# University of Nebraska - Lincoln DigitalCommons@University of Nebraska - Lincoln

Papers in Plant Pathology

Plant Pathology Department

4-1-2006

# Phycodnaviruses: A peek at genetic diversity

David Dunigan University of Nebraska-Lincoln, ddunigan2@unl.edu

Lisa A. Fitzgerald University of Nebraska-Lincoln, lisa.fitzgerald@nrl.navy.mil

James L. Van Etten University of Nebraska-Lincoln, jvanetten1@unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/plantpathpapers

Part of the Plant Pathology Commons

Dunigan, David; Fitzgerald, Lisa A.; and Van Etten, James L., "Phycodnaviruses: A peek at genetic diversity" (2006). *Papers in Plant Pathology*. 92.

https://digitalcommons.unl.edu/plantpathpapers/92

This Article is brought to you for free and open access by the Plant Pathology Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Papers in Plant Pathology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Published in Virus Research 117,:1 (April 2006), pp. 119-132: Comparative Genomics and Evolution of Complex Viruses. doi:10.1016/j.virusres.2006.01.024 Copyright © 2006 Elsevier B.V. Used by permission. <u>http://www.sciencedirect.com/science/journal/01681702</u>

Published online March 6, 2006.

# Phycodnaviruses: A peek at genetic diversity

David D. Dunigan<sup>a, c</sup>, Lisa A. Fitzgerald<sup>b</sup>, and James L. Van Etten<sup>a, c</sup>

<sup>a</sup>Department of Plant Pathology, University of Nebraska-Lincoln, Lincoln, NE 68583-0722 <sup>b</sup>Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE 68588-0304 <sup>c</sup>Nebraska Center for Virology, University of Nebraska-Lincoln, Lincoln, NE 68583-0722

Abstract: The family *Phycodnaviridae* encompasses a diverse collection of large icosahedral, dsDNA viruses infecting algae. These viruses have genomes ranging from 160 to 560 kb. The family consists of six genera based initially on host range and supported by sequence comparisons. The family is monophyletic with branches for each genus, but the phycodnaviruses have evolutionary roots that connect with several other families of large DNA viruses, referred to as the nucleocytoplasmic large DNA viruses (NCLDV). The genomes of members in three genera in the *Phycodnaviridae* have recently been sequenced and the purpose of this manuscript is to summarize these data. The viruses have diverse genome structures, some with large regions of non-coding sequence and others with regions of single-stranded DNA. Typically, phycodnaviruses have the coding capacity for hundreds of genes. The genera of the phycodnavirses sequenced to date. Thus, the gene diversity far exceeds the number of so-called "core" genes. Little is known about the replication of these viruses, but the consequences of these infections of the phytoplankton have global affects, including altered geochemical cycling and weather patterns.

Keywords: Algae, Genome, Phycodnavirus, Genetic diversity, Evolution, Nucleocytoplasmic large DNA virus

## **1. Introduction**

Members and prospective members of the family *Phy*codnaviridae constitute a genetically diverse, but morphologically similar, group of viruses with eukaryotic algal hosts from both fresh and marine waters. The family name derives from two distinguishing characteristics: (i) "phyco" from their algal hosts, and (ii) "dna" because all of these viruses have ds-DNA genomes (Wilson et al., 2005b). The phycodnaviruses are among the virioplankton recently recognized as important ecological elements in aqueous environments (Chen et al., 1996). More than 50% of the carbon dioxide fixed on the planet is by phytoplankton which includes cyanobacteria and eukaryotic microalgae. The phycodnaviruses, along with other viruses, play important roles in the dynamics of algal blooms, nutrient cycling, algal community structure, and possibly gene transfer between organisms.

The discovery phase of aquatic viruses, including the phycodnaviruses, is just beginning with new viruses continually being discovered as more environmental samples are examined. Ongoing metagenomic studies involving massive DNA sequencing indicate a greater viral diversity than could have been imagined just a few years ago (Hambly and Suttle, 2005 and Wommack and Colwell, 2000). The ease with which new viruses are found suggests that aquatic viruses may represent the greatest source of un-characterized genetic diversity on Earth (Hambly and Suttle, 2005). The genetic diversity that exists in the phycodnaviruses, albeit with only a few completed sequences, indicates that the limits of diversity are nowhere in sight. Our appreciation for this diversity is limited, in part, because phycodnaviruses have only been identified in about 0.1% of the ~40,000 known eukaryotic algal species (Guiry and Rindi, 2005). Paradoxically, it may turn out that both biological and genetic diversity are the unifying themes of the phycodnaviruses.

The phycodnaviruses are probably ancient. Accumulating genetic evidence indicates that the phycodnaviruses together with the poxviruses, iridoviruses, African swine fever virus (ASFV), and the recently discovered Mimivirus (Raoult et al., 2004) have a common evolutionary ancestor, perhaps arising at the point of eukaryogenesis, variously reported to be ca. 2.0–2.7 billion years ago (Han and Runnegar, 1992, Feng et al., 1997, Brocks et al., 1999 and Glansdorff, 2000). All of these viruses share nine gene products and 33 more gene

Genus <sup>a</sup>	Type species <sup>a</sup>	Known host range <sup>a</sup>	Source	Particle diameter (nm)	Genome size (kbp) and conformation	Latent period (h)	Burst size
Chlorovirus	Paramecium bursaria chlorella virus 1 (PBCV-1)	Chlorella NC64A Chlorella Pbi	FW	190	313–370 Closed linear dsDNA, hairpin termini	6–8	200-350 <sup>b</sup>
Coccolithovirus	Emiliania huxleyi virus 86 (EhV-86)	Emiliania huxleyi	MW	160-200	407–415 Circular	4-6	400-1000
Phaeovirus	Ectocarpus siliculousus virus 1 (EsV-1)	Phaeophyceae Ectocarpus siliculousus Ectocarpus fasciculatus Feldmannia simplex Feldmannia irregularis Feldmannia species Hincksia hincksiae Myriotrichia clavaeformis Pilayella littoralis	MW	130–200	160–340 Open linear, single stranded regions	ND	>1.10 <sup>6</sup>
Prasinovirus	Micromonas pusilla virus SP1 (MpV-SP1)	Micromonas pusilla Pyramimonas orientalis	MW	115-200	200–560	7–14	200-1000
Prymnesiovirus	Chrysochomulina brevifilum virus PW1 (CbV-PW1)	Haptophyceae (aka Prymnesiophyceae) Chrysochomulina brevifilum Chrysochomulina strobilus Chrysochomulina globosa	MW	120–160	485–510	12–19	400-4100
Raphidovirus	<i>Heterosigma akashiwo</i> virus 01 (HaV01)	Heterosigma akashiwo	MW	202	294	30-33	770

Table 1 Taxonomy and general characteristics of some phycodnaviruses

FW, fresh water; MW, marine/coastal water; ND, not determined.

<sup>a</sup> Data abstracted from http://www.ncbi.nlm.nih.gov/ICTVdb/Ictv/index.htm.

<sup>b</sup> PFU/cell.

products are present in at least two of these five viral families (Iyer et al., 2001 and Raoult et al., 2004). Collectively, these viruses are referred to as nucleocytoplasmic large DNA viruses (NCLDV) (Iyer et al., 2001).

The purpose of this review is to present the common and distinguishing characteristics of the *Phycodnaviridae* family. The genomes of a few members of the family have been sequenced and this information is contributing to our understanding of the relationships of virus members within the family, as well as with other large DNA viruses and cellular organisms. We have limited most of our discussion to those viruses whose genomes have been sequenced.

## 2. General characteristics and taxonomy

Phycodnaviruses are big (mean diameter of  $160 \pm 60$  nm) icosahedrons, that encapsidate large (160–560 kb) dsDNA genomes. Where known, the viruses have an internal membrane that is required for infection. Phylogenetic analyses of their  $\delta$ -DNA polymerases indicate that they are more closely related to each other than to other dsDNA viruses and that they form a monophyletic group, consistent with a common ancestor (Wilson et al., 2005b). However, the viruses fall into six clades which correlate with their hosts and each has been given genus status. Often the genera can be distinguished by additional properties, e.g., lytic versus lysogenic life styles or linear versus circular genomes (Wilson et al., 2005b). Members of the genus *Chlorovirus* infect fresh water algae; whereas, members of the other five genera (*Coccolithovirus*, *Phaeovirus*, *Prasinovirus*, *Prymnesiovirus*, and *Raphidovirus*) infect marine algae. The general characteristics of the six genera were reviewed recently (Brussaard, 2004), and Table 1 summarizes selected properties of genera. Viruses in three genera have been sequenced and are described below, followed by a comparison of their genes.

# 2.1. Chloroviruses

# 2.1.1. Biology

Chloroviruses are large, plaque-forming dsDNA viruses that infect certain, unicellular, chlorella-like green algae (see reviews, Kang et al., 2005, Van Etten, 2003, Van Etten et al., 2002 and Van Etten et al., 1991). The chlorella hosts are normally endosymbionts with protists including the protozoan *Paramecium bursaria* and the coelenterate *Hydra viridis*. When the symbiotic chlorella (also called zoochlorella) are in the paramecium, they are resistant to virus infection because the chlorella are enclosed in individual host derived vacuoles that exclude the viruses. Fortunately, some zoochlorella can be grown in the laboratory free of the paramecium. Chloroviruses are ubiquitous in nature and have been isolated from fresh water collected throughout the world. Typically, the virus titers in native waters are 1–100 plaque-forming units (PFU)/ml, but titers as high as 100,000 PFU/ml of native water have been obtained. Titers fluctuate with the seasons with the highest titers occurring in the spring.

Several hundred plaque-forming chloroviruses have been characterized to various degrees. They infect either *Chlorella* NC64A cells (NC64A viruses), an endosymbiont of *P. bursaria* isolated from North America, or *Chlorella* Pbi cells (Pbi viruses) that are endosymbiotic with a paramecium isolated in Europe. The most studied chlorella virus and the first phycodnavirus to be sequenced was *P. bursaria* chlorella virus 1 (PBCV-1) which infects *Chlorella* NC64A (Li et al., 1997).

### 2.1.2. Virion structure

Chlorella virus particles are large (molecular weight  $\sim 1 \times 10^9$  Da) and complex. The PBCV-1 virion contains more than 110 different virus-encoded proteins (Skrdla et al., 1984; Dunigan et al., manuscript in preparation). The PBCV-1 54kDa major capsid protein is a glycoprotein and comprises ~40% of the total virus protein. The major capsid protein consists of two 8-stranded, antiparallel β-barrel, jelly-roll domains related by pseudo six-fold rotation (Nandhagopal et al., 2002). This structure resembles the major coat proteins from some other dsDNA viruses which infect all three domains of life including bacteriophage PRD1, human adenoviruses, and a virus STIV infecting the Archaea, Sulfolobus solfataricus. This finding led to the suggestion that these three viruses may also have a common evolutionary ancestor with the NCLDVs, even though there is no significant amino acid sequence similarity among their proteins (Benson et al., 2004).

