

February 2007

Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella Pbi*

Lisa A. Fitzgerald
University of Nebraska-Lincoln, fitz918@msn.com

Michael V. Graves
University of Massachusetts-Lowell, Michael_Graves@uml.edu

Xiao Li
University of Massachusetts-Lowell

Tamara Feldblyum
The Institute for Genomic Research, Rockville, MD

James Hartigan
Agencourt Bioscience Corporation, Beverly, MA

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/virologypub>

 Part of the [Virology Commons](#)

Fitzgerald, Lisa A.; Graves, Michael V.; Li, Xiao; Feldblyum, Tamara; Hartigan, James; and Van Etten, James L., "Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella Pbi*" (2007). *Virology Papers*. 1.
<http://digitalcommons.unl.edu/virologypub/1>

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

Authors

Lisa A. Fitzgerald, Michael V. Graves, Xiao Li, Tamara Feldblyum, James Hartigan, and James L. Van Etten

Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella Pbi*

Lisa A. Fitzgerald^a, Michael V. Graves^b, Xiao Li^b, Tamara Feldblyum^c,
James Hartigan^d, and James L. Van Etten^{e, f, *}

^aDepartment of Chemistry, University of Nebraska–Lincoln, Lincoln, NE 68588-0304

^bDepartment of Biological Sciences, University of Massachusetts–Lowell, Lowell, MA 01854

^cThe Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850

^dAgencourt Bioscience Corporation, 500 Cummings Center, Suite 2450, Beverly, MA 01915

^eDepartment of Plant Pathology, University of Nebraska–Lincoln, Lincoln, NE 68583-0722

^fNebraska Center for Virology, University of Nebraska, Lincoln, NE 68588-0666

*Corresponding author. Email: jvanetten@unlnotes.unl.edu

Abstract: Viruses MT325 and FR483, members of the family Phycodnaviridae, genus *Chlorovirus*, infect the fresh water, unicellular, eukaryotic, chlorella-like green alga, *Chlorella Pbi*. The 314,335-bp genome of MT325 and the 321,240-bp genome of FR483 are the first viruses that infect *Chlorella Pbi* to have their genomes sequenced and annotated. Furthermore, these genomes are the two smallest chlorella virus genomes sequenced to date, MT325 has 331 putative protein-encoding and 10 tRNA-encoding genes and FR483 has 335 putative protein-encoding and 9 tRNA-encoding genes. The protein-encoding genes are almost evenly distributed on both strands, and intergenic space is minimal. Approximately 40% of the viral gene products resemble entries in public databases, including some that are the first of their kind to be detected in a virus. For example, these unique gene products include an aquaglyceroporin in MT325, a potassium ion transporter protein and an alkyl sulfatase in FR483, and a dTDP–glucose pyrophosphorylase in both viruses. Comparison of MT325 and FR483 protein-encoding genes with the prototype chlorella virus PBCV-1 indicates that approximately 82% of the genes are present in all three viruses.

Keywords: Chlorella viruses, Phycodnaviridae, Virus MT325, Virus FR483, Genome sequence

Introduction

Members and prospective members of the family Phycodnaviridae consist of a genetically diverse, but morphologically similar, group of large dsDNA-containing viruses (170–560 kb) that infect eukaryotic algae from both fresh and marine waters (Kang *et al.*, 2005 and Wilson *et al.*, 2005b). The phycodnaviruses are among the virioplankton recognized as important ecological elements in aqueous environments (Suttle, 2005). They, along with other viruses, play significant 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.

The phycodnaviruses, together with the poxviruses, iridoviruses, asfarviruses, and the 1.2-Mb Mimivirus are believed to have a common evolutionary ancestor (Iyer *et al.*, 2001, Iyer *et al.*, 2006 and Raoult *et al.*, 2004). All of these viruses share 9 gene products and at least two of these viral families encode an additional 41 homologous gene products (Iyer *et al.*, 2006). Collectively, these viruses are referred to as nucleocytoplasmic large DNA viruses (NCLDV) (Iyer *et al.*, 2001).

Currently, the phycodnaviruses are grouped into 6 genera, based initially on host range and subsequently supported by sequence comparison of their DNA polymerases (Wilson *et al.*, 2005b). Members of the genus *Chlorovirus* (chlorella viruses) infect fresh water algae, whereas members of the other five genera (*Coccolithovirus*, *Phaeovirus*, *Prasinovirus*, *Prymnesiovirus*, and *Raphdovirus*) infect marine algae. The genomes

of members of three Phycodnaviridae genera have been sequenced (Delaroque *et al.*, 2001, Li *et al.*, 1997 and Wilson *et al.*, 2005a). Comparative analysis of the three genomes has revealed more than 1000 unique genes with only 14 genes in common among the three genera (Dunigan *et al.*, 2006). Thus the gene diversity in the phycodnaviruses is enormous.

The chlorella viruses infect certain unicellular, eukaryotic chlorella-like green algae that normally exist as endosymbionts in various protists, such as *Paramecium bursaria* (Kawakami and Kawakami, 1978 and Van Etten *et al.*, 1982), *Hydra viridis* (Meints *et al.*, 1981 and Van Etten *et al.*, 1981) and *Acanthocystis turfacea* (Bubeck and Pfitzner, 2005). *P. bursaria* chlorella virus (PBCV-1) is the type member of the group and has a 331-kb genome that was sequenced about 10 years ago. The virus contains 366 putative protein-encoding genes and a polycistronic gene that encodes 11 tRNAs (Li *et al.*, 1997).

To investigate the diversity of the chlorella viruses, we are sequencing the genomes of several additional family members. The proceeding manuscript described the sequence and annotation of the 369-kb genome from virus NY-2A and the 345-kb genome from virus AR158, which, like PBCV-1, infect *Chlorella* NC64A (NC64A viruses). The current manuscript describes the sequence and annotation of the 314-kb genome from virus MT325 and the 321-kb genome from virus FR483 that infect another chlorella species, *Chlorella* Pbi (Pbi viruses). Virus MT325 was sequenced first and its sequence revealed a gene encoding a water channel, an aquaglyceroporin (AQPV), and a gene encoding a potassium ion channel protein (Kcv) whose predicted structure differed from Kcv proteins encoded by the NC64A viruses (Gazzarrini *et al.*, 2006). Hybridization of the MT325 *aqpv* and *kcv* genes to DNA from 47 additional Pbi viruses produced signals with 46 of them for both genes. However, a very weak signal was obtained for both genes with virus FR483 DNA, suggesting that either the DNA sequences were very different for both genes, or that they might be absent in FR483. Therefore, FR483 was chosen for sequencing. As reported here, FR483 lacks both *aqpv* and *kcv* genes but it does encode a putative potassium ion transporter protein.

Results and discussion

As part of the chlorella virus genome sequencing effort, a project website has been created at <http://greengene.uml.edu/>. This site contains the genomic DNA sequence assemblies as

well as the predicted amino acid sequences of all virus-encoded ORFs and is viewable in text format or through a graphical genome browser. This database also contains the complete annotation for each chlorella virus-encoded ORF. The supplemental data files referenced below are also available at this site.

Description of the viral genomes

The MT325 and FR483 genomes were assembled into contiguous sequences of 314,335 bp and 321,240 bp (Table 1), respectively, which agrees with their predicted sizes determined by pulse-field gel electrophoresis (unpublished results). Since the presumed hairpin termini were not sequenced, the left most nucleotide of the assembled sequences was designated 1.

To orient the MT325 and FR483 genomes relative to the prototype virus PBCV-1, plots of PBCV-1 proteins and either the MT325 or the FR483 proteins were compared. Unlike the dot plots that compared the three *Chlorella* NC64A-infecting viruses (Fitzgerald *et al.*, 2007), there is only slight co-linearity between the prototype PBCV-1 and the two Pbi genomes (Fig. 1). However, there is a high degree of co-linearity between MT325 and FR483.

The average G + C content of the MT325 and FR483 genomes is 45.3% and 44.6%, respectively, higher than those in the NC64A viruses, PBCV-1, NY-2A, and AR158 (40–41%) (Fitzgerald *et al.*, 2007).

