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

Dissertations and Theses in Biological Sciences | Biological Sciences, School of

Fall 12-4-2009

Transcription analysis of the chlorovirus Paramecium bursaria chlorella virus-1

Giane M. Yanai
University of Nebraska at Lincoln, gianeyanai@yahoo.com

Follow this and additional works at: http://digitalcommons.unl.edu/bioscidiss

 Part of the Genomics Commons, and the Virology Commons

Yanai, Giane M., "Transcription analysis of the chlorovirus Paramecium bursaria chlorella virus-1" (2009). *Dissertations and Theses in Biological Sciences*. 5.
http://digitalcommons.unl.edu/bioscidiss/5

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

TRANSCRIPTION ANALYSIS OF THE CHLOROVIRUS *PARAMECIUM BURSARIA* CHLORELLA VIRUS-1

by

Giane M. Yanai

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Biological Sciences

Under the Supervision of Professor James L. Van Etten

Lincoln, Nebraska

December, 2009

TRANSCRIPTION ANALYSIS OF THE CHLOROVIRUS *PARAMECIUM BURSARIA* CHLORELLA VIRUS-1

Giane M. Yanai, Ph.D.

University of Nebraska, 2009

Adviser: James Van Etten

Paramecium bursaria chlorella virus (PBCV-1), a member of the family Phycodnaviridae, is a large dsDNA, plaque-forming virus that infects the unicellular green alga *Chlorella* NC64A. The 331 kb PBCV-1 genome is predicted to encode 365 proteins and 11 tRNAs. To follow global transcription during PBCV-1 replication, a microarray containing 50-mer probes to the PBCV-1 365 protein-encoding genes (CDS) was constructed. Competitive hybridization experiments were conducted employing cDNAs from poly A-containing RNAs obtained from cells at seven time points after virus infection. The results led to the following conclusions: i) the PBCV-1 replication cycle is temporally programmed and regulated; ii) 360 (99%) of the arrayed PBCV-1 CDSs were expressed at some time in the virus life cycle in the laboratory; iii) 227 (62%) of the CDSs were expressed before virus DNA synthesis begins; iv) these 227 CDSs were grouped into two classes: 127 transcripts disappeared prior to initiation of virus DNA synthesis (considered early) and 100 transcripts were still detected after virus DNA synthesis begins (considered early/late); v) 133 (36%) of the CDSs were expressed after virus DNA synthesis begins (considered late); vi) expression of most late CDSs is

inhibited by adding the DNA replication inhibitor, aphidicolin, prior to virus infection. This study provides the first comprehensive evaluation of virus gene expression during the PBCV-1 lifecycle.

To my parents, Izumi and Teruyo Yanai, for the past and present.

To my husband, Rob Balser, for the present and future.

ACKNOWLEDGMENTS

First and foremost, I would like to express my most sincere gratitude to my advisor, Dr. James L. Van Etten, for accepting me in his lab and for introducing me to the fascinating world of Chlorella viruses. Thank you for your guidance, help, support, and constant encouragement. Your enthusiasm and passion for your research have undoubtedly changed my perspective of science and career. You are an admirable professional and an extraordinary person and I hope you know how grateful I am to have worked for you.

Second, I would like to thank my committee members, Dr. Steven Harris, Dr. Roy French, Dr. Rosevelt Pardy, and Dr. Andrew Benson for their support and suggestions that greatly improved my research. Thank you, Dr. Harris and Dr. Benson, for serving as readers as well.

In special, I thank Dr. Garry Duncan, who started this work before I joined the lab. He graciously allowed me to take over the project and encouraged me through all of the steps in the development of this research.

I consider myself very lucky to have had the opportunity to work with many amazing scientists in the Van Etten lab. I cannot thank enough James Gurnon, Dr. Irina Agarkova, and Dr. David Dunigan for all their help, advice, support, and their great friendship. I also thank Gentry Lewis for her help when we took classes together and her friendship, and Dr. Janet Rowe for being so friendly and helpful during the preparation of this dissertation. To Dr. Les Lane, Jason Vitek, and many past members, Timo Greiner,

Swetha Tati, Cristina Brovont, Casey Holmes, Aberlardo Vernaza, Dr. Kwanhong Lee, Dr. Yuanzheng Zhang, Dr. Paola Valbuzzi, Parakkottil Madhu, Andrea (Van Etten) McClymont, thanks for your help and friendship. It was a pleasure to work with Dr. Lisa Fitzgerald and Phil Boucher and I thank them for their time and support. Thank you all for contributing for an enjoyable working atmosphere.

My appreciation goes to Dr. James Eudy, Justin Jackson, Tong Ying, and other members of the Microarray Core facility, for all their help. I also thank Dr. Dong Wang for the statistical guidance. I am thankful for Susan Kuczmarski, Eva Bachman, Xiao Li, Dr. Yuannan Xia, and Dr. Jean-Jack Riethoven for their valuable help.

To the many members of the Plant Pathology department, in particular Dr. James Steadman, Dr. Anne Vidaver, and Dr. Amit Mitra, thank you for the financial support and constant encouragement. I also thank Margaret Denning and Mary Jo Haverkamp for their help and assistance. To all the students in Plant Pathology, especially Camile Semighini, Bill Rittenour, Leny Galvez, and John Nopsa, thank you for a great time in graduate school. Special thanks to the person responsible for brightening my day early in the morning, Karen Dettman. I would like to thank the many friends I have made in the Morrison building, in particular Leticia da Silva, Chris Bohl, Sung-Hwan Kang, Jerry McEntee, and Michele Malchow.

I am so glad for meeting new friends while living in Lincoln. I would like to thank my dearest friend Tina (Fang Tian). She filled my life with laughter, help, hope, perspective, and support. I also want to thank the many Brazilian friends living in Lincoln, who always brought a little bit of “Brazil” to the United States.

Words are not enough to express how thankful I am for my parents, Izumi and Teruyo Yanai. They are a continuous source of love, support, and encouragement. Thank you for being so supportive when I decided to come to the United States to continue my education. Thank you for the many hours over the phone making sure that I am doing well. Thank you for always believing and trusting me to make my own decisions. I am so proud of you in so many ways and I will live the rest of my life trying to follow your steps. Although I miss you greatly every single day, I know distance is no barrier for our strong bond – I carry you both in my heart! I also would like to thank my two brothers, Henrique (and his wife, Cristina) and Tercio, for making me laugh like always and more importantly, for taking care of our parents for me. I could not forget to thank my two lovely grandmothers, Maria Murakami and Yukiko Iwata, for being always so sweet over the phone (even though they cannot hear me!) and especially for being such an example of life. I am so lucky to have a wonderful family, including many aunts, uncles, cousins, and little nieces and nephews. Each of you have helped me in so many different ways along my journey and I thank you all from the bottom of my heart.

I want to express my deepest appreciation to my new family, especially my father and mother-in-law, Steve and Joann Petersen. Thank you for making me feel so loved and welcome in your family. I am so grateful for your advice, your help, and particularly for being so positive about my future. I also thank my new sisters-in-law, Melissa and Kristin (and their husbands), and my new brother-in-law, David, for making me feel like part of the family and giving me such adorable nieces and nephews. I thank Mark and Sandy Scheller as well for being so special and for bringing their kids into my life. To

Grandpa (*in memoriam*) and Grandma Storch, thank you for building such a wonderful family that I am now proud to be part of. I miss you, Grandpa!

Lastly, I want to thank the most important person during my years as a graduate student, my husband and best friend, Rob. He has been my endless source of love, support, and encouragement. With his amazing ability of making me laugh every single day, he has made my days easier and more enjoyable. Thank you for moving to Nebraska for me. I know one year became four years, but I think our love only grew stronger in this time. Thank you for being in my life.

TABLE OF CONTENTS

CHAPTER II

MICROARRAY ANALYSIS OF CHLORELLA VIRUS *PARAMECIUM BURSARIA* CHLORELLA VIRUS-1 TRANSCRIPTION

CHAPTER III

PUTATIVE GENE PROMOTER SEQUENCES IN THE CHLORELLA VIRUSES

APPENDIX II

CHLORELLA VIRUSES ENCODE MOST, IF NOT ALL, OF THE MACHINERY TO GLYCOSYLATE THEIR GLYCOPROTEINS INDEPENDENT OF THE ENDOPLASMIC RETICULUM AND GOLGI

TABLE OF FIGURES AND TABLES

CHAPTER I

CHAPTER II

CHAPTER III

APPENDIX I

APPENDIX II

CHAPTER I

INTRODUCTION AND BACKGROUND

1. OVERVIEW

The traditional concept that viruses are strictly submicroscopic entities with minimal genomes has changed. To date, many large DNA viruses infecting a wide range of hosts have been characterized: they are ubiquitous in nature, exhibit extraordinary diversity in genome structure and composition, and in some instances, their genomes overlap those of the smallest bacteria (e.g., ~1.2-Mb amoeba virus vs. ~0.5-Mb *Mycoplasma genitalium*) [1, 2]. Among these viruses, the *Phycodnaviridae* family encompasses a morphologically similar but genetically diverse group of icosahedral double-stranded DNA (dsDNA) viruses that infect eukaryotic algal hosts from marine and freshwater environments [3]. Members of this group, as implied by the family name, are strictly DNA viruses that infect certain types of algae [4]. The phycodnaviruses are now recognized as key elements in the ecology of aqueous environments, playing a significant role in the dynamics of algal blooms, nutrient cycling, algal community structure, and possibly gene transfer between organisms [3, 5, 6, 7, 8, 9].

Current estimates indicate that there are approximately 100,000 to several million species of algae; however only ~40,000 have been identified [10]. In addition, at any one time, as many of 20% of the photosynthetic microorganisms in the ocean (including prokaryotes) are infected with a virus [11]. From this perspective, if one considers that

only 150 phycodnaviruses have been partially characterized and approximately 100 others have been mentioned in the literature, it indicates there is a vast array of algal viruses yet to be discovered [10]. At present, the phycodnaviruses are divided into six genera based on their host range: *Chlorovirus*, which includes the type-species *Paramecium bursaria* chlorella virus-1 (PBCV-1), *Coccolithovirus*, *Phaeovirus*, *Prasinovirus*, *Prymnesiovirus*, and *Raphidovirus* [3, 4]. These divisions are supported by phylogenetical analysis using the DNA polymerase gene and fragments and whole gene sequences of the major capsid protein as markers (Fig. 1) [12]. The *Chloroviruses* are the only viruses that infect freshwater algae, while other members of the family are found in marine environments. The genomes of ten phycodnaviruses, represented by four genera (*Chlorovirus*, *Coccolithovirus*, *Phaeovirus,* and *Prasinovirus*) have been sequenced and their genomes range in size from 154 kb to 406 kb [10, 13, 14, 15]. However, this range is even larger since a virus infecting the microalga *Pyramimonas orientalis* with a genome of ~ 560 kb has been described [16]. Genome comparison between *Chlorovirus, Coccolithovirus* and *Phaeovirus* members revealed that from a pool of approximately 1000 genes, only 14 are shared among them. This suggests extraordinary viral diversity within the group and also indicates that group divergence was not a recent event [10].

The phycodnaviruses are closely related to other large DNA viruses that infect a wide range of eukaryotic hosts including protists (mimivirus) and metazoans (poxviruses, African swine fever virus, and iridoviruses). Collectively, these viruses are referred to as Nucleo Cytoplasmic Large DNA Viruses (NCLDVs) with respect to their large size and replication site. The genome size of members of this group ranges from 100 kb to 1.2 Mb. They have nine common genes (core genes) involved in metabolic processes and