Cryo-electron microscopy and three-dimensional image reconstruction of the PBCV-1 virion indicate that the outer capsid is icosahedral and covers a lipid bilayered membrane (Yan et al., 2000). This membrane is required for infection because the virus loses infectivity after exposure to organic solvents. The outer diameter of the viral capsid ranges from 1650 Å along the two- and three-fold axes to 1900 Å along the fivefold axis. The capsid shell consists of 1680 donut-shaped trimeric capsomers plus 12 pentameric capsomers at each icosahedral vertex. The trimeric capsomers are arranged into 20 triangular facets (trisymmetrons, each containing 66 trimers) and 12 pentagonal facets (pentasymmetrons, each containing 30 trimers and one pentamer at the icosahedral vertices). Assuming all the trimeric capsomers are identical, the outer capsid of the virus contains 5040 copies of the major capsid protein. The virus has a triangulation number (T number) of 169 (note: PBCV-1 is not the largest phycodnavirus; Phaeocystis pouchetti virus (PpV01) has an icosahedral capsid with a triangulation number of 219 (Yan et al., 2005)).

Structural proteins of many viruses, such as herpesviruses, poxviruses, and paramyxoviruses, as well as the chlorella viruses, are glycosylated. Typically, viral proteins are glycosylated by host-encoded glycosyltransferases located in the endoplasmic reticulum (ER) and golgi and then transported to a host membrane (Olofsson and Hansen, 1998). Nascent viruses acquire the glycoprotein(s) and only become infectious by budding through the membrane, usually as they are released from the cell. Consequently, the glycan portion of virus glycoproteins is host specific.

However, glycosylation of PBCV-1 major capsid protein differs from this paradigm. Accumulating evidence indicates that PBCV-1 encodes most, if not all, of the enzymes involved in constructing the complex oligosaccharides attached to its major capsid protein and that the process occurs independently of the ER and golgi (Markine-Goriaynoff et al., 2004). Furthermore, five of six putative PBCV-1 encoded glycosyltransferases are predicted to be located in the cytoplasm.

### 2.1.3. Life cycle

PBCV-1 infects its host by attaching rapidly, specifically, and irreversibly to the external surface of the algal cell wall (Meints et al., 1984). Attachment always occurs at a virus vertex, possibly with hair-like appendages (Van Etten et al., 1991), and is followed by degradation of the host wall at the attachment point. The determinants for host range are associated with attachment. NC64A viruses do not attach to *Chlorella* Pbi cells and Pbi viruses do not attach to *Chlorella* NC64A cells. The PBCV-1 virion-associated protein A140/ 145R is conserved in the NC64A viruses and a homolog from the chlorovirus CVK2 interacts with the host cell wall (Onimatsu et al., 2004). Thus, the A140/145R protein and homologs may be responsible for facilitating virion attachment. No homologs of A140/145R exist in the other phycodnaviruses.

Electron micrographic studies reveal that the cell wall is degraded at the point of virus attachment (Meints et al., 1984). The chloroviruses are unique among the phycodnaviruses in that they are the only genera known to encode enzymes involved in polysaccharide degradation, which may be involved in cell wall digestion. Following host cell wall degradation, the internal membrane of the virus probably fuses with the host membrane resulting in entry of the viral DNA and virion-associated proteins into the cell, leaving an empty capsid on the surface. Infection results in rapid depolarization of the host membrane (Mehmel et al., 2003 and Frohns et al., 2006) and we hypothesize that this rapid depolarization is caused by a virus-encoded potassium ion channel (called Kcv) located in the internal membrane of the virions. Presumably, Kcv is activated when the virus membrane fuses with the host membrane and this depolarization may aid in the release of DNA into the cell and/or limit subsequent infection by additional viruses.

Circumstantial evidence indicates that the viral DNA and suspected DNA-associated proteins quickly move to the nucleus where early transcription is detected within 5–10 min p.i. (Schuster et al., 1986). However, experimental data have shown that within minutes of infection, host chromosomal DNA begins to be degraded, possibly due to virion-associated site-specific endonucleases (Agarkova, et al., manuscript in preparation). This degradation could aid in the recycling of nucleotides from the host DNA into newly replicating viral DNA, as well as inhibit host transcription.

In the immediate-early phase of infection, the host is reprogrammed to transcribe viral RNAs. Very little is known as to how this occurs, but chromatin remodeling may be involved. PBCV-1 encodes a 119 amino acid SET domain containing protein (referred to as vSET) that di-methylates Lys27 in histone 3 (Manzur et al., 2003). vSET is packaged in the PBCV-1 virion and accumulating evidence indicates that vSET could be involved in repression of host transcription following PBCV-1 infection (Manzur et al., unpublished results).

Viral DNA replication begins 60-90 min after infection and is followed by transcription of late virus genes (Schuster et al., 1986 and Van Etten et al., 1984). Ultrastructural studies of PBCV-1 infected chlorella suggest that the nuclear membrane remains intact, at least during the early stages of virus replication (Meints et al., 1986). However, a functional host nucleus is not required for virus replication since PBCV-1 can replicate, albeit poorly and with a small burst size, in UV-irradiated cells (Van Etten et al., 1986). Approximately 2-3 h p.i., assembly of virus capsids begins in localized regions in the cytoplasm, called virus assembly centers, which become prominent at 3-4 h p.i. (Meints et al., 1986). By 5 h p.i., the cytoplasm is filled with infectious progeny virus particles (~1000 particles/cell (Van Etten et al., 1983)) and by 6-8 h p.i. localized lysis of the host cell releases progeny. Of the progeny released, 25-50% of the particles are infectious, i.e., each infected cell yields ~350 PFUs (Van Etten et al., 1983). Intact infectious PBCV-1 particles accumulate inside the host 30-40 min before release. Other chlorella viruses have longer replication cycles than PBCV-1. For example, NC64A virus NY-2A requires approximately 18 h for replication and consequently forms smaller plaques.

## 2.1.4. PBCV-1 genome

The PBCV-1 genome is a linear 330 kb, nonpermuted ds-DNA with 35 nucleotide-long, partially base-paired, covalently closed hairpin termini (Girton and Van Etten, 1987 and Zhang et al., 1994). A 2221 bp inverted repeat region is adjacent to each termini (Strasser et al., 1991); however, the rest of the genome consists primarily of single copy DNA. PBCV-1 has ~700 open reading frames (ORFs) of 65 codons or larger, of which ~370 are probably protein encoding. The putative protein-encoding genes are evenly distributed on both strands and the intergenic space is minimal, 275 ORFs are separated by less than 100 nucleotides. One exception is a 1788-bp sequence near the middle of the genome. This region contains 11 tRNA genes. In this review, we use "ORF" to mean a predicted protein-coding gene, as well as the noted tRNA genes.

Approximately 50% of the 370 PBCV-1 gene products have been tentatively identified; they consist of both prokaryotic- and eukaryotic-like proteins, many of which have not been associated with viruses before. Eighty-four ORFs have paralogs within PBCV-1, forming 26 groups. The size of these groups ranges from two to six members. The PBCV-1 genome contains three types of introns: a self-splicing intron in a transcription factor TFIIS-like gene, a spliceosomal-processed intron in a DNA polymerase gene, and a small intron in one of the tRNA genes.

Three additional chlorella virus genomes have been sequenced recently. The largest, the 370 kb NY-2A virus, contains ~400 protein encoding genes (Fitzgerald et al., manuscript in preparation). Most common genes in viruses PBCV-1 and NY-2A, which infect the same host chlorella, are co-linear. However, there is almost no co-linearity between common genes in Pbi virus MT325 and those in PBCV-1 and NY-2A, suggesting a high degree of plasticity in these genomes.

One unusual feature of PBCV-1 DNA, as well as other chlorella virus DNAs, is that they contain methylated bases. Genomes from 37 chlorella viruses contain 5-methylcytosine (5 mC) in amounts varying from 0.12 to 47.5% of the total cytosines. Twenty-four of the 37 viral DNAs contain  $N^6$ -methyladenine (6 mA) ranging from 1.5 to 37% of the total adenines (Van Etten et al., 1991). This led to the discovery that many chlorella viruses encode multiple DNA methyltransferases, as well as site-specific endonucleases (Nelson et al., 1998).

## 2.1.5. PBCV-1 encoded proteins

Many PBCV-1 encoded enzymes are either the smallest or among the smallest proteins of their class. In addition, homologous genes in the chloroviruses can differ in nucleotide sequence by as much as 50%, which translates into amino acid differences of 30-40%. Therefore, comparative gene sequence analyses can identify conserved amino acids in proteins as well as regions that tolerate amino acid changes. The small sizes and the finding that many virus-encoded proteins are "user friendly" have resulted in the biochemical and structural characterization of several PBCV-1 enzymes. Examples include: (i) the smallest eukaryotic ATP-dependent DNA ligase (Ho et al., 1997), which is the subject of intensive mechanistic and structural studies ((Sriskanda and Shuman, 2002) and references cited therein). (ii) The smallest type II DNA topoisomerase (Lavrukhin et al., 2000). This PBCV-1 enzyme cleaves dsDNAs about 30 times faster than the human type II DNA topoisomerase (Fortune et al., 2001); consequently, the virus enzyme is being used as a model enzyme to study the topoisomerase II DNA cleavage process. (iii) An RNA guanylyltransferase (Ho et al., 1996) that was the first enzyme of its type to have its crystal structure resolved (Hakansson et al., 1997 and Hakansson and Wigley, 1998). (iv) A small prolyl-4-hydroxylase that converts Pro-containing peptides into hydroxyl-Pro-containing peptides in a sequence-specific fashion (Eriksson et al., 1999). (v) The smallest protein (94 amino acids) to form a functional K<sup>+</sup> channel (Plugge et al., 2000). These minimalist enzymes may represent precursors of contemporary proteins, but it is also possible that they are products of evolutionary optimization during viral evolution.

Some of the chloroviruses are also unusual because they encode enzymes involved in sugar metabolism. Three PBCV-1 encoded enzymes glutamine:fructose-6-phosphate amin-dotransferase (GFAT), UDP-glucose dehydrogenase (UDP-GlcDH), and hyaluronan synthase are involved in the synthesis of hyaluronan, a linear polysaccharide composed of alternating  $\beta$ -1,4-glucuronic acid and  $\beta$ -1,3-*N*-acetylglucosamine res-

idues (DeAngelis et al., 1997 and Landstein et al., 1998). All three genes are transcribed early in PBCV-1 infection and hyaluronan accumulates on the external surface of the infected chlorella cells (Graves et al., 1999). The predicted amino acid sequences of the PBCV-1-encoded GFAT and UDP-GlcDH enzymes most closely resemble bacterial enzymes, whereas the amino acid sequence of the PBCV-1-encoded hyaluronan synthase has the strongest resemblance to vertebrate enzymes. These observations suggest that the viruses may have acquired the GFAT and UDP-GlcDH genes separately from the hyaluronan synthase gene.

Two PBCV-1-encoded enzymes, GDP-D-mannose dehydratase and fucose synthase, comprise a three step pathway that converts GDP-d-mannose to GDP-l-fucose (Tonetti et al., 2003). The function of this putative pathway is unknown. However, fucose, a rare sugar, is present in the glycans attached to the major capsid protein.

PBCV-1 encodes four enzymes involved in polyamine biosynthesis: ornithine decarboxylase (ODC), homospermidine synthase, agmatine iminohydrolase, and *N*-carbamoylputrescine amidohydrolase (Van Etten, 2003). Ornithine decarboxylase catalyzes the decarboxylation of ornithine to putrescine, which is the first and the rate limiting enzymatic step in the polyamine biosynthetic pathway. The PBCV-1-encoded ODC is the smallest characterized ODC (Morehead et al., 2002). The PBCV-1 enzyme is also interesting because it decarboxylates arginine better than ornithine (Shah et al., 2004).