Genes

A putative protein-coding region, or open-reading frame (ORF), 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 64 or more additional codons. Using this criterion, 845 ORFs were identified in the 314-kb genome of MT325 and 849 ORFs were identified in the 321-kb genome of FR483. The ORF names were based on three criteria. First, the MT325 ORF name begins with either an “M” for a major ORF (predicted to be a protein-encoding gene) or an “m” for a minor ORF (unlikely to be a protein-encoding gene). The names for the FR483 ORFs begin with either an “N” for a major ORF or an “n” for a minor ORF. Second, the ORFs were numbered consecutively in the order in which they appeared in the genome after alignment with the PBCV-1 genome. Third, the letter R or L following the ORF number indicates that the putative transcript runs either left-to-right or right-to-left, respectively. The letters “M” or “m” were cho-

Table 1
MT325 and FR483 genomes compared to the prototype chlorella virus, PBCV-1

Genome	General Characteristics				Similarity to PBCV-1		Similarity to MT325		Similarity to FR483	
	Size (bp)	Genes	tRNA Genes	G + C (%)	Homologous Genes ^a	Average a.a. Identity ^b	Homologous Genes ^a	Average a.a. Identity ^b	Homologous Genes ^a	Average a.a. Identity ^b
PBCV-1	330,743	366	11	40.0			82%	50%	84%	50%
MT325	314,335	331	10	45.3	81%	49%			95%	86%
FR483	321,240	335	9	44.6	79%	49%	94%	86%		

^a Percentage of protein-coding genes that have at least one homolog (blastp E-value $\leq 10^{-5}$) in the indicated virus genome.

^b Average amino acid identity (%) between homologous protein-coding genes.

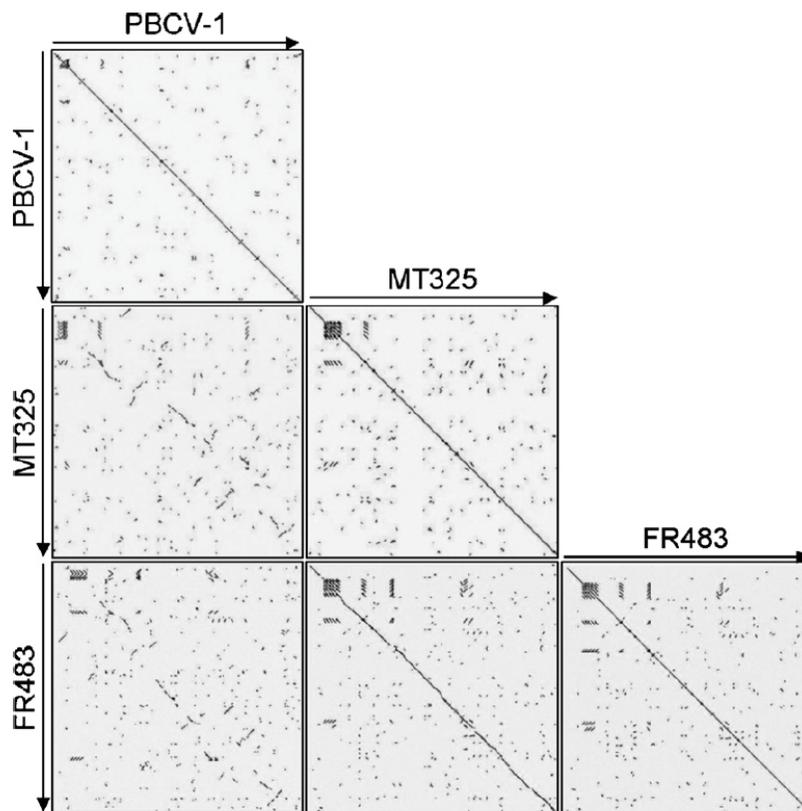


Figure 1. Comparison of three sequenced chlorella viruses, PBCV-1, MT325, and FR483, major ORFs with blastp dot plots. The dots represent ORF homology between two viruses with an *E*-value of less than 0.001.

sen to name the MT325 ORFs, which was the first Pbi virus genome sequenced and “N” or “n” was chosen to name the FR483 ORFs, which was the second Pbi virus genome sequenced, thus avoiding confusion between the different chlorella virus genomes. The letters serve to distinguish these virus ORFs from those that infect *Chlorella* NC64A [*i.e.*, PBCV-1, NY-2A, and AR158 (designated with upper and lower case “A,” “B,” and “C,” respectively)] (Fitzgerald *et al.*, 2007).

The 845 MT325 ORFs and 849 FR483 ORFs were classified as either major or minor ORFs based on the following criteria. When an ORF, of either the same or opposite polarity, resided within or significantly overlapped another ORF, the larger ORF was classified as a major ORF; the smaller ORFs were classified as minor. All of the ORFs were analyzed using the non-redundant, Pfam, and COG databases and ORFs predicted to encode a functional protein were classified as major. These conditions led to the prediction that 331 of the 845 MT325 ORFs and 335 of the 849 FR483 ORFs probably encode proteins.

The NC64A viruses have three types of introns in their protein-encoding genes. PBCV-1 and NY-2A have a self-splicing intron in a transcription factor TFIIS-like gene (Fitzgerald *et al.*, 2007, Li *et al.*, 1997 and Yamada *et al.*, 1994). A splicosomal-processed intron is present in the DNA polymerase gene from all three sequenced NC64A viruses (Grabherr *et al.*, 1992 and Zhang *et al.*, 2001) and an 81-nucleotide splicosomal-processed intron exists in the pyrimidine dimer-specific glycosylase gene from some of the viruses (Fitzgerald *et*

al., 2007 and Sun *et al.*, 2000). However, initial analyses of the MT325 and FR483 genomes did not reveal any introns in their protein-encoding genes including the DNA polymerase, TFIIS, and pyrimidine dimer-specific glycosylase gene homologs. Like PBCV-1, one of the MT325 and one of the FR483 tRNA genes are predicted to contain an intron (see below). No inteins were detected in the two Pbi viruses.

GCG software was used to determine several general characteristics and properties for each ORF, including the nucleotide composition of the ORF, the A + T content of the 50 nucleotides upstream of the ORF that is likely to contain promoter elements, the frame in which the putative protein is encoded, the number of amino acids in the encoded protein, the predicted protein molecular weights, and the isoelectric points. These properties are listed in Supplements 1 and 2. Fig. 2 reports some general characteristics of both the MT325 and FR483 major ORFs, including the relative orientation of the ORFs (Figs. 2A, B). The directions in which the MT325 ORFs are encoded are slightly skewed in the reverse (54%) orientation. In contrast, the FR483 genes are evenly distributed on both strands. The average size of all putative MT325 proteins is 284 amino acids (Fig. 2C), while the average size of all putative FR483 proteins is 290 amino acids (Fig. 2D); nearly half of these proteins are in the 65 to 200-amino-acid size range. The predicted isoelectric points of the proteins are depicted in Figs. 2E, F. Despite a trend for the proteins to have a *pI* in the 10–11 pH range, a peak also occurs at pH 4.5. Basic proteins are probably associated with the virion where they pre-

