Genome Sequence of *Ostreococcus tauri* Virus OtV-2 Throws Light on the Role of Picoeukaryote Niche Separation in the Ocean[▽]

Karen D. Weynberg,^{1,2} Michael J. Allen,¹ Ilana C. Gilg,³ David J. Scanlan,² and William H. Wilson^{1,3*}

Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, United Kingdom¹; School of Life Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom²; and Bigelow Laboratory for Ocean Sciences, 180 McKown Point, West Boothbay Harbor, Maine 04575-0475³

Received 8 October 2010/Accepted 26 January 2011

Ostreococcus tauri, a unicellular marine green alga, is the smallest known free-living eukaryote and is ubiquitous in the surface oceans. The ecological success of this organism has been attributed to distinct low- and high-light-adapted ecotypes existing in different niches at a range of depths in the ocean. Viruses have already been characterized that infect the high-light-adapted strains. *Ostreococcus tauri* virus (OtV) isolate OtV-2 is a large double-stranded DNA algal virus that infects a low-light-adapted strain of *O. tauri* and was assigned to the algal virus family *Phycodnaviridae*, genus *Prasinovirus*. Our working hypothesis for this study was that different viruses infecting high- versus low-light-adapted *O. tauri* strains would provide clues to propagation strategies that would give them selective advantages within their particular light niche. Sequence analysis of the 184,409-bp linear OtV-2 genome revealed a range of core functional genes exclusive to this low-light genotype and included a variety of unexpected genes, such as those encoding an RNA polymerase sigma factor, at least four DNA methyltransferases, a cytochrome *b₅*, and a high-affinity phosphate transporter. It is clear that OtV-2 has acquired a range of potentially functional genes from its host, other eukaryotes, and even bacteria over evolutionary time. Such piecemeal accretion of genes is a trademark of large double-stranded DNA viruses that has allowed them to adapt their propagation strategies to keep up with host niche separation in the sunlit layers of the oceanic environment.

Ostreococcus tauri is the smallest free-living eukaryote described to date, with a size of less than 1 μm (11). The cellular organization of *O. tauri* is very simple, with only a single chloroplast, a single mitochondrion, a single Golgi body, and a very reduced cytoplasmic compartment (22). *O. tauri* also lacks flagella, and there is no cell wall surrounding the cell membrane.

The *Ostreococcus* genus includes distinct genotypes physiologically adapted to high- or low-light environments, providing evidence of niche adaptation in eukaryotic picophytoplankton (39). Such adaptation has been well characterized in recent studies on the diversity and ecophysiology of the cyanobacterium *Prochlorococcus*. The global success of this abundant prokaryotic primary producer has been attributed in part to distinct low- and high-light-adapted ecotypes existing in different niches and utilizing different resources (38). Strains of *O. tauri* have been isolated in geographically different locations and depths and were shown to be genetically (based on 18S rRNA and internal transcribed spacer [ITS] sequencing) and physiologically (light-limited growth rates) different from one another (39). The growth rates of strains isolated from deep in the euphotic zone were reported to display severe photoinhibition at high light intensities (and are thus commonly referred to as low-light-adapted strains), while strains isolated from surface waters have very slow growth rates at the lowest light intensities (and are thus commonly referred to as high-light-adapted strains). The genetic distances between isolates ap-

pear to result from the contrast in both light and nutrient conditions experienced by surface and deep isolates which drives their genetic divergence (7, 39, 44).

Another factor that has not been considered in determining niche separation in *Ostreococcus* spp. is the role that viruses play. There are two primary mechanisms that viruses use to shape the diversity and magnitude of microbial populations. The first is simply killing cells, leading to host-specific lysis. Here, viruses exert an important influence on the biogeochemistry of the oceans, as nutrients are shunted between the particulate and dissolved phases (20, 51). A second and arguably more important function that viruses play is their role in horizontal gene transfer (HGT). Viruses can simply be seen as vectors that facilitate gene shuttling, a role that has been poorly described in marine systems. However, genes transferred between hosts and viruses can give selective advantages in growth (for the host) or propagation (for the virus) in particular environmental niches.

Information on virus propagation strategies and HGT events can be inferred and deduced, respectively, from genome sequence information. *Ostreococcus* spp. are an excellent model system since there are two host genomes, both of which are high-light-adapted species (15, 32), and two virus genomes (14, 50) that have already been sequenced. All grow or propagate in high-light-adapted systems. Our working hypothesis for this study was that different viruses infecting high- versus low-light-adapted *O. tauri* strains would provide clues to propagation strategies that would give them selective advantages within their particular light niche. Here, we report the genomic sequence of a virus (*O. tauri* virus [OtV-2]) that infects a low-light-adapted strain of *O. tauri*, and we compare the inferred

* Corresponding author. Mailing address: Bigelow Laboratory for Ocean Sciences, 180 McKown Point, West Boothbay Harbor, ME 04575-0475. Phone: (207) 633-9666. Fax: (207) 633-9641. E-mail: wwilson@bigelow.org.

[▽] Published ahead of print on 2 February 2011.

functionality of its coding sequences with those of high-light counterparts.

MATERIALS AND METHODS

Virus isolation. The virus OtV-2 was isolated from surface seawater collected on 5 July 2007 at the L4 sampling station in the western English Channel (coordinates are 50°15'N, 04°13'W). The OtV-2 host, *Ostreococcus tauri* strain RCC 393, was grown in Keller (K) medium (25) at 20°C under a 16:8-h light/dark cycle at irradiance of 30 $\mu\text{mol m}^{-2} \text{s}^{-1}$ in a Sanyo MLR-350 incubator. In order to obtain clonal virus stocks, OtV-2 was purified to extinction by serial dilution, as the host strain failed to grow successfully on agarose solid-bottom plates, preventing the use of plaque purification techniques. Briefly, virus lysate was obtained by adding 100 μl of concentrated seawater from station L4 to exponentially growing *O. tauri*. Once clearing of the host culture was observed, the lysate was passed through a 0.2- μm polyvinylidene difluoride (PVDF) filter (Durapore; Millipore).

Filtered lysate was used to infect exponentially growing phytoplankton cells in an 8-step series of 10-fold-dilution steps from 1 to 10^{-7} . Aliquots (100 μl) of each dilution were added to 3 wells of a 24-well assay plate, each well containing 200 μl of exponentially growing *O. tauri* RCC 393 culture. Cell lysis was recorded as the appearance of a virus group and a decline in cell numbers on a FACScan analytical flow cytometer (Becton Dickinson, Oxford, United Kingdom) equipped with a 15-mW laser exciting at 488 nm and with a standard filter setup. Phytoplankton abundance estimates were analyzed at a high flow rate ($\sim 70 \mu\text{l min}^{-1}$) and were discriminated by differences in their forward or right angle light scatter (FALS and RALS, respectively) and chlorophyll fluorescence. Samples for viral abundance analysis were fixed with glutaraldehyde (0.5% final concentration) for 30 min at 4°C, snap-frozen in liquid nitrogen, and stored at -80°C . Samples were subsequently defrosted at room temperature and diluted 500-fold with TE buffer (10 mmol liter $^{-1}$ Tris-HCl, pH 8, 1 mmol liter $^{-1}$ EDTA), stained with SYBR green 1 (Molecular Probes) (28a) at a final dilution of the commercial stock of 5×10^{-5} , incubated at 80°C for 10 min in the dark, and then allowed to cool for 5 min before flow cytometric analysis. Samples were analyzed for 2 min at a flow rate of $\sim 35 \mu\text{l min}^{-1}$, and virus groups were discriminated on the basis of their RALS versus green fluorescence. Data files were analyzed using WinMDI software, version 2.8 (Joseph Trotter [http://facs.scripps.edu]). Algal lysate from the most dilute step was filtered through a 0.2- μm PVDF filter, and the procedure repeated a further two times. The clonal virus sample obtained was filtered and stored at 4°C in the dark.