## 2.1.6. Diversity of Chlorovirus genomes

Not all PBCV-1 genes are required for virus replication in the laboratory. For example, four spontaneously derived PBCV-1 mutants were isolated that contain 27- to 37-kb deletions at the left end of the 330-kb genome (Landstein et al., 1995). Two of these mutants have deletions beginning at nucleotide coordinates 4.9 or 16 kb and ending at 42 kb. In total, these two deleted regions, which probably resulted from recombination, encode 28 putative proteins. The other two mutants, which probably arose from nonhomologous recombination, lack the entire left terminal 37-kb of the PBCV-1 genome, including the 2.2 kb terminal inverted-repeat region. The deleted left terminus was replaced by the transposition of an inverted 7.7- or 18.5-kb copy from the right end of the PBCV-1 genome. These regions encode 26 single-copy ORFs, of which 23 are common to those deleted in the first two mutant viruses. Taken together, ~40 kb of single-copy DNA encoding 31 ORFs at the left end of the genome, or 12% of the PBCV-1 genome, is unnecessary for PBCV-1 replication in the laboratory. However, replication of the PBCV-1 deletion mutants is attenuated, i.e., their burst sizes are about half of wildtype virus.

The deletion mutants also indicate that the size of the inverted terminal repeats in PBCV-1 can vary. This conclusion is consistent with the finding that the size and sequence of the inverted repeat region is not conserved among chlorella viruses (Strasser et al., 1991 and Yamada and Higashiyama, 1993). This lack of conservation is somewhat surprising because one predicts that the DNA termini might be essential for either virus DNA replication and/or DNA packaging. These results also indicate that the virus DNA packaging machinery tolerates significant differences in genome size, e.g., the largest deletion in PBCV-1 creates a genome of ~302 kb whereas, chlorella virus NY-2A has a genome of ~370 kb. Similar large deletions occurred in the left terminus of the chlorella virus CVK1 genome (Songsri et al., 1997).

Large insertions also occur in the chlorella viruses. Comparison of PBCV-1 and CVK2 genomes revealed that an ~15 kb region in the PBCV-1 left terminal region is absent in the CVK2 genome. However, CVK2 contains a 22.2 kb insert in this region that contains five gene copies of a homolog of PBCV-1 glycoprotein Vp260; this 22.2 kb sequence is absent in the PBCV-1 genome (Chuchird et al., 2002 and Nishida et al., 1999b).

The sizes and locations of the deletions and transpositions in the chlorella viruses resemble poxviruses (Turner and Moyer, 1990) and ASFV (Blasco et al., 1989) deletion mutants. Like PBCV-1, poxviruses and ASFV genomes have inverted terminal repeats and covalently closed hairpin ends. Models to explain the generation of deletions and deletion/ transpositions in the poxvirus genomes (Shchelkunov and Totmenin, 1995 and Turner and Moyer, 1990) may be relevant to the chlorella viruses.

Several other observations reflect the diversity of chlorella virus genomes. (i) The NC64A virus NY-2A genome is about 40 kb larger than the PBCV-1 330-kb genome and contains ~30 more genes (Fitzgerald et al., manuscript in preparation). Some of these additional NY-2A genes encode ubiquitin, chitin synthase, N-acetylglucosaminyl transferase, 6 transposases, and 28 homing endonucleases. Furthermore, inteins exist in two of the NY-2A gene products, the  $\alpha$ -subunit of ribonucleotide reductase and a putative helicase. (ii) Southern hybridization and DNA sequence analyses indicate that not all PBCV-1 genes exist in all of the NC64A virus isolates. Also an extra ORF is often inserted between co-linear genes. The insertion of extra genetic elements, referred to as "morons" (for more DNA), between adjacent genes also occurs in related lambda phages (Hendrix et al., 2000). (iii) PBCV-1 carries a polycistron of 11 tRNAs genes with less than 33 intergenic nucleotides; other NC64A viruses carry polycistronic tRNAs of up to 16 tRNA genes (Cho et al., 2002 and Nishida et al., 1999a). Clustering of tRNAs is common in bacteria and rare in eukaryotes. Some of the chlorella virus tRNA genes probably reflect gene duplications. Also some tRNA genes contain introns as large as 1 kb (Nishida et al., 1999a). (iv) Sequence analyses of a gene (pdg) encoding a UV-specific DNA repair enzyme from 42 NC64A viruses revealed that 15 of them contain a 98nucleotide spliceosomal-processed intron that is 100% conserved; four other viruses contain an identically positioned 81-nucleotide intron that is nearly 100% identical (Sun et al., 2000). In contrast, the nucleotides in the pdg coding regions (exons) from the intron-containing viruses are 84-100% identical. The 100% identity of the 98-nucleotide intron sequence in 15 viruses and the near 100% identity of an 81-nucleotide intron sequence in another four viruses imply that either the intron was acquired recently or that there is strong selective pressure to maintain the DNA sequence of the intron once it is in the pdg gene. However, the abilities of intron-containing and intron-lacking viruses to repair UV-damaged DNA in the dark were indistinguishable (Sun et al., 2000). These findings contradict the dogma that intron sequences are more variable than exon sequences. (v) Yamada and his colleagues (Nishida et al., 1998 and Yamada et al., 1994) reported that 8% of the NC64A viruses isolated in Japan contain a self-splicing group I intron. This intron is inserted in the gene encoding either transcriptional elongation factor TFIIS (~60% of the viruses) or an unidentified ORF encoding a 14.2 kDa polypeptide (~40% of the viruses); however, in a few viruses the intron is in the major capsid protein gene. Yamada et al. (1994) suggested that the self-splicing intron might function as a mobile genetic element, e.g., transferring genes between algae and protists.

To summarize, considerable variation occurs in the chlorovirus genomes and the total number of genes in the chlorovirus gene pool exceeds that of a single isolate. The different sizes of the chlorovirus genomes as well as the large deletions and insertions, suggest that dynamic and frequent rearrangements of virus genomes occur in natural environments. The fact that the left end of the chlorella virus genome is tolerant to deletions/insertions/rearrangements suggests that a recombinational "hotspot" in this region allows viruses to exchange genes among themselves and possibly with their host(s).

# 2.2. Coccolithoviruses

## 2.2.1. Biology

The marine coccolithophorid *Emiliania huxleyi* is a unicellular alga found throughout the world. It is best known for its immense coastal and mid-oceanic blooms at temperate latitudes which cover 10,000 km<sup>2</sup> or more. The size and intensity of these blooms make *E. huxleyi* important in nutrient and  $CO_2$  cycling, as well as biogenic sulfur production (in the form of dimethylsulphide (DMS) (Malin et al., 1994)). Consequently, *E. huxleyi* is a key species in global biogeochemical cycles and climate modeling (e.g., Charlson et al., 1987, Westbroek et al., 1993 and Westbroek et al., 1994).

*E. huxleyi* blooms are visible with low-orbit satellite imagery due to the reflectance of calcium carbonate scales, called coccoliths, which cover the algal surface. During a bloom, the alga constantly produces and sheds coccoliths into the surrounding water. However, the bloom is most visible in satellite images when the *E. huxleyi* cells die and shed their remaining coccoliths, ultimately resulting in a large movement of coccoliths to the sea-floor. This process represents one of the largest long-term sinks of inorganic carbon on Earth. One of the primary mechanisms for terminating *E. huxleyi* blooms is viral lysis (Bratbak et al., 1993, Bratbak et al., 1996 and Wilson et al., 2002). The fact that the coccolithoviruses play a major role in the demise of these blooms has only recently been appreciated.

## 2.2.2. Virion structure

Morphologically, the coccolithoviruses have an icosahedral capsid ranging from 160 to 200 nm that encapsidates a single copy of circular dsDNA. SDS-PAGE of sucrose gradient purified virions reveals at least nine proteins with a 54 kDa protein being the major capsid protein (Castberg et al., 2002). Given the high number of proteins detected in chlorovirus virions and that the coccolithoviruses are predicted to replicate in the cytoplasm (see below), the coccolithovirus virions probably have many more proteins.

## 2.2.3. Life cycle

Very little is known about the replication cycle of these viruses. Infection is host specific, such that the prototype virus EhV-86 infects only *E. huxleyi* strains that are low producers of DMS (Schroeder et al., 2002). Electron micrographs of virus-infected cells suggest that virus capsids remain on the cell surface and may bind to the calcium carbonate coccoliths. However, this results in a distance of several nanometers between the cell surface and the bound particle. Another possibility is that the virus infects *E. huxleyi* cells that are shedding coccoliths.

One property that clearly distinguishes EhV-86 from other phycodnaviruses is that it encodes six RNA polymerase subunits; in contrast, neither PBCV-1 nor EsV-1 encodes a recognizable RNA polymerase component. Thus, like the poxviruses, EhV-86 may carry out its entire life cycle in the cytoplasm.

The latent period for EhV-86 is 4–6 h (William Wilson, personal communication) with a burst size of 400–1000 particles per cell (Castberg et al., 2002). Though EhV-86 lacks genes encoding cell wall degrading enzymes, nascent coccolithoviruses may be released from their host by utilizing cellular degradative pathways, e.g., those stimulated by the apoptotic-promoting function of the ceramide synthesis pathway (Wilson et al., 2005a).

## 2.2.4. EhV-86 genome

The circular EhV-86 407 kb genome is predicted to have 472 protein-encoding genes (Wilson et al., 2005a). Only 66 (14%) of these genes match a sequence in GenBank. The virus encodes several unexpected genes never found in a virus before, including four genes involved in sphingolipid biosynthesis (sterol desaturase, serine palmitoyltransferase, transmembrane fatty acid elongation protein, lipid phosphate phosphatase) and two additional genes which encode desaturases. EhV-86 also has eight genes encoding proteases.

The EhV-86 genome is 40% G + C and has three distinct repeat families of non-coding and coding sequences (Allen et al., 2005). One family is predicted to contribute to gene regulation, another is thought to involve calcium binding that may contribute to virus adsorption and/or release, and the third family may be involved in the origin of DNA replication.

# 2.3. Phaeoviruses

# 2.3.1. Biology

Marine filamentous brown algae in the class Phaeophyceae, including *Ectocarpus* species, *Feldmannia* species, *Hincksia hincksiae*, *Myriotrichia claveformis*, and *Pilayella littoralis*, are often infected with viruses from the *Phaeovirus* genus (Müller et al., 1998). The *Ectocarpus siliculosus* virus (EsV-1) is the type member of this genus. EsV-1 is pandemic in *E. siliculosus* populations on the coasts of all oceans in temperate climate zones (Müller et al., 2000 and Sengco et al., 1996), with viral infections also detected in subarctic and subantarctic coasts (Müller and Stache, 1992). The pathology associated with the phaeoviruses appears to vary. For example, EsV infection of *E. siliculosus* has no detrimental impact on photosynthesis or growth of the alga (del Campo et al., 1997). However, photosynthesis in a virus infected *Feldmannia* species is reduced significantly, implying that virus infection contributes to the stunted growth observed in these virus-infected brown algae (Robledo et al., 1994).