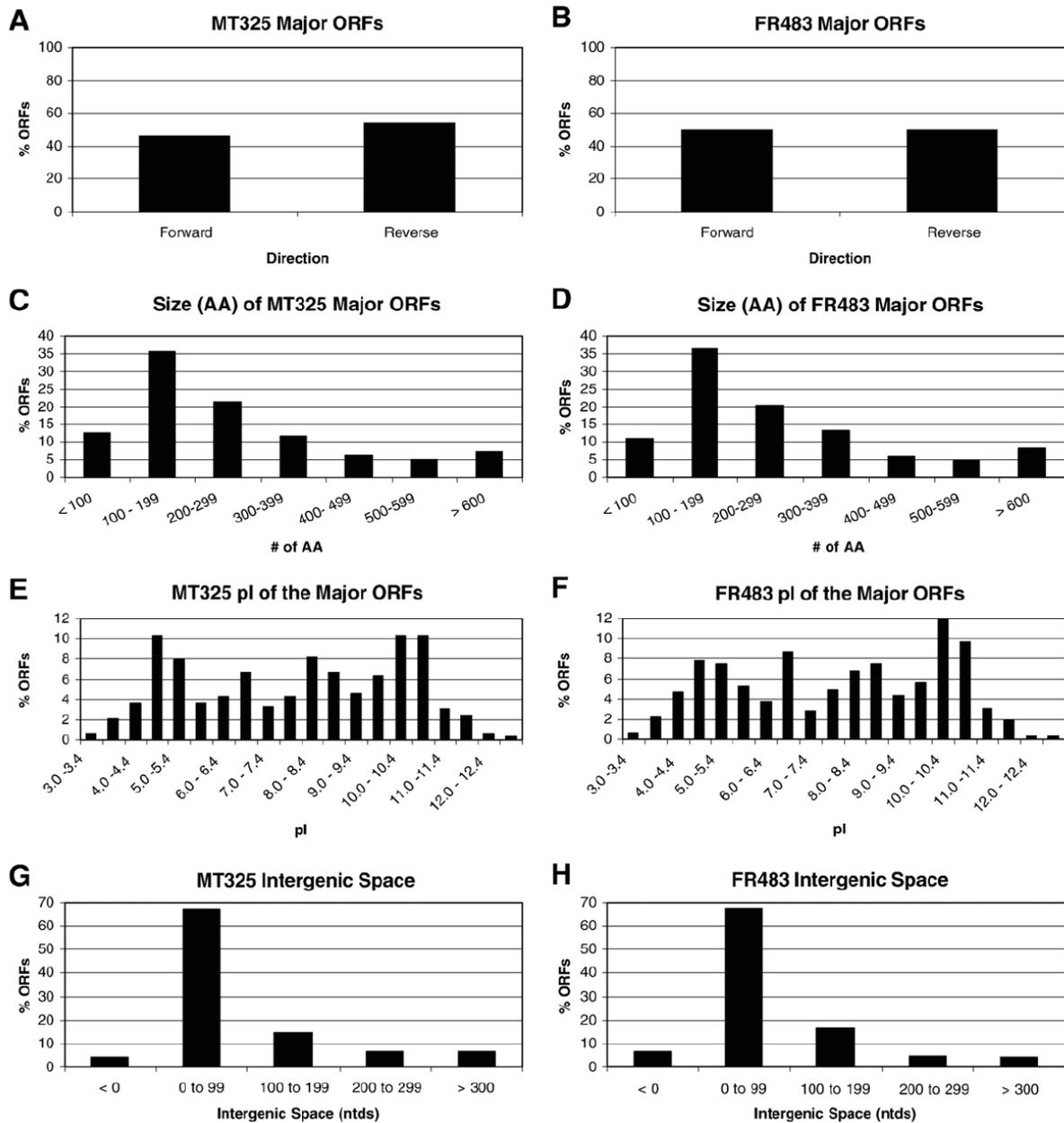


Figure 2. General characteristics of MT325 and FR483 major ORFs. Direction of the MT325 (A) and FR483 (B) ORFs. Size of the major MT325 (C) and FR483 (D) ORFs. Predicted isoelectric points of the major MT325 (E) and FR483 (F) ORFs. Intergenic space between major MT325 (G) and FR483 (H) ORFs.

sumably help neutralize the positively charged genomic DNA. However, the functions of the proteins that have *pI*s in the 4.5 range vary [e.g., the small subunit of ribonucleotide reductase (M653L and N651L), a Ser/Thr protein kinase (M221L and N223L), and a β -1,3 glucanase (M124L and N119L)]. Figs. 2G and H indicate the intergenic space between the major ORFs. In both MT325 and FR483, approximately 70% of the ORFs are separated by less than 100 nucleotides.

Annotation of the MT325 and FR483 genomes

Every ORF was compared with the non-redundant database at NCBI using the criteria described in the Materials and methods section. The Pfam and COG databases were used to

identify conserved domains and proteins within the MT325 and FR483 ORFs (Supplements 1 and 2). Gene maps of the MT325 and FR483 genomes illustrate the location of the putative genes (Fig. 3) and some of the ORFs are listed by their predicted metabolic function (Table 2). With the exception of the MT325-encoded potassium ion channel protein and the aqaglyceroporin protein (see below), none of the MT325 and FR483 proteins have been tested for function. However, we assume that any of the MT325- and FR483-encoded proteins that have functional PBCV-1 homologs are probably functional.

Seventy-nine to eighty-four percent of the major ORFs are homologous between any two of the Pbi viruses and the prototype virus, PBCV-1. However, approximately 95% of the genes are homologous between MT325 and FR483. This find-

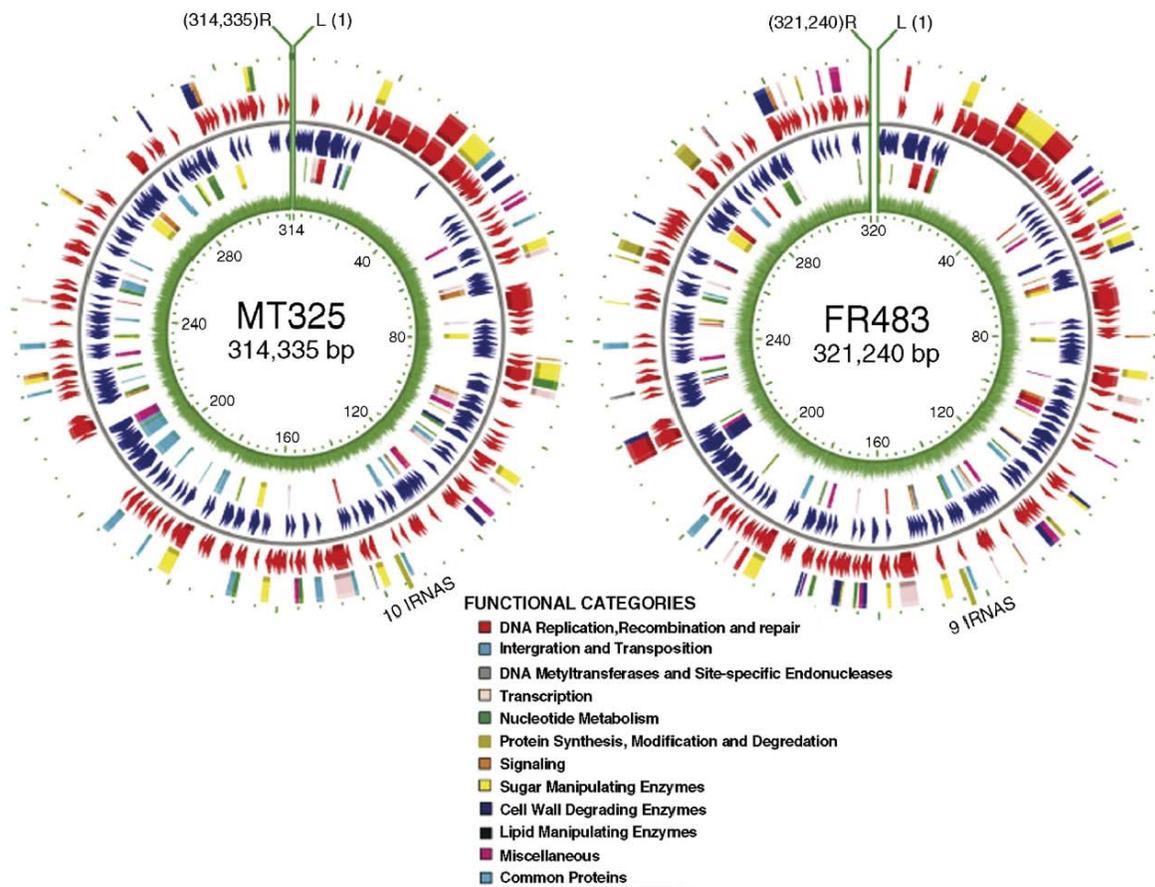


Figure 3. Map of the MT325 and FR483 genomes arranged in a circle. However, the genomes are linear molecules and the ends are depicted at the top of the figures as green lines (L and R represent the left and right ends of the genomes, respectively). The predicted ORFs are shown in both orientations and colored according to their functional category. The inner green circle represents the percent A + T across the genome using a 25-bp window.

ing suggests that most of the major ORFs from the chlorella viruses are essential for virus replication in nature and the genes are more conserved between the Pbi viruses. The average amino acid identity between homologous proteins from PBCV-1 and either MT325 or FR483 is 50%. There is an average of 86% amino acid identity between MT325 and FR483 homologs. MT325 and FR483 contain the 9 genes that are shared by all the NCLDV viruses (Iyer *et al.*, 2001 and Iyer *et al.*, 2006).

DNA replication and repair-associated proteins

MT325 and FR483 have several ORFs that are involved in either DNA replication, recombination, or repair, such as δ -DNA polymerase (M019L and N023L), superfamily III helicase (M674R and N674R), two ORFs resembling sliding clamp processivity factor (PCNA) proteins (M014R, M697L, N019R, and N687L), DNA primase (M664L and N662L), RNase H (M570L and N575L), type II DNA topoisomerase (M546R and N549R), replication factor C protein (M430L and N441L), and pyrimidine-specific glycosylase (M627L and N629L). Interestingly, both MT325 and FR483 lack an ATP-dependent ligase that is present in the NC64A viruses (Fitzgerald *et al.*, 2007).

