. *PLoS Biol.* **4**, e368 (2006).
- Rosario, K., Duffy, S. & Breitbart, M. Diverse circovirus-like genome architectures revealed by environmental metagenomics. *J. Gen. Virol.* **90**, 2418–2424 (2009).
- Nelson, D. M., Treguer, P., Brzezinski, M. A., Leynaert, A. & Queguiner, B. Production and dissolution of biogenic silica in the ocean: revised global estimates, comparison with regional data and relationship to biogenic sedimentation. *Global Biogeochem. Cycle* **9**, 359–372 (1995).
- Tomaru, Y., Fujii, N., Oda, S., Toyoda, K. & Nagasaki, K. Dynamics of diatom viruses on the western coast of Japan. *Aquat. Microb. Ecol.* **63**, 223–230 (2011).
- Tomaru, Y. et al. First evidence for the existence of pennate diatom viruses. *ISME J.* **6**, 1445–1448 (2012).
- Tomaru, Y. & Nagasaki, K. in *The diatom world, cellular origin, life in extreme habitats and astrobiology vol. 19* (eds Seckbach, J. & Kocielek, J.P.) 211–225 (Springer, Dordrecht, 2011).
- Tomaru, Y., Shirai, Y., Suzuki, H., Nagumo, T. & Nagasaki, K. Isolation and characterization of a new single-stranded DNA virus infecting the cosmopolitan marine diatom *Chaetoceros debilis*. *Aquat. Microb. Ecol.* **50**, 103–112 (2008).
- Tomaru, Y. et al. Isolation and characterization of a single-stranded DNA virus infecting *Chaetoceros lorenzianus* Grunow. *Appl. Environ. Microbiol.* **77**, 5285–5293 (2011).
- Nagasaki, K. et al. Previously unknown virus infects marine diatom. *Appl. Environ. Microbiol.* **71**, 3528–3535 (2005).
- Toyoda, K. et al. Isolation and characterization of a single-stranded DNA virus infecting the marine planktonic diatom *Chaetoceros* sp. (strain TG07-C28). *Plankton Benthos Res.* **7**, 20–28 (2012).
- Tomaru, Y., Shirai, Y., Toyoda, K. & Nagasaki, K. Isolation and characterisation of a single-stranded DNA Virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Aquat. Microb. Ecol.* **64**, 175–184 (2011).
- Phan, T. G. et al. The fecal viral flora of wild rodents. *PLoS Pathog.* **7**, e1002218 (2011).
- Todd, D. et al. in *Virus taxonomy, classification, and nomenclature of viruses, 7th report* (eds van Regenmortel, M. H. V. et al.) 299–303 (Academic Press, San Diego, 2000).
- Aasayama, T. Maturation process of the granulosis virus of the diamondback moth, *Plutella xylostella*. *Jpn. Soc. Appl. Entom. Zool.* **19**, 149–156 (1975).
- Shalla, T. A. Assembly and aggregation of tobacco mosaic virus in tomato leaflets. *J. Cell Biol.* **21**, 253–264 (1964).
- Greber, R. S. Some characteristics of solanum nodiflorum mottle virus— a beetle-transmitted isometric virus from Australia. *Aus. J. Biol. Sci.* **34**, 369–378 (1981).
- Eiessler, Y., Wang, K., Chen, F., Wommack, E. & Coats, W. Ultrastructural characterization of the lytic cycle of an intranuclear virus infecting the diatom *Chaetoceros cf. wighamii* (bacillariophyceae) from Chesapeake Bay, USA. *J. Phycol.* **45**, 787–797 (2009).
- Tomaru, Y., Takao, Y., Suzuki, H., Nagumo, T. & Nagasaki, K. Isolation and characterization of a single-stranded RNA virus infecting the bloom forming diatom *Chaetoceros socialis*. *Appl. Environ. Microbiol.* **75**, 2375–2381 (2009).
- Tomaru, Y., Tanabe, H., Yamanaka, S. & Nagasaki, K. Effects of temperature and light on stability of microalgal viruses, HaV, HcV and HcRNAV. *Plankton Biol. Ecol.* **52**, 1–6 (2005).
- Lawrence, J. E., Chan, A. M. & Suttle, C. A. A novel virus (HaNIV) causes lysis of the toxic bloom-forming alga *Heterosigma akashiwo* (Raphidophyceae). *J. Phycol.* **37**, 216–222 (2001).
- Shirai, Y. et al. Isolation and characterization of a single-stranded RNA virus infecting the marine planktonic diatom *Chaetoceros tenuissimus* Meunier. *Appl. Environ. Microbiol.* **74**, 4022–4027 (2008).
- Gutierrez, C. Geminivirus DNA replication. *Cell. Mol. Life Sci.* **56**, 313–329 (1999).
- Ikari, J. On some *Chaetoceros* of Japan. I. *The Botanical Magazine* **40**, 517–534 (1926).
- Imai, I., Itakura, S., Matsuyama, Y. & Yamaguchi, M. Selenium requirement for growth of a novel red tide flagellate *Chattonella verruculosa* (Raphidophyceae) in culture. *Fisheries Sci.* **62**, 834–835 (1996).



28. Tomaru, Y., Tarutani, K., Yamaguchi, M. & Nagasaki, K. Quantitative and qualitative impacts of viral infection on *Heterosigma akashiwo* (Raphidophyceae) population during a bloom in Hiroshima Bay, Japan. *Aquat. Microb. Ecol.* **34**, 227–238 (2004).
29. Suttle, C. A. in *Handbook of methods in aquatic microbial ecology* (eds Kemp, P. F., Sherr, E. & Cole, J. J.) 121–137 (Lewis Publishers, Boca Raton, 1993).
30. Mizumoto, H., Tomaru, Y., Takao, Y., Shirai, Y. & Nagasaki, K. Intraspecies host specificity of a single-stranded RNA virus infecting a marine photosynthetic protist is determined at the early steps of infection. *J. Virol.* **81**, 1372–1378 (2007).
31. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
32. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
33. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).

## Acknowledgments

This study was partially supported by Grants-in-Aid for Young Scientists (A) (22688016) from the Ministry of Education, Science, and Culture of Japan.

## Author contributions

Y.T. wrote the main manuscript text and prepared the figures, H.S. and T.N. identified diatom species, and K.T., K.K. and Y.T. prepared Figures 4–6. All authors reviewed the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Tomaru, Y. *et al.* New single-stranded DNA virus with a unique genomic structure that infects marine diatom *Chaetoceros setoensis*. *Sci. Rep.* **3**, 3337; DOI:10.1038/srep03337 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>