Phaeoviruses are the only viruses in the family Phycodnaviridae reported to infect more than one species, e.g., EsV infects two additional algae. When EsV infects Feldmannia simplex, it initially induces pathological symptoms, which later subside. Because EsV is not able to produce additional virus particles in this host, this may be a dead-end for this virus (Müller et al., 1996a and Müller and Parodi, 1993). However, EsV is able to infect Kuckuckia kylinii and complete an infection cycle. The nascent EsV particles produced in K. kylinii still infect its original host, E. siliculosus (Müller, 1992). Ectocarpus fasciculatus virus (EfasV) is another example of a phaeovirus that is able to infect more than one host. EfasV infects E. siliculosus (Müller et al., 1996b) and Myriotrichia clavaeformis (Maier et al., 1997). In both cases, there are morphological changes associated with infection, however, no progeny virions are produced (Müller et al., 1998). The ability of a virus to infect more than one host provides a conduit for transferring genes between different algal species.

### 2.3.2. Virion structure

Phaeoviruses are icosahedrons with diameters ranging from 120 nm for *F. simplex* virus 1 (FlexV-1) (Friess-Klebl et al., 1994) to 180 nm for *M. clavaeformis* virus 1 (MclaV-1) (Kapp et al., 1997). The particles have one or two electron dense layers surrounding an electron dense core which may represent a lipid membrane (Kapp et al., 1997). The internal membrane appears to be derived from the endoplasmic reticulum as capsids bud from the cisternae (Wolf et al., 1998 and Wolf et al., 2000). SDS-PAGE of proteins from four phaeoviruses revealed significant differences (Kapp et al., 1997 and Klein et al., 1995). However, one common feature in viruses which infect *F. simplex* or *E. siliculosus*, is the presence of glycoproteins (Friess-Klebl et al., 1994, Klein et al., 1995 and Lanka et al., 1993).

# 2.3.3. Life cycle

EsV initiates its life cycle by infecting free-swimming, wallless gametes of its host. Virus particles enter the cell by fusion with the host plasma membrane and release a nucleoprotein core particle into the cytoplasm, leaving remnants of the capsid at the surface. The viral core moves to the nucleus within 5 min p.i. (Maier et al., 2002). One important feature that distinguishes the EsV-1 life cycle from the other phycodnaviruses is that the viral DNA is integrated into the host genome and is transmitted mitotically to all cells of the developing alga (Brautigam et al., 1995, Delaroque et al., 1999 and Müller, 1991b). Not surprisingly, EsV-1 encodes a putative integrase. The viral genome remains latent in vegetative cells until it is expressed in the algal reproductive cells, the sporangia or gametangia. Massive replication of viral DNA occurs in the nuclei of these reproductive cells, followed by nuclear breakdown and viral assembly that continues until the cell becomes densely packed with virus particles (Lanka et al., 1993, Müller et al., 1998, Wolf et al., 1998 and Wolf et al., 2000). Virus release is stimulated by the same factors that induce discharge of gametes from the host, i.e., changes in temperature, light, and water composition (Müller, 1991a). This synchronization facilitates interaction of viruses with their susceptible host cells.

# 2.3.4. EsV-1 genome

The EsV-1 sequence was reported in 2001, the largest virus genome to be sequenced at the time. The 335 kb genome has a 52% G + C content with low levels of methylated cytosines (1% 5 mC) and adenines (3% 6 mA) (Delaroque et al., 2001 and Lanka et al., 1993). Tandem repeats are located in several regions of the genome, which make up  $\sim 12\%$  of the genome (Delaroque et al., 2001). In addition to these repeats, the EsV-1 genome has the unusual property of containing several single stranded DNA regions of various lengths (Klein et al., 1994 and Lanka et al., 1993). Studies on the structure of the genome have produced an enigma. Pulsed field electrophoresis, DNA restriction mapping, and electron microscopy experiments indicate that the EsV-1 genome is a circular molecule (Kapp, 1998, Lanka et al., 1993 and Müller et al., 1998). However, the DNA sequencing results produced one 335-kb linear contig with termini containing almost perfect inverted repeats of 1.8 and 1.6 kb (Delaroque et al., 2001). Attempts to link the termini (i.e., to complete a circle) by PCR using primers corresponding to sequences within the inverted repeat regions were unsuccessful. Thus, the inverted repeats apparently mark the ends of a linear molecule. These conflicting results can be resolved if the complementary inverted terminal repeats anneal with each other to form a cruciform structure that effectively closes the DNA circle (Delaroque et al., 2001).

Analysis of the genome revealed 231 major ORFs of 65 codons or larger. Forty-eight percent of the major ORFs had sequence similarities to another protein in the public databases. EsV-1 differs from other algal viruses by encoding six hybrid His-protein kinases, including one that is located in the virion internal membrane (Delaroque et al., 2001). Hybrid His-protein kinases are members of a large family of two component signaling systems that serve as stimulus-response coupling mechanisms in many organisms (Stock et al., 2000). Due to the life style of EsV-1 it is not surprising that the genome also encodes proteins involved in integration/transposition (Delaroque et al., 2001). In addition to these predicted proteins, EsV-1 also encodes several proteins involved in replication, nucleotide metabolism, transcription, and sugar/lipid metabolism (Fig. 1).

## 2.3.5. Diversity of Phaeovirus genomes

Among the phycodnaviruses, the phaeovirus genomes exhibit the greatest range in size. FsV-1 is another well-studied phaeovirus which infects an unclassified *Feldmannia* species. Although the FsV-1 180-kb genome sequence is incomplete, enough genomic fragments have been sequenced to allow comparisons with the 335 kb EsV-1 genome: (i) like EsV-1, FsV-1 contains numerous repeat regions, however, the sequences of these repeat regions differ from EsV-1. (ii) The similarity between FsV-1 and the EsV-1 genes is low. For ex-

Genus	Virus	Host	Genome size (bp)	% G+C content	ORFs	tRNAs	MCP (aa)
Chlorovirus	PBCV-1	Chlorella NC64A	330,743	40.0	366	11	437
	NY-2A	Chlorella NC64A	368,683	40.7	394	7	437
	AR158	Chlorella NC64A	344,690	40.8	357	6	437
	MT325	Chlorella Pbi	314,335	45.3	333	10	433
Phaeovirus	EsV-1	Ectocarpus siliculousus	335,593	51.7	231	0	476
	FirrV-1 <sup>a</sup>	Feldmannia irregularis	191,667	ND	ND	ND	435
Coccolithovirus	EhV-86	Emiliania huxleyi	407,339	40.2	472	5	533

Table 2 Genome data of sequenced phycodnaviruses

ND, not determined.

<sup>a</sup> Genomic sequencing and annotation incomplete.

ample, the EsV-1 DNA polymerase only has 45% amino acid identity to the FsV-1 enzyme (Delaroque et al., 2001).

Recently, another phaeovirus genome was sequenced. The FirrV-1 (endemic to Feldmannia irregularis) genome consists of a 180 kb linear dsDNA plus smaller fragments of 10-170 kb which are probably degradation products of the 180 kb fragment (Delaroque et al., 2003). Although the FirrV-1 genome is approximately half the size of the EsV-1 genome, there is some sequence similarity between these viruses, including genes involved in nucleic acid metabolism, sugar metabolism, integration, and transduction. EsV-1 has several genes with redundant functions; whereas, FirrV-1 has very few redundant genes. One-third of the FirrV-1 genes have no orthologs in the EsV-1 genome. EsV-1 has 75 more genes than FirrV-1. These related viruses may have evolved from a more complex common ancestor via loss-of-genes and recombination mechanisms, resulting in highly divergent genome structures (Delaroque et al., 2003).

### 3. Comparison of EhV-86, EsV-1, and PBCV-1 genomes

PBCV-1, EsV-1, and EhV-86 genomes each have distinguishing structural and functional characteristics (Table 2). One common characteristic is their large dsDNA genomes. The PBCV-1 genome is a linear molecule with hairpin ends, EsV-1 has a linear genome with apparent sticky ends, and the EhV-86 genome is circular. PBCV-1 and EhV-86 replicate autonomously as episomes, while EsV-1 integrates into a host chromosome and reactivates its genome in response to developmental cues.

## [Table 2.]

Common features exist between two of the three phycodnaviruses. For example (i) EsV-1 and PBCV-1 genome sizes are similar (330–335 kb) whereas EhV-86 has a larger 407 kb genome. (ii) PBCV-1 and EsV-1 have linear genomes containing inverted terminal repeats. The PBCV-1 inverted repeat regions (2.2 kb) are 100% identical; whereas, the EsV-1 terminal repeats consist of short non-identical regions. (iii) Methylated bases are common in chlorella virus genomes; whereas, EsV-1 DNA contains a low level of methylated bases (Lanka et al., 1993). The methylation status of the EhV-86 genome is unknown.

There are three common features between EhV-86 and PBCV-1. Both EhV-86 and PBCV-1 encode tRNAs; whereas, EsV-1 does not. The 11 PBCV-1 tRNAs are clustered together

centrally within the genome. EhV-86 encodes five tRNAs; four of them are located in a 4 kb region within the first 10% of the genome and one is located in the middle of the genome. Second, the 40% G + C content of the PBCV-1 and EhV-86 genomes, which is lower than the 52% G + C content of the EsV-1 genome. Third, the PBCV-1 and EhV-86 viruses contain introns. The PBCV-1 genome contains three types of introns, EhV-86 has two introns, and EsV-1 appears to lack introns.

There is one common feature in the two marine algal viruses, EsV-1 and EhV-86. Large regions of repeated sequences occur in both genomes and are in addition to the EsV-1 terminal repeats. The EsV-1 repeat regions occupy  $\sim$ 12% of the genome. An additional 22% of the EsV-1 genome apparently lacks coding potential; thus, only  $\sim$ 67% of the EsV-1 genome contains protein encoding genes. As mentioned above, the EhV-86 genome contains three families of distinctly different repeat sequences that are found throughout its genome (Allen et al., 2005).

## 3.1. Transposition and horizontal gene flow

One unexpected outcome of the sequencing projects was the discovery of several virus encoded gene products which could contribute to DNA rearrangements and/or transposition. PBCV-1 encodes two proteins that resemble bacterial transposases. In addition, PBCV-1 encodes 12 ORFs that contain motifs found in homing endonucleases (Kowalski et al., 1999). As mentioned previously, the recently sequenced chlorovirus NY-2A encodes six putative transposases and 28 homing endonucleases (Fitzgerald et al., manuscript in preparation). EsV-1 encodes two ORFs that resemble bacterial transposases and two ORFs with a homing endonuclease motif, while EhV-86 encodes one ORF that resembles a homing endonuclease.

The EsV-1 genome has three large dispersed repeat regions, termed R1, R2, and R3, that contain ORFs which currently are not considered to be genes because they lack upstream AT-rich sequence elements. These ORFs do not match any proteins in the databases. Interestingly, the R1 and R2 repeats are located downstream of genes encoding putative bacterial IS4 family transposases (IS, insertion sequence; (Rezsohazy et al., 1993)). The R1- and R2-transposase units are bracketed by imperfect inverted repeats of 20 bp that resemble the ends of bacterial IS4 elements. Furthermore, R1 and R2 frame a genomic ~29 kb segment forming a transposonlike structure. Like bacterial transposons, the EsV-1 transposon could confer advantages to the host. In summary, the phycodnaviruses encode several proteins that could be involved in DNA rearrangements either within or between virus genomes. These enzymes could contribute to the plasticity that appears to exist in the virus genomes. In the laboratory, the chlorella virus genomes are quite stable with few spontaneous mutants observed. It should be noted, however, that none of the putative phycodnavirus transposases and homing endonucleases have been tested for function.