The MT325 and FR483 type II DNA topoisomerases have approximately 40% amino acid identity with type II topoi-

somerases from several eukaryotic organisms. This enzyme is ATP-dependent and functions by threading a double-stranded DNA segment through a transient double-stranded break in the DNA (Roca, 1995). PBCV-1 encodes one of the smallest known type II DNA topoisomerases of 1061 amino acids (Lavrukhin *et al.*, 2000). It cleaves double-stranded DNAs approximately 30 times faster than the human type II DNA topoisomerase (Fortune *et al.*, 2001). However, the smallest characterized topoisomerase II enzyme is encoded by Pbi virus CVM-1 (Dickey *et al.*, 2005). The CVM-1 type II enzyme is 1058 amino acids in length and displays DNA cleavage activity that is approximately 50-fold faster than the human topoisomerase II. The MT325 and FR483 type II DNA topoisomerases are the same size as the CVM-1 enzyme (1058 amino acids) and have 96% and 67% amino acid identities to the CVM-1 and PBCV-1 topoisomerases, respectively.

Like PBCV-1, MT325 and FR483 encode two proteins that resemble PCNA-like proteins from other organisms. The MT325 and FR483 proteins are more similar to their homologs from other organisms than they are to each other. This finding suggests that the viral PCNA genes did not arise recently by gene duplication. PCNA interacts with proteins involved in not only DNA replication but also in DNA repair and post-replicative processing, such as DNA methylases and

Table 2

Putative ORFs group by their functional categories for MT325, FR483 and the prototype PBCV-1 virus

DNA Replication, Recombination and Repair						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
δ -DNA polymerase	A185R	913	M019L	907	N023L	937
Archaeo-eukaryotic primase	A468R	443	M664L	431	N662L	429
PCNA	A193L	262	M014R	264	N019R	264
	A574L	264	M697L	276	N687L	259
Replication factor C	A417L	429	M430L	360	N441L	360
RNase H	A399R	194	M570L	164	N575L	164
Helicase-Superfamily III	A456L	654	M674R	653	N674R	653
DNA Topoisomerase II	A583L	1061	M546R	1058	N549R	1058
ATP-dependent DNA ligase	A544R	298				
ATPase (PP-loop)	A554/556/557L	498	M389R	499	N404R	513
ATPase (DNA packaging)	A392R	258	M586L	255	N589L	255
Pyrimidine dimer-specific glycosylase	A050L	141	M627L	144	N629L	144
Exonuclease	A166R	268	M215R	279	N218R	288
Transcription						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Transcription factor TFIIB	A107L	290	M139L	312	N135L	312
Transcription factor TFIID	A552R	270	M266R	273	N271R	273
Transcription factor TFIIS	A125L	180	M163L	180	N165L	180
VLTF2-type transcription factor	A482R	215	M635R	215	N640R	215
Superfamily II helicase	A153R	459	M201R	454	N206R	454
	A241R	725	M225L	715	N227L	712
	A363R	811	M372R	1131	N383R	1124
mRNA guanylyltransferase	A103R	330	M133R	319	N128R	319
RNA triphosphatase	A449R	193	M399L	183	N416L	183
Histone H3, Lys 27 methylase	A612L	119	M727L	119	N719L	119
SWI/SNF chromatin remodeling complex	A189/192R	1299	M015L	1286	N020L	1290
SWI/SNF helicase	A548L	458	M272L	459	N277L	555
RNase III	A464R	275	M672L	266	N670L	266
Cytosine deaminase	A200R	118	M010L	119	N014L	119
Sugar Manipulation						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
D-lactate dehydrogenase	A053R	363	M026L	346		
GDP-D-mannose dehydratase	A118R	345				
Fucose synthase	A295L	317				
UDP-glucose 6-dehydrogenase	A609L	389	M719L	389	N712L	389
Glucosamine synthetase	A100R	595	M037R	593	N035R	644
Hyaluronan synthase	A098R	568	M128R	561	N124R	558
Cellulase precursor			M354R	438	N364R	438
Glycosyltransferase	A064R	638	M186R	520	N191R	520
	A111/114R	860	M467R	849	N472R	848
	A219/222/226R	677	M491R	382	N500R	382
	A473L	517	M721L	604	N715L	629
	A546L	321				
dTDP glucose pyrophosphorylase			M174L	458	N177L	458

Table 2 (continued)

Lipid Manipulation						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Acetyltransferase	A254R	158				
N-acetyltransferase	A654L	197	M758R	197	N747R	197
Glycerophosphoryl diesterase	A049L	219				
Lipoprotein lipase	A402R	227	M564L	228	N570L	229
Lysophospholipase	A271L	159				
Patatin-like phospholipase	A173L	288	M219L	274	N222L	274
Alkyl sulfatase					N003L	659
Cell Wall Degradation						
	PBCV-1		MT325		AR158	
	ORF	AA	ORF	AA	ORF	AA
Chitinase	A181/182R	830	M085R	523	N779R	805
	A260R	484	M791R	520		
Chitosanase	A292L	328	M091R	343	N087R	343
β & α 1,4 linked glucuronic lyase	A215L	321	M289R	368	N293R	364
β -1,3-glucanase	A094L	364	M124L	348	N119L	348
Nucleotide Metabolism						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Aspartate transcarbamylase	A169R	323				
Ribo. Reductase (small subunit)	A476R	324	M653L	326	N651L	348
Ribo. Reductase (large subunit)	A629R	771	M777L	765	N766L	765
Deoxynucleoside kinase	A416R	188	M425R	186	N438R	187
dCMP deaminase	A596R	142	M530L	144	N546L	144
dUTP pyrophosphatase	A551L	141	M264L	140	N269L	139
Thymidylate synthase X	A674R	216	M034L	214	N034L	214
Glutaredoxin	A438L	78	M241L	87	N241L	87
			M423R	86	N436R	86
Thioredoxin	A427L	119	M445L	116	N453L	116
			M448L	112	N457L	112
			M449L	104	N458L	104
Integration and Transposition						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Transposase	A625R	433				
Thr 6Fp DNA mobile protein	A121R	97	M150R	104	N152R	104
GIY-YIG endonuclease	A134L	165	M032L	229	N103R	311
	A287R	251	M111R	239	N186L	157
	A315L	246	M180L	158	N380R	230
	A351L	358	M288L	352	N387R	305
	A495R	221	M370R	267	N391R	248
	A539R	173	M460R	251	N471L	293
	A651L	230	M465L	287	N524R	168
			M502L	245	N559R	293
			M517R	168	N616L	381
			M683L	234		
HNH endonuclease	A087R	456	M069R	432	N187L	355
	A267L	314	M093L	377	N482R	318
	A354R	237	M578R	208	N584R	208
	A422R	342	M622L	349	N609L	354
	A478L	310			N706R	357
	A490L	310			N815L	355

Table 2 (continued)

DNA Restriction/Modification						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Adenine methyltransferase	A251R	326			N751L	458
	A581R	265				
Cytosine methyltransferase	A517L	344	M359L	342	N369L	342
	A530R	335			N748L	349
	A683L	367				
DNA restriction endonuclease	A252R	342				
	A579L	183				
Protein Synthesis, Modification and Degradation						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Translation elongation factor-3	A646L	141	M375R	179	N733R	901
	A666L	918	M742R	901		
Prolyl-4-hydroxylase	A085R	242	M401R	224	N418R	251
Thiol oxidoreductase	A465R	118	M670R	116	N668R	116
Protein disulfide isomerase	A448L	106	M403R	106	N420R	106
SKP-1 protein	A039L	151	M807R	155	N799R	155
Ring finger ubiquitin ligase	A481L	224	M629L	232	N633L	232
Ubiquitin C-terminal hydrolase	A105L	284	M137L	279	N132L	279
Zn metalloproteinase	A521L	392	M496L	205	N506L	205
	A604L	134				
Initiation factor 2			M488L	249	N496L	271
			M489R	180	N497R	202
Signalling						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Aquaglyceroporin			M030R	270		
Potassium channel protein	A250R	94	M183R	95		
Potassium ion transporter					N110R	660
Ligand-gated channel protein	A163R	433				
Glutamate receptor	A162L	411				
Ca ²⁺ transporting ATPase			M535L	871		
Dual specificity phosphatase	A305L	204	M313L	185	N318L	185
Serine/Threonine protein kinase	A034R	308	M143L	605	N140L	618
	A248R	308	M221L	300	N223L	300
	A277L	303	M543R	272	N548R	271
	A278L	610	M729L	583	N720L	561
	A282L	569	M733R	316	N724R	317
	A289L	283	M794R	270	N784R	270
	A614L	577				
	A617R	321				
Miscellaneous						
	PBCV-1		MT325		FR483	
	ORF	AA	ORF	AA	ORF	AA
Ornithine/Arginine decarboxylase	A207R	372	M307L	372	N312L	372
Agmatine iminohydrolase	A638R	359	M766L	363		
N-carbamoylput. amidohydrolase	A078R	298	M103L	296	N095L	296
Homospermidine synthase	A237R	518	M233L	502	N232L	504
Histidine decarboxylase	A598L	363	M601L	359	N600L	359
Monoamine oxidase	A217L	394	M283R	386	N289R	386
Amidase	A284L	279	M101R	279	N093R	279
Cu/Zn-superoxide dismutase	A245R	187	M099R	169	N091R	179
O-methyltransferase	A061L	209				
FkbM Methyltransferase			M177L	470	N119L	398
ABC transporter protein	A445L	462	M404R	463	N421R	463
ATPase (AAA+ class)	A044L	599				
Fibronectin binding protein	A180R	108	M789L	110	N777L	110