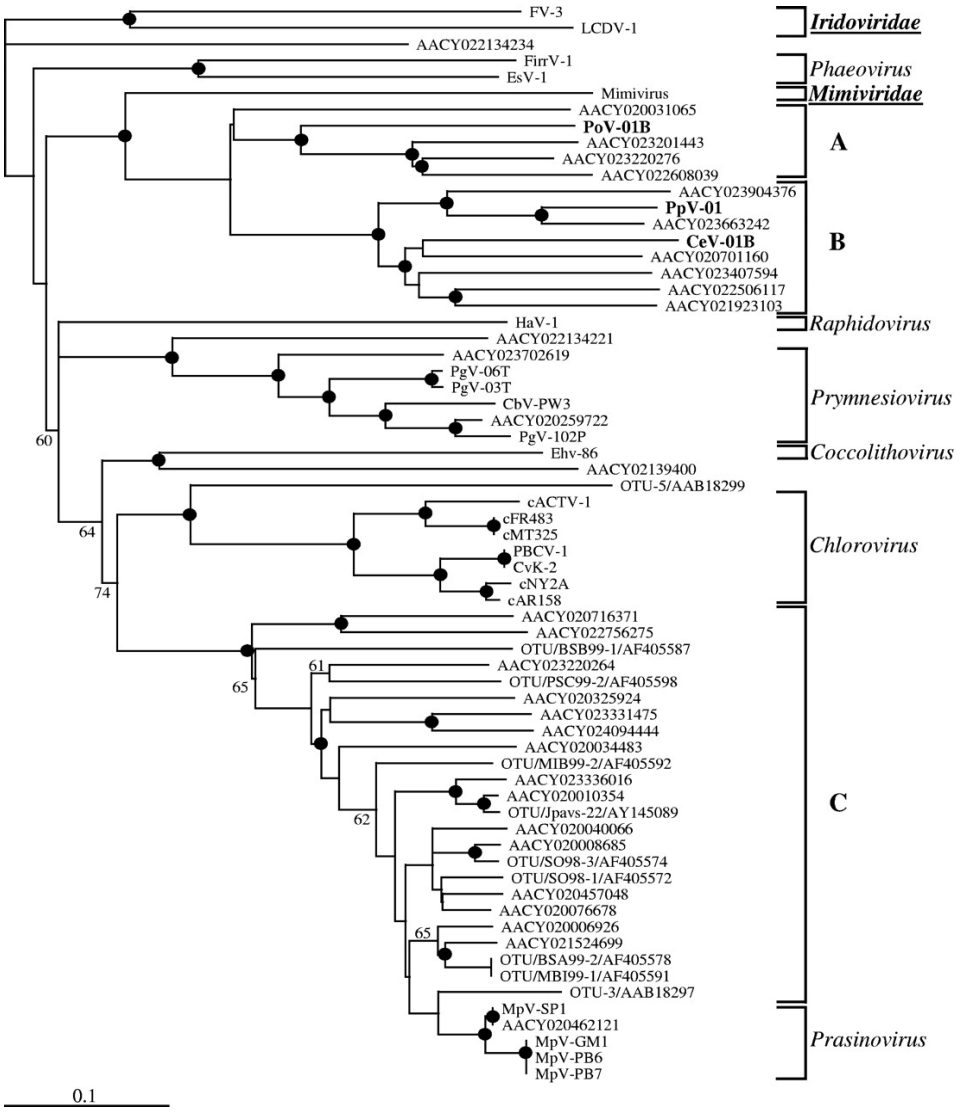


Figure 1. Phylogenetic tree based on alignment of DNA PolB protein sequences from the Phycodnaviridae family and mimivirus*. Only the conserved region of the protein was used. The tree was rooted using homologous sequences from two iridoviruses FV-3 and LCDV-1. Group A, clade of virus sequences with unknown hosts; group B, clade containing viruses infecting Pyramimonas orientalis; group C, different clusters of sequences from the metagenome survey Previous identified operational taxonomic unit*

(OTUs) are indicated by name and GenBank accession number, while sequences obtained from the Sargasso Sea metagenome are indicated by accession number alone. Sequences reported in this study are marked in bold. Nodes with >75% bootstrap values are marked by black dots, while actual bootstrap values are shown for nodes with values between 50 and 75%. Currently assigned genus names are shown to the right of the tree. The scale bar indicates the number of amino acid substitutions per residue. Adapted from [12].

virus structure and they are thought to be monophyletic [17]. In addition, NCLDVs carry a diverse assemblage of not only prokaryotic and eukaryotic, but also archaeal, gene homologs. This high degree of genome diversity suggests that NCLDVs have an ancient origin, possibly predating the divergence of the three kingdoms of life [18]. Interestingly, sequence analysis, where available, indicates that NCLDV genomes are only distantly related to their host genome, contradicting the concept that frequent horizontal gene exchange occurs between virus and host.

Starting with a short description of the algal host and the discovery of algal viruses, this chapter will then focus on the virus *Paramecium bursaria* chlorella virus 1 (PBCV-1), the subject of this research. PBCV-1 general characteristics, the protein-encoding genes (CDSs), and advances in the understanding of the infection cycle are reviewed. A comparison of PBCV-1 with other sequenced chloroviruses and a brief overview of other algal host systems will conclude the first chapter.

2. THE ALGAL HOST *CHLORELLA* NC64A

Algae are one of the most important primary producers on Earth with a dominant impact on global productivity and biogeochemical cycling [19]. They are common in nature and can be found mostly in aquatic environments but also in terrestrials' habitats. Algae constitute a diverse group of photosynthetic organisms, ranging from the most simple cell structure to large spreading seaweeds belonging to different phylogenetic groups and many taxonomic divisions. Algae constitute part of the phytoplankton responsible for the release of large amounts of oxygen into the atmosphere and account for over 50% of the fixed carbon on the planet. Because most algae require only light, simple salts, and a nitrogen source for growth, they are potential producers of proteins, carbohydrates, and fats for feeding humans and animals [20]. Furthermore, algae are a source of useful compounds such as β-carotene, alginic acid, carrageen, and agar. They are also used to remove heavy metals from wastewaters [21]. In addition, there is also a growing economical interest to use algae as a source for renewable biofuels as an alternative to fossil fuel [22, 23].

Algae are currently classified in the Kingdom Protist. Cyanobacteria, or blue-green algae, were traditionally considered to be algae; however, they were removed from this group because of their prokaryotic cell structure. The diversity within this group of organisms can be seen with: (i) the structure and composition of the cell wall (e.g., some algae have predominantly cellulosic cell walls and others have chitinaceous or silicacious wall; (ii) the characteristics of stored polysaccharides and lipids (e.g., starch, oil droplets); (iii) and the pigment molecules associated with light-harvesting antenna

complexes that drive photosynthetic electron transport (e.g., green algae have chlorophylls a and b and various carotenoids, red algae have phycobiliproteins, and diatoms and dinoflagellates have xanthophylls) [19].

The genus *Chlorella* (Phylum Chlorophyta, Class Trebouxiophyceae) consists of small, spherical, unicellular green algae. Algae originally included in this genus (>100) are very heterogeneous; the GC composition of the genomic DNA from different species ranges from 43 to 79% [24]. The cell wall composition, containing a variety of polysaccharides and a small amount of protein and lipids, varies as well. This diversity is confirmed by recent molecular taxonomy analysis showing that many of the isolates originally classified as *Chlorella* species are in fact dispersed over two classes of chlorophytes, the Trebouxiophyceae and the Chlorophyceae [26, 27]. One such study suggests that only three *Chlorella* species should be included in the "true" *Chlorella* genus, i.e., *C. sorokiniana, C. lobophoro,* and *C. vulgaris* str. Beijerinck [28].

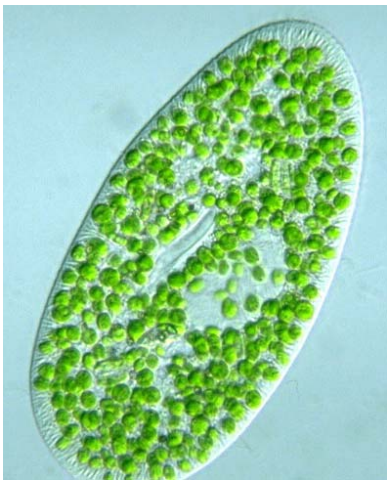


Figure 2. **P. bursaria** ***with endosymbiotic algae.***

Most *Chlorella* species are free-living, however, *Chlorella* sp. NC64A is a hereditary endosymbiont of the ciliate *Paramecium bursaria* (Fig. 2). Although they both retain the ability to grow independently, algal-free *P. bursaria* are rare in natural environments [29]. "Green" *P. bursaria* have several hundred *Chlorella* cells enclosed individually in the perialgal vacuole in the cytoplasm of the protozoan. Interestingly, timing in cell divisions for both algae and host is well coordinated and synchronized according to environmental conditions. The successful interaction is established by mutual symbiosis: (i) the host provides nitrogen components and CO_2 to the algae and offers protection against viruses; (ii) the algae supply the host not only with photosynthetic products, maltose and oxygen, but also have a photoprotective role. In addition, *P. bursaria* grows better in the presence of algae and photosynthetic products of *Chlorella* are related to expression of circadian rhythms in the host [29].

The genome of *Chlorella* sp. NC64A has recently been sequenced (G. Blanc et al., manuscript in preparation). The 46.2 Mb genome is predicted to encode 9, 791 protein-encoding genes (CDSs) that are presumably distributed among 13 chromosomes. *Chlorella* NC64A has a nuclear genome with a high GC content (67%). It also contains a circular chloroplast DNA (~124 kb, 34% GC) and a partially sequenced circular mitochondrial DNA (~68 kb, 29% GC). Furthermore, *Chlorella* NC64A as well as other *Chlorella* species were previously thought to be non-motile and only reproduce asexually. However, gene annotation reveals the presence of a significant subset of proteins found in flagella and all the known meiosis-specific proteins. As expected, the genome content also shows adaptation to endosymbiosis with the presence of many cell

transporters and several families of proteins containing protein-protein interaction domains (Blanc et al., manuscript in preparation).

3. THE DISCOVERY OF ALGAL VIRUSES

The first report of virus infecting algae occurred in the Russian literature between 1950s and 1966, where the authors describe the lysis of a free-living green alga, *Chlorella pyrenoidosa* [30, 31, 32, 33]. However, they acknowledged the fact that cells were contaminated with bacteria, and in fact, many of the reported viruses may have been bacteriophages. In the 1970s, several independent reports described the presence of viruses or virus-like particles (VLP) in algae: (i) viruses infecting *Sirodotia tenuissima*, a red algae [34]; (ii) viruses infecting *Oedogonum sp*., a green alga [35]; (iii) viruses infecting *Chorda tomentosa*, a brown alga [36], and (iv) viruses infecting the coccolithophore *Emiliania huxleyi* [37]. The first virus to eventually be classified as a phycodnavirus was isolated from the marine unicellular alga *Micromonas pusilla* in 1979 [38]. These early observations were not further pursued by virologists for several reasons that include: (i) evidence for the presence of viruses in algae was solely microscopic and many times it corresponded to a single observation; (ii) in some cases, algae were field-collected and they were no longer available for study; (iii) cultures were not axenic; (iv) technical problems regarding virus isolation (e.g., low concentration); (v) the inability to culture the host [20].

In 1978, Kawakami and Kawakami [39] described virus-like particles in *Chlorella* isolated from the protozoan *Paramecium bursaria*. However, the virus particles

were neither isolated nor characterized. The beginning of algal virology as a field occurred in the early 1980s, when Van Etten and Meints [40] characterized a large virus with polyhedron structure present in the isolated symbiotic alga *Chlorella* isolated from *Hydra viridis*. They reported that the virus lysed the whole algal population in 24 h and virus particles were absent in endosymbiotic *Chlorella.* Further characterization determined that the virus possessed a 250 kb dsDNA with at least 19 structural proteins, including a 46 kDa major capsid protein. Six other hydra *Chlorella* viruses were isolated within a short time concurrently with the isolation of a virus from symbiotic *Chlorella* from *Paramecium bursaria*, named *Paramecium bursaria* chlorella virus-1 (PBCV-1) [41, 42]. Initial restriction fragment analysis of all the chloroviruses revealed four different types of isolates and show no correlation between genomic composition and geographical origin of the viruses. Hydra *Chlorella* virus showed a very different DNA pattern from PBCV-1. Taking advantage of the fact that paramecium free-*Chlorella* can be grown in culture along with technical difficulties in isolating *Chlorella* from *Hydra,* made the laboratory concentrate all the efforts in the study of the virus PBCV-1 [20].

After testing many symbiotic and free-living *Chlorella*, three isolates supported PBCV-1 growth: NC64A, ATCC 30562, and N1a. Koch's postulates for the virus were fulfilled and plaque-assay in lawns of NC64A was developed providing a sensitive bioassay for studies and is still used today. Subsequently, many viruses infecting *Chlorella* NC64A, called NC64A viruses, have been discovered from freshwater collected around the globe.

To date, viruses infecting endosymbiotic *Chlorella* (also referred to as chloroviruses) in association with *Paramecium bursaria, Hydra viridis,* and *Acanthocystis turfacea* have been reported [43]. *Chlorella* NC64A and *Chlorella* Pbi are symbionts of *P. bursaria*, and their viruses are called NC64A viruses and Pbi viruses, respectively. Sequenced NC64A viruses are PBCV-1 and NY-2A, and sequenced Pbi viruses are MT325 and FR383 [44, 45]. Recently, a chlorovirus infecting *Chlorella* isolated from the helizoon *A. turfacea* was sequenced and characterized and is referred to as ATCV-1 [46]. A comparative analysis of PBCV-1 and these other chloroviruses will be described later in this chapter.

4. *PARAMECIUM BURSARIA* CHLORELLA VIRUS -1

Paramecium bursaria chlorella virus-1 (PBCV-1) is the prototype of the *Phycodnaviridae* family, a group of large, icosahedral, plaque-forming, dsDNA viruses that infect certain freshwater chlorella-like green algae. PBCV-1 infects *Chlorella* NC64A, a natural endosymbiont green alga of the protozoa *Paramecium bursaria*. PBCV-1 has a linear 330,744 bp non-permutated dsDNA genome with a GC content of approximately 40%. The ends of the PBCV-1 genome form a hairpin loop-like structure consisting of a 35 nucleotide-long region that is covalently closed and unevenly base-paired [47]. The 35-nucleotides are complementary when the sequences are inverted. Adjacent to the hairpin ends are identical 2221 bp repeats that possibly house the initiation point for DNA replication site [48]. The PBCV-1 genome also contains

methylated nucleotides (1.9% of the cytosines are m5C and 1.5% of the adenines are m6A) [20, 49, 50].