## 4. Comparison of gene content

Collectively, the PBCV-1, EsV-1, and EhV-86 genomes have in excess of 1000 unique ORFs based on sequence similarity to public databases (Delaroque et al., 2001, Li et al., 1997 and Wilson et al., 2005a). Many of these ORFs have no significant similarity to any of the database entries and therefore, are classified as unknowns. Interestingly, we find there are only 14 ORFs that are homologous between these three viral genomes (note: using other methods Iyer et al. (2006) have identified a greater number of common genes in the phycodnaviruses, which include the set we present here). One hundred twenty-three putative ORFs are arranged by their metabolic domains in Fig. 1. Some ORFs encode for the same function; however, their evolution and origin are unclear. For example, replication factor C (large subunit) is found in both PBCV-1 and EsV-1, yet the sequence similarity falls below the threshold of E = 0.001 in a blastp search analysis. Nevertheless, Fig. 1 represents functional grouping, and in most cases the relationships are homologous. The similarities between the different genomic ORFs were determined by using blastp (E < 0.001, blosum62 matrix) (Altschul et al., 1990).

There are 12 (~10%) ORFs of common functionality encoded by all three phycodnaviruses, of which 10 are homologous between these three viruses. It is not surprising that several of these 12 ORFs are involved in DNA replication, such as δ-DNA polymerase, large and small subunits of ribonucleotide reductase, PCNA, superfamily II and III helicases, and the newly recognized archaeo-eukaryotic primases (Iyer et al., 2005). In addition to the 10 homologs that have been assigned a putative function, there are three homologous ORFs that are not found in public databases, suggesting these genes are unique to the phycodnaviruses. These unknown homologs have amino acid sequence identities of 19-33%. We assume these homologs are involved in important processes associated with the phycodnavirus life cycles otherwise they would have been eliminated. Another common ORF among the three phycodnaviruses is the major capsid protein, which has amino acid sequence identities of 26-30%. Overall, the amino acid identity between the common protein homologs ranges from 19 to 53%.

PBCV-1 and EsV-1 have 28 homologs that are absent in EhV-86. Seven are listed in Fig. 1 and the remaining 21 have unknown functions. The average amino acid identity between the unknown ORFs is 30% with a range of 22–37%. PBCV-1 and EhV-86 have 10 homologs that are absent in EsV-1. Interestingly, most of the PBCV-1 homologs found in EsV-1 are unknown; whereas, only one of the 10 PBCV-1 homologs in EhV-86 is unknown.

There is no co-linearity of the common genes in the three viruses. However, this is not surprising because very little co-

linearity exists between two viruses that infect different isolates of chlorella (Fitzgerald et al., manuscript in preparation).

# 5. Phycodnavirus genes are probably very old

As mentioned in Section 1, accumulating evidence indicates that at least some of the phycodnavirus genes, and by implication the viruses themselves, have a long evolutionary history, possibly dating back to the time that prokaryotic and eukaryotic organisms separated, ca. 2.0-2.7 billion years ago. This evidence includes: (i) phylogenetic analysis of  $\delta$ -DNA polymerases place the phycodnaviruses near the origin of all the eukaryotic δ-DNA polymerases (Villarreal and DeFilippis, 2000). (ii) Phylogenetic analyses of several other PBCV-1 gene products place the virus gene products near the base of the trees. These products include ornithine decarboxylase, potassium ion channel protein, and GDP-d-mannose-4,6 dehydratase. (iii) Many of the PBCV-1 encoded proteins, including histone H3-lysine<sup>27</sup> dimethylase, ornithine decarboxylase, DNA topoisomerase type II, and the potassium ion channel protein are among the smallest of their type and are possibly the ancestor of their larger relatives. Traditionally, one thinks that viruses acquire genes from their host and that these gene products become less complex with time. However, the reverse seems possible, i.e., the precursor proteins were simpler and that during evolution the proteins added domains to fill their specialized roles in more complex organisms (e.g., gaining regulatory capabilities). (iv) Some PBCV-1-encoded enzymes are more flexible than those from higher eukaryotic organisms. For example, some virus enzymes carry out two functions whereas more 'advanced' organisms require two separate enzymes to accomplish the same tasks. One interpretation is that these virus proteins may be progenitor enzymes and thus more precocious than their highly evolved homologs in eukaryotes, where two separate enzymes carry out the function of one PBCV-1 enzyme. This dual functionality in the PBCV-1 enzymes does not result from gene fusion. Examples include: (a) ornithine decarboxylase which decarboxylates arginine more efficiently than ornithine (Shah et al., 2004). (b) dCMP deaminase which also deaminates dCTP and dCDP, as well as the expected dCMP (Zhang et al., manuscript in preparation). The catalytic site of the PBCV-1 dCMP deaminase appears to be more flexible than the enzyme from other organisms (note: the Km of the virus enzyme is similar for dCMP and dCTP but it is higher than dCMP deaminases from other organisms). (c) GDP-d-mannose 4,6 dehydratase not only catalyzes the formation of GDP-4-keto-6-deoxy-dmannose, which is an intermediate in the synthesis of GDP-1-fucose, the enzyme can also reduce the same intermediate to GDP-d-rhamnose (Tonetti et al., 2003). We expect that other phycodnavirus enzymes will have dual functions, consistent with their presumed ancient origin. (v) Indeed, phycodnavirus genomes are mosaics of prokaryotic and eukaryotic genes. The phycodnaviruses encode an amazing variety of proteins, some of which are prokaryotic-like, e.g., the DNA methyltransferases and restriction endonucleases. Other proteins are very eukaryotic-like, e.g., the mRNA capping enzymes. The genes encoding these proteins could have been acquired from

DNA Replication, Recombinat	ion and Repa	air		DNA Res	triction/Modification	
Description	PBCV-1	EsV-1	EhV-86	Description	PBCV-1	EsV-1 EhV-f
Archaeo-eukarvotic primase				Adenine DNA methylase	2*	
ATPasal		2		Cytosine DNA methylase	3*	
DNA ligner	- C. C		7777	DNA restriction and anuclease	2+4	
DNA ngase				DNA restriction endonuclease	2.04	
DNA topoisomerase II	The second secon	-				
DNA-binding protein	10		1111		ranscription	
Exonuclease <sup>b</sup>	*	2		Description	PBCV-1	EsV-1 EhV-8
Helicase			2	DNA-dependent RNA polymerase	II largest	
Helicase-Rec BCD-like				DNA-directed RNA polymerase II	subunit	
Helicase-Superfamily III				DNA-directed RNA polymerase sul	bunit	3
Nuclaia agid binding protain	_			Uistona U2 Lug 27 mathulaga	**	
DCNAs	10000			misione ris, Lys 27 methylase		
PUNA	444		4	mRNA guanyiyitransterase	~	
Pyrimidine dimer-specific glycosylase	×			Oligoribonuclease		
Replication factor C (lg subunit)				RNA triphosphatase	*	
Replication factor C (sm subunit)		4		RNase III	*	
RNaseH				Superfamily II helicaseg	003-00	
δ DNA polymerase				SWI/SNF helicase	4	
o britt polymenase				Transcription factor TEIIB		
Destain Conthesis Medifection	and Deserved			Transcription factor TFIID		
Protein Synthesis, Modification,	and Degrada	tion	1	Transcription factor TFIID		
Description	PBCV-1	EsV-1	EhV-86	Transcription factor TFIIS		
ATPase (AAA+ class)				VLTF2-type transcription factor		
ATP-dependent protease proteolytic subunit						a 13
Esterase				Cell	Wall Degradation	
Hydrolase				Description	PBCV-1	EsV-1 EbV-f
Prolyl 4-hydroxylasa	4			Chitinase	24	L31-1 Env-0
Protocol Control of Co	-			Chinase	201	
Protease-Cysteine	-			Chitosanase	**	
Protease-OTU like cysteine			2	β&α1,4 linked glucuronic lyase	*	
Protease-Serine			5	β-1,3-glucanase	*	
Protein disulfide isomerase	4					
SKP1 protein				Nucle	otide Metabolism	
Thial axidaraduataca	4			Description	PRCV 1	E-VI E-VI
Timot oxidoreductase	T			Description	FBCV-1	ESV-I ERV-0
Translation elongation factor-3		-		Aspartate transcarbamylase		
Ubiquitin C-terminal hydrolase				Cytosine deaminase		
Zn metallopeptidase	and the second second			dCMP deaminase	*	
tRNAs <sup>d</sup>	11.		5	Deoxynucleoside kinase		
				dUTP pyrophosphatase	*	
Integration and Trans	osition			Glutaredoxin	*	
Description	PRCV 1	E-V.1	ELV OC	NTD purperhambahudralasa		
Description	PBUV-1	ESV-1	Env-80	NTP pyrophosphonydrofase		
Anti-repressor of lysogeny	-	_		Nucleic acid independent nucleosid	e	
Endonuclease	_		3	Ribo, reductase (large subunit)		
Homing endonuclease GIY-YIG	7	1.1.1.1.1.1.1.1.1	A CONTRACTOR OF	Ribo, reductase (small subunit)		
Homing endonuclease HNHe	00600	2.00		Ribonuclease		
Integrase	1000		2.0.0.0.1	Thermonuclease		
Protelomense				Thioredoxin		2
Paperson of lungagery	-			Thumidulata kinaca		
Repressor of tysogeny				Invmidylate kinase		
Transposase			1	Thymidylate synthase X	*	
				Thymidylate synthase-bifunct, dihy	drofolate	
Sugar & Lipid Manip	ulation					
Description	PBCV-1	EsV-1	EhV-86	N	Aiscellaneous	
Alginate mannuronan epimerase	1			Description	PBCV-1	EsV-1 EhV-8
D-lactate dehydrogenase				ABC transporter protein	4	
Eatta A aid deseturase	-		2	A amoting iminghudrologg	*	
Fatty Acid desaturase	-	_	- 4	Agmatine iminonydrofase		
Fructose-2,6 bisphosphatase		_		Amidase	×	
Fucose synthetase	*			Cu/Zn-superoxide dismutase	**	
GDP-D-mannose dehydratase	*			Fibronectin binding protein		
Glucosamine synthetase	*			Histidine decarboxylase		
Glycerophosphoryl diesterase	All successions of			Homospermidine synthase	*#	
Glycosyltransferase	6			Monoamine oxidase	1	
Understand without	-			N and an interview in the second		
riyaluronan synthase	*		11111	rx-carbamoyiput amidohydrolase	×	
Lipase	-			O-methyltransferase		
Lipid phosphate phosphatase				Ornithine decarboxylase	*	
Lysophospholipase				Pathogenesis-related protein	· · · · · · · · · · · · · · · · · · ·	
N-acetyltransferase				Calcium binding protein		
Patatin phospholinase				Collagen-like protein	<u> </u>	
Salidaça	-			Thoumatin like protein		
Sanuase Codes estadoute esta				Thaumann-like protein		
Senne paimitoyitransferase		-	-	Lectin protein		2
Sterol desaturase				Longevity-assurance family protein		
Sugar lyase				Major facilitator		
Transmembrane fatty acid elongation protein			1	Phosphate permease		
UDP-glucose dehydrogenase	*	1111		Phosphoglycerate mutase		
and growers well-urogeniase				- mosphore for and indiase		
Olar - Var				D		
Signaling				Present in all 3 viruses		
Description	PBCV-1	EsV-1	EhV-86	PBCV-1 + EsV-1		
Dual specificity phosphatase	#			DDCUL 4 DDCUL		
Glutamate receptor				PBCV-1 + EhV-86		
Hybrid histidine kinase		6		EsV-1 + EbV-86		
Ligand gatad channel protain			<u> </u>	L37-1 - L117-00		
Ergano-gated channel protein				Unique to one virus		
Discontractor						
Phosphoshuttle				Functional Enzyme	*	
Phosphoshuttle Potassium channel protein	*4			Functional Enzyme	*	

their hosts over time, but another possibility is that the genes existed at the time that prokaryotic and eukaryotic organisms separated. This implies that the progenitor organism(s) lost genes as it evolved into a virus. (vi) Finally, one of the earliest eukaryotic cells could have resembled a single celled alga (Yoon et al., 2004), and viruses of their progenitors may have co-evolved with their hosts. Certainly, the suggestion of the antiquity of the phycodnavirus proteins is speculative (although not without support), but additional viral genomes are being sequenced which may help resolve any ambiguities resulting from the current studies.