DNA transposases (Warbrick, 2000). Because the chlorella viruses encode proteins involved in both DNA repair and DNA methylation, their two PCNAs may serve different functions in their respective viral life cycles.

Transcription-associated proteins

No recognizable RNA polymerase or RNA polymerase components have been detected in any of the chlorella viruses that have been sequenced to date, including MT325 and FR483. This observation supports the idea that infectious viral DNAs are targeted to the nucleus and that host RNA polymerase(s) initiates viral transcription, possibly in conjunction with virion-packaged transcription factors. MT325 and FR483 encode at least four putative transcription factor-like elements: TFIIB (M139L and N135L), TFIID (M266R and N271R), TFIIS (M163L and N165L), and VLT2-type transcription factor (M635R and N640R). However, none of these factors are packaged in the PBCV-1 virion (Dunigan *et al.*, manuscript in preparation) and are unlikely to be packaged in the MT325 or FR483 virions. MT325 and FR483 encode two proteins that are involved in creating a mRNA cap structure, a mRNA guanylyltransferase (M133R and N128R) and an RNA triphosphatase (M399L and N416L). These two viruses also encode an RNase III (M672L and N670L) that presumably is involved in the processing of viral mRNAs and/or tRNAs.

In the immediate-early phase of infection, the host is re-programmed to transcribe viral RNAs, which in the case of PBCV-1, begins 5–10 min p.i. (Schuster *et al.*, 1986). It is not known how this process occurs, but histone methylation may be involved in inhibiting host transcription. PBCV-1 encodes a 119-amino acid protein with a SET domain (named 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 may be involved in repressing host transcription after PBCV-1 infection (Manzur *et al.*, manuscript in preparation). MT325 and FR483 each contain a PBCV-1 vSET gene homolog (M727L and N719L). In addition to the histone methyltransferase, MT325 and FR483 encode a putative SWF/SNF family helicase (M272L and N277L) and a SWI/SNF chromatin-remodeling complex subunit, OSA2 (M015L and N020L). Both proteins are also implicated in chromatin remodeling (*e.g.*, Kim and Clark, 2002).

Finally, MT325 and FR483, as well as all the chlorella viruses, encode a putative cytidine deaminase (M010L and N014L). This observation suggests that either some of the viral transcripts or host transcripts may undergo some post-transcriptional editing (Gerber and Keller, 2001).

Protein synthesis, modification and degradation

The chlorella viruses are the first viruses to encode a translation elongation factor (EF) (Yamada *et al.*, 1993). The viral proteins have approximately 45% amino acid identity to an EF-3 protein from fungi (Belfield and Tuite, 1993 and Chakraborty, 2001). The fungal protein stimulates EF-1 GTP-dependent binding of an amino acyl-tRNA to the A site of the ribosome. Like fungal EF-3 proteins, the chlorella virus pro-

teins have an ABC transporter family signature and two ATP/GTP-binding site motifs.

MT325 contains two ORFs that encode proteins homologous to EF-3 from chlorella viruses, whereas the FR483 genome contains one EF-3-like ORF. The 901-amino-acid MT325 and FR483 EFs (M742R and N733R) are similar in size and have 65% amino acid identity to their PBCV-1 homolog. MT325 also encodes a smaller EF (M375R, 179 amino acids) that is similar to the N-terminus of a chlorella virus CVK2 EF-3 homolog (39% amino acid identity) (Yamada *et al.*, 1993). This small ORF does not resemble any other proteins in the databases, including the MT325 M742R ORF and its function is unknown.

MT325 and FR483 have genes that encode proteins similar to those involved in post-translational modification, including prolyl-4-hydroxylase (M401R and N418R), protein kinases (see below), and glycosyltransferases (see below). MT325 and FR483 also encode a protein disulfide isomerase (M403R and N420R), a SKP-1 protein (M807R and N799R) and a thiol oxidoreductase (M670R and N668R). Additionally, the two viruses encode proteins involved in protein degradation including an ubiquitin C-terminal hydrolase (M137L and N132L), a ring finger ubiquitin ligase (M629L and N633L), and a Zn metalloproteinase (M496L and N506L).

tRNAs

The MT325 and FR483 genomes were analyzed for tRNAs using the tRNAscan-SE program (Lowe and Eddy, 1997). MT325 is predicted to encode 10 tRNAs: 2 for Asn and 1 each for Arg, Gly, Ile, Leu, Lys, Phe, Thr, and Tyr (Table 3). These 10 tRNAs are clustered in a region of the MT325 genome, nucleotide sequence 134,513 to 135,567. FR483 is predicted to encode 9 tRNAs: 2 for Asn and 1 each for Arg, Gly, Ile, Leu, Lys, Thr, and Tyr (Table 3). These 9 tRNAs are also clustered in the FR483 genome, nucleotide sequence 141,848 to 142,809. Presumably, the tRNAs are transcribed as a large polycistronic RNA and processed via intermediates to mature RNAs as they are in chlorella virus CVK2 (Nishida *et*

al., 1999). MT325 and FR483, with the exception of tRNA^{Phe} in MT325, code for the same tRNAs which are present in the same order. Only seven of the eleven tRNAs encoded by PBCV-1 are found in both MT325 and FR483. Therefore, three unique tRNAs are encoded by the Pbi viruses; tRNA^{Gly} and tRNA^{Thr} in MT325 and FR483 and one tRNA unique to MT325, tRNA^{Phe} (Table 3). Even though the orientation of the tRNA genes is the same in all five sequenced chlorella genomes, their order varies between the viruses. None of the tRNAs have a CCA sequence at the 3' end of the acceptor stem. Typically, these three nucleotides are added post-transcriptionally.

One tRNA, tRNA^{Tyr}, encoded by both the MT325 and FR483 genomes, contains a predicted 13-nucleotide intron from nucleotide 135,213 to 135,225 and 142,453 to 142,465, respectively. The insertion of a small intron in the tyrosine tRNA (anti-codon GTA) also occurs in NC64A viruses, PBCV-1 and NY-2A; however, tRNA^{Tyr} is absent from the AR158 genome (Fitzgerald *et al.*, 2007). Codon usage analyses of a few viral-encoded proteins indicate a strong correlation between the abundance of the viral-encoded tRNAs and their usage in viral proteins.