DNA preparation and sequencing. For preparation of large quantities of viruses for genome sequencing, 10-liter volumes of exponentially growing *O. tauri* culture were inoculated with 10 ml of OtV-2 at a multiplicity of infection (MOI) of one. Lysed cultures were passed sequentially through 0.8- μm and 0.2- μm filters to remove large cellular debris. Virus filtrates were concentrated by ultrafiltration to ~ 50 ml using a Quixstand benchtop system and hollow-fiber cartridges with a 30,000 molecular-weight cutoff (MWC) (GE Healthcare Amersham Biosciences). The 50-ml concentrate was then further concentrated to ~ 10 ml using a Mid-Gee benchtop system and hollow-fiber cartridge with a 30,000 MWC (GE Healthcare Amersham Biosciences). Aliquots (3 ml) of the concentrated OtV-2 lysate were adjusted with CsCl to densities of 1.1, 1.2, 1.3, and 1.4, and gradients from 1.1 to 1.4 were formed by ultracentrifugation at $100,000 \times g$ at 22°C for 2 h in a SW40-Ti Beckman rotor. Virus bands were removed with a syringe and dialyzed 4 times against 1-liter volumes of filtered sterile seawater.

Sequence assembly and finishing. Complete sequencing of the OtV-2 genome was performed at the NERC Biomolecular Analysis Facility based at the University of Liverpool, United Kingdom, using a GS-FLX 454 genome sequencer. The resulting reads were then *de novo* assembled with the 454's own Newbler assembler software, version 1.1.03.24. The resulting contigs were next screened according to coverage depth to filter out the low-coverage algal host contamination and bacterial contamination from the viral DNA. The remaining contigs were subsequently ordered and oriented using the *Ostreococcus virus 5* (OsV5) complete genome (NC_010191) as a reference sequence, using MUMmer 3.2. The OtV-2 genome was sequenced to an average of 10-fold coverage. Primers were designed to fill gaps between contigs, and the resultant Sanger sequence data were merged with the 454-generated contigs to form the completed genome sequence. Putative open reading frames (ORFs) were then identified, *de novo*, with glimmer3 using the provided g3-iterated.csh script. These ORFs were subjected to BLAST analysis (2a) against the reference sequence to identify those ORFs which may have been split due to frameshift errors caused by the 454 sequencer. Where found, relevant putative coding sequence (CDS) regions were joined and annotated accordingly. Coding regions of less than 65 amino acids were excluded from subsequent analysis. All remaining coding regions were then

preliminarily annotated after BLAST search against the reference to identify similarity. Putative tRNA genes were identified with the aid of tRNAscan-SE, version 1.23. The annotation was then checked and supplemented manually within the Artemis software tool (release 11) (41).

Sequence annotation. Whole-genome sequences of OtV-2 were analyzed and annotated using the software program Artemis (release 11) (41), with CDSs being generated based on predicted ORFs, correlation scores for each potential reading frame, G+C content, and codon usage indices. A CDS was defined as a continuous stretch of DNA that translates into a polypeptide that is initiated by an ATG translation start codon and extends for 65 or more additional codons. Similarities of putative CDSs were detected by using BLAST, and any homologous sequences were recorded. CDSs were assigned putative functions and were color coded based on their function. Each CDS was used in a search for homologues using the protein-protein BLAST (BLASTP) program (2). Comparative genomic analysis was conducted between the genomic data obtained in this study and similar virus genomes in the database using the software program Artemis Comparison Tool (ACT), release 8 (9).

Phylogenetic analysis. Phylogenetic analyses were performed as described previously (50) with the amino acid sequences of the DNA polymerase gene and high-affinity phosphate transporter genes obtained from the GenBank database. DNA polymerase amino acid sequences from a range of representative viruses were multiply aligned with the ClustalW program on the region spanning the highly conserved regions I and IV of the DNA *pol* genes. The entire sequences of high-affinity phosphate transporter genes were aligned.

Nucleotide sequence accession number. The OtV-2 sequence has been deposited in the GenBank database (accession no. FN600414).

RESULTS AND DISCUSSION

Virus isolation. *Ostreococcus tauri*-specific virus isolate OtV-2 was isolated from surface seawater collected on 5 July 2007 at the L4 sampling station in the western English Channel (coordinates 50°15'N, 04°13'W). Despite being isolated from surface seawater, host range analysis revealed that it did not infect the high-light-adapted *Ostreococcus* strains OTH 95, RCC 501, and RCC 356 (results not shown). The OtV-2 host, *O. tauri* strain RCC 393, belongs to clade B, based on analysis of the ITS region of the small subunit (SSU) rRNA operon (21, 39). RCC 393 is currently one of four characterized clade B *Ostreococcus* strains and was isolated at a depth of 90 m from the Mediterranean Sea (39). *O. tauri* strain RCC 393 displays a different pigment composition than high-light strains, with a higher chlorophyll *b*/chlorophyll *a* ratio and the additional presence of the chlorophyll pigment Chl *c* cs170, which is absent in high-light-adapted strains of the species (39). The latter chlorophyll pigment most likely plays a light-harvesting function in the blue-green region, which dominates at low levels of ambient light.

Description of the OtV-2 genome. Sequence analysis of the OtV-2 genome (accession number FN600414) revealed a linear genome of 184,409 bp. Features of the OtV-2 genome sequence include (i) a nucleotide composition of 42.15% G+C, (ii) a total of 237 predicted coding sequences (CDSs) that were defined as having a start codon followed by at least 65 additional codons prior to a stop codon, (iii) CDSs equally distributed on both strands (53% on the positive strand and 47% on the negative strand), (iv) an average gene length of 725 bp, and (v) a coding density of 1.285 genes per kbp (Table 1). The OtV-2 genome was oriented with the genomes of both OtV-1 and OtV-5, and a high degree of colinearity was observed among all three genomes (data not shown).