# 6. Perspectives

Sequence analyses of three phycodnaviruses suggest that this family may have more sequence diversity than any other virus family. There are in excess of 1000 unique ORFs, and only 14 homologous ORFs are common in these three viruses. To begin to appreciate the potential diversity in the phycodnaviruses world-wide, one estimates that there are  $10^3$  to  $10^5$ phycodnaviruses per ml of water (this number is derived from estimates of 10<sup>7</sup> virus-like particles per ml of seawater and that 0.01-1% of these particles are phycodnaviruses (Bergh et al., 1989 and Proctor and Fuhrman, 1990)). Given the current estimates of fresh and oceanic waters having a volume of 10<sup>24</sup> ml, and assuming recombination and mutation frequencies similar to their eukaryotic host ( $\sim 10^{-6}$  nucleotide changes per replication cycle), genome sizes of 10<sup>5</sup> kb, infection incidences of 10%, a 10% sequence change being enough to create a new virus, and one infection cycle per day, then we estimate that 10<sup>19</sup> new phycodnaviruses could be formed each day. Much of the genetic change would be selected against or result in defective viruses; nevertheless, these estimations suggest that the phycodnaviruses may be exploring sequence space more extensively than any other eukaryotic virus. This phenomenon is primarily due to the vast numbers of hosts and viruses constantly engaged in fresh and marine/coastal waters.

Despite the large genetic diversity in the three sequenced phycodnaviruses (~1000 unique ORFs), phylogenetic analyses

of  $\delta$ -DNA polymerases (Chen et al., 1996, Villarreal and De-Filippis, 2000 and Wilson et al., 2005b) and the superfamily of archeao-eukaryotic primases (Iyer et al., 2005) indicate that the phycodnaviruses group into a monophyletic clade within the NCLDVs. A recent analysis using eight concatenated core NCLDV genes also indicates that the phycodnaviruses cluster together and are members of the NCLDV "superfamily" (Allen et al., 2006). However, it is obvious to us that only the first steps have been taken in identifying phycodnaviruses. Metagenomic studies, such as DNA sequences from the Sargasso Sea samples (Venter et al., 2004), indicate that many gene products resemble those of the phycodnaviruses. For example, the PBCV-1 major capsid protein has 213 BLAST hits when using blastp against the non-redundant and environmental nonredundant databases at NCBI. One hundred of these hits are to the unknown environmental sequences from the Sargasso Sea project. Many of these "unknowns" are more similar to the major capsid protein of PBCV-1 than the major capsid proteins from other phycodnaviruses. While the PBCV-1 major capsid protein matches well with the protein coded by the Heterosigma akashiwa virus 01 (Table 1), 14 "unknown" sequences are more similar to the PBCV-1 protein than the next protein from a known phycodnavirus, FirrV-1. These results are interesting because PBCV-1 infects a fresh water alga, whereas the Sargasso Sea is located in the mid-Atlantic Ocean.

Phylogenetic studies with individual phycodnavirus genes also suggest that the viruses, or at least some of their genes, are evolutionary very old. Thus, studies on the phycodnaviruses, as well as other NCLDV members, may reveal information on the evolution of genes and genomes. A few evolutionary biologists have suggested that large dsDNA viruses, like the phycodnaviruses, may be the origin of the nucleus in eukaryotic cells (e.g., Bell, 2001, Pennisi, 2004, Villarreal, 2004 and Villarreal and DeFilippis, 2000); whereas, another biologist has suggested that the NCLDVs may be members of a fourth domain of life (Raoult et al., 2004). Even if these intriguing hypotheses prove to be incorrect, the study of the phycodnaviruses will continue to produce many unexpected and exciting phenomena.

Figure 1. Selected ORFs in the PBCV-1, EsV-1, and EhV-86 genomes are arranged by their metabolic domains. If a genome encodes a putative protein more than once, the value in the box indicates the number of genes of this type per genome. Color-coding is indicated on the figure and is used to depict the relationship between viruses. Red indicates proteins that are encoded by all three viruses; yellow indicates proteins that are encoded by PBCV-1 and EsV-1, but not EhV-86; green indicates proteins that are encoded by PBCV-1 and EhV-86, but not EsV-1; orange indicates proteins that are encoded by EsV-1 and EhV-86, but not PBCV-1; and blue indicates there are no shared homologs. Solid colored boxes indicate that the putative proteins are homologs. A diagonally-striped box indicates that the putative proteins are non-homologous, and a checkered box indicates that the putative proteins are a mix of homologous, non-homologous, or unique ORFs. In this case, a footnote has been added to clarify the specific differences; in parentheses the ORFs has been defined by the gene number and any ORF beginning with an "A" is from the PBCV-1 genome, an "EsV" is from the EsV-1 genome and an "EhV" is from the EhV-86 genome. Proteins known to be functional are indicated with a star (\*) and proteins known to be associated with the virion are indicated with a plus sign (+). <sup>a</sup>ATPase, one homolog between all three viruses (A392R, EsV-26, and EhV072), one homolog between PBCV-1 and EsV-1 (A565R and EsV-171) and two PBCV-1 ATPases which have no homologs in EsV-1 or EhV-86 (A561L and A554/556/557L). bExonuclease, one homolog between PBCV-1 and EsV-1 (A166R and EsV-64) and one unique to EsV-1 (EsV-126). °PCNA, one homolog between all three viruses (A193L, EsV-132, and EhV020), one homolog between PBCV-1 and EhV-86 (A574L and EhV020). However, EhV440, another EhV-86 encoded PCNA, has no homologs in PBCV-1 or EsV-1. dtRNAs, PBCV-1 has 11 tRNA genes, encoding AA Leu, Ile (2), Asn (2), Lys (3), Tyr, Arg, and Val. EhV-86 has five tRNA genes, encoding AA Leu, Ile, Gln, Asn, and Arg. "Homing endonuclease HNH, one homolog between two viruses (A422R and EhV087). PBCV-1 (A87R) and EsV-1 (EsV-119) are homologous. Four other HNH endonculeases are unique to PBCV-1 (A267L, A354R, A478R, and A490L). fSer/Thr protein kinase, four PBCV-1, two EsV-1, and one EhV-86 S/T kinase ORFs are homolgous (A248R, A277L, A282L, A289L, EsV-82, EsV-111, and EhV451). The remaining S/T kinase ORFs from the three genomes (A34R, A278L, A614L, A617R, EsV-104, EsV-156, and EhV402) are unique. <sup>g</sup>Superfamily II helicase, PBCV-1 (A153R), and EsV-1 (EsV-66) are homologous. The two additional PBCV-1 encoded helicases (A241R and A363R) and EhV104 are unique.

### Acknowledgements

Research at our laboratory was supported by Public Health Service Grant GM32441 from the National Institute of General Medical Sciences and the National Science Foundation Grant EF-0333197. Partial support was provided to DDD through a grant from the National Institutes of Health, CO-BRE Grant P20RR015635. We wish to thank several members of the algal virus scientific community for sharing information and preliminary results from their very interesting viruses, especially Gunnar Bratbak, Corina Brussaard, Nicholas Delaroque, Claire Evans, Gil Malin, Keizo Nagasaki and Willie Wilson. We also would like to thank Dr. L. Aravind for sharing preliminary results, also reported in this special issue on DNA viruses.

## References

- Allen, M.J., Schroeder, D.C., Holden, M.T.G., Wilson, W.H., 2006. Evolutionary history of the Coccolithoviridae. Mol. Biol. Evol. 23, 86–92.
- Allen, M.J., Schroeder, D.C., Wilson, W.H., 2005. Preliminary characterisation of repeat families in the genome of EhV-86, a giant algal virus that infects the marine microalga *Emiliania huxleyi*. Arch. Virol., doi:10.1007/s00705-005-0647-1.
- Altschul, S.F., Gish, W., Miller, W., Meyers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
- Bell, P.J., 2001. Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? J. Mol. Evol. 53, 251–256.
- Benson, S.D., Bamford, J.K., Bamford, D.H., Burnett, R.M., 2004. Does common architecture reveal a viral lineage spanning all three domains of life? Mol. Cell 16, 673–685.
- Bergh, O., Borsheim, K.Y., Bratbak, G., Heldal, M., 1989. High abundance of viruses found in aquatic environments. Nature 340, 467–468.
- Blasco, R., de la Vega, I., Almazan, F., Aguero, M., Vinuela, E., 1989. Genetic variation of African swine fever virus: variable regions near the ends of the viral DNA. Virology 173, 251–257.
- Bratbak, G., Egge, J.K., Heldal, M., 1993. Viral mortality of the marine alga *Emiliania huxleyi (Haptophyceae)* and termination of algal blooms. Mar. Ecol.: Prog. Ser. 93, 39–48.
- Bratbak, G., Heldal, M., Thingstad, T.F., Tuomi, P., 1996. Dynamics of virus abundance in coastal seawater. FEMS Microbiol. Ecol. 19, 263–269.
- Brautigam, M., Klein, M., Knippers, R., Müller, D.G., 1995. Inheritance and meiotic elimination of a virus genome in the host *Ectocarpus siliculosus* (*Phaeophyceae*). J. Phycol. 31, 823–827.
- Brocks, J.J., Logan, G.A., Buick, R., Summons, R.E., 1999. Archean molecular fossils and the early rise of eukaryotes. Science 285, 1033–1036.
- Brussaard, C.P.D., 2004. Viral control of phytoplankton populations—a review. J. Eukaryot. Microbiol. 51, 125–138.
- Castberg, T., Thyrhaug, R., Larsen, A., Sandaa, R.-A., Heldal, M., Van Etten, J.L., Bratbak, G., 2002. Isolation and characterization of a virus that infects *Emiliania huxleyi* (Haptophyta). J. Phycol. 38, 767–774.
- Charlson, R.J., Lovelock, J.E., Andraeae, M.O., Warren, S.G., 1987. Oceanic phytoplankton, atmosphere sulfur, cloud albedo and climate. Nature 326, 655–661.
- Chen, F., Suttle, C.A., Short, S.M., 1996. Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. Appl. Environ. Microbiol. 62, 2869–2874.
- Cho, H.H., Park, H.H., Kim, J.O., Choi, T.J., 2002. Isolation and characterization of Chlorella viruses from freshwater sources in Korea. Mol. Cells 14, 168–176.