Nucleotide metabolism

MT325 and FR483 encode eleven enzymes involved in nucleotide metabolism. These enzymes are important given that the DNA concentration in PBCV-1-infected cells increases at least four-fold following infection (Van Etten *et al.*, 1984). Therefore, large quantities of dNTPs must be synthesized to support viral DNA replication. MT325 and FR483 encode the small (M653L and N651L) and large (M777L and M766L) subunits of ribonucleotide reductase, three thioredoxins (M445L, M448L, M449L, N453L, N457L, and N458L), two glutaredoxins (M241L, M423R, N241L, and N436R), dUTP pyrophosphatase (M264L and N269L), deoxycytidylate (dCMP) deaminase (M530L and N546L), and thymidylate synthase X (M034L and N034L). Interestingly, the two Pbi viruses MT325 and FR483 lack the gene for aspartate transcar-

Table 3
Comparison of the PBCV-1, MT325, and FR483 tRNA genes

tRNA	Anticodon	PBCV-1			MT325			FR483		
		tRNA # ^a	Start	End	tRNA # ^a	Start	End	tRNA # ^a	Start	End
Ile	TAT	2	165,022	165,094	1	134,513	134,585	1	141,848	141,920
Leu	TAA	3	165,118	165,200	2	134,610	134,693	2	141,945	142,028
Phe	GAA	–			3	134,718	134,790	–		
Arg	TCT	10	165,777	165,851	4	134,814	134,886	3	142,053	142,125
Gly	TCC	–			5	134,890	134,960	4	142,129	142,199
Asn	GTT	4	165,224	165,297	6	134,984	135,057	5	142,223	142,296
Asn	GTT	6	165,398	165,471	7	135,080	135,151	6	142,320	142,393
Tyr ^b	GTA	8	165,572	165,658	8	135,174	135,259	7	142,416	142,501
Lys	CTT	7	165,547	165,547	9	135,262	135,334	8	142,504	142,576
Thr	CGT	–			10	135,496	135,567	9	142,738	142,809
Leu	CAA	1	164,916	164,996	–			–		
Lys	CTT	5	165,301	165,373	–			–		
Lys	TTT	9	165,681	165,753	–			–		
Val	AAC	11	165,885	165,957	–			–		

^a Order of the tRNA genes in the genome.

^b tRNA gene contains an intron.

bamylase that is present in the three NC64A viruses (Fitzgerald *et al.*, 2007).

Two MT325 and FR483-encoded enzymes, dUTP pyrophosphatase and dCMP deaminase, produce dUMP, the substrate for thymidylate synthetase. The chlorella viruses, including MT325 and FR483, lack a traditional thymidylate synthetase A. Instead, they encode a protein that is a member of a newly recognized family of flavin-dependent thymidylate synthetases, ThyX (Graziani *et al.*, 2004 and Myllykallio *et al.*, 2002).

Unlike the NC64A viruses that encode a single glutaredoxin and thioredoxin, MT325 and FR483 each have two glutaredoxins, with 28–30% amino acid identity between the two glutaredoxins encoded by the same virus; however, M423R and N436R have 93% amino acid identity while M241L and N241L have 95% amino acid identity to one another. The two Pbi viruses each encode three thioredoxins with 26–34% amino acid identities between the thioredoxins encoded by the same virus. However, there is 98% amino acid identity between MT325 and FR483 thioredoxin homologs.

Protein kinases, phosphatases, and channel proteins

Both MT325 and FR483 encode several Ser/Thr protein kinases (Table 2) and a protein that resembles a dual-specificity phosphatase (M313L and N318L). The large number of viral-encoded proteins involved in phosphorylation/dephosphorylation suggests that they are involved in one or more signal transduction pathways that are vital for viral replication.

The algal viruses are the first viruses to encode K⁺ channel proteins (called Kcv) (Kang *et al.*, 2004). The MT325 Kcv (M183R) conducts a current when expressed in *Xenopus* oocytes (Gazzarrini *et al.*, 2006). It is interesting, however, that the predicted structure of the 95-amino acid MT325 Kcv differs from the 94-amino acid PBCV-1 Kcv (Gazzarrini *et al.*, 2006), yet both are capable of conducting currents in heterologous systems. FR483 lacks a Kcv homolog. However, FR483 encodes a putative potassium ion transporter protein (N110R).

MT325 is unique among the chlorella viruses because it also encodes a functional aquaglyceroporin channel (called AQPV, M030R) (Gazzarrini *et al.*, 2006). Genes encoding AQPV are absent in the three sequenced NC64A viruses and in Pbi virus FR483. MT325 also encodes an 871-amino-acid Ca²⁺-transporting ATPase (M535L). To date, a gene encoding this protein has only been detected in one NC64A virus, AR158 (Fitzgerald *et al.*, 2007).

Sugar- and lipid-manipulating proteins

MT325, and FR483 encode several proteins that have high identities to enzymes involved in manipulating sugars, synthesizing polysaccharides, or transferring sugars to proteins (Yamada *et al.*, 2006). Therefore, phycodnaviruses are unusual in that, of the five chlorella viruses sequenced, they all encode enzymes involved in sugar metabolism, but not always the same enzymes. Like PBCV-1, MT325 and FR483 each encode three enzymes, glucosamine synthetase (M037R and N035R), UDP-glucose 6-dehydrogenase (M719L and N712L),

and hyaluronan synthase (M128R and N124R), which are involved in the synthesis of hyaluronan, a linear polysaccharide composed of alternating β -1,4 glucuronic acid and β -1,3-*N*-acetylglucosamine residues (DeAngelis *et al.*, 1997 and Landstein *et al.*, 1998). MT325 and FR483 also have four putative glycosyltransferases (Table 2) that are presumably involved in glycosylating their major capsid proteins.

MT325 and FR483 encode one enzyme, dTDP–glucose pyrophosphorylase (M174L and N177L), that is absent in the three NC64A viruses. MT325 and FR483 lack the two enzymes involved in converting GDP-d-mannose to GDP-l-fucose, GDP-d-mannose 4, 6-dehydratase and GDP-4-keto-6-deoxy-d-mannose epimerase/reductase that exist in the three NC64A viruses (Fitzgerald *et al.*, 2007). Furthermore, a d-lactate dehydrogenase is encoded by MT325 (M026L) but is absent in FR483.

Finally, MT325 and FR483 encode 3 enzymes that are predicted to be involved in lipid metabolism, *N*-acetyl-transferase (M758R and N747R), patatin phospholipase (M219L and N222L) and lipoprotein lipase (M564L and N570L). These same enzymes are encoded by PBCV-1; however, none of them have been investigated for enzymatic activity. FR483 encodes an alkyl sulfatase (N003L), which is not present in the other four sequenced chlorella virus genomes.

Cell wall-degrading enzymes

MT325 has five ORFs and FR483 has four ORFs that may be involved in degrading *Chlorella* Pbi cell walls either during virus infection or virus release. These proteins include two chitinases (M085R and M791R) in MT325 and one chitinase (N779R) in FR483, a chitosanase (M091R and N087R), a β -1,3-glucanase (M124L and N199L), and a novel polysaccharide lyase that cleaves chains of either β - or α -1,4 linked glucuronic acids (M289R and N293R). All of the cell wall-degrading enzymes have functional homologs in PBCV-1 (Yamada *et al.*, 2006). Unlike most algae, *Chlorella* Pbi contains chitin in its cell wall (Kapaun *et al.*, 1992 and Kapaun and Reisser, 1995).

Restriction–modification enzymes

Chlorella viruses contain different levels of 5-methylcytosine (5 mC) and N⁶-methyladenine (6 mA) in their genomes (Van Etten *et al.*, 1991). Therefore, it is not surprising that these viruses encode 5 mC and 6mA DNA methyltransferases, *e.g.*, NY-2A and PBCV-1 encode 18 and 5 DNA methyltransferases, respectively (Fitzgerald *et al.*, 2007). MT325 and FR483 only encode a few DNA methyltransferases. MT325 encodes one methyltransferase (M359L), which methylates cytosines. FR483 encodes three methyltransferases; one (N751L) is a 6 mA methyltransferase and two (N369L and N748L) are 5 mC methyltransferases.

Integration and transposition enzymes

Homing endonucleases are rare DNA-cleaving enzymes that are typically encoded by introns and inteins and cleave DNA infrequently. Homing endonucleases are classified into

four families (Belfort and Roberts, 1997). MT325 encodes 14 homing endonucleases; ten are members of the GIY-YIG family and four are members of the HNH family. FR483 encodes 15 homing endonucleases; nine are members of the GIY-YIG family and six are members of the HNH family (Table 2). It is unknown if the putative MT325- and FR483-encoded homing endonucleases are functional and/or if they have an essential role in the viral life cycle. Unlike the NC64A viruses, the two Pbi viruses do not encode transposases.