4.1. PBCV-1 structure

PBCV-1 has a multilayer structure, with an outer glycoprotein capsid that surrounds a lipid bilayer membrane and a dsDNA genome [51, 52]. Vp54, a glycoprotein and the major component of the capsid, represents about 40% by weight of the total protein content in the virus. The crystal structure of Vp54 showed that the subunit consists of two 8-stranded, antiparallel β-barrel, "jelly-roll" domains. The monomeric VP54 forms a trimeric capsomer (a "major capsomer"), which has pseudo-6-fold symmetry, relating the jelly-roll structures [52].

Using cryoelectron microscopy (cryoEM) and assuming icosahedral symmetry [51], it has been shown that the virion has a diameter ranging from 1.650 Å, measured along the 2-fold and 3-fold axes, and 1,900 Å, measured along the 5-fold axes. The PBCV-1 glycoprotein shell is composed of 20 triangular units, or "trisymmetrons", and 12 pentagonal caps, or "pentasymmetrons" at the 5-fold vertices. Each trisymmetron and pentasymmetron consists of a pseudohexagonal array of 66 and 30 trimeric casomers, respectively. The triangulation number (T) is 169d (h = 7, k = 8), where T represents the number of jelly rolls in the icosahedral asymmetric unit [53].

Recent 5-fold symmetry, 3D reconstruction analyses indicate that PBCV-1 has a quasiicosahedral structure, which means it does not have a perfect icosahedral

conformation. Instead, it has a unique vertex composed of special capsomers (Fig. 3) [53]. The unique vertex contains a pocket on the inside and a cylindrical spike structure on the outside of the capsid (Fig. 3). While this pocket is predicted to house enzymes to be used in the initial stages of the infection (i.e., cell wall degrading enzymes), the spike might be used to puncture the host cell, similar to the ones described in some bacteriophages [53]. The spike is too narrow for DNA to pass through it.

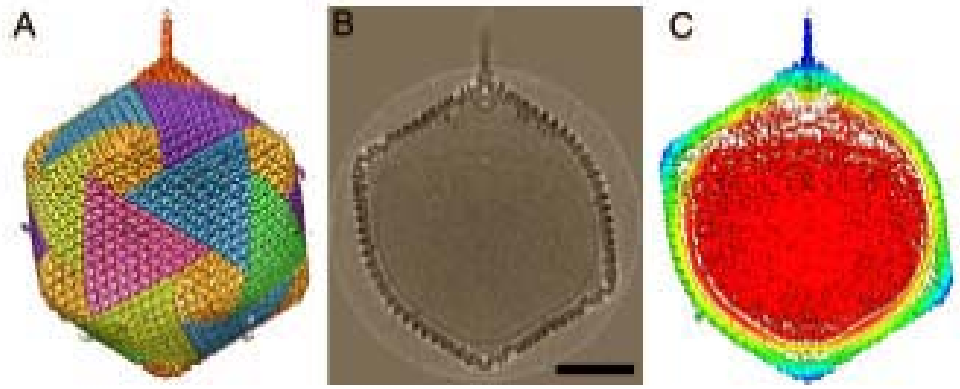


Figure 3. The 5-fold-averaged cryoEM structure of PBCV-1 viewed down a quasi-2-fold axis. *(A) Hexagonal arrays of major capsomers form trisymmetrons and pentasymmetrons (yellow). The unique vertex with its spike structure is at the top. Capsomers in neighboring trisymmetrons are related by a 60° rotation, giving rise to the boundary between trisymmetrons. (B) Central cross-section of the cryoEM density. (Scale bar: 500 Å.) (C) The same view as in B but colored radially, with red density being within 680 Å, yellow between 680 and 745 Å, green between 745 and 810 Å, light blue between 810 and 880 Å, and dark blue greater than 880 Å. Note the typical lipid low-density gap surrounding the red nucleocapsid density. Figure from [53].*

4.2. PBCV-1 CDSs

Initial sequencing of the PBCV-1 genome revealed 697 potential open reading frames (ORFs), of which 373 were predicted to be protein-encoding genes (CDSs) [54, 55, 56, 57, 58]. A CDS is defined here as a gene sequence flanked by start and stop codons that presumably codes for a protein, as opposed to tRNAs genes. The following criteria were used to define a CDS: (i) minimal size of 65 codons or more initiated by an ATG site; (ii) the largest CDS was chosen when CDSs overlapped; (iii) the putative promoter region (50 nucleotides upstream of the start codon) is AT-rich (>70%). The total number of CDSs from PBCV-1 has changed since the first publication describing the sequence because of sequencing mistakes. Now, the following CDSs are considered to be a single CDS: A92/93L, A111/114R, A122/123R, A140/145R, A181/182, A189/192, A219/222/226R, and A554/556/557L. The former annotation number is maintained to avoid confusion. It is possible that other CDSs will be detected with the upcoming re-sequencing of PBCV-1 and other chloroviruses. In addition, sequence comparisons with other chloroviruses help to identify sequencing mistakes. Currently, PBCV-1 is predicted to encode 365 CDSs [59]. The CDS density is 1.11 CDS/kb, which is similar to that reported for other large DNA viruses (e.g., *Ostreococcus tauri* virus-5 is 1.45 CDS/kb and *Feldmania sp*. virus-1 is 0.968 CDS/kb) [13, 15].

PBCV-1 CDSs are evenly distributed on both strands of the virus genome and intergenic space is minimal (>100 nucleotides). The exception is a 1788 nucleotide region near the middle of the genome that contains many stop codons and harbors 11 tRNA genes. In addition, this area has 900 nucleotides that are GC rich (>60%) and the

observed-to-expected ratio of CpG dinucleotides is 1.45, characteristics of a CpG island. The 2221 bp inverted terminal repeat region contains 4 duplicated CDSs [55, 60].

Approximately 35% of the PBCV-1 CDSs have a match in the public databases and these can be classified into a cluster of orthologous groups (COG) according to their putative functions. The CDSs resemble genes from all three domains of life, with the majority being bacterial-like. Prokaryotic CDSs tend to be clustered in islands towards the extremity of the virus genome and are co-localized with bacterial-like insertion sequences [18]. The functions of at least 44 PBCV-1 recombinant proteins have been characterized and show the expected activity [61]. Interestingly, a few PBCV-1 proteins revealed unusual characteristics. For example, some proteins are bifunctional (e.g., dCMP deaminase can also function as a dCTP deaminase) [62]; other virus-encoded enzymes have extremely high activity (e.g., virus topoisomerase II activity is 30x faster than human topoisomerase II) [63, 64]; and a few PBCV-1 proteins were the first characterized from a virus (e.g. K^+ ion channel protein) [65]. A detailed description of PBCV-1 CDSs according to their COG as well as highlights of current research will be presented. A review of PBCV-1 infection cycle will follow.

DNA replication, recombination, and repair

It is assumed that early PBCV-1 CDSs are transcribed in the nucleus, translated in the cytoplasm, and the proteins then return to the nucleus to initiate virus DNA synthesis (60-90 min p.i.) [66]. Large DNA viruses are known to code for most of the machinery

needed for viral DNA replication. Thus, PBCV-1 encodes 18 CDSs involved in DNA replication, recombination, and repair.

For the DNA replication process, PBCV-1 encodes a DNA polymerase δ family protein (A185R) that contains a proof-reading 3'-5' exonuclease domain [67]. This gene has a 101-nucleotide (nt) intron that is spliceosomal-processed [68]. This presence of this intron was recently confirmed by the transcriptome sequencing of infected *Chlorella* NC64A at 20 min p.i. (G. M. Yanai-Balser, unpublished results). PBCV-1 also encodes two sliding clamp processivity factor proteins (PCNA) (A193L and A574L). It is unlikely that they are a result of gene duplication as they have low similarity at the amino acid level (26%) and more closely resemble PCNAs from other organisms than with each other. PCNAs not only interact with enzymes involved in DNA replication but also with enzymes involved in DNA repair and post-replicative processing such as DNA methyltransferases and DNA transposases [69]. Since PBCV-1 encodes proteins involved with DNA repair, DNA methylation, and transposases, it is possible that the two PCNAs serve different roles in PBCV-1 replication.

PBCV-1 also encodes a replication factor C (RFC) (A417L), a protein complex required for the ATP-dependent loading of PCNA onto DNA in eukaryotes and archaea [70, 71]. RFCs from eukaryotes and the corresponding protein (called γ-complex) in bacteria are heteropentamers, whereas RFC from archaea are heterodimers [72, 73]. PBCV-1's RFC resembles the archaea protein.

Other proteins possibly involved directly in DNA replication include an ATP-dependent DNA ligase (A544R), an RNase H (A399R), and a type II DNA

topoisomerase (A583L). The PBCV-1 topoisomerase II is smaller than other characterized topoisomerase II enzymes and it has a 30x greater DNA cleavage processivity than human topoisomerase IIa [64]. The viral enzyme lacks a nuclear localization signal and PSORT analysis predicted that the virus topoisomerase II resides in the cytoplasm. Assuming this is true and viral DNA replication occurs in the host cell nucleus, this suggests the enzyme has a role in DNA packaging which does occur in the cytoplasm [61].

PBCV-1 also encodes an archaeo-like DNA primase (A468R) that initiates synthesis of new DNA strands by synthesizing short RNA oligomers on ssDNA [74]. In addition, this primase displays amino acid sequence similarity to a specific class of trifunctional primase/helicase/origin recognition proteins (M. Griep, personal communication). The helicase portion of the viral protein resembles the superfamily III helicases. A DNA binding protein (A437L) is located in the virion and is predicted to aid in neutralizing virus dsDNA.

PBCV-1 encodes a 5'-3' exonuclease homolog (A166R), which is probably involved in DNA recombination and it is known to be functional. In a study using temperature-sensitive mutants, it was established that genetic recombination occurs in PBCV-1 at a frequency of ~2% [75]. Finally, the virus has a CDS encoding a homolog of the bacteriophage T4 pyrimidine dimer-specific glycosylase (PDG). This is a well characterized DNA repair enzyme involved in pyrimidine photodimer excision [76]. Comparative studies revealed that PBCV-1 PDG cleaves both *cys-syn* and *trans-syn II* cyclobutane pyrimidine dimers, whereas the T4 enzyme only cleaves the *cis-syn* form.

PBCV-1 PDG also excises two other monomeric products induced by UV radiation or hydroxylradicals, 4,6-diamino-5-formamidopyrimidine and 2,6 diamino-4-hydroxy-5-formamidopyrimidine [77]. Moreover, PBCV-1 PDG is more processive than the T4 enzyme and is the first *trans-syn II* specific glycosylase identified to date.

DNA restriction/modification

PBCV-1 and other chloroviruses genomes have varying amounts of methylated nucleotides. Cytosine methylation (m5C) can range from 0.12% to 47.5% of the total cytosine and adenine methylation (m6A) varies from 1.5% to 37%. The finding that methylated bases occur at specific DNA sequences led to the discovery of virus encoded restriction/modification (R/M) systems [49, 50]. DNA methyltransferases (MTases) protect the DNA from cleavage by methylating one of the bases in the same nucleotide sequence, and DNA restriction endonuclease (REases) recognize and cleave specific nucleotide sequences. The number of virus-encoded MTases and REases varies among the different chloroviruses. About 25% of the virus-encoded DNA MTases have a companion REases.

PBCV-1 has two R/M systems: (i) the CviAI R/M system, which consists of the REase R.CviAI (A252R) (cleaves /GATC, but not $G^{m}ATC$ sites) and the adenine MTase M.CviAI (A251R) (methylates GATC sequences) [78], (ii) the CviAII system consists of R.CviAII (A579L) (cleaves C/ATG, but not $C^{m}ATG$) and the adenine MTase M.CviAII (A581R) (methylates CATG sequences) [79]. In addition to these two R/M systems,

PBCV-1 also encodes three other MTases (A517L, A530R, A683L) but they lack a companion REase [61].

Until recently, the biological function of the virus encoded R/M systems was unknown. However, analysis of host chromosomal DNA in PBCV-1-infected cells using pulsed-field gel electrophoresis revealed that host chromosomes were degraded within minutes after PBCV-1 infection (2-5 min p.i.). The DNA degradation was impervious to protein synthesis inhibitors or UV inactivation of virus particles, indicating that a virion packaged protein, probably is the cause. Mass spectrometry analysis of PBCV-1 virion revealed that the two PBCV-1 REases were present in the virion but not their companion MTases (D. Dunigan, manuscript in preparation). Nuclease activities, including those of the two known REases and an uncharacterized general nuclease(s), were detected in disrupted PBCV-1 particles. This nuclease(s) degraded host and virus DNA *in vitro*, however, it did not degrade virus DNA *in vivo*, suggesting differential intracellular trafficking of the virion-associated nucleases [80].

Nucleotide metabolism

Host chromosome degradation starts 2-5 min after PBCV-1 infection [80]. Virus DNA replication begins 60-90 min p.i. and by 4 h p.i., the total DNA in the cell increases at least four to five-fold [61]. With a ~330 kb genome, PBCV-1 requires a large quantity of dNTPs for replication that cannot be accounted for by simply recycling deoxynucleotides from degraded host DNA. To guarantee a supply of deoxynucleotides

in non-proliferating host cells, large DNA viruses frequently encode nucleotide synthesis enzymes. PBCV-1 genome encodes 11 CDSs involved in nucleotide metabolism and many of them have been functionally characterized.

Most PBCV-1 nucleotide enzymes are directed to the *de novo* dTTP synthesis pathway (Fig. 4) by producing key intermediates in dTTP formation [62]. PBCV-1 encodes: (i) the two subunits for ribonucleotide reductase (A476R and A629R), which is responsible for the reduction of UDP or UTP to dUDP and dUTP, respectively (Fig 4A); (ii) the first known bifunctional dCMP deaminase (A596R) that deaminates dCMP into dUMP and also deaminates dCTP to dUTP (Fig. 4B and C, respectively) [62]; (iii) the deoxyuridine triphosphatase (dUTPase) (A551L) that catalyzes the hydrolysis of dUTP to dUMP and PPi. PBCV-1 dUTPase forms a trimer and has properties similar to dUPTases from eukaryotes, bacteria and other viruses (Fig 4D) [81]; (iv) a FAD-dependent thymidylate synthase X (ThyX) (A674R) (Fig 4E). This was unexpected because most organisms have the homodimeric thymidylate A (ThyA) enzyme. However, the PBCV-1 enzyme resembles the homotetramer ThyX, common in pathogenic bacteria. While ThyA uses the substrate methylenetrtrahydrofolate as both methylene donor and reductant, ThyX uses it only as methylene donor. The PBCV-1 recombinant protein converts dUMP to dTMP more efficiently than other characterized ThyXs [82].

Recombinant proteins from most of these enzymes involved in nucleotide synthesis were functionally characterized, except for ribonucleotide reductase. It is not surprising that many PBCV-1 nucleotide CDSs encode enzymes involved in the three pathways that lead to dUMP production, possibly as a means to compensate for

differences in DNA content of the host (PBCV-1 genome is 40% GC and host *Chlorella* NC64A nuclear genome is 67% GC.

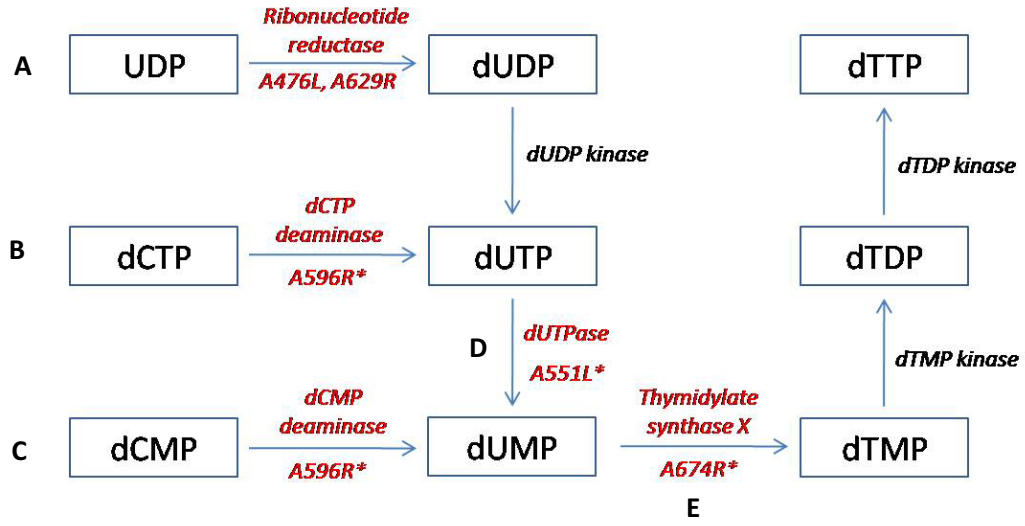


Figure 4. The three known pathways for synthesis of dUMP, an intermediate in dTTP synthesis. *The enzymes shown in the diagram in red are encoded by chlorovirus PBCV-1. Asterisks indicate enzymes known to be functional. Adapted from [62].*

In addition, PBCV-1 encodes an aspartate transcarbamylase (ATCase) (A169R), a key regulatory enzyme that controls the *de novo* biosynthetic pathway of pyrimidines. PBCV-1 ATCase resembles several plant ATCases including the catalytic domain; however, it lacks the regulatory domain [83]. Recently, the PBCV-1 glutaredoxin (A438L) was functionally characterized and it is predicted to provide the reducing power

to ribonucleotide reductase [84]. PBCV-1 glutaredoxin contains an unusual monothiol active site rather than the typical dithiol active site.

Other PBCV-1 enzymes predicted to be involved in nucleotide metabolism include deoxynucleoside kinase (A416R), thioredoxin (A427L), cytosine deaminase (A200R), and NTP pyrophosphohydrolase (A326L).

Transcription

A whole genome transcription analysis using microarrays has been done for PBCV-1 virus and details about the time of expression of each CDS will be discussed in Chapter II. PBCV-1 gene expression is temporally regulated and the first viral transcripts are detected 5-10 min p.i. and chloroplast, but not cytoplasmic, rRNA degradation begins about 30 min p.i. [85]. Virus infection rapidly inhibits most host transcription. The virus has at least two ways to inhibit host transcription: (i) by host chromosome degradation using viral encoded restriction endonucleases, and (ii) by methylating histone H3 at Lys27 using a viral encoded vSET (A625L) [86]. PBCV-1 vSET is a small 120 amino acid proteins that contains a nuclear localization signal and is known to trigger gene silencing in eukaryotes. Both restriction endonuclease(s) and vSET are packaged in the PBCV-1 virion (Dunigan, manuscript in preparation).

A few PBCV-1 CDSs can overlap or possibly form polycistronic transcripts. The virus can also post-transcriptionally modify mRNAs since introns are present in at least

two PBCV-1 CDSs (DNA polymerase [A185R] and TFIIS [A125L]). Transcripts also often extend beyond the translational stop codon [87, 88, 89] and this is confirmed by the preliminary analysis of a recent transcriptome-sequencing project (Yanai-Balser, unpublished). Recently, three putative promoter sequences were identified for PBCV-1 CDSs and their relationship with time of expression will be reported in Chapter III.

PBCV-1 encodes many genes involved in transcription; however, it still depends on host transcription machinery because it does not encode a recognizable RNA polymerase or a subunit of it. No RNA polymerase activity was detected in purified PBCV-1 virions. The lack of a virus-encoded RNA polymerase suggests that infecting viral DNA must go the cell nucleus and a host RNA polymerase initiates viral transcription, possibly in conjunction with virus-packaged transcription factors. Consistent with this possibility, PBCV-1 encodes at least three transcription factor-like elements: (i) TFIIB (A107L); (ii) TFIID (A552R); (iii) TFIIS (A125L).

Other PBCV-1 CDSs involved in transcription are two enzymes involved in forming the mRNA cap structure, RNA triphosphatase (A449R) [90] and RNA guanylyltransferase (A103R) [91]. The size, amino acid sequence, and biochemical properties of the PBCV-1 capping enzymes resemble yeast-capping enzymes more than poxvirus and African swine fever virus multifunctional RNA capping enzymes [92, 93].

PBCV-1 also encodes an active RNAse III (or ribonuclease III) (A464R), which is a multifunctional, ubiquitous enzyme present in prokaryotic and eukaryotic organisms. RNAse III is involved in the processing of cellular and viral precursor RNAs to produce

mature functional rRNAs, mRNA, and tRNAs. It is also involved in the degradation of specific mRNAs and recognizes dsRNA structures and cleaves them at specific sites. Orthologs of this enzyme also participate in gene inactivation by RNA interference and in gene regulation during development. PBCV-1 RNAse III resembles bacterial enzymes both in size and its activity is presumably responsible for processing either virus mRNAs or tRNAs [94]. According to Nishida et al. (1999) [95] tRNAs from the chlorovirus CVK2 are transcribed as a polycystronic message at early and late times after infection and are aminoacetylated *in vivo* [95].

Lastly, PBCV-1 encodes three proteins that contain sequence elements of superfamily II helicases (A153R, A241R, 363R) that are probably involved in transcription. A CDS for a SWI/SNF helicase (A548L) and a SWI/SNF chromatin remodeling complex (A189/192R) are also present in the PBCV-1 genome.

Protein synthesis, modification, and degradation

PBCV-1 infection inhibits host protein synthesis and some early virus proteins appear within 15 min p.i. Viral proteins are synthesized on the cytoplasmic ribosomes and not on organellar ribosomes since cycloheximide, not chloramphenical, inhibits viral replication [96]. The mechanisms which allow the virus to take over the host translational machinery forcing it to translate virus mRNAs are not known.

PBCV-1 codon usage is biased to codons ending in A/U (63%) over those ending in C/G (37%). This bias is expected since PBCV-1 DNA is 40% GC, whereas host nuclear DNA is 67% GC (Blanc et al, manuscript in preparation). Therefore, finding that PBCV-1 encodes 11 tRNA genes is not surprising, as viral tRNAs might overcome codon usage barriers that exist between virus and its host ensuring virus maximal replication potential. Some chloroviruses encode as many as 16 tRNAs [95, 97]. There is a strong correlation between the abundance of virus-encoded tRNAs and the virus gene codon usage. The virus encoded tRNAs contain internal A and B boxes characteristic of RNA polymerase III promoter elements, suggesting that tRNAs might be transcribed individually by RNA polymerase III. However, as mentioned previously, chlorovirus tRNA genes are transcribed as a large precursor RNA and processed via intermediates to mature tRNA at both early and late stages of virus replication. Some, if not all, of the tRNAs are aminoacetylated *in vivo*, suggesting they probably function in viral protein synthesis [95].

The chloroviruses were the first viruses reported to encode a translation elongation factor enzyme (EF-3) [98]. The putative protein from PBCV-1 (A666L) is similar to EF-3 from fungi, which is known to stimulate EF-1α-depending binding of aminoacyl-tRNA to the A site of the ribosome. Like fungal EF-3 proteins, the PBCV-1 protein has an ABC transporter family motif and two ATP/GTP-binding site motifs. Possibly, the virus-encoded EF-3 in combination with the viral tRNAs alter the host protein synthetic machinery to preferentially translate viral mRNAs.

Including glycosylation (discussed in the *Sugar Metabolism* section), PBCV-1 encodes several enzymes involved in post-translational modification. Like several large dsDNA viruses, PBCV-1 encodes an ERV/ALR protein that functions as a protein thiol oxireductase (465R) [99]. PBCV-1 also encodes a putative protein disulfide isomerase (A448L) and a prolyl 4-hydroxylase (A85R) that converts Pro-containing peptides into hydroxyl-Pro-containing peptides and was functionally characterized [100]. As many as 8 putative protein kinases are present in the PBCV-1 genome and will be described in the *Signaling* section.

PBCV-1 encodes two putative proteins that interact with ubiquitin, a ubiquitin C-terminal hydrolase (A105L) and a Skp1 protein (A39L). This last protein belongs to the SFC-E3 ubiquitin ligase family that target cell cycle and other regulatory factors for degradation. Other PBCV-1 CDSs possibly involved with protein degradation are a zinc metallopeptidase (A604L) and a ATPase (AAA+ class) (A44L).

Signaling

PBCV-1 was the first virus discovered to encode a functional K^+ channel protein. The 94-amino acid protein (called Kcv) produces a K^+ selective and slightly voltage-sensitive conductance in *Xenopus* oocytes [65], mammalian HEK293 (human embryonic kidney cells) [101] and Chinese-hamster ovary cells [102]. Kcv has minimal structure: the channel monomer consists of two transmembrane domains (TM), a pore loop, and a 12 amino acid-long cytoplasmic N-terminus. Like other K^+ channels, PBCV-1 Kcv forms

a functional tetramer and electrophysiology experiments established that it has many properties similar to more complex K channels, including gating, selectivity, sensitivity to voltage, and susceptibility to channel blockers.

The biological function of PBCV-1 Kcv is hypothesized to be involved in host membrane depolarization upon infection [103]. Virus attaches and degrades the *Chlorella* host cell wall, which is followed by fusion of the viral internal membrane and the host membrane. This facilitates entry of the viral DNA and virion-associated proteins into the cell, leaving an empty virus capsid on the outside. Rapid depolarization of the host membrane occurs and results in an immediate loss of K^+ from the host, a process which energetically favors ejection of viral DNA into the host [104]. Virus competition assays indicated that host membrane depolarization might also prevent the infection of the host by another virus [105].

Additional PBCV-1 signaling proteins include eight *Ser/Thr* protein kinases (PK) (A34R, A248R, A277L, A278L, A282L, A289L, A614L, A617R). A *Ser/Thr* PK (A248R) is adjacent to the Kcv CDS (A250R). Some of the PKs are clustered in the PBCV-1 genome and recombinant proteins from all of them have been biochemically tested and shown to be functional (P. Valbuzzi, unpublished results).

Cell wall degradation

Recent genomic analysis of virus the host, *Chlorella* NC64A, indicates it encodes a complex array of synthesizing and modifying polysaccharides (Blanc et al., manuscript

in preparation). Surprisingly, genes involved in the synthesis of cellulose and hemi-cellulose, major components of the primary cell wall of green plants, are absent. Instead, paralogs for chitin synthase and 25 paralogs for chitin deacetylase (which converts chitin into chitosan) are present. Two chitinase- encoding genes (plant and bacterial types) and four chitosanase genes for the degradation of chitin and chitosans are also present. As mentioned previously, the cell wall of *Chlorella* species, including *Chlorella* NC64A, contains glucosamine polymers.

Consequently, to be able to penetrate the host, PBCV-1 and other chloroviruses have CDSs involved with chitin and chitosan degradation, including two chitinases (A181/182R, A260R), a chitosanase (A292L), a β-1,3-glucanase (A94L), and β and α 1,4 linked glucuronic lyase (A215L). Recombinant proteins produced from these five genes have the expected enzyme activities [106, 107, 108, 109, 110, 111, 112]. All of these proteins, except for the β-1,3-glucanase, are expressed late; however, none of them are packaged in the virion (Dunigan, manuscript in preparation). The chitosanase CDS was studied in another chlorovirus, CVK2, and it is transcribed late and encodes two proteins with chitinase activity (37 kDa and 65 kDa). The larger protein is packaged in nascent virions, whereas the smaller one occurs only in infected cells [112]. Interestingly, PBCV-1 contains the same chitinase CDS flanked by two CDSs common to CVK2.

Phylogenetic analysis of chitosanase and chitin synthase from the host *Chlorella* NC64A and PBCV-1 indicate the sequences are evolutionary related. However, no direct evidence for gene exchange was obtained. The fact that higher plants do not have gene families encoding enzymes involved in either chitin or chitosanase synthesis suggests that

the host genes involved in chitin metabolism might have come from the viruses, contradicting the concept that viruses always acquire genes from their hosts.

Sugar metabolism

The PBCV-1 genome encodes several enzymes involved in nucleotide sugar metabolism including two enzymes from the highly conserved pathway that converts GDP-D-mannose to GDP-L-fucose: (i) GDP-D-mannose 4,6 dehydratase (GMD) (A118R) and (ii) the bifunctional GDP-4-keto-6-deoxy-D-mannose epimerase/reductase (GMER) (A295L). Experiments with recombinant GMD and GMER established that both virus proteins have properties comparable to human and bacterial enzymes, however PBCV-1 GMD is more stable than GMDs from other organisms. In addition to the dehydratase activity, the PBCV-1 GMD also catalyzes NADPH-dependent reduction of the intermediate GDP-4-keto-6-deoxy-D-mannose, forming GDP-D-rhamnose. As a consequence, *in vitro* reconstruction of the biosynthetic pathway using PBCV-1 GMD and GMER produces both GDP-L-fucose and GDP-D-rhamnose in the presence of NADPH [113].

Fucose and rhamnose are present in the glycans attached to the virus major capsid protein, Vp54. However, uninfected *Chlorella* NC64A cell walls contain rhamnose and trace amounts of fucose [114], suggesting that the virus-encoded pathway could circumvent a limited supply of GDP-sugars by the host. Other genes involved in sugar

metabolism are D-lactate-dehydrogenase (A53R) and fructose-2,6-biphosphatase (A297L) but their functions have not been tested.

- Protein glycosylation

Structural proteins of many viruses, including herpesviruses, poxviruses, and paramyxoviruses, as well chloroviruses, 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 [115, 116]. 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.

Several observations indicate that chloroviruses differ from the above scenario: (i) PBCV-1 Vp54, the major capsid protein located on the viral surface, is one of the three glycosylated viral proteins. The glycan portion of Vp54 contains several neutral sugars that contribute to the protease resistance and antigenicity of the virus [117]. Polyclonal antibodies to wild-type PBCV-1 and its antigenic variants do not react with many other independent virus isolates that infect *Chlorella* NC64A. (ii) Vp54 crystal structure revealed four *N*-linked glycosylation sites and two *O*-linked oligosaccharides. However, the four glycosylated Asn resides are not located in consensus sequences typical of eukaryotic *N*-linked glycans [52]. (iii) PBCV-1 encodes most, if not all, glycosyltransferases involved in constructing the glycans attached to Vp54 (A64R,

A111/114R, A219/222/226R, A473L, and A546L). None of these have an identifiable signal peptide that would target them to the ER. Furthermore, cellular protein localization analysis using PSORT predicts that all of these proteins, except for A473L which is predicted to localize in a membrane, are cytoplasmic.

- Host cell surface changes

PBCV-1 is unusual because it encodes the enzyme hyaluronan synthase (HAS) (A98R) that synthesizes hyaluronan (or hyaluronic acid) [118, 119]. Hyaluronan is a ubiquitous constituent of the extracellular matrix in vertebrates, and consists of ~20,000 alternating β-1,4-glucuronic acid and β-1,3-N-acetylglucosamine residues [120]. Hyaluronan synthesis occurs at the plasma membrane and is simultaneously extruded through the membrane to the exterior of the cell. Previously, hyaluronan was thought to occur only in vertebrates and a few pathogenic bacteria, where it forms an extracellular capsule, presumably to avoid the immune system [120, 121].

PBCV-1 also encodes two enzymes involved in the biosynthesis of hyaluronan precursors, glutamine:fructose-6-phosphate amidotransferase also known as glucosamine synthase (A100R) and UDP-glucose-dehydrogenase (UDP-GlcDH) (A609L) [121]. All three CDSs are expressed early during PBCV-1 infection. These results let to the discovery that hyaluronan lyase-sensitive hair-like fibers begin to accumulate on the surface of PBCV-1 infecting host cells by 15 min p.i. and by 4 h p.i. the infected cells are covered with a dense fibrous hyaluronan network [119].

HAS is present in many, but not all, chloroviruses isolated from diverse geographical regions [119]. Surprisingly, many chloroviruses that lack HAS have a gene encoding a functional chitin synthase (CHS). Furthermore, cells infected with these viruses produce chitin fibers on their external surface [122]. Chitin, an insoluble linear homopolymer of β-1,4-linked-N-acetyl-glucosamine residues, is a commom component of insect exoskeletons, shells of crustaceans, fungal and algal cell walls [123]. A few chloroviruses contain both HAS and CHS CDSs, and form both hyaluronan and chitin on the surface of their infected cells [122, 124]. Finally, a few chloroviruses probably lack both genes because no extracellular polysaccarides are formed on the surface of infected cells [119]. The fact that many chloroviruses require a huge expenditure of ATP for their synthesis, are important in the virus life cycle. However, at present this function(s) is unknown.

Lipid metabolism

PBCV-1 encodes six enzymes involved in lipid metabolism. A patatin-like protein (A173L) is present in the virus genome. Patatins are plant storage glycoproteins that have lipolytic activity and they are thought to serve as a defense against plant parasites and they have been shown to participate in plant signal transduction. Many bacteria also encode patatin-like proteins [125]. In addition, two acetyltransferases (A254R and A654L), a glycerophosphoryl diesterase (A49L), a lysophospholipase (A271L), and a lipoprotein lipase (A402R) are also encoded by PBCV-1.

Integration and Transposition

PBCV-1 has a large number of CDSs involved with genome integration and transposition: seven GIY-YIG endonucleases (A539R, A351L, A134L, A287R, A315L, A495R, A651L), six HNH endonucleases (A478L, A354R, A422R, A490L, A267L, A87R), and two transposases (A625R, A366L). Currently, the functions of these CDSs are unknown.

Structural proteins

SDS-PAGE mass-spectrometry revealed ~130 proteins associated with the PBCV-1 virion (Dunigan et al., unpublished results). At least 9 CDSs are predicted to serve a structure role in the virion. They include the major capsid protein (A430L), A140/145R, 189/192R, A363R, A384L, A540L, A558L, A561L, A622L. Homologs of all of these proteins are present in the virions of other chloroviruses.

Polyamine biosynthetic enzymes

Polyamines are small cationic organic molecules that are known to have an important role in the replication of many viruses. They are present (as putrescine, spermidine and spermine) in large amounts in the capsid of some viruses, where they can neutralize up to 50% of the viral nucleic acid. In some instances, virus infection can result in an increase in polyamines in the host cell and inhibition of polyamine synthesis

can interfere with virus replication. In contrast, host polyamines can also confer resistance to some viruses. All of these reports have assumed that polyamines are made by the host cells. PBCV-1 is unusual because it encodes four enzymes for the polyamine biosynthesis: (i) ornithine decarboxylase (ODC) (A207R), which converts ornithine to putrescine, but can also use arginine as a substrate. Because of its dual function and because it shows higher activity with arginine, currently the virus ODC is reclassified as ODC/ADC (A207R) [126]. The virus also encodes two enzymes necessary to produce putrescine from agmatine: (ii) agmatine iminohydrolase (A638R) and (iii) *N*-carbamoylputrescine amidohydrolase (A78R). In addition, PBCV-1 encodes (iv) homospermidine synthase (HSS) (A237R) [127] that synthesizes the rare polyamine homospermidine from two molecules of putrescine. Because spermidine synthase is lacking, the end product of the pathway in the virus is likely homospermidine. All of these four recombinant polyamine enzymes were functional [126, 127, 128].

The importance of polyamines in PBCV-1 pathogenesis remains unclear since the host cell also produces polyamines. Little change occurs in either polyamine concentration or in composition (putrescine, cadaverine, spermidine, and homospermidine are present) during the first hour of virus infection [127]. However, by 240 min p.i., the concentration of putrescine increases about 3.5-fold, whereas the other polyamines decrease. The net result is that polyamine concentration decreases slightly during virus replication. Also, it is unlikely that polyamines play a role in neutralizing virus DNA because the number of polyamine molecules per PBCV-1 virion is so low that they can only neutralize ~0.2% of the virus phosphate residues. Furthermore, the physiological significant of the polyamines in the PBCV-1 particles must be limited

because they are only loosely associated with the virions (they can be replaced by Tris or displaced by washing the particles in a polyamine-free buffer without affecting virus infectivity [127].

4.3. Current infection cycle

PBCV-1 binds to the host cell and specifically attaches to the cell wall probably at a unique vertex [53]. Following attachment, virus punctures a hole in the cell wall and cell membrane using a spike-like structure. It has been suggested that the spike is then released because of the lower density linking the spike and the unique vertex [53]. Host cell wall degrading enzymes possibly located in the virus "pocket" [53] at the unique vertex are released and the host cell wall is degraded at the point of contact (Fig. 4A).

The virus internal membrane fuses with host plasma membrane, causing a K^+ efflux possibly mediated by viral K^+ ion channel (Kcv) [104]. H^+ flux and H_2O flux occur as well causing host membrane depolarization. It is suspect that this depolarization might prevent other viruses from attaching to the cell [105]. Secondary active transport of the cells is also rapidly inhibited by infection [129]. Finally, viral DNA along with DNA-associated proteins are ejected into the cell leaving an empty capsid outside (Fig. 4B).

Host chromosome degradation caused by restriction endonucleases packaged in the virion occurs within 2-5 min p.i. [80]. We predict that the DNA and virus-associated proteins goes quickly to the nucleus and commandeer the host transcription machinery; the first viral transcripts can be detected at 5-10 min p.i.

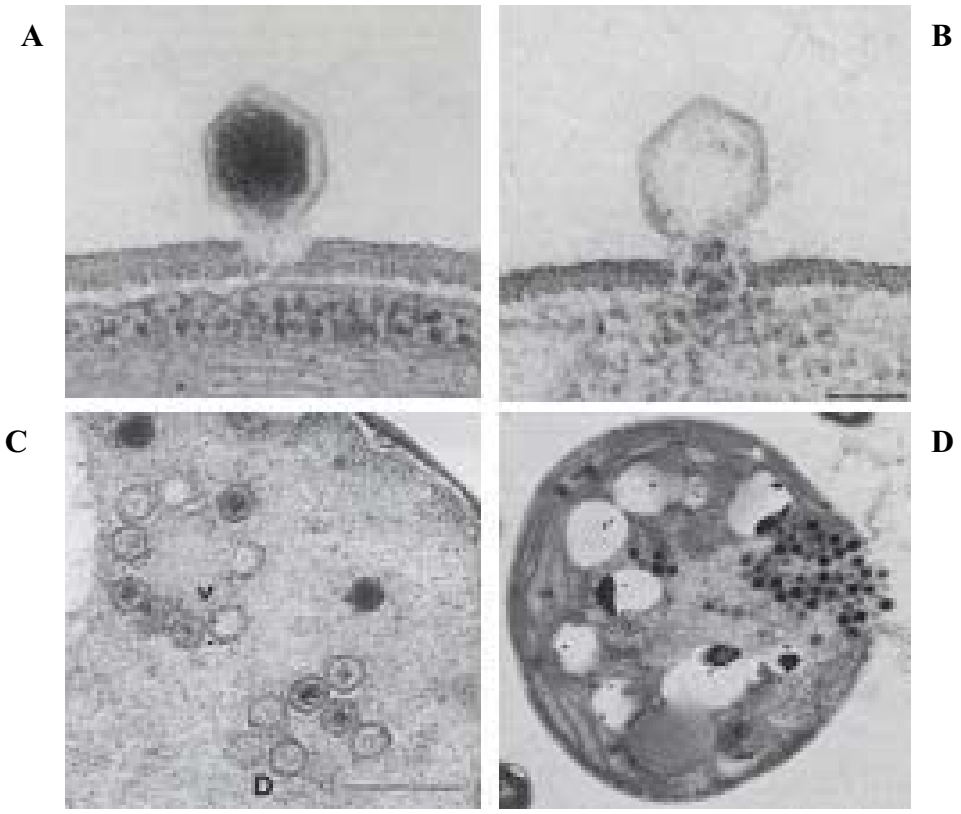


Figure 5. Infection of Chlorella strain NC64A by PBCV-1. *A. Digestion of the wall at the point of attachment (1-3 min p.i.) .* ***B.*** *The contents of the virus are released into the cell (3-5 min p..i).* ***C.*** *Cytoplasmic virus assembly centers at 3-5 h p.i.* ***D.*** *Release of progeny virus by lysis of host cell at 6-8 h p.i.*

Virus DNA replication begins between 60-90 min p.i. At 240 min, cells are covered with hyaluronan [120]. Virus assembly occurs at 3-5 h p.i. in localized regions of the cytoplasm called virus assembly centers (Fig 4C). Finally, at 6-8h p.i., host cell lysis occurs and release of progeny virions (Fig. 4D). Typically, about 1000 particles per cell are released and about ~25% of them are infectious.

5. OTHER CHLOROVIRUSES

Five additional chlorovirus genomes have been sequenced: NY-2A and AR158 that infect *Chlorella* NC64A (same host for PBCV-1); MT325 and FR483 that infect *Chlorella* Pbi, and ATCV-1, that infects *Chlorella* SAG 3.83. Both NC64A and Pbi are normal endosymbionts of the protozoa *P. bursaria,* while SAG 3.83 is an endosymbiont of the heliozoon *Acanthocystis turfacea* [44, 45, 46]. Chlorovirus genomes have CDSs distributed evenly on both strands and have minimal intergenic space; the GC content is host-specific. The main genomic features for each new chlorovirus can be seen in TABLE 1 in comparison to the prototype PBCV-1.

Table 1. *Comparison of sequenced chlorella virus genomes*[a]

Virus	Host[b]	Genome (bp)	CDSs	tRNA genes	GC (%)
PBCV-1	NC64A	330,743	365	11	40
NY-2A	NC64A	368,683	404	7	40.7
AR158	NC64A	344,690	360	6	40.7
MT325	Pbi	314,335	331	10	45.3
FR483	Pbi	321,240	335	9	44.6
ATCV-1	SAG 3.83	288,047	329	11	49.4

[a] Adapted from [46]

[b] Specific *Chlorella* strains

Most CDSs involved in the main metabolic functions (i.e., DNA replication and nucleotide biosynthesis) are conserved among chloroviruses (~80%). Despite the strong conservation each virus has unique features. For example, (i) NY-2A has the largest genome and it is also heavily methylated: 45% of cytosines are 5-methylcytosine (5mC)

and 37% of the adenines are N^6-methyladenine (6mA) compared to 1.9% 5mC and 1.5% 6mA in PBCV-1. In addition, two NY-2A CDSs encode inteins (the first detected in chloroviruses), the large subunit of ribonucleotide reductase and a superfamily II helicase. (ii) AR158 is the only NC64A virus that does not encode a full-length K^+ ion channel protein (Kcv). (iii) MT325 is the first virus to encode a water channel, an aquaglyceroporin (AQV). The K^+ ion channel (Kcv) from MT325 differs in structure from PBCV-1 Kcv [130]. (iv) The virus FR483 lacks both AQV and Kcv, but it encodes a putative potassium ion transporter protein and an alkyl sulfatase [45]. (v) Like PBCV-1, ATCV-1 also encodes GDP-D-mannose 4,6 dehydratase (GMD) and functional studies revealed it has the expected dehydratase activity but lacks the additional reductase activity present in the PBCV-1 encoded enzyme [131]. (vi) The ATCV-1-Kcv is the smallest functional K^+ ion channel reported to date (82 amino acids) and is more permeable to K^+ than Rb^+, which differs from PBCV-1 Kcv [132].

NC64A viruses have three types of introns in their CDSs. PBCV-1 and NY2-A have a self-splicing intron in a transcription factor TFIIS-like CDS [44, 56]. A spliceosomal-processed intron is present in the DNA polymerase gene and a 81-nucleotide spliceossomal-processed intron exists in the pyrimidine dimmer-specific glycosylase gene from some viruses [111]. However, initial analyses of Pbi virus genomes did not detect any intron or inteins in their CDSs, although a single tRNA in MT235 and FR483 contains an intron [45].

Considerable variation occurs in chlorovirus genomes and the total number of genes in the chlorovirus gene pool exceeds that of a single isolate. In addition, not all

PBCV-1 genes are required for virus replication, since PBCV-1 mutants with large deletions (27-37 kb) [133] can replicate in *Chlorella*. Large insertions also occur in some chloroviruses, for example, the virus CVK2 contains an extra 22 kb insert [95, 134]. The different sizes and diversity of the chlorella virus genomes suggest that dynamic and frequent rearrangements of virus genomes occur in natural environments.

6. NATURAL HISTORY OF CHLOROVIRUSES

Even after almost 30 years of study, the natural history of chlorella viruses is still poorly understood. At present, chloroviruses have been isolated from three different *Chlorella:* (i) *Chlorella* NC64A (host for the sequenced PBCV-1, NY-2A, and AR158), (ii) *Chlorella* Pbi (host for MT235 and FR383), and (iii) *Chlorella* SAG 3.93 (host for ATCV-1). *Chlorella* NC64A and Pbi are natural endosymbionts of the protozoan *Paramecium bursaria*. *Chlorella* SAG 3.83 is a natural endosymbiont of the heliozoon *Acanthocystis turfacea*.

Chlorella NC64A and Pbi were originally isolated from American and European strains of *P. bursaria*, respectively. Although NC64A and Pbi viruses have similar morphology and biological and biochemical properties, viruses infecting *Chlorella* NC64A cannot infect or attach to *Chlorella* Pbi. The component sugars in the cell walls of the two *Chlorella* differ considerably [135] and it was predicted that host receptor for the viruses could also be the recognition factor for the paramecia. However, this is not true since both *Chlorella* can establish symbiotic relationships with paramecium isolated

from American and European waters. *A. turfacea* virus was isolated from *Chlorella* from a German freshwater pond. Furthermore, viruses of the more recently discovered *Chlorella* SAG 3.83 endosymbiont of *A. turfacea* also only attach to and infect this *Chlorella* isolate.

Chlorella NC64A viruses have been isolated from freshwater collected all over the world: the United States [136, 137, 138], Japan, Brazil [139], Australia (Van Etten and J Rohozinski, unpublished), China [140], Argentina and Israel (Y Zhang, M Nelson, JL Van Etten, unpublished), South Korea, and Italy [20]. *Chlorella* Pbi viruses were originally collected in Europe [141], and more recently in Australia, Canada, and the northern United States (Minnesota, Wisconsin, and Montana) or at higher elevations in the western United States (M. Nelson, unpublished results). Some water samples from Nebraska (United States) and Australia contained both NC64A and Pbi viruses (J. Van Etten, and J. Rohozinski, unpublished; M. Nelson, unpublished). The most important factors influencing the distribution of NC64A and Pbi viruses are probably latitude and altitude.

Chlorella virus titers in indigenous water vary from 1-100 plaque-forming units (PFU) per ml, but titer as high as 1 x 10^5 PFU/ml have been obtained. The concentration of chlorella viruses fluctuates with the seasons, with the highest titers found in late spring [139, 142]. A water sample collected through ice had a titer of >1000 PFU/ml (M. Nelson, unpublished).

It is not known whether *Chlorella* NC64A and Pbi viruses replicate exclusively in the *Chlorella* free of the paramecium or whether they have another host(s). In fact, it is

not even known if the symbiotic *Chlorella* are free living in natural environments. A single paramecium can contain up to 1000 *Chlorella* cells; if each infected *Chlorella* can produce ~250 infectious viruses, then, it is predicted that *Chlorella* cells released from a single paramecium could produce up to 250,000 virus particles. This could explain the high virus titers occasionally observed.

The occasional high titers in indigenous water are also surprising because viruses are constantly exposed to solar radiation, that damages viral DNA and inactivates viruses. Inactivation of bacteriophages and cyanophages occurs at rates of 0.4 to 0.8 per hour in full sunlight [143, 144]. Chlorella viruses have adapted to solar radiation by having access to two independent DNA repair systems [76]. PBCV-1 uses a (i) pyrimidine dimmer-specific glycosylase that initiates UV-induced thymidine dimer repair, which functions in the light and in the dark; and (ii) host photolyase to repair UV-induced thymidine dimers (also called host cell reactivation) [76]. PBCV-1 also encodes a putative Cu/Zn superoxide dismutase that could help protect DNA from reactive oxygen species.

Plaque-forming viruses are specific to certain species of *Chlorella*, since attempts to find viruses infecting other *Chlorella* hosts or the green alga, e.g., *Chlamydomonas reinardt,* have been unsuccessful. It is possible that some viruses might exist in a lysogeny state with their hosts, but no evidence for lysogeny has been detected for chloroviruses.

7. OTHER ALGAL VIRUSES

Perhaps the next best studied algal virus is the coccolithovirus that infects the unicellular free-living calcifying *Emiliania huxleyi*, which is best known for forming large costal and open ocean algal blooms. Much of the appreciation for aquatic viruses came with the identification of the *Emiliania huxleyi virus* (EhV-1), partially responsible for the termination of *E. huxleyi* blooms. EhV-1 has a 407 kb circular genome (40% GC) that encodes 472 protein-encoding genes, of which 14% have a match in the public database [145]. The virus particle is a 160-200 nm icosahedron and the genome encodes for some unusual genes involved in sphingolipid biosynthesis, two desaturases, and eight proteases, but more importantly, encodes for 6 subunits of the RNA polymerase, which is lacking in PBCV-1. Upon infection, EhV-1 capsid remains at the cell surface, and it is speculated that binding occurs in the underlying plasma membrane of the host [146]. Interestingly, shell-less algae (haploid) are resistant to virus infection [147] and during release, EhV-1 buds through the membrane in an animal-like cell extrusion [146].

Another well characterized phycodnavirus belongs to the genus *Phaeovirus*. To date, eight classes of brown algae (macroalgae) are reported to be infected by viruses [15], however only two phaeovirus genomes have been sequenced: the 335 kb *Ectocarpus siliculosus* virus (EsV-1) [13] and the smaller genome (158 kb) of *Feldmania sp.* virus (FsV) [15]. The viruses in this group are probably the most diverse in terms of genome size (ranging from 150-335 kb) and they can infect more than one host from different families. This last factor probably reflects the unusual life style of these viruses that unlike the other phycodnavirus members that are lytic, integrates its genome into the

host genome and live in lysogeny. EsV-1 is the type-species of the group and perhaps the best characterized *Phaeovirus*. EsV-1 infects only free-swimming wall-less gametes or spores and after infection, virus DNA integrates to the host genome. The host exhibits no symptoms of infection, except for a partial or total inhibition of sexual reproduction, and the virus genome can be transmitted by mitosis. EsV-1 controls the host by only expressing genes in reproductive organ cells, where the virus disintegrates the organelles and forms a dense pack of virus particles. Upon certain environmental stimuli, EsV-1 induces lysis which releases spores and gametes, guaranteeing synchronous infection [13]. The FsV has two genome sizes, 158 kb and 178 kb that co-occur. Although the two genomes present similar restriction fragments and integration sites, they integrate in distinct regions in the host genome.

EsV-1 also encodes for a K^+ ion channel that structurally resembles PBCV-1 Kcv and eukaryotic K^+ channels. Despite their similarity, the EsV-1 protein (named Kesv), is sorted to the mitochondria whereas PBCV-1 Kcv goes to the plasma membrane. Mutation analysis revealed that the N-terminus and structural features in the second transmembrane domain are responsible for the mitochondrial signal in Kesv [148].

Recently, two viruses infecting the green alga *Ostreococcus tauri*, (OtV-5 and OtV-1), currently unclassified in the *Phycodaviridae* family, were isolated and their genomes sequenced [14, 149]. The host *O. tauri* is the smallest free-living, photosynthetic marine eukaryote, whose genome has also been sequenced [150]. OtV-5 and OtV-1 have a linear dsDNA genome (186 kb and 191 kb, respectively) and show high level of colinearity. Their genomes are predicted to encode 268 and 232 CDSs,

respectively. *O. tauri* lacks a cell wall and when it is attacked by high amounts of OtV-5 (high multiplicity of infection) not all cells become infected, suggesting either host resistance or a lack of virus receptor in the host cell membrane. This indicates that attachment is specific and a crucial step during infection. A 50 nm space can be microscopically observed between the virus and the host cell membrane and may indicate the presence of an envelope around host plasma membrane. The OtV-5 burst size (25 virus particles/cell) is smaller than that observed in other phycodnaviruses. OtV-5 encodes for DNA methyltransferases but not the companion restriction endonuclease as identified in PBCV-1, and during the lytic infection cycle the host chromosome remains intact [14].

It seems that algal virus diversity is even more spectacular than previously thought as shown by additional virus characterizations. In 2003, Tai and colleagues [151] reported the first single-stranded RNA (ssRNA) virus infecting the raphytophyte *Heterosigma akashiwo* and in 2004, another ssRNA virus infecting the marine diatom *Rhisozolenia sp.* [152]. In the same year, a double-stranded RNA (dsRNA) virus (24.6 kb) infecting the prasinophyte *Micromonas pusilla* was reported by Brussaard et al. (2004) [153]. Later in 2005, a virus infecting *Chaeotocerus salsugenum*, a marine diatom, has an intriguing genome type: the genome is composed of a 6-kb closed circle ssDNA with a 1-kb linear, partially paired ssDNA [154]. Finally, in 2008, Tomaru et al. [155] reported a ssDNA virus that seems to have a fragmented genome in the marine diatom *Chaetocerus debilis*. Furthermore, ssRNA and dsDNA viruses infecting dinoflagellates (Dinophyceae), and additional dsDNA viruses infecting prymnesiophytes (Haptophyceae) and Pelagophyceae were recently identified [156].

Collectively, algal viruses infect a wide range of hosts and presents a remarkable variety in size (25 to above 220 nm), genome type (dsDNA, ssDNA, dsRNA, ssRNA), genome structure (linear, circular, fragmented), and genome size (4.4 to above 500 kb) [156]. The diversity of algal viruses described so far is impressive considering this field is only in the beginning.

8. REFERENCES

1. 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* 306:1344-50.
2. Glass J.I., Assad-Garcia N., Alperovich N., Yooseph S., Lewis M.R., Maruf M., Hutchison C.A. 3rd, Smith H.O., Venter J.C. 2006. Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A*, 103:425-30.
3. Dunigan D.D., Fitzgerald L.A., Van Etten J.L. 2006. Phycodnaviruses: a peek at genetic diversity. *Virus Res*. 117:119-132.
4. Wilson, W.H., Van Etten, J.L., Schroeder, D.S., Nagasaki, K., Brussaard, C., Delaroque, N., Bratbak, G., and Suttle, C. 2005. Phycodnaviridae, p.163-175. *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., and Ball, L. A. (eds.), Virus Taxonomy: Classification and Nomenclature of Viruses. Elsevier Academic Press, San Diego, CA.
5. Chen, F., and Suttle, C. A. 1996. Evolutionary relationships among large doublestranded DNA viruses that infect microalgae and other organisms as inferred from DNA polymerase genes. *Virology* 219, 170-178.
6. Westbroek, P., Van Hinte, J. E., Brummer, G. J., Veldhuis, M., Brownlee, C., Green, J. C., Harris, R., and Heimdal, B. R. 1994. *Emiliania huxleyi* as a key to biosphere-geosphere interactions, p. 321-334. *In* J. C. Green (ed.), The Haptophyte Algae, Clarendon Press, Oxford.
7. Westbroek, P., Brown, C. W., Van Bleijswijk, J., Brownlee, C., Brummer, G. J., Conte, M., Egge, J., Fernadex, E., Jordan, R., Knappertsbusch, M., Stefels, J.,

Veldhuis, M., Van der Val, P., and Young, J. 1993. A model system approach to biological climate forcing--the example of *Emiliania huxleyi*, p.27-46. *In* Global Planetary Change, Elsevier, Amsterdam.

8. Charlson, R. J., Lovelock, J. E., Andreae, M. O., and Warren, S. G. 1987. Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. *Nature* 326: 655-661.

9. Malin, G., Liss, P. S., and Turner, S. M. 1994. Dimethyl sulfide production and atmospheric consequences, p. 303-320. *In* J.C. Green (ed.), The Haptophyte Algae, Clarendon Press, Oxford.

10. Wilson WH, Van Etten JL, Allen MJ. 2009. The Phycodnaviridae: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol*. 328:1-42.

11. Suttle C.A. 2007. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol.* 5:801-12.

12. Larsen J.B., Larsen A., Bratbak G., Sandaa R.A. 2008. Phylogenetic analysis of members of the Phycodnaviridae virus family, using amplified fragments of the major capsid protein gene. *Appl Environ Microbiol*. 74:3048-57.

13. 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.

14. Derelle E., Ferraz C., Escande M.L., Eychenié S., Cooke R., Piganeau G., Desdevises Y., Bellec L., Moreau H., Grimsley N. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri. PLoS One.* 3(5):e2250.

15. Schroeder D.C., Park Y., Yoon H.M., Lee Y.S., Kang S.W., Meints R.H., Ivey R.G., Choi T.J. 2009. Genomic analysis of the smallest giant virus--Feldmannia sp. virus 158. *Virology* 384:223-32.

16. Sandaa, R.A., Heldal M., Castberg T., Thyrhaug R., and Bratbak G. 2001. Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prasinophyceae). *Virology* 290:272-280.

17. Iyer L. M., Balaji S., Koonin E. V., Aravind L. 2006. Evolutionary genomics of Nucleo-Cytoplasmic Large DNA Viruses. *Virus Research* 117:156-84.

18. Filée J. 2009. Lateral gene transfer, lineage-specific gene expansion and the evolution of Nucleo Cytoplasmic Large DNA viruses. *J Invertebr Pathol* 101:169-71.

19. Grossman A.R. 2007. In the grip of algal genomics. *Adv Exp Med Biol* 616:54-76.

20. Van Etten J.L., Lane L.C., Meints R.H. 1991. Viruses and viruslike particles of eukaryotic algae. *Microbiol Rev* 55:586-620.

21. Rajamani S., Siripornadulsil S., Falcao V., Torres M., Colepicolo P., Sayre R. 2007. Phycoremediation of heavy metals using transgenic microalgae. *Adv Exp Med Biol* 616:99-109.

22. Harwood J.L., Guschina I.A. 2009. The versatility of algae and their lipid metabolism. *J Ind Microbiol Biotechnol* 36:269-74.

23. Li Y., Horsman M., Wu N., Lan C.Q., Dubois-Calero N. 2008. Biofuels from microalgae. *Biotechnol Prog* 24:815-20.

24. 5. Kessler E. 1976. Comparative physiology, biochemistry, and the taxonomy of Chlorella (Chlorophyceae). *Plant Systematics and Evolution* 125:129-138.

25. Takeda H. 1988. Classification of Chlorella strains by cell wall sugar composition. *Phytochemistry* 27:3823-3826.

26. Yamada T., Sakaguchi K. 1982. Comparative studies on Chlorella cell walls: Induction of protoplast formation. *Archives of Microbiology* 132:10-13.

27. Huss V.A.R., Frank C., Hartmann E.C., Hirmer M., Kloboucek A., Seidel B.M., Wenzeler P., Kessler E. 1999. Biochemical taxonomy and molecular phylogeny of the genus Chlorella sensu lato (Chlorophyta). *Journal of Phycology* 35:587-598.

28. Krienitz L., Hegewald E.H., Hepperle D., Huss V.A.R., Rohr T., Wolf T.M. 2004. Phylogenetic relationship of Chlorella and Parachlorella gen. nov.(Chlorophyta, Trebouxiophyceae). *Phycologia* 43: 529-542.

29. Kodama Y. and Fujishima M. 2009. Infection of *Paramecium bursaria* by symbiotic Chlorella species. *In* M. Fujishima (ed.), Endosymbionts in Paramecium, Microbiology Monographs 12, Springer-Verlag, Berlin Heidelberg.

30. Tikhonenko A.S., and Zavarzina N.B. 1966. Morphology of the lytic agents of Chlorella pyrenoidosa. *Mikrobiologiya* 35:850-852.

31. Zavarzina, N.B. 1961. A lytic agent in cultures of Chlorella pyrenoidosa Pringh. Dokl. Akad. Nauk SSSR Ser. Biol. 137: 435-437.

32. Zavarzina N.B. 1964. Lysis of Chlorella cultures in the absence of bacteria. *Mikrobiologiya* 33:561-564.

33. Zavarzina N.B., A.E. Protsenko. 1958. The lysis of Chlorella pyrenoidosa Pringh cultures. *Dokl. Akad. Nauk SSSR Ser. Biol.* 122:936-939.

34. Lee, R.E. 1971. Systemic viral material in the cells of the freshwater red alga *Sirodotia tenuissima* (Holden) Skuja. *J. Cell Sci*. 8:623-631.

35. Pickett-Heaps, J.D. 1972. A possible virus infection in the green alga *Oedogonium*. *J. Phycol.* 8:44-47.

36. Toth, R., Wilce R. T. 1972. Viruslike particles in the marine alga *Chorda tomentosa* Lyngbye (Phaeophyceae). *J. Phycol*. 8:126-130.

37. Manton, I., Leadbeater B. S. C. 1974. Fine-structural observations on six species of *Chrysochromulina* from wild Danish marine nanoplankton, including a discription of *C. campanulifera* sp. nov. and a preliminary summary of the nanoplankton as a whole. *Dan. Vidensk. Sel. Biol. Skr*. 20:1-26.

38. Mayer, J.A., Taylor F.J.R. 1979. A virus which lyses the marine nanoflagellate *Micromonas pusilla*. *Nature* 281:299–301.

39. Kawakami, H. and Kawakami N. 1978. Behavior of a virus in a symbiotic system, *Paramecium bursaria*-zoochlorella. *J. Protozool*. 25, 217-225.

40. Meints, R. H., Van Etten, J. L., Kuczmarski, D., Lee, K., Ang B. 1981. Viral infection of the symbiotic chlorella-like alga present in *Hydra viridis*. *Virology* 113:698-703.

41. 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* 113:704-11.

42. 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. USA* 79:3867-3871.

43. Bubeck, J.A., Pfitzner A.J. 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.

44. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Nierman W.C., Van Etten J.L. 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A. *Virology* 358:472-484.

45. Fitzgerald L.A., Graves M.V., Li X., Feldblyum T., Hartigan J., Van Etten J.L. 2007. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect Chlorella Pbi. *Virology* 358:459-471.

46. Fitzgerald L.A., Graves M.V., Li X., Hartigan J., Pfitzner A.J., Hoffart E., Van Etten J.L. 2007. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. *Virology* 362:350-361.

47. Rohozinski, J., Girton L.E., Van Etten J.L. 1989. Chlorella viruses contain linear nonpermuted double-stranded DNA genomes with covalently closed hairpin ends. *Virology* 168:363–369.

48. Yamada T., Higashiyama T. 1993. Characterization of the terminal inverted repeats and their neighboring tandem repeats in the chlorella CVK1 virus genome. *Mol. Gen. Genet.* 241:554–63.

49. Nelson M., Burbank D.E., Van Etten J.L. 1998. Chlorella viruses encode multiple DNA methyltransferases. *Biol. Chem.* 379:423–28.

50. Nelson M., Zhang Y., Van Etten J.L. 1993. DNAmethyltransferases and DNA site-specific endonucleases encoded by chlorella viruses, p. 186-211. *In* J.P. Jost, H.P. Saluz (eds.), DNA Methylation: Molecular Biology and Biological Significance, Basel: Birkhauser, Verlag.

51. 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.

52. Nandhagobal, 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 USA* 99:14758-14763.

53. Cherrier, M. V., V. A. Kostyuchenko, C. Xiao, V. D. Bowman, A. J. Battisti, X. Yan, P. R. Chipman, T. S. Baker, J. L. Van Etten, and M. G. Rossmann. 2009. An icosahedral algal virus has a complex unique vertex decorated by a spike. *Proc. Natl. Acad. Sci. USA* 106:11085-11090.

54. Kutish, G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L. 1996. Analysis of 76 kb of the chlorella virus PBCV-1 330- kb genome: map positions 182 to 258. *Virology* 223:303–317.

55. Li, Y., Lu Z., Burbank D.E., Kutish G.F., Rock D.L., Van Etten J.L. 1995. Analysis of 43 kb of the chlorella virus PBCV-1 330 kb genome: map position 45 to 88. *Virology* 212:134–50.

56. 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* 237:360–77.

57. Lu, Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L. 1996. Analysis of 94 kb of the chlorella virus PBCV-1 330–kb genome: map positions 88 to 182. *Virology* 216:102–23.

58. Lu, Z., Li Y., Zhang Y., Kutish G.F., Rock D.L., Van Etten J.L. 1995. Analysis of 45 kb of DNA located at the left end of the chlorella virus PBCV-1 genome. *Virology* 206:339–52.

59. Yanai-Balser G.M., Duncan G.A., Eudy J.D., Wang D., Li X., Agarkova I.V., Dunigan D.D., Van Etten J.L. 2010. Microarray analysis of chlorella virus PBCV-1 transcription. *J Virol. in press*.

60. 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* 180:763–69.

61. Van Etten, J.L. 2003. Unusual life style of giant chlorella viruses. *Annu Rev Genet.* 37:153-95.

62. Zhang Y., Maley F., Maley G.F., Duncan G., Dunigan D.D., Van Etten J.L. 2007. Chloroviruses encode a bifunctional dCMP-dCTP deaminase that produces two key intermediates in dTTP formation. *J Virol.* 81:7662-71.

63. 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*. 275:6915–21.

64. 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.

65. Plugge B., Gazzarrini S., Nelson M., Cerana R., Van Etten J.L., Derst C, DiFrancesco D., Moroni A., Thiel G. 2000. A potassium channel protein encoded by chlorella virus PBCV-1. *Science* 287: 1641–1644.

66. 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.

67. Hubscher, U., Nasheuer H.P., Syvaoja J.E. 2000. Eukaryotic DNA polymerases, a growing family. *Trends Biochem. Sci.* 25:143–147.

68. 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.

69. Warbrick, E. 2000. The puzzle of PCNA's many partners. *BioEssays* 22:997–1006 .

70. Ellison, V, Stillman B. 2001. Opening of the clamp: an intimate view of an ATP driven biological machine. *Cell* 106: 655–60.

71. Mossi, R., Hubscher U. 1998. Clamping down on clamps and clamp loaders—the eukaryotic replication factor C. *Eur. J. Biochem*. 254:209–216.

72. Kelman, Z., Hurwitz J. 2000. A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon *Methanobacterium thermoatuotrophicum* delta H. *J. Biol. Chem*. 275:7327–36.

73. Pisani, F.M., DeFelice M., Carpentieri F., Rossi M. 2000. Biochemical characterization of a clamp-loader complex homologous to eukaryotic replication factor C from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Mol. Biol.* 301:61–73.

74. Arezi, B., Kuchta R.D. 2000. Eukaryotic DNA primase. *Trends Biochem. Sci.* 25:572–576.

75. Tessman, I. 1985. Genetic recombination of the DNA plant virus PBCV-1 in a chlorella-like alga. *Virology* 145:319–322.

76. Furuta, M., Schrader J.O., Schrader H.S., Kokjohn T.A., Nyaga S., McCullough A.K., Lloyd R.S., Burbank D.E., Landstein D., Lane L., Van Etten J.L. 1997. Chlorella virus PBCV-1 encodes a homolog of the bacteriophage T4 UV DNA repair gene denV. *Appl. Environ. Microbiol*. 63:1551–1556.

77. Jaruga, P., Jabil R., McCullough A.K., Rodriguez H., Dizdaroglu M., Lloyd R.S. 2002. Chlorella virus pyrimidine dimmer glycosylase excises ultraviolet radiation and hydroxyl radical-induced products 4,6-diamino-5-formamidopyrimidine and 2,6-diamino-4-hydroxy-5- formamidopyrimidine from DNA. *Photochem. Photobiol*. 75:85–91.

78. Xia, Y., Van Etten J.L. 1986. DNA methyltransferase induced by PBCV-1 virus infection of a chlorella-like green alga. *Mol. Cell. Biol*. 6:1440–445.

79. Zhang, Y., Nelson M., Van Etten J.L. 1992. A single amino acid change restores DNAcytosine methyltransferase activity in a cloned chlorella virus pseudogene. *Nucleic Acids Res*. 20:1637–1642.

80. Agarkova, I.V., Dunigan D.D., Van Etten J.L. 2006. Virion-associated restriction endonucleases of chloroviruses. *J Virol*. 2006 80:8114-8123.

81. Zhang, Y., Moriyama H., Homma K., Van Etten J.L. 2005. Chlorella virus-encoded deoxyuridine triphosphatases exhibit different temperature optima. *J Virol.* 79:9945-9953.

82. 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.

83. Landstein, D., Mincberg M., Arad S., Tal J. 1996. An early gene of the Chlorella virus PBCV-1 encodes a functional aspartate transcarbamylase. *Virology* 221:151-158.

84. Fitzgerald, L.A., Zhang Y., Lewis G., Van Etten J.L. 2009. Characterization of a monothiol glutaredoxin encoded by Chlorella virus PBCV-1. *Virus Genes* 39:418-426.

85. 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.

86. Mujtaba, S., Manzur K.L., Gurnon J.R., Kang M., Van Etten J.L., Zhou M.M. 2008. Epigenetic transcriptional repression of cellular genes by a viral SET protein. *Nat Cell Biol*. 10:1114-22.

87. Graves, M.V, Meints R.H. 1992. Characterization of the gene encoding the most abundant in vitro translation product from virus-infected chlorella-like algae. *Gene* 113:149–155.

88. Graves, M.V., Meints R.H. 1992. Characterization of the major capsid protein and cloning of its gene from algal virus PBCV-1. *Virology* 188:198–207.

89. Schuster, A.M., Graves M., Korth K., Ziegelbein M., Brumbaugh J., et al. 1990. Transcription and sequence studies of a 4.3 kbp fragment from a dsDNA eukaryotic algal virus. *Virology* 176:515–23.

90. Ho, C.K., Gong C., Shuman S. 2001. RNA triphosphatase component of the mRNA capping apparatus of Paramecium bursaria Chlorella virus 1. *J. Virol*. 75:1744–1750.

91. Ho, C.K., Van Etten J.L., Shuman S. 1996. Expression and characterization of an RNA capping enzyme encoded by chlorella virus PBCV-1. *J. Virol.* 70: 6658–6664.

92. Gong, C., Shuman S. 2002. Chlorella virus RNA triphosphatase: mutational analysis and mechanism of inhibition by tripolyphosphate. *J. Biol. Chem*. 277: 15317–15324.

93. Shuman, S. 2000. Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog. Nucleic Acid Res. Mol. Biol.* 66:1–40.

94. Zhang, Y., Calin-Jageman I., Gurnon J.R., Choi T.J., Adams B., Nicholson A.W., Van Etten J.L. 2003. Characterization of a chlorella virus PBCV-1 encoded ribonuclease III. *Virology* 317:73-83.

95. Nishida, K., Kawasaki T., Fujie M., Usami S., Yamada T. 1999. Aminoacylation of tRNAs Encoded by Chlorella Virus CVK2. *Virology* 263:220-229.

96. 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.

97. 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. Cell* 14:168–176.

98. 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* 97:742–750.

99. Senkevich, T.G., White C.L., Koonin E.V., Moss B. 2000. A viral member of the ERV1/ALR protein family participates in a cytoplasmic pathway of disulfide bond formation. *Proc. Natl. Acad. Sci. USA* 97:12068–12073.

100. 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.

101. Moroni, A., Viscomi C., Sangiorgio V., Pagliuca C., Meckel T., et al. 2002. The short N-terminus is required for functional expression of the virus-encoded miniature KC channel Kcv. *FEBS Lett.* 530:65–69.

102. Gazzarrini, S., Severino M., Lombardi M., Morandi M., DiFrancesco D., Van Etten J.L., Thiel G., Moroni A. 2003. The viral potassium channel Kcv: structural and functional features. *FEBS Lett*. 552:12-16.

103. 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.

104. Neupärtl, M., Meyer C., Woll I., Frohns F., Kang M., Van Etten J.L., Kramer D., Hertel B., Moroni A., Thiel G. 2008. Chlorella viruses evoke a rapid release of K+ from host cells during the early phase of infection. *Virology* 372:340-348.

105. Greiner, T., Frohns F., Kang M., Van Etten J.L., Käsmann A., Moroni A., Hertel B., Thiel G. 2009. Chlorella viruses prevent multiple infections by depolarizing the host membrane. *J Gen Virol*. 90:2033-2039.

106. Hiramatsu, S., Fujie M., Usami S., Sakai K., Yamada T. 2000. Two catalytic domains of chlorella virus CVK2 chitinase. *J. Biosci. Bioeng*. 89:252–257.

107. Hiramatsu, S., Ishihara M., Fujie M., Usami S., Yamada T. 1999. Expression of a chitinase gene and lysis of the host cell wall during chlorella virus CVK2 infection. *Virology* 260:308–315.

108. Suda, K., Tanji Y., Hori K., Unno H. 1999. Evidence for a novel chlorella virus encoded alginate lyase. *FEMS Microbiol. Lett*. 180:45–53.

109. Sugimoto, I., Hiramatsu S., Murakami D., Fujie M., Usami S., Yamada T. 2000. Algal-lytic activities encoded by chlorella virus CVK2. *Virology* 277: 119–126.

110. Sun, L., Adams B., Gurnon J.R., Ye Y., Van Etten J.L. 1999. Characterization of two chitinase genes and one chitosanase gene encoded by chlorella virus PBCV-1. *Virology* 263:376–387.

111. Sun, L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L. 2000. Characterization of a β-1,3-glucanase encoded by chlorella virus PBCV-1. *Virology* 276:27–36.

112. Yamada, T., Hiramatsu S., Songsri P., Fujie M. 1997. Alternative expression of a chitosanase gene produces two different proteins in cells infected with chlorella virus CVK2. *Virology* 230:361–68.

113. 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.

114. Meints, R.H., Burbank D.E., Van Etten J.L., Lamport D.T.A. 1988. Properties of the chlorella receptor for the virus PBCV-1. *Virology* 164:15–21.

115. Doms, R.W., Lamb R.A., Rose J.K., Helenius A. 1993. Folding and assembly of viral membrane proteins. *Virology* 193:545–562.

116. Olofsson, S., Hansen J.E.S. 1998. Host cell glycosylation of viral glycoproteins: a battlefield for host defence and viral resistance. *Scand. J. Infect Dis*. 30:435–440.

117. Wang, I.-N., Li Y., Que Q., Bhattacharya M., Lane L.C., et al. 1993. Evidence for virus encoded glycosylation specificity. *Proc. Natl. Acad. Sci. USA* 90:3840–3844.

118. DeAngelis, P.L., Jing W., Graves M.V., Burbank D.E., Van Etten J.L. 1997. Hyaluronan synthase of chlorella virus PBCV-1. *Science* 278:1800–1803.

119. 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.

120. DeAngelis, P.L. 1999. Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cell Mol. Life Sci*. 56: 670–682.

121. DeAngelis, P.L. 2002. Evolution of glycosaminoglycans and their glycosyltransferases: Implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. *Anat. Rec.* 268:317–326.

122. Kawasaki, T., Tanaka M., Fujie M., Usami S., Sakai K., Yamada T. 2002. Chitin synthesis in chlorovirus CVK2–infected chlorella cells. *Virology* 302:123–131.

123. Gooday G.W., Humphreys A.M., McIntosh W.H. 1986. *In* R. Muzzarelli, C. Jeuniaus, G.W Gooday (eds.), p 83-91, Role of chitinases in fungal growth, in Chitin in Nature and Technology, Plenum Press, New York.

124. Yamada, T., Kawasaki T. 2005. Microbial synthesis of hyaluronan and chitin: New approaches. *J Biosci Bioeng*. 99:521-528.

125. Banerji S., Flieger A. 2004. Patatin-like proteins: a new family of lipolytic enzymes present in bacteria? *Microbiology* 150:522-525.

126. 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.* 279:35760-7.

127. Kaiser, A., Vollmert M., Tholl D., Graves M.V., Xing W., et al. 1999. Chlorella virus PBCV-1 encodes a functional homospermidine synthase. *Virology* 263:254–62.

128. Baumann, S., Sander A., Gurnon J.R., Yanai-Balser G.M., Van Etten J.L., Piotrowski M. 2007. Chlorella viruses contain genes encoding a complete polyamine biosynthetic pathway. *Virology* 360:209-217.

129. Agarkova, I., Dunigan D., Gurnon J., Greiner T., Barres J., Thiel G., Van Etten J.L. 2008. Chlorovirus-mediated membrane depolarization of Chlorella alters secondary active transport of solutes. *J Virol.* 82:12181-12190.

130. 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 USA* 103:5355-5560.

131. Fruscione, F., Sturla L., Duncan G., Van Etten J.L., Valbuzzi P., De Flora A., Di Zanni E., Tonetti M. 2008. Differential role of NADP+ and NADPH in the activity and structure of GDP-D-mannose 4,6-dehydratase from two chlorella viruses. *J Biol Chem.* 283:184-193.

132. Gazzarrini, S., Kang M., Abenavoli A., Romani G., Olivari C., Gaslini D., Ferrara G., Van Etten J.L., Kreim M., Kast S.M., Thiel G., Moroni A. 2009. Chlorella virus ATCV-1 encodes a functional potassium channel of 82 amino acids. *Biochem J*. 420:295-303.

133. 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* 214:413–20.

134. 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* 295:289–98.

135. Kapaun, E., Loos E., Reisser W. 1992. Cell wall composition of virus-sensitive symbiotic chlorella species. *Phytochemistry* 31:3103–3104.

136. Schuster, A. M., Burbank, D. E., Meister, B., Skrdla, M. P., Meints, R. H., Hattman, S., Swinton, D., Van Etten, J. L. 1986. Characterization of viruses infecting a eukaryotic Chlorella-like green alga. *Virology* 150:170-177.

137. Van Etten, J. L., Burbank, D. E., Schuster, A. M., Meints, R. H. 1985. Lytic viruses infecting a Chlorella-like alga. *Virology* 140:135-143.

138. Van Etten, J. L., Van Etten, C. H., Johnson, J. K., Burbank, D. E. 1985. A survey for viruses from freshwater that infect a eukaryotic chlorella-like green alga. *Appl. Environ. Microbiol.* 49: 1326-1328.

139. Yamada, T., Higashiyama, T., Fukuda, T. 1991. Screening of natural waters for viruses which infect chlorella cells. Appl. Environ. Microbiol. 57, 3433.

140. Zhang, Y., Burbank, D. E., Van Etten, J. L. 1988. Chlorella viruses isolated in China. *Appl. Environ. Microbiol.* 54:2170.

141. Reisser W., Burbank D.E., Meints S.M., Meints R.H., Becker B., Van Etten J.L. 1988. A comparison of viruses infecting two different chlorella-like green algae. *Virology* 167:143–49.

142. van Hulten, M.C., Witteveldt J., Peters S., Kloosterboer N., Tarchini R., et al. 2001. The white spot syndrome virus DNA genome sequence. *Virology* 286:7–22.

143. Suttle, C.A., Chan A.