- Chuchird, N., Nishida, K., Kawasaki, T., Fujie, M., Usami, S., Yamada, T., 2002. A variable region on the chlorovirus CVK2 genome contains five copies of the gene for Vp260, a viral-surface glycoprotein. Virology 295, 289–298.
- DeAngelis, P.L., Wei, J., Graves, M.V., Burbank, D.E., Van Etten, J.L., 1997. Hyaluronan synthase of chlorella virus PBCV-1. Science 278, 1800–1803.
- del Campo, E., Ramazonov, Z., Garcia-Reina, G., Müller, D.G., 1997. Photosynthetic responses and growth performance of virus-infected and noninfected *Ectocarpus siliculosus (Phaeophyceae)*. Phycologia 36, 186–189.
- Delaroque, N., Boland, W., Müller, D.G., Knippers, R., 2003. Comparisons of two large phaeoviral genomes and evolutionary implications. J. Mol. Evol. 57, 613–622.
- Delaroque, N., Maier, I., Knippers, R., Müller, D.G., 1999. Persistent virus integration into the genome of its algal host, *Ectocarpus siliculosus (Phaeophyceae*). J. Gen. Virol. 80, 1367–1370.
- Delaroque, N., Müller, D.G., Bothe, G., Pohl, T., Knippers, R., Boland, W., 2001. The complete DNA sequence of the *Ectocarpus siliculosus virus* EsV-1 genome. Virology 287, 112–132.
- Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., Kivirikko, K.I., 1999. Evidence for 4-hydroxyproline in viral proteins. Characterization of a viral prolyl 4-hydroxylase and its peptide substrates. J. Biol. Chem. 274, 22131–22134.
- Feng, D.F., Cho, G., Doolittle, R.F., 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. U.S.A 94, 13028–13033.
- Fortune, J.M., Lavrukhin, O.V., Gurnon, J.R., Van Etten, J.L., Lloyd, R.S., Osheroff, N., 2001. Topoisomerase II from Chlorella virus PBCV-1 has an exceptionally high DNA cleavage activity. J. Biol. Chem. 276, 24401–24408.
- Friess-Klebl, A.-K., Knippers, R., Müller, D.G., 1994. Isolation and characterization of a DNA virus infecting *Feldmannia simplex (Phaeophyceae)*. J. Phycol. 30, 653–658.
- Frohns, F., K¨asmann, A., Kramer, D., Sch¨afer, B., Mehmel, M., Kang, M., Van Etten, J.L., Gazzarrini, S., Moroni, A., Thiel, G., 2006. Potassium channels of chlorella viruses cause rapid depolarization of host cells during infection. J. Virol. 80, 2437–2444.
- Girton, L.E., Van Etten, J.L., 1987. Restriction site map of the Chlorella virus PBCV-1 genome. Plant Mol. Biol. 9, 247–257.
- Glansdorff, N., 2000. About the last common ancestor, the universal life-tree and lateral gene transfer: a reappraisal. Mol. Microbiol. 38, 177–185.
- Graves, M.V., Burbank, D.E., Roth, R., Heuser, J., DeAngelis, P.L., Van Etten, J.L., 1999. Hyaluronan synthesis in virus PBCV-1-infected chlorella-like green algae. Virology 257, 15–23.
- Guiry, M.D., Rindi, F., 2005. AlgaeBase version 3.0. http://www. algaebase. org/. World-wide electronic publication, National University of Ireland, Galway.
- Hakansson, K., Doherty, A.J., Shuman, S., Wigley, D.B., 1997. X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. Cell 89, 545–553.
- Hakansson, K., Wigley, D.B., 1998. Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. Proc. Natl. Acad. Sci. U.S.A. 95, 1505–1510.
- Hambly, E., Suttle, C.A., 2005. The viriosphere, diversity, and genetic exchange within phage communities. Curr. Opin. Microbiol. 8, 444–450.
- Han, T.M., Runnegar, B., 1992. Megascopic eukaryotic algae from the 2.1billion-year-old negaunee iron-formation. Mich. Sci. 257, 232–235.
- Hendrix, R.W., Lawrence, J.G., Hatfull, G.F., Casjens, S., 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8, 504–508.
- Ho, C.K., Van Etten, J.L., Shuman, S., 1997. Characterization of an ATPdependent DNA ligase encoded by Chlorella virus PBCV-1. J. Virol. 71, 1931–1937.

- Ho, C.K., VanEtten, J.L., Shuman, S., 1996. Expression and characterization of an RNA capping enzyme encoded by Chlorella virus PBCV-1. J. Virol. 70, 6658–6664.
- Iyer, L.M., Aravind, L., Koonin, E.V., 2001. Common origin of four diverse families of large eukaryotic DNA viruses. J. Virol. 75, 11720–11734.
- Iyer, L.M., Koonin, E.V., Leipe, D.D., Aravind, L., 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palmdomain proteins: structural insights and new members. Nucleic Acids Res. 33, 3875–3896.
- Iyer, M.L., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 117, 156–184.
- Kang, M., Dunigan, D.D., Van Etten, J.L., 2005. Chlorovirus: a genus of Phycodnaviridae that infects certain chlorella-like green algae. Mol. Plant Pathol. 6, 213–224.
- Kapp, M., 1998. Viruses infecting marine brown algae. Virus Genes 16, 111–117.
- Kapp, M., Knippers, R., Müller, D.G., 1997. New members of a group of DNA viruses infecting brown algae. Phycol. Res. 45, 85–90.
- Klein, M., Lanka, S., Müller, D., Knippers, R., 1994. Single-stranded regions in the genome of the *Ectocarpus siliculosus* virus. Virology 202, 1076–1078.
- Klein, M., Lanka, S.T.J., Knippers, R., Müller, D.G., 1995. Coat protein of the *Ectocarpus siliculosus* virus. Virology 206, 520–526.
- Kowalski, J.C., Belfort, M., Stapleton, M.A., Holpert, M., Dansereau, J.T., Pietrokovski, S., Baxter, S.M., Derbyshire, V., 1999. Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. Nucleic Acids Res. 27, 2115–2125.
- Landstein, D., Burbank, D.E., Nietfeldt, J.W., Van Etten, J.L., 1995. Large deletions in antigenic variants of the Chlorella virus PBCV-1. Virology 214, 413–420.
- Landstein, D., Graves, M.V., Burbank, D.E., DeAngelis, P., Van Etten, J.L., 1998. Chlorella virus PBCV-1 encodes functional glutamine: fructose-6-phosphate amidotransferase and UDP-glucose dehydrogenase enzymes. Virology 250, 388–396.
- Lanka, S.T.J., Klein, M., Ramsperger, U., Müller, D.G., Knippers, R., 1993. Genome structure of a virus infecting the marine brown alga *Ectocarpus siliculosus*. Virology 193, 802–811.
- Lavrukhin, O.V., Fortune, J.M., Wood, T.G., Burbank, D.E., Van Etten, J.L., Osheroff, N., Lloyd, R.S., 2000. Topoisomerase II from Chlorella virus PBCV-1: characterization of the smallest known type II topoisomerase. J. Biol. Chem. 275, 6915–6921.
- Li, Y., Lu, Z., Sun, L., Ropp, S., Kutish, G.F., Rock, D.L., Van Etten, J.L., 1997. Analysis of 74 kb of DNA located at the right end of the 330-kb chlorella virus PBCV-1 genome. Virology 237, 360–377.
- Maier, I., Müller, D.G., Katsaros, C., 2002. Entry of the DNA virus, *Ectocarpus fasciculatus* virus type 1 (Phycodnaviridae), into host cell cytosol and nucleus. Phycol. Res. 50, 227–231.
- Maier, I., Rometsch, E., Wolf, S., Kapp, M., Mueller, D.G., Kawai, H., 1997. Passage of a marine brown algal DNA virus from *Ectocarpus fasciculatus (Ectocarparpales, Phaeophyceae)* to *Myriotrichia clavaeformis* (*Dictyosiphonales, Phaeophyceae*): infection symptoms and recovery. J. Phycol. 33, 838–844.
- Malin, G., Liss, P.S., Turner, S.M., 1994. Dimethyl sulfide production and atmospheric consequences. In: Green, J.C. (Ed.), The Haptophyte Algae, vol. 51. Clarendon Press, Oxford, pp. 303–320.
- Manzur, K.L., Farooq, A., Zeng, L., Plotnikova, O., Koch Alexander, W., Sachchidanand, Zhou, M.-M., 2003. A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. Nat. Struct. Biol. 10, 187–196.
- Markine-Goriaynoff, N., Gillet, L., Van Etten, J.L., Korres, H., Verma, N., Vanderplasschen, A., 2004. Glycosyltransferases encoded by viruses. J. Gen. Virol. 85, 2741–2754.