Polyamine biosynthetic enzymes

PBCV-1 was the first virus to encode polyamine biosynthetic enzymes, including two pathways for the synthesis of putrescine. Three of the four PBCV-1 enzymes, which are known to be functional (Van Etten, 2003), are also encoded by MT325 and FR483. These enzymes are an ornithine/arginine decarboxylase (ODC) (M307L and N312L), an N-carbamoylputrescine amidohydrolase (M103L and N095L), and a homospermidine synthase (M233L and N232L). However, agmatine iminohydrolase (M766L) is only encoded by MT325. ODC catalyzes the first and rate-limiting step in polyamine biosynthesis, the decarboxylation of ornithine to putrescine (Davis *et al.*, 1992). The PBCV-1 ODC is the smallest ODC characterized to date (372 amino acids) (Morehead *et al.*, 2002). Unexpectedly, the PBCV-1 enzyme decarboxylates arginine more efficiently than ornithine (Shah *et al.*, 2004). MT325 and FR483 each encode a 372-amino-acid ODC that has 63% amino acid identity to its PBCV-1 homolog. The product of arginine decarboxylation is agmatine. PBCV-1 and MT325 each encode two enzymes, agmatine iminohydrolase and N-carbamoylputrescine amidohydrolase, which convert agmatine to putrescine (Nakada *et al.*, 2001) (Baumann *et al.*, submitted for publication). FR483 encodes N-carbamoylputrescine amidohydrolase, but lacks agmatine iminohydrolase.

Homospermidine synthase synthesizes homospermidine, a rare polyamine, from two molecules of putrescine (Kaiser *et al.*, 1999). The MT325 and FR483 homospermidine synthase homologs have a 69% amino acid identity to the functional PBCV-1-encoded enzyme.

Both PBCV-1 and the two Pbi viruses, MT325 and FR483, encode two other putative enzymes involved in amine metabolism, monoamine oxidase (M283R and N289R) and histidine decarboxylase (M601L and N600L). The finding that all of the *Chlorella* viruses sequenced to date encode five of these six proteins suggests that they must serve an important role(s) in the viral life cycle.

Miscellaneous proteins

MT325 and FR483 also encode several other putative proteins, including a Cu/Zn superoxide dismutase (M099R and N091R), an amidase (M101R and N093R), a fibronectin-binding protein (M789L and N777L), and an ABC transporter protein (M404R and N421R). The Cu/Zn superoxide dismutases have 76% amino acid identity with a PBCV-1-encoded homolog. This enzyme functions to convert superoxide radicals into molecular oxygen and hydrogen peroxide (Bannister *et*

al., 1987). Presumably, the MT325- and FR483-encoded enzyme reduces light-induced superoxide accumulation. Finally, MT325 and FR483 each encode an enzyme (M177L and N180L) that has some similarity to an enzyme from several *Mycobacterium* species that methylates rhamnose (Jeevarajah *et al.*, 2002).

MT325 and FR483 structural proteins

MT325 ORF M463L and FR483 ORF N470L have the highest amino acid identities (96–99%) and are approximately the same size as the characterized major coat protein, Vp49, from Pbi virus CVG-1 (Plugge *et al.*, 1999). Therefore, we assume that M463L and N470L are the MT325 and FR483 major capsid proteins, respectively. Four additional MT325 ORFs (M078R, M269L, M381R, and M748L) and 5 FR483 ORFs (N074R, N254L, N274L, N395R, and N737L) have 31–43% amino acid identities to Vp49 from CVG-1.

Identification of gene families

A total of 68 of the MT325 ORFs resemble 1 or more other MT325 ORFs based on a blastp search with an *E*-value of less than 10^{-10} , suggesting that they might be either gene families or gene duplications. This number is somewhat misleading since some of these ORFs are grouped as families because they contain a common conserved domain, *e.g.*, ankyrin repeats or a PAPK repeat, even though the amino acid sequence similarity of the rest of the protein is small. A total of 13 families have two members, 3 families have three members, 1 family has five members, 1 family has six members, 2 families have seven members, and 1 family has eight members.

A similar analysis indicates that 68 of the FR483 ORFs resemble 1 or more other FR483 ORFs. A total of 10 families have two members, 3 families have three members, 2 families have four members, 1 family has six members, 1 family has seven members, 1 family has eight members, and 1 family has ten members.

Conclusions

The 314,335-bp MT325 genome and the 321,240-bp FR483 genome, the smallest *Chlorella* viral genomes sequenced to date and the first viruses to be sequenced that infect *Chlorella* Pbi, are predicted to encode 331 and 335 proteins as well as 10 and 9 tRNAs, respectively. The putative protein-coding genes are relatively evenly distributed on both strands and intergenic space is minimal. Approximately 40% of the gene products have been identified; some have prokaryotic characteristics whereas other resemble eukaryotic proteins. Approximately 82% of these Pbi protein-coding genes have homologs in PBCV-1, suggesting that these proteins are important in viral replication. However, some of the gene products are unique to MT325 (*e.g.*, aquaglyceroporin), unique to FR483 (*e.g.*, alkyl sulfatase and potassium ion transporter) or unique to the Pbi viruses (*e.g.*, dTDP-glucose pyrophosphorylase). Finally, genes encoding several proteins in the three NC64A viruses are absent in the two Pbi viruses in-

cluding, an ATP-dependent DNA ligase, aspartate transcarbamylase, GDP-d-mannose dehydratase, fucose synthase, glycerophosphoryl diesterase and lysophospholipase.

Materials and methods

Viral DNA isolation and sequencing

Plaque-forming viruses MT325 and FR483 were isolated from fresh-water samples collected in Montana state (August, 1996) and France (June, 1997), respectively. The MT325 and FR483 host, *Chlorella* Pbi, was grown on FES medium (Reisser *et al.*, 1986). The MT325 and FR483 viruses were produced, purified, and the viral DNAs were isolated using methods and protocols developed for PBCV-1 (Van Etten *et al.*, 1981, Van Etten *et al.*, 1983 and Van Etten *et al.*, 1991). MT325 genomic DNA was sequenced to 8-fold coverage and assembled at The Institute for Genomic Research (TIGR). The sequence of the FR483 genome was determined and assembled at Agencourt Biosciences Co., Beverly, MA. The FR483 genome was sequenced to a 8-fold coverage with a minimum quality value of 20 per base.

Genomic sequence analysis

A potential protein-encoding region was defined as a continuous stretch of DNA that translated into a polypeptide initiated by an ATG translation start codon and extended for 64 or more codons using the standard genetic code. The ORF Finder program (http://bioinformatics.org/sms/orf_find.html) was used to identify all potential ORFs that met this criterion. The ORFs were numbered consecutively starting at the beginning of the genome (as determined by alignment with the PBCV-1 genome). The letter R or L following the number indicates that the orientation of the putative ORF is either left-to-right or right-to-left, respectively.

Dot plots of the virus major ORFs were created to determine the orientation of the MT325 and FR483 genomes relative to the PBCV-1 genome. Every major ORF was individually plotted against the PBCV-1 major ORFs using blastp (protein vs. protein). Similarities between the two ORFs with *E*-values of $< 10^{-3}$ are presented. Transfer RNA genes were identified using the tRNAscan-SE program developed by Lowe and Eddy at Washington University School of Medicine (Lowe and Eddy, 1997). Gene families were identified when a major ORF had an *E*-value of less than 10^{-10} to another ORF within the same genome.

Analysis with public databases

Each ORF identified was used in a search for homologs using the protein-protein BLAST (blastp) program (Altschul *et al.*, 1990) against the non-redundant (NR) protein databases at NCBI. The criterion used to search the NR database was as follows: Scoring matrix = blosum62. Each putative identified ORF was scanned for potential functional attributes using Pfam version 18.0 (Finn *et al.*, 2006). Every identified ORF

was additionally scanned to determine if it belonged to a particular COG. In each of the analyses the top 10 results were recorded regardless of the *E*-values.

Nucleotide sequence accession number

The MT325 and FR483 DNA sequences have been deposited in the GenBank database (accession numbers DQ491001 and DQ890022, respectively) and the sequences can also be found at <http://greengene.uml.edu/>

Acknowledgments

We thank James Gurnon for preparing the MT325 DNA and Ming Kang for preparing the FR483 DNA. This investigation was supported in part by National Science Foundation grant EF-0333197 (M.G. and J.V.E.), by National Institutes of Health grant GM32441 (J.V.E.) and by the Center of Biomedical Research Excellence program of the National Center for Research Resources Grant P20-RR15635 (J.V.E.).

Supplementary data

Supplementary data associated with this article is archived in this repository as 4 separate files: Appendices A–D. Each document, in spreadsheet format, shows Gene Name, Genome Position, A.A. length, Peptide Mw, pI, CDD Hit Number, COGs, COG Definition, Bit Score, *E*-value, % Identity, % Positive, Query from-to, Hit from-to, BLASTp Hit Number, Hit Accession, BLASTp Definition, Bit Score, *E*-value, % Identity, % Positive, Query from-to, and Hit from-to.