Of the 237 CDSs identified in the OtV-2 genome, 165 (69.6%) encode putative proteins to which no function can be attributed (Table 1). The remaining 72 CDSs (30.4%) have

TABLE 1. Comparison of the general characteristics of the genomes of OtV-2, OtV-1, and OtV-5^a

Characteristic	OtV-1	OtV-2	OtV-5
Genome size (bp)	191,761	184,409	185,373
% G+C	45.36	42.15	45.29
No. of tRNAs	4	5	5
No. of predicted CDSs (≥65 aa)	232	237	268
Mean gene size (bp/gene)	750	725	692
Coding density (genes/kb)	1.199	1.285	1.424
% of genes with:			
Known function	31	30.4	22
Hypothetical function	13	18.5	20
No known function	56	69.6	77

^a References: OtV-2 (this study), OtV-1 (50), and OtV-5 (14).

homology to known proteins and were split into 9 functional groups (Table 2) based on their predicted metabolic functions. At least 10 CDSs can be assigned putative functions involved in DNA replication, recombination, and repair, 7 in nucleotide metabolism and transport, 8 in transcription, 15 in protein and lipid synthesis, modification, and degradation, 1 in signaling, 5 in DNA methylation, and 9 in sugar metabolism, 8 encode capsid proteins, and a further 8 have other miscellaneous functions.

OtV-2 phylogeny. BLASTP searches identified a putative DNA polymerase (OtV-2_207). Phylogenetic analysis of highly conserved regions of the DNA polymerase of OtV-2 and other viral DNA polymerases showed close similarities with the DNA polymerases of other members of the *Phycodnaviridae* (Fig. 1). As the three currently described OtV viruses cluster with strong bootstrap support with the *Micromonas pusilla* viruses and other phycodnaviruses, OtV-2 can be putatively assigned to the *Phycodnaviridae* family.

TABLE 2. Putative proteins encoded within the OtV-2 genome grouped by function

Function and protein ^a	CDS	Function and protein ^a	CDS
DNA and RNA replication, recombination, and repair		DNA restriction/methylation	
DNA polymerase	OtV-2_207	Adenine-specific methyltransferase	OtV-2_083
ATP-dependent DNA ligase	OtV-2_169	DNA-cytosine methyltransferase ^c	OtV-2_050
DNA topoisomerase I	OtV-2_161	DNA-cytosine methyltransferase ^c	OtV-2_003
DNA topoisomerase II	OtV-2_211	Methyltransferase FkbM	OtV-2_010
PCNA	OtV-2_100	DNA methylase	OtV-2_078
Exonuclease	OtV-2_002	Protein and lipid synthesis/modification	
Exonuclease	OtV-2_118	Prolyl 4-hydroxylase alpha-subunit	OtV-2_106
RNase H	OtV-2_156	2OG-Fe(II) oxygenase ^c	OtV-2_165
VV A18 Helicase	OtV-2_056	2OG-Fe(II) oxygenase ^c	OtV-2_198
NTPase/helicase	OtV-2_107	2OG-Fe(II) oxygenase ^c	OtV-2_200
Nucleotide transport and metabolism		Procollagen-lysine 2-oxoglutarate 5-dioxygenase	OtV-2_158
Ribonucleotide reductase, large subunit	OtV-2_123	Ubiquitin C-terminal hydrolase	OtV-2_146
Ribonucleotide reductase, small subunit	OtV-2_140	ATP-dependent protease proteolytic subunit	OtV-2_171
dUTP pyrophosphatase	OtV-2_194	FtsH2 metalloprotease	OtV-2_009
ATPase (VV A32 virion packaging)	OtV-2_084	Aminotransferase family protein	OtV-2_032
Thymidine kinase	OtV-2_188	Acetolactate synthase	OtV-2_030
CMP/dCMP deaminase, Zinc-binding	OtV-2_190	<i>N</i> -Myristoyltransferase	OtV-2_076
FAD-dependent thymidylate synthase	OtV-2_051	TPR domain-containing protein	OtV-2_027
Transcription		Asparagine synthetase	OtV-2_022
RNA transcription factor TFIIB	OtV-2_142	Aspartyl/asparaginyl beta-hydroxylase	OtV-2_166
RNA transcription factor TFIIS	OtV-2_019	3-Methyl-2-oxobutanoate hydroxymethyltransferase	OtV-2_029
RNase III	OtV-2_117	Capsid	
RNA polymerase sigma factor	OtV-2_202	Capsid protein 1	OtV-2_057
SWI/SNF helicase domain protein	OtV-2_119	Capsid protein 2	OtV-2_062
TATA box binding protein	OtV-2_130	Capsid protein 3	OtV-2_066
mRNA capping enzyme ^b	OtV-2_079	Capsid protein 4	OtV-2_085
mRNA capping enzyme ^b	OtV-2_145	Capsid protein 5	OtV-2_086
Sugar manipulation		Capsid protein 6	OtV-2_154
GDP-D-mannose dehydratase	OtV-2_007	Capsid protein 7	OtV-2_206
Glycosyltransferase family 1	OtV-2_008	Capsid protein 8	OtV-2_210
Glycosyltransferase family 2	OtV-2_028	Miscellaneous	
Glycosyltransferase family 25	OtV-2_163	Virus inclusion body protein	OtV-2_109
Glycosyltransferase family 1	OtV-2_216	Rhodanese domain-containing protein	OtV-2_203
6-Phosphofructokinase	OtV-2_159	ABC1 domain protein	OtV-2_052
NAD-dependent epimerase/dehydratase	OtV-2_031	Fibronectin-binding protein	OtV-2_199
dTDP-D-glucose 4,6-dehydratase	OtV-2_033	Tail fiber assembly protein	OtV-2_065
6-Phosphogluconate dehydrogenase	OtV-2_167	Cytochrome <i>b₅</i>	OtV-2_201
Signaling		Phosphate starvation inducible protein	OtV-2_024
Serine/threonine protein kinase	OtV-2_135	High-affinity phosphate transporter	OtV-2_222

^a Shading indicates proteins exclusive to OtV-2 (i.e., not present in either OtV-1 or OtV-5).^b Genes are related but with different functions.^c Genes are paralogous.

FIG. 1. Neighbor-joining tree of DNA pol fragments. The species designations and GenBank accession numbers of DNA polymerase sequences used for phylogenetic analysis were as follows: alcelaphine herpesvirus (AAK00812); AcNPV, *Autographa californica* nuclear polyhedrosis baculovirus (P18131); BmNPV, *Bombyx mori* nucleopolyhedral virus (NP047469); CIV, chilo iridescent virus (AF303741); CbV-PW1, *Chrysochromulina brevifilum* virus PW1 (AAB49739); CeV-01, *Chrysochromulina ericina* virus (ABU23716); EsV-1, *Ectocarpus siliculosus* virus (NP_077578); EhV-86, *Emiliania huxleyi* virus-86 (AJ890364); equid herpesvirus (YP_053075.1); FsV-1, *Feldmannia species* virus-1 (AAB67116); FirrV-1, *Feldmannia irregularis* virus-1 (AY225133.1); FowPV, fowlpox virus (NP_039057); HaV-01, *Heterosigma akashiwo* virus-01 (AB194136.1); human herpesvirus (CAA28464); LdNPV, *Lymantria dispar* virus (AAC70269); MpV-SP1, *Micromonas pusilla* virus SP1 (AAB66713); MpV-SG1, *Micromonas pusilla* virus SG1 (AAB49746); MpV-PL1, *Micromonas pusilla* virus PL1 (AAB49747); mimivirus, *Acanthamoeba polyphaga* mimivirus (YP_142676); MOCV, *Molluscum contagiosum* virus (NP043990); OtV5, *Ostreococcus tauri* virus 5 (EU304328); OtV-1, *Ostreococcus tauri* virus 1 (FN386611); PBCV-1, *Paramecium bursaria Chlorella* virus (AAC00532); PBCV NY-2A, *Paramecium bursaria Chlorella* virus NY-2A (AAA88827); PgV, *Phaeocystis globosa* virus 1 (ABD65727); PoV01, *Pyramimonas orientalis* virus (ABU23717); *Sulfurisphaera ohwakuensis* (BAA 93703); VACV, *Vaccinia* virus (A24878); and OtV-2, *Ostreococcus tauri* virus 2 (FN600414). The scale bar indicates a distance of 0.2 fixed mutations per amino acid position.

The OtV-2 low-light genotype. CDSs exclusive to OtV-2 are highlighted in Tables 2 and 3. Of particular note, OtV-2 is the only one of the three OtV viruses that encodes a cytochrome *b₅* (OtV-2_201), RNA polymerase sigma factor (OtV-2_202), and a high-affinity phosphate transporter (OtV-2_222) (Table 2) (discussed in detail below). These three CDSs share homology with proteins encoded in the host genus, *Ostreococcus* (Table 3) (15, 32). A notable feature of most of these exclusive OtV-2 hostlike genes is their position in the OtV-2 genome. They

occur predominantly in the final quarter of the genome, between bp 147332 (start of OtV-2_183) and bp 180613 (end of OtV-2_229). Most intriguing is the clustering of the genes encoding a putative cytochrome *b₅* (OtV-2_201) and a putative RNA polymerase sigma factor (OtV-2_202), which are adjacent to one another. Moreover, in this same region of the OtV-2 genome, four genes encoding predicted proteins with no known function were identified as having high similarity to proteins encoded within the *Ostreococcus* host. This clustering

TABLE 3. Features of low-light (OtV-2) versus high-light (OtV-1 and OtV-5) genotypes

Putative function of CDS at indicated light level	Presence in genome of:			
	OtV-1	OtV-2	OtV-5	Host ^a
Low light				
Cytochrome <i>b</i> ₅		✓		✓
RNA polymerase sigma factor		✓		✓
High-affinity phosphate transporter		✓		✓
DNA methylase		✓		
Procollagen-lysine2-oxoglutarate5 dioxygenase		✓		
3-Methyl-2-oxobutanoate hydroxymethyltransferase		✓		
Tail fiber assembly protein		✓		
CMP/dCMP deaminase, zinc binding		✓		
6-Phosphogluconate dehydrogenase		✓		
High light				
ATP-dependent protease proteolytic subunit	✓			
33-kDa <i>in vitro</i> translation peptide	✓		✓	
3-Dehydroquinate synthase	✓		✓	
Echinonectin	✓			
Proline dehydrogenase	✓		✓	
Patatin phospholipase	✓		✓	
Oxo-acyl carrier protein dehydrogenase	✓		✓	

^a The host was *Ostreococcus tauri* RCC 745 (15).

of islands of hostlike genes raises the possibility of a “hot spot” region within the OtV-2 genome with a propensity to acquire genes from the host. This may be driven by environmental selection pressure, such as growth/propagation in low-light environments. Several hostlike genes have been detected localized to a region of a number of cyanomyovirus genomes, indicating that this hyperplastic region may be site-specifically associated with the acquisition of hostlike genes (29). The accretion of host genes is believed to play a role in the evolution of viruses (24). The evolution of bacteriophages is be-

lieved to be driven in part by the acquisition of more and more foreign DNA, the so-called “moron” hypothesis (23). The existence of several hostlike genes in the OtV-2 genome provides strong evidence of HGT events, and functional analysis of the genes involved should enable a greater insight into this close interaction.

Novel functionality of the OtV-2 genome. Arguably the most environmentally relevant acquisition of a host gene by OtV-2 is the presence of a putative high-affinity phosphate transporter gene (OtV-2_222). This is in addition to a gene encoding a putative phosphate starvation-inducible protein of the PhoH protein family found in all 3 *Ostreococcus tauri* viruses (OtV-2_024). A database BLASTP search for other viral proteins with homology to the high-affinity phosphate transporter (OtV-2_222) in OtV-2 gave only one other match, to a putative phosphate-repressible phosphate permease in the coccolithovirus EhV-86. The high-affinity phosphate transporter gene reported in OtV-2 encodes a PHO4 family protein. With an amino acid identity of 57% to the host protein, it is likely the virus has acquired this gene from its *Ostreococcus* host. This gene has been identified and characterized in other phytoplankton species, such as *Tetraselmis chui* (12). Phylogenetic analysis of the OtV-2 high-affinity phosphate transporter (Fig. 2) shows that the viral protein forms a distinct cluster with the *Ostreococcus* host proteins (bootstrap value of 100% for 1,000 trials). These results suggest that the gene encoding this putative high-affinity phosphate transporter has been acquired by the virus through horizontal transfer from the host. Indeed, a ClustalW alignment of the putative high-affinity phosphate transporter in the virus and the equivalent proteins in the host species *O. tauri* and *Ostreococcus lucimarinus* shows a relatively high degree of conservation among the three proteins (Fig. 3). However, the virus version is missing 54 amino acid (aa) residues in the center of the protein (a result of a 162-bp in-frame deletion in the middle area of the gene), as well as several residues at the N terminus found in both host versions.

A putative cytochrome *b*₅ gene (OtV-2_201) was identified

FIG. 2. Neighbor-joining tree of high-affinity phosphate transporter sequences. The species designations and GenBank accession numbers of high-affinity phosphate transporter sequences used for phylogenetic analysis were as follows: *Alteromonas macleodii* (NC_011138.1); *Chlamydomonas reinhardtii* (XP_001695776.1); *Emiliana huxleyi* (AF334403.1); *Emiliana huxleyi* virus-86 (YP_293871.1); *Neurospora crassa* (XP_959489.1); *Ostreococcus lucimarinus* (XP_001416412.1); *Ostreococcus tauri* (CAL52329); *Saccharomyces cerevisiae* (CBK39373.1); *Tetraselmis chui* (AAO47330.1); and *Thalassiosira pseudonana* (XP_002281698.1). The scale bar indicates a distance of 0.1 fixed mutations per amino acid position.

Otv-2		-----MLQQQFEWIG
<i>O. lucimarinus</i>	MNFLPSLDGSLDWNQSRLLSRTNREFVPPQGFSESCLFKFPRGNMRGRQRNTSRVSSRRTTMSSTDTLAQASRHLLAGATETKR....	
<i>O. tauri</i>	-----MTDESRTAKYFSRSKSRKEYSLIFSEYFTFLRPGDADGVTSRRALPVSMGSGEALAHAGRALLMAGGEATKN.....	: *****
Otv-2	VTGAIFAFLASFGIGANDVANAFATSVGSGALKIKQAVPLAAVCEFGSGLFMGSHVVKTIRKGISDQDCFVDD-----PGLLMYGCLCVI	
<i>O. lucimarinus</i>	.C.SFL..F.A.....T.N.V...IF..C.A.M...G.....AN.K..AGTGGAND.....	
<i>O. tauri</i>	.I.GFL..F.A.....S..S.V..G.F..C.A.M...G...E.M...VT..K..NGKSGSHD.V.....	* *,:*:*,*****,*.* **,: * *,*:***,***,:***,:*.* *
Otv-2	LSVAMWLIVASYLEMPVSTTHSCIGGMIGMTMVARGSSCVTIWAKSDQFPYVKGVVAIIISWLLSPIISGLFASALFFVVRSLILRSDDS	
<i>O. lucimarinus</i>	FA..I..VI..AF..A.....V.....L....E..I.TK.A.E.....V.....V...A..FVF.VTL.T.VM...H.	
<i>O. tauri</i>	FA.GV..V...A.....V..L...L...DD..V.SK.A.E.....A..VV..V...V...C.SF.M.LIL.T.VM...N.	:*:***:* *:*****:*:****:****,* * *:*****,:*:***:*** *: :*: :*:****:
Otv-2	YKRIQYGFPPVLVAGTFIINTFFIVYKGAKFLKLDDTPLWKACAISFGIGGGTGIFSYYFFINPIIFN-----	
<i>O. lucimarinus</i>	.S.TAVA...L.C.L..I.....E....VGT...A....C..IA...VT.Y.LKTTDELFEKQQOLEKAERGSGKKAE	
<i>O. tauri</i>	FS.VS.V...FGA.L..V.....SPH.G.K...VGT...AA..V...A..L...L.S.Y.VRTEELYAQEQAEQAERGSGKKVEE	:* ****,*:** *****: * ****: *** :*** ***:*** **:***:*. *
Otv-2	-----DEERRIEEFCHTDSDVSQSIHDNAEVFDTKTEHSRMRYLQILTACDSFAHGANDVANSIGP	
<i>O. lucimarinus</i>	KVVRQPPEYPVGVGAPRRMWYALQDHLESSLAHKA.DILDE.MA.LA..E..K..E..LC.....	
<i>O. tauri</i>	TVVRQPPEYPRGVFGAPTRVWYAIQDHLETSLAVKASDLIEE.MA.NA...K..E..L.....A.....	: : : . * : * :*:*** ** *** ,*****:*****
Otv-2	FAAIYAIYKSGTVS--KNAEMGNDAYWILSLGAGGIVVGLATYGYKILEALGTKMAKLTPSRGICIELGAAVILIILGSRLGWPLSTTHCQ	
<i>O. lucimarinus</i>	..SMVVVF..K.--.E...D.S...G...A...C...L.....H...I.....C..M.....	
<i>O. tauri</i>	..T.VLV..A.K..TK.KDP..D.....L.....L...R..H...L.....C.....M.....	**:: :*:** * *: *:***** ** ** ** **,**,*****:*****:***:***:*****
Otv-2	VGATVGVALFEGTGGVNWKLLYKTIAGWLLTLVVVGSTTAFLFAQAYAPMKYPSYVINA-----	
<i>O. lucimarinus</i>	L.RK.I..FIIG..VF..II..I..FS.AF.....KS..C.ITGSC-----	
<i>O. tauri</i>	...M...ML.RK.I..FI.A.V...II.....FS.SAF....F...H..C.IAATAKDSISEALCTI	***:***:*** ** *: :*:***** ** *:***** ** *: * :

FIG. 3. Amino acid alignment of high-affinity phosphate transporter sequences of OtV-2, *O. tauri*, and *O. lucimarinus*. Dashes (-) represent gaps. Dots (.) underneath residues represent consensus sequences. Asterisks (*) underneath residues represent conserved amino acids.

in the OtV-2 genome which shares homology with a gene carried by the host *O. tauri*. Cytochrome *b*₅ is a small hemoprotein and a ubiquitous electron transport carrier found in animals, yeasts, and plants. The role cytochrome *b*₅ plays in the synthesis of unsaturated fatty acids has been well characterized in animals and plants (45). Cytochrome *b*₅ consists of two domains, namely, a hydrophobic tail that anchors the protein to the membrane and a hydrophilic portion, the heme-binding domain, which is active in redox reactions (48).

BLASTP analysis of the OtV-2 protein against public databases resulted in a best match (an amino acid identity of 60% and an E-value of $2e^{-29}$) to a cytochrome b_5 in *O. tauri*. An alignment of the putative cytochrome b_5 in OtV-2 and the *O. tauri* homologue shows that the virus protein aligns with the C-terminal region of the host protein. Across most of this alignment, there is a high level of conservation between the two proteins. However, much of the remainder of the host protein is missing in the viral version (Fig. 4). The core of the heme-binding domain is formed by the characteristic cytochrome b_5 motif His-Pro-Gly-Gly (45). This diagnostic cytochrome b_5 motif 2 of the heme-binding domain was identified both in the OtV-2 cytochrome b_5 , at residues 36 to 39 of the 91-aa polypeptide, and in the *O. tauri* host cytochrome b_5 , at residues 545 to 548 of the 586-aa polypeptide. From the evidence presented, there is a possibility the gene encoding this cytochrome b_5 in OtV-2 was acquired from the algal host, particularly as the adjacent gene encodes a hostlike RNA polymerase sigma factor. A putative cytochrome b_5 gene has also

been identified in *Acanthamoeba polyphaga* mimivirus (36). OtV-2 is, therefore, only the second virus reported to encode a putative cytochrome *b₅*.

Nucleotide transport and metabolism. The OtV-2 genome encodes a putative deoxycytidylate deaminase (dCD) (OtV-2_190) that is not found in OtV-1 or OtV-5. This enzyme converts dCMP to dUMP (28) and is a major supplier of the substrate for thymidylate synthase, an important enzyme in DNA synthesis (8). This is of significance as the OtV-2 virus encodes a thymidylate synthase, ThyX (OtV-2_051), that is also found in OtV-1 and OtV-5. Both enzymes have been shown to be simultaneously elevated in rapidly dividing cells and have minimal activity in nondividing cells (28). The enzyme dCD is present in most eukaryotes and bacteria, with those present in humans and T4-bacteriophage being the most extensively studied. Viruses known to encode a dCD include certain bacteriophages, e.g., T4, the chloroviruses, and mimivirus (18, 19, 36). Members of the cytidine deaminase superfamily of enzymes have been investigated extensively in eukaryotes, as they play a role in antibody production in the immune system. As dCD is a major provider of dUMP and thymidylate synthase is the only *de novo* source of dTMP in most biological systems, these enzymes have also become potential targets for anticancer therapy (40).

Transcription. The OtV-2 genome encodes eight CDSs involved in transcription (Table 2). A putative RNA polymerase sigma factor (OtV-2_202) has been identified in the OtV-2 genome which is not found in OtV-1 or OtV-5. Genes encod-

FIG. 4. Alignment of the cytochrome b_5 proteins encoded in *Ostreococcus tauri* and the virus OtV-2. The conserved heme-binding domain is the shaded region.

ing RNA polymerase sigma factors have previously only been reported in bacteriophages, and this is believed to be the first finding of such a gene in a virus infecting eukaryotes. These factors assist the polymerase in binding selectively to the promoter and initiating transcription (6). The initiation of transcription from promoter elements is triggered by the reversible association of sigma factors with the complex to form a holoenzyme. The sigma-70 (σ^{70}) factors are known as the primary or major sigma factors and can initiate the transcription of a wide variety of genes. Members of the σ^{70} family are components of the RNA polymerase holoenzyme that direct bacterial or plastid core RNA polymerase to specific promoter elements that are situated 10 and 35 base pairs upstream of transcription initiation points (31). The primary family σ factor, which is essential for general transcription in exponentially growing cells, is reversibly associated with RNA polymerase. The σ^{70} family members have four conserved regions, the highest conservation being found in regions 2 and 4, which are involved in binding to RNA polymerase-recognizing promoters and separating DNA strands (DNA "melting"). Bacteriophages, such as T4, encode a σ^{70} factor (47).

BLASTP analysis revealed that the protein product of CDS OtV-2_202 shared homology (amino acid identity of 42% and an E-value of $1e^{-09}$) with an RNA polymerase σ^{70} factor encoded in the host *O. tauri* genome. All members of the σ^{70} factor superfamily contain region 2, the most conserved domain of this protein. Region 2 of this protein is highly con-

served between organisms, as it contains both the -10 promoter recognition helix and the primary core RNA *pol*-binding determinant. Analysis of the putative OtV-1_202 protein using the Pfam database indicated that this protein contains a recognizable region 2 (Fig. 5), as does the host protein, thus confirming this as a putative σ^{70} factor. An alignment of this conserved domain of host and virus proteins and those of the closest homologues was performed (Fig. 5). The virus gene has deletions at both ends when compared to the host gene, indicating that the virus has undergone significant changes in its sequence and, possibly, function over time. Moreover, the virus gene has a G+C content of 37.95% but the host gene has a G+C content of 60%. Therefore, if this gene was acquired by the virus from the host, it is unlikely to have been a recent event. An alternative source of this gene is from a bacterium. Aside from the closest BLASTP hit to a small region of the host protein, all of the closest BLASTP hits are with equivalent proteins from bacterial species, e.g., *Rickettsia* and *Synechococcus*. The acquisition of genes by phycodnaviruses from a bacterium has been reported previously (16, 17). A more convoluted possibility for the transfer of this gene to the virus may have been from a bacterial source to the host, followed by a second HGT event from host to virus. Both BLASTP and Pfam searches of the host protein gave the closest hits to similar proteins in cyanobacterial species, with approximate E-values of $3e^{-31}$ and amino acid identities of 33%. The only close hit to a similar protein in a eukaryote species is to one in *Mi-*

FIG. 5. Amino acid alignment of the highly conserved region 2 of the putative RNA polymerase sigma factor gene in OtV-2 and its closest matching homologues.

cromonas pusilla, which is also a prasinophyte species. This finding suggests an alternative hypothesis to all those previously outlined. The *Prasinophyceae* diverged early at the base of the *Chlorophyta* (5). As members of the *Prasinophyceae*, both *Micromonas* and, particularly, *Ostreococcus*, with its small cell size (less than 1 μm), lack of flagella, and simple cellular organization, hold key phylogenetic positions in the eukaryotic tree of life. As a primitive species, *Ostreococcus* may have acquired this gene from a bacterial evolutionary ancestor. The absence of this gene from the OtV-1 and OtV-5 genomes indicates that OtV-2 only has undergone a further HGT event that has not occurred in the two viruses infecting the high-light host strain. The functionality of the gene encoding a putative RNA polymerase sigma factor is unconfirmed at present, but such features of the OtV-2 genome may indicate past HGT events between the virus and bacterial/phage sources.

Protein and lipid synthesis/modification. The OtV-2 genome encodes a putative procollagen-lysine 2-oxoglutarate (2OG) 5-dioxygenase (PLOD) (OtV-2_158), an enzyme which exists in eukaryotes but has not been reported in prokaryotes or viruses, with the exception of the mimivirus (36). This enzyme has been extensively characterized in animals, where its role is to hydroxylate lysine residues in collagens. The resulting hydroxylysines serve as attachment sites for carbohydrate units and are essential for the stability of intermolecular collagen cross-links (26). The putative PLOD protein in OtV-2 shares approximately 30% amino acid identity with a similar protein in several eukaryotic organisms. The function of this gene in the OtV-2 genome may be to ensure the stability of carbohydrate units in structures such as the capsid protein. Without functional analysis, this hypothesis remains speculative.

A putative 3-methyl-2-oxobutanoate hydroxymethyltransferase is encoded by OtV-2 (OtV-2_029). This enzyme is the first in the pantothenate biosynthesis pathway and catalyzes the constraining step in the synthesis of pantothenate, or vitamin B5 (46). Pantothenate is a necessary precursor to coenzyme A and phosphopantetheine, the prosthetic group of the acyl carrier protein, both of which are vital to a multitude of metabolic processes. Coenzyme A assists in transferring fatty acids from the cytoplasm to mitochondria (30). Pantothenate is synthesized by microorganisms, i.e., bacteria and fungi, and plants, but not animals, which require it as part of their diet. This biosynthesis pathway has been well studied in bacteria, such as *Escherichia coli* (27).

The putative 3-methyl-2-oxobutanoate hydroxymethyltransferase CDS in the OtV-2 genome is most similar to a gene carried by several bacterial species (amino acid identity of approximately 45% and E-value of e^{-60}). Although it is not known if this gene is functional in OtV-2, its protein product is predicted to exhibit a Rossmann fold (35), which is highly conserved in these enzymes, and the active site domains are also conserved.

The OtV-2 genome encodes three putative 2OG-Fe(II) oxygenases (OtV-2_165, OtV-2_198, and OtV-2_200) and a putative prolyl 4-hydroxylase (OtV-2_106). These enzymes are related, as their family contains members of the 2-oxoglutarate and Fe-dependent oxygenase superfamily and includes the C terminus of the prolyl 4-hydroxylase alpha subunit. The enzyme 2OG-Fe(II) oxygenase belongs to a class of enzymes that are widespread in eukaryotes and bacteria and catalyze a va-

riety of reactions, typically involving the oxidation of an organic substrate using a dioxygen molecule (34). An extensively characterized reaction involving this enzyme is the hydroxylation of proline and lysine side chains in collagen (3). The presence of a putative prolyl-4-hydroxylase, three putative 2OG-Fe(II) oxygenases, and a procollagen-lysine,2-oxoglutarate 5-dioxygenase in the OtV-2 genome indicates that this virus may code for the stabilization of complex structures, such as carbohydrate units, during its replication. Both the OtV-1 and OtV-5 genomes encode only one putative 2OG-Fe(II) oxygenase each. Therefore, this begs the question of why a virus infecting a low-light strain would require three of these genes. Perhaps the assembly and stabilization requirements of the virus capsid necessitate the involvement of these enzymes.

Of note, OtV-2 does not encode two enzymes involved in lipid metabolism, namely, a patatin-like phospholipase and an oxo-acyl carrier dehydrogenase or a 3-dehydroquinase synthase, which are found encoded in the viruses infecting the high-light host strain. This may be due to the OtV-2 virus utilizing host genes involved in these metabolic processes, thus negating the requirement for these genes in the virus genome. Alternatively, the infection process of OtV-2 may differ somewhat from the infection processes of the other OtV viruses and, thus, exclude the need for genes involved in lipid synthesis.

Sugar manipulation enzymes. Several proteins encoded within the OtV-2 genome have close identities to enzymes involved in sugar manipulation, polysaccharide synthesis, and the transfer of sugars to proteins. Some viruses can affect the expression of host glycosyltransferases, and a few express their own glycosyltransferases. The OtV-2 genome encodes at least four glycosyltransferases: two from family 1 and one each from families 2 and 25 (Table 2). This is two fewer than are encoded by the OtV-1 genome, which also encodes an alpha galactosyltransferase not encoded by OtV-5 or OtV-2. All three genomes encode a dTDP-D-glucose 4,6-dehydratase. The OtV-2 genome also contains the *gnd* gene, encoding a putative 6-phosphogluconate dehydrogenase (OtV-2_167), not found in viruses infecting the high-light *O. tauri* strain. This enzyme catalyzes the decarboxylating reduction of 6-phosphogluconate into ribulose 5-phosphate in the presence of NADP (52). This reaction is part of the hexose monophosphate shunt and pentose phosphate pathways. Virus DNA can be synthesized from the intermediates of the reductive pentose phosphate pathway during photosynthesis, from the intermediates of the oxidative pentose phosphate pathway, and also, from nucleotide precursors degraded from host DNA (43). The oxidative pentose phosphate pathway metabolizes glucose-6-phosphate to ribose-5-phosphate, which is necessary for the *de novo* biosynthesis of purine and pyrimidine nucleotides of viral DNA (42). Prokaryotic and eukaryotic 6-phosphogluconate dehydrogenases are proteins of approximately 470 amino acids with highly conserved sequences. The OtV-2 protein has approximately 30% amino acid identity and E-values of approximately $2e^{-50}$ with those of several bacterial species. It is possible the virus acquired the *gnd* gene from a bacterial source. The only other virus reported to encode this enzyme is cyanophage syn9 (49), and this is therefore only the second report of this enzyme in a virus. Presumably, the majority of viruses, which do not contain the *gnd* gene, hijack the host's biosynthetic pathway for

nucleotide synthesis during infection. The OtV-2 genome may possess this gene due to the niche its host occupies, thus making a viral-encoded version a beneficial acquisition.

DNA methylation. The OtV-2 genome encodes a number of putative methyltransferases (Table 2). DNA methyltransferases are rare among viruses but are reported most commonly in bacteriophages and, also, in the chloroviruses (1) and phaeoviruses (13, 33). Of particular note is the presence of a putative 6-adenine methyltransferase encoded in the OtV-2 genome. DNA adenine methyltransferase (DAM) genes are found in the genomes of many fungi, bacteria, protists, and archaea (10). However, the methylation of this base is a rare modification in eukaryotes (37). Restriction/modification systems in prokaryotes help protect cells against invading phages and plasmids (4). The presence of genes encoding methyltransferases in the OtV-2 genome may indicate coevolution enabling the protection of viral DNA from the host restriction/modification system during infection.

Miscellaneous. OtV-2 has several genes encoding other putative proteins, including a rhodanese domain-containing protein (OtV-2_203) and an ABC domain protein (OtV-2_052). The product of a large CDS (OtV-2_109) resembles a virus-like inclusion body protein encoded by the protist *Trichomonas vaginalis*. The OtV-2 CDS is 2,037 amino acids in length, with a G+C content of 41.75%. This gene in OtV-2 aligns with three consecutive CDSs in the reference genome, OtV-5, and a single large CDS (3,398 amino acids long and G+C content of 43.76%) in the OtV-1 genome. Resequencing of the large genes in OtV-1 and OtV-2 confirmed that these genes are single CDSs. No conserved domains were identified within these CDSs, and further functional work is required.

A putative tail fiber assembly protein (OtV-2_065) is encoded by OtV-2 but not seen in the other OtV genomes. PSI-BLAST of this protein against the public databases resulted in close hits to bacteriophage tail fiber assembly proteins, with an average amino acid identity of 30% and E-value of $2e^{-32}$.

Ostreococcus tauri virus isolate OtV-2 is a virus that specifically infects a low-light-adapted *O. tauri* strain (RCC 393). Our original working hypothesis was that different viruses infecting high- versus low-light-adapted *O. tauri* strains would provide clues to propagation strategies that would give them selective advantages within their particular light niche. It is clear that OtV-2 has acquired a range of metabolically diverse genes from its host, other eukaryotes, and even bacteria over evolutionary time. Such piecemeal accretion of genes is a trademark of large double-stranded DNA viruses that has allowed them to adapt their propagation strategies in order to respond to selection pressures, including host niche separation in the surface layers of the oceanic environment. As more viruses are isolated and their genomes sequenced, we will start to see a clearer picture of what core genes are necessary to function in a particular oceanic niche. This study essentially represents a starting point in defining this low-light core virus genotype. In addition, it may help researchers establish a function for many of the unknown genes exclusive to a low-light core genotype (or indeed any other niche genotype). Crucially, the collection of environmental metadata will become a necessary component to help establish the functionality of genes of organisms isolated from a particular niche.

ACKNOWLEDGMENTS

This research was supported by a standard Ph.D. studentship (reference no. NER/S/A/2005/13204) awarded to W.H.W. and D.J.S., a small projects grant (grant MGF196) awarded to M.J.A. from the Natural Environment Research Council (NERC), and a National Science Foundation grant (grant EF0949162) awarded to W.H.W.

We would like to acknowledge technical help from Margaret Hughes, Kevin Ashelford, and Neil Hall at the NERC Biomolecular Analysis Facility at the University of Liverpool.

REFERENCES

1. Agarkova, I. V., D. D. Dunigan, and J. L. Van Etten. 2006. Virion-associated restriction endonucleases of chloroviruses. *J. Virol.* **80**:8114–8123.
2. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- 2a. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
3. Aravind, L., and E. V. Koonin. 2001. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**:RESEARCH0007.
4. Bertani, G., and J. J. Weigle. 1953. Host controlled variation in bacterial viruses. *J. Bacteriol.* **65**:113–121.
5. Bhattacharya, D., and L. Medlin. 1998. Algal phylogeny and the origin of land plants. *Plant Physiol.* **116**:9–15.
6. Burgess, R., A. Travers, J. Dunn, and E. Bautz. 1969. Factor stimulating transcription by RNA polymerase. *Nature* **221**:43–46.
7. Cardol, P., et al. 2008. An original adaptation of photosynthesis in the marine green alga *Ostreococcus*. *Proc. Natl. Acad. Sci. U. S. A.* **105**:7881–7886.
8. Carreras, C. W., and D. V. Santi. 1995. The catalytic mechanism and structure of thymidylate synthase. *Annu. Rev. Biochem.* **64**:721–762.
9. Carver, T. J., et al. 2005. ACT: the Artemis comparison tool. *Bioinformatics* **21**:3422–3423.
10. Cheng, X. D. 1995. Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **24**:293–318.
11. Chretiennotdinet, M. J., et al. 1995. A new marine picoeukaryote—*Ostreococcus tauri* gen. et sp. nov. (Chlorophyta, Prasinophyceae). *Phycologia* **34**:285–292.
12. Chung, C.-C., S.-P. L. Hwang, and J. Chang. 2003. Identification of a high-affinity phosphate transporter gene in a Prasinophyte alga, *Tetraselmis chui*, and its expression under nutrient limitation. *Appl. Environ. Microbiol.* **69**:754–759.
13. Delarouge, N., et al. 2001. The complete DNA sequence of the Ectocarpus siliculosus virus EsV-1 genome. *Virology* **287**:112–132.
14. Derelle, E., et al. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. *PLoS One* **3**:e2250.
15. Derelle, E., et al. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc. Natl. Acad. Sci. U. S. A.* **103**:11647–11652.
16. Filee, J., N. Pouget, and M. Chandler. 2008. Phylogenetic evidence for extensive lateral acquisition of cellular genes by nucleocytoplasmic large DNA viruses. *BMC Evol. Biol.* **8**:320.
17. Filee, J., P. Signier, and M. Chandler. 2007. I am what I eat and I eat what I am: acquisition of bacterial genes by giant viruses. *Trends Genet.* **23**:10–15.
18. Fitzgerald, L. A., et al. 2007. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella* Pbi. *Virology* **358**:459–471.
19. Fitzgerald, L. A., et al. 2007. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. *Virology* **362**:350–361.
20. Fuhrman, J. A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**:541–548.
21. Guillou, L., et al. 2004. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist* **155**:193–214.
22. Henderson, G. P., L. Gan, and G. J. Jensen. 2007. 3-D ultrastructure of *O. tauri*: electron cryotomography of an entire eukaryotic cell. *PLoS One* **2**:e749.
23. Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens. 2000. The origins and ongoing evolution of viruses. *Trends Microbiol.* **8**:504–508.
24. Iyer, L. A., S. Balaji, E. V. Koonin, and L. Aravind. 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* **117**:156–184.
25. Keller, M. D., R. C. Selvin, W. Claus, and R. R. L. Guillard. 1987. Media for the culture of marine ultraphytoplankton. *J. Phycol.* **23**:633–638.
26. Kivirikko, K. I., R. Myllyla, and T. Pihlajaniemi. 1992. Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins, p. 1–51. *In* J. J. Harding and M. J. C. Crabbe (ed.), *Focus on post-translational modifications of proteins*. CRC Press, Boca Raton, FL.

27. **Lobley, C. M. C., et al.** 2003. Structural insights into the evolution of the pantothenate-biosynthesis pathway. *Biochem. Soc. Trans.* **31**:563–571.
28. **Maley, F., and G. F. Maley.** 1999. Structure, function analysis of T4-phage deoxycytidylate deaminase and its role in the phage metabolic pathway. *Paths Pyrimidines* **7**:1–7.
- 28a. **Marie, D., C. P. D. Brussaard, R. Thyrhaug, G. Bratbak, and D. Vault.** 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl. Environ. Microbiol.* **65**:45–52.
29. **Millard, A. D., K. Zwirgmaier, M. Downey, N. H. Mann, and D. J. Scanlan.** 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanism of cyanophage evolution. *Environ. Microbiol.* **11**:2370–2387.
30. **Ottenhof, H. H., et al.** 2004. Organisation of the pantothenate (vitamin B-5) biosynthesis pathway in higher plants. *Plant J.* **37**:61–72.
31. **Paget, M., and J. Helmann.** 2003. The sigma70 family of sigma factors. *Genome Biol.* **4**:203.
32. **Palenik, B., et al.** 2007. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc. Natl. Acad. Sci. U. S. A.* **104**:7705–7710.
33. **Park, Y., G. D. Kim, and T. J. Choi.** 2007. Molecular cloning and characterization of the DNA adenine methyltransferase gene in *Feldmannia* sp. virus. *Virus Genes* **34**:177–183.
34. **Prescott, A. G.** 1993. A dilemma of dioxygenases: or where molecular biology and biochemistry fail to meet. *J. Exp. Botany* **44**:849–861.
35. **Rao, S., and M. G. Rossmann.** 1973. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **76**:241–256.
36. **Raoult, D., et al.** 2004. The 1.2-megabase genome sequence of mimivirus. *Science* **306**:1344–1350.
37. **Ratel, D., J. L. Ravanat, F. Berger, and D. Wion.** 2006. N6-methyladenine: the other methylated base of DNA. *Bioessays* **28**:309–315.
38. **Rocap, G., et al.** 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**:1042–1047.
39. **Rodriguez, F., et al.** 2005. Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). *Environ. Microbiol.* **7**:853–859.
40. **Rose, M. G., M. P. Farrell, and J. C. Schmitz.** 2002. Thymidylate synthase: a critical target for cancer chemotherapy. *Clin. Colorectal Cancer* **1**:220–229.
41. **Rutherford, K., et al.** 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944–945.
42. **Sindelar, L.** 1986. The content of ATP, ADP, AMP, PI, the activity of enzymes involved in the glycolytic pathway and some problems of its regulation and energy balance in tobacco plants infected with potato virus Y. *Biol. Plant.* **28**:449–459.
43. **Sindelar, L., M. Sindelarova, and L. Burketova.** 1999. Changes in activity of glucose-6-phosphate and 6-phosphogluconate dehydrogenase isozymes upon potato virus Y infection in tobacco leaf tissues and protoplasts. *Plant Physiol. Biochem.* **37**:195–201.
44. **Six, C., et al.** 2008. Contrasting photoacclimation costs in ecotypes of the marine eukaryotic picoplankter *Ostreococcus*. *Limnol. Oceanogr.* **53**:255–265.
45. **Smith, M. A., A. K. Stobart, P. R. Shewry, and J. A. Napier.** 1998. Cytochrome b₅ and polyunsaturated fatty acid biosynthesis, p. 181–188. *In* P. R. Shewry, J. A. Napier, and P. J. Davis (ed.), *Engineering crop plants for industrial end uses*. Portland Press, London, United Kingdom.
46. **Teller, J. H., S. G. Powers, and E. E. Snell.** 1976. Ketopantoate hydroxymethyltransferase. I. Purification and role in pantothenate biosynthesis. *J. Biol. Chem.* **251**:3780–3785.
47. **Travers, A. A.** 1969. Bacteriophage sigma factor for RNA polymerase. *Nature* **223**:1107–1110.
48. **Vergeres, G., and L. Waskell.** 1995. Cytochrome b₅, its functions, structure and membrane topology. *Biochimie* **77**:604–620.
49. **Weigle, P. R., et al.** 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ. Microbiol.* **9**:1675–1695.
50. **Weynberg, K. D., M. J. Allen, K. Ashelford, D. J. Scanlan, and W. H. Wilson.** 2009. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environ. Microbiol.* **11**:2821–2839.
51. **Wilhelm, S. W., and C. A. Suttle.** 1999. Viruses and nutrient cycles in the sea—viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* **49**:781–788.
52. **Wood, T.** 1986. Distribution of the pentose phosphate pathway in living organisms. *Cell Biochem. Funct.* **4**:235–240.