- Mehmel, M., Rothermel, M., Meckel, T., Van Etten, J.L., Moroni, A., Thiel, G., 2003. Possible function for virus encoded K+ channel Kcv in the replication of chlorella virus PBCV-1. FEBS Lett. 552, 7–11.
- Meints, R.H., Lee, K., Burbank, D.E., Van Etten, J.L., 1984. Infection of a Chlorella-like alga with the virus, PBCV-1: ultrastructural studies. Virology 138, 341–346.
- Meints, R.H., Lee, K., Van Etten, J.L., 1986. Assembly site of the virus PBCV-1 in a Chlorella-like green alga: ultrastructural studies. Virology 154, 240–245.
- Morehead, T.A., Gurnon, J.R., Adams, B., Nickerson, K.W., Fitzgerald, L.A., Van Etten, J.L., 2002. Ornithine decarboxylase encoded by chlorella virus PBCV-1. Virology 301, 165–175.
- Müller, D.G., 1991a. Marine virioplankton produced by infected *Ectocarpus siliculosus (Phaeophyceae)*. Mar. Ecol.: Prog. Ser. 76, 101–102.
- Müller, D.G., 1991b. Mendelian segregation of a virus genome during host meiosis in the marine brown alga *Ectocarpus siliculosus*. J. Plant Physiol. 137, 739–743.
- Müller, D.G., 1992. Intergenic transmission of a marine plant DNA virus. Naturwissenschaften 79, 37–39.
- Müller, D.G., Brautigam, M., Knippers, R., 1996a. Virus infection and persistence of foreign DNA in the marine brown alga *Feldmannia simplex* (*Ectocarpales, Phaeophyceae*). Phycologia 35, 61–63.
- Müller, D.G., Kapp, M., Knippers, R., 1998. Viruses in marine brown algae. Adv. Virus Res. 50, 49–67.
- Müller, D.G., Parodi, E., 1993. Transfer of a marine DNA virus from *Ecto-carpus* to *Feldmannia (Ectocarpales, Phaeophyceae)*—aberrant symptoms and restitution of the host. Protoplasma 175, 121–125.
- Müller, D.G., Sengco, M., Wolf, S., Brautigam, M., Schmid, C.E., Kapp, M., Knippers, R., 1996b. Comparison of two DNA viruses infecting the marine brown algae *Ectocarpus siliculosus* and *E. fasciculatus*. J. Gen. Virol. 77, 2329–2333.
- Müller, D.G., Stache, B., 1992. Worldwide occurrence of virus-infections in filamentous marine brown algae. Helgolander Meeresuntersuchungen 46, 1–8.
- Müller, D.G., Westermeier, R., Morales, J., Garcia-Reina, G., del Campo, E., Correa, J.A., Rometsch, E., 2000. Massive prevalence of viral DNA in *Ectocarpus (Phaeophyceae, Ectocarpales)* from two habitats in the North Atlantic and South Pacific. Bot. Marina 43, 157–159.
- Nandhagopal, N., Simpson, A.A., Gurnon, J.R., Yan, X., Baker, T.S., Graves, M.V., Van Etten, J.L., Rossmann, M.G., 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc. Natl. Acad. Sci. U.S.A. 99, 14758–14763.
- Nelson, M., Burbank, D.E., Van Etten, J.L., 1998. Chlorella viruses encode multiple DNA methyltransferases. J. Biol. Chem. 379, 423–428.
- Nishida, K., Kawasaki, T., Fujie, M., Usami, S., Yamada, T., 1999a. Aminoacylation of tRNAs encoded by Chlorella virus CVK2. Virology 263, 220–229.
- Nishida, K., Kimura, Y., Kawasaki, T., Fujie, M., Yamada, T., 1999b. Genetic variation of Chlorella viruses: variable regions localized on the CVK2 genomic DNA. Virology 255, 376–384.
- Nishida, k., Suzuki, S., Kimura, Y., Nomura, N., Fujie, M., Yamada, T., 1998. Group I introns found in Chlorella viruses: biological implications. Virology 242, 319–326.
- Olofsson, S., Hansen, J.E., 1998. Host cell glycosylation of viral glycoproteins—a battlefield for host defence and viral resistance. Scand. J. Infect. Dis. 30, 435–440.
- Onimatsu, H., Sugimoto, I., Fujie, M., Usami, S., Yamada, T., 2004. Vp130, a chloroviral surface protein that interacts with the host Chlorella cell wall. Virology 319, 71–80.
- Pennisi, E., 2004. Evolutionary biology. The birth of the nucleus. Science 305, 766–768.
- Plugge, B., Gazzarrini, S., Nelson, M., Cerana, R., Van Etten, J.L., Derst, C., DiFrancesco, D., Moroni, A., Thiel, G., 2000. A potassium channel pro-

tein encoded by chlorella virus PBCV-1. Science 287, 1641-1644.

- Proctor, L.M., Fuhrman, J.A., 1990. Viral mortality of marine bacteria and cyanobacteria. Nature 343, 60–62.
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.M., 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.
- Rezsohazy, R., Hallet, B., Delcour, J., Mahillon, J., 1993. The IS4 family of insertion sequences: evidence for a conserved transposase motif. Mol. Microbiol. 9, 1283–1295.
- Robledo, D.R., Sosa, P.A., Garciareina, G., Müller, D.G., 1994. Photosynthetic performance of healthy and virus-infected *Feldmannia irregularis* and *F. simplex (Phaeophyceae)*. Eur. J. Phycol. 29, 247–251.
- Schroeder, D.C., Oke, J., Malin, G., Wilson, W.H., 2002. Coccolithovirus (Phycodnaviridae): characterisation of a new large dsDNA algal virus that infects *Emiliania huxleyi*. Arch. Virol. 147, 1685–1698.
- Schuster, A.M., Girton, L., Burbank, D.E., Van Etten, J.L., 1986. Infection of a Chlorella-like alga with the virus PBCV-1: transcriptional studies. Virology 148, 181–189.
- Sengco, M.R., Brautigam, M., Kapp, M., Müller, D.G., 1996. Detection of virus DNA in *Ectocarpus siliculosus* and *E. fasciculatus* (Phaeophyceae) from various geographic areas. Eur. J. Phycol. 31, 73–78.
- Shah, R., Coleman, C.S., Mir, K., Baldwin, J., Van Etten, J.L., Grishin, N.V., Pegg, A.E., Stanley, B.A., Phillips, M.A., 2004. *Paramecium bursaria* chlorella virus-1 encodes an unusual arginine decarboxylase that is a close homolog of eukaryotic ornithine decarboxylases. J. Biol. Chem. 279, 35760–35767.
- Shchelkunov, S.N., Totmenin, A.V., 1995. Two types of deletions in orthopoxvirus genomes. Virus Genes 9, 231–245.
- Skrdla, M.P., Burbank, D.E., Xia, Y., Meints, R.H., Van Etten, J.L., 1984. Structural proteins and lipids in a virus, PBCV-1, which replicates in a Chlorella-like alga. Virology 135, 308–315.
- Songsri, P., Hiramatsu, S., Fujie, M., Yamada, T., 1997. Proteolytic processing of Chlorella virus CVK2 capsid proteins. Virology 227, 252–254.
- Sriskanda, V., Shuman, S., 2002. Role of nucleotidyl transferase motif V in strand joining by chlorella virus DNA ligase. J. Biol. Chem. 277, 9661–9667.
- Stock, A.M., Robinson, V.L., Goudreau, P.N., 2000. Two-component signal transduction. Annu. Rev. Biochem. 69, 183–215.
- Strasser, P., Zhang, Y.P., Rohozinski, J., Van Etten, J.L., 1991. The termini of the Chlorella virus PBCV-1 genome are identical 2.2-kbp inverted repeats. Virology 180, 763–769.
- Sun, L., Li, Y., McCullough, A.K., Wood, T.G., Lloyd, R.S., Adams, B., Gurnon, J.R., Van Etten, J.L., 2000. Intron conservation in a UV-specific DNA repair gene encoded by chlorella viruses. J. Mol. Evol. 50, 82–92.
- Tonetti, M., Zanardi, D., Gurnon, J.R., Fruscione, F., Armirotti, A., Damonte, G., Sturla, L., De Flora, A., Van Etten, J.L., 2003. *Paramecium bursaria* Chlorella virus 1 encodes two enzymes involved in the biosynthesis of GDP-l-fucose and GDP-d-rhamnose. J. Biol. Chem. 278, 21559–21565.
- Turner, P.C., Moyer, R.W., 1990. The molecular pathogenesis of poxviruses. Curr. Top. Microbiol. Immunol. 163, 125–151.
- Van Etten, J.L., 2003. Unusual life style of giant chlorella viruses. Annu. Rev. Genet. 37, 153–195.
- Van Etten, J.L., Burbank, D.E., Joshi, J., Meints, R.H., 1984. DNA synthesis in a Chlorella-like alga following infection with the virus PBCV-1. Virology 134, 443–449.
- Van Etten, J.L., Burbank, D.E., Meints, R.H., 1986. Replication of the algal virus PBCV-1 in UV-irradiated Chlorella. Intervirology 26, 115–120.
- Van Etten, J.L., Burbank, D.E., Xia, Y., Meints, R.H., 1983. Growth cycle of a virus, PBCV-1, that infects Chlorella-like algae. Virology 126, 117–125.
- Van Etten, J.L., Graves, M.V., Müller, D.G., Boland, W., Delaroque, N., 2002. Phycodnaviridae—large DNA algal viruses. Arch. Virol. 147, 1479–1516.

- Van Etten, J.L., Lane, L.C., Meints, R.H., 1991. Viruses and viruslike particles of eukaryotic algae. Microbiol. Rev. 55, 586–620.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.H., Smith, H.O., 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304, 66–74.
- Villarreal, L.P., 2004. Are viruses alive? Sci. Am. 291, 100-105.
- Villarreal, L.P., DeFilippis, V.R., 2000. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J. Virol. 74, 7079–7084.
- Westbroek, P., Brown, C.W., Van Bleijswijk, J., Brownlee, C., Brummer, G.J., Conte, M., Egge, J., Fernadez, E., Jordan, R., Knappertsbusch, M., Stefels, J., Veldhuis, M., Van der Val, P., Young, J., 1993. A model system approach to biological climate forcing—the example of *Emiliania huxleyi*. Global Planetary Change 8, 27–46.
- Westbroek, P., Van Hinte, J.E., Brummer, G.J., Veldhuis, M., Brownlee, C., Green, J.C., Harris, R., Heimdal, B.R., 1994. *Emiliania huxleyi* as a key to biosphere-geosphere interactions. In: Green, J.C. (Ed.), In: The Haptophyte Algae, vol. 51. Clarendon Press, Oxford, pp. 321–334.
- Wilson, W.H., Schroeder, D.C., Allen, M.J., Holden, M.T., Parkhill, J., Barrell, B.G., Churcher, C., Hamlin, N., Mungall, K., Norbertczak, H., Quail, M.A., Price, C., Rabbinowitsch, E., Walker, D., Craigon, M., Roy, D., Ghazal, P., 2005a. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. Science 309, 1090–1092.
- Wilson, W.H., Tarran, G., Zubkov, M.V., 2002. Virus dynamics in a coccolithophore-dominated bloom in the North Sea. Deep-Sea Research Part II-Topical Studies in Oceanography 49, 2951–2963.
- Wilson, W.H., Van Etten, J.L., Schroeder, D.S., Nagasaki, K., Brussaard, C., Delaroque, N., Bratbak, G., Suttle, C., 2005b. Phycodnaviridae. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Virus Taxonomy: Classification and Nomenclature of Viruses, Eighth Report of the International Committee on the Taxonomy of Viruses. Elsevier Academic Press, San Diego, pp. 163–175.
- Wolf, S., Maier, I., Katsaros, C., Müller, D.G., 1998. Virus assembly in *Hinck-sia hincksiae (Ectocarpales, Phaeophyceae)* an electron and fluorescence microscopic study. Protoplasma 203, 153–167.
- Wolf, S., Müller, D.G., Maier, I., 2000. Assembly of a large icosahedral DNA virus, MclaV-1, in the marine alga *Myriotrichia clavaeformis* (*Dictyosiphonales, Phaeophyceae*). Eur. J. Phycol. 35, 163–171.
- Wommack, K.E., Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114.
- Yamada, T., Higashiyama, T., 1993. Characterization of the terminal inverted repeats and their neighboring tandem repeats in the Chlorella CVK1 virus genome. Mol. Genet. 241, 554–563.
- Yamada, T., Tamura, K., Aimi, T., Songsri, P., 1994. Self-splicing group I introns in eukaryotic viruses, Nucleic Acids Symposium Series 31, pp. 199–200.
- Yan, X., Chipman, P.R., Castberg, T., Bratbak, G., Baker, T.S., 2005. The marine algal virus PpV01 has an icosahedral capsid with T = 219 quasi-symmetry. J. Virol. 79, 9236–9243.
- Yan, X., Olson, N.H., Van Etten, J.L., Bergoin, M., Rossmann, M.G., Baker, T.S., 2000. Structure and assembly of large lipid-containing dsDNA viruses. Nat. Struct. Biol. 7, 101–103.
- Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., Bhattacharya, D., 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. 21, 809–818.
- Zhang, Y., Strasser, P., Grabherr, R., Van Etten, J.L., 1994. Hairpin loop structure at the termini of the chlorella virus PBCV-1 genome. Virology 202, 1079–1082.