Appendix A: Gene names m002R through m843L

Appendix B: Gene names M001L through M807R

Appendix C: Gene names n001L through n849R

Appendix D: Gene names N003L through N847R

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Bannister, J.V., Bannister, W.H., Rotilio, G., 1987. Aspects of the structure, function, and applications of superoxide dismutase. *CRC Rev. Biochem.* 22, 111–180.
- Baumann, S., Sander, A., Gurnon, J.R., Yanai-Balser, G., Van Etten, J.L., Piotrowski, M., submitted for publication. *Chlorella* viruses encode a complete polyamine biosynthesis pathway. *Virology*.
- Belfield, G.P., Tuite, M.F., 1993. Translation elongation factor 3: a fungus-specific translation factor? *Mol. Microbiol.* 9, 411–418.
- Belfort, M., Roberts, R.J., 1997. Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* 25, 3379–3388.
- Bubeck, J.A., Pfizner, A.J.P., 2005. Isolation and characterization of a new type of chlorovirus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. *J. Gen. Virol.* 86, 2871–2877.

- Chakraborty, K., 2001. Translational regulation by ABC systems. *Res. Microbiol.* 152, 391–399.
- Davis, R.H., Morris, D.R., Coffino, P., 1992. Sequestered end products and enzyme regulation: the case of ornithine decarboxylase. *Microbiol. Rev.* 56, 280–290.
- 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.
- Delaroque, N., Muller, 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.
- Dickey, J.S., Choi, T.-J., Van Etten, J.L., Osheroff, N., 2005. Chlorella virus Marburg topoisomerase II: high DNA cleavage activity as a characteristic of chlorella virus type II enzymes. *Biochemistry* 44, 3899–3908.
- Dunigan, D.D., Fitzgerald, L.A., Van Etten, J.L., 2006. Phycodnaviruses: a peek at genetic diversity. *Virus Res.* 117, 119–132.
- Finn, R.D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S.R., Sonnhammer, E.L., Bateman, A., 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* 34, D247–D251.
- Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Nierman, W.C., Van Etten, J.L., 2007. The sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A. *Virology* 358, 472–484.
- Fortune, J.M., Lavruchin, 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.
- Gazzarrini, S., Kang, M., Epimashko, S., Van Etten, J.L., Dainty, J., Thiel, G., Moroni, A., 2006. Chlorella virus MT325 encodes water and potassium channels that interact synergistically. *Proc. Natl. Acad. Sci. U.S.A.* 103, 5355–5360.
- Gerber, A.P., Keller, W., 2001. RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.* 26, 376–384.
- Grabherr, R., Strasser, P., Van Etten, J.L., 1992. The DNA polymerase gene from chlorella viruses PBCV-1 and NY-2A contains an intron with nuclear splicing sequences. *Virology* 188, 721–731.
- Graziani, S., Xia, Y., Gurnon, J.R., Van Etten, J.L., Leduc, D., Skouloubris, S., Myllykallio, H., Liebl, U., 2004. Functional analysis of FAD-dependent thymidylate synthase ThyX from *Paramecium bursaria* chlorella virus-1. *J. Biol. Chem.* 279, 54340–54347.
- 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., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* 117, 156–184.
- Jeevarajah, D., Patterson, J.H., McConville, M.J., Billman-Jacobe, H., 2002. Modification of glycopeptidolipids by an O-methyltransferase of *Mycobacterium smegmatis*. *Microbiology* 148, 3079–3087.
- Kaiser, A., Vollmert, M., Tholl, D., Graves, M.V., Gurnon, J.R., Xing, W., Lisec, A.D., Nickerson, K.W., Van Etten, J.L., 1999. Chlorella virus PBCV-1 encodes a functional homospermidine synthase. *Virology* 263, 254–262.
- Kang, M., Moroni, A., Gazzarrini, S., DiFrancesco, D., Thiel, G., Severino, M., Van Etten, J.L., 2004. Small potassium ion channel proteins encoded by chlorella viruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5318–5324.
- 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.
- Kapaun, E., Loos, E., Reisser, W., 1992. Cell wall composition of virus-sensitive symbiotic chlorella species. *Phytochemistry* 31, 3103–3104.
- Kapaun, E., Reisser, W., 1995. A chitin-like glycan in the cell wall of a *Chlorella* sp. (Chlorococcales, Chlorophyceae). *Planta* 197, 577–582.
- Kawakami, H., Kawakami, N., 1978. Behavior of a virus in a symbiotic system, *Paramecium bursaria*-zoochlorella. *J. Protozool.* 25, 217–225.
- Kim, Y., Clark, D.J., 2002. SWI/SNF-dependent long-range remodeling of yeast HIS3 chromatin. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15381–15386.
- 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.
- Lavruchin, 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.
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
- Manzur, K.L., Farooq, A., Zeng, L., Plotnikova, O., Koch, A.W., Zhou, M.-M., 2003. A dimeric viral SET domain methyltransferase specific to Lys27 of histone H3. *Nat. Struct. Biol.* 10, 187–196.
- Meints, R.H., Van Etten, J.L., Kuczmariski, D., Lee, K., Ang, B., 1981. Viral infection of the symbiotic chlorella-like alga present in *Hydra viridis*. *Virology* 113, 698–703.
- 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.
- Myllykallio, H., Lipowski, G., Leduc, D., Filee, J., Forterre, P., Liebl, U., 2002. An alternative flavin-dependent mechanism for thymidylate synthesis. *Science* 297, 105–107.
- Nakada, Y., Jiang, Y., Nishijyo, T., Itoh, Y., Lu, C.D., 2001. Molecular characterization and regulation of the *aguBA* operon, responsible for agmatine utilization in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* 183, 6517–6524.
- Nishida, K., Kawasaki, T., Fujie, M., Usami, S., Yamada, T., 1999. Aminoacylation of tRNAs encoded by chlorella virus CVK2. *Virology* 263, 220–229.
- Pluge, B., Becker, B., Wolf, A.H., 1999. Several genes in chlorella virus strain CVG-1 encode putative virion components. *J. Gen. Virol.* 80, 1067–1072.

- 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.
- Reisser, W., Becker, B., Klein, T., 1986. Studies on ultrastructure and host range of a chlorella attacking virus. *Protoplasma* 135, 162–165.
- Roca, J., 1995. The mechanisms of DNA topoisomerases. *Trends Biochem. Sci.* 20, 156–160.
- 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.
- 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.
- 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.
- Suttle, C.A., 2005. Viruses in the sea. *Nature* 437, 356–361. Van Etten, J.L., 2003. Unusual life style of giant chlorella viruses. *Annu. Rev. Gen.* 37, 153–195.
- Van Etten, J.L., Meints, R.H., Burbank, D.E., Kuczmarski, D., Cuppels, D.A., Lane, L.C., 1981. Isolation and characterization of a virus from the intracellular green alga symbiotic with *Hydra viridis*. *Virology* 113, 704–711.
- Van Etten, J.L., Meints, R.H., Kuczmarski, D., Burbank, D.E., Lee, K., 1982. Viruses of symbiotic chlorella-like algae isolated from *Paramecium bursaria* and *Hydra viridis*. *Proc. Natl. Acad. Sci. U.S.A.* 79, 3867–3871.
- 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., 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., Lane, L.C., Meints, R.H., 1991. Viruses and viruslike particles of eukaryotic algae. *Microbiol. Rev.* 55, 586–620.
- Warbrick, E., 2000. The puzzle of PCNA's many partners. *BioEssays* 22, 997–1006.
- 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., Rabinowitsch, 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., 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.
- Yamada, T., Fukuda, T., Tamura, K., Furukawa, S., Songsri, P., 1993. Expression of the gene encoding a translational elongation factor 3 homolog of chlorella virus CVK2. *Virology* 197, 742–750.
- Yamada, T., Tamura, K., Aimi, T., Songsri, P., 1994. Self-splicing group I introns in eukaryotic viruses. *Nucleic Acids Res.* 22, 2532–2537.
- Yamada, T., Onimatsu, H., Van Etten, J.L., 2006. Chlorella viruses. In: Maramorosch, K., Shatkin, A. (Eds.), *Advances in Virus Research*, vol. 66. Elsevier Inc., pp. 293–336.
- Zhang, Y., Adams, B., Sun, L., Burbank, D.E., Van Etten, J.L., 2001. Intron conservation in the DNA polymerase gene encoded by chlorella viruses. *Virology* 285, 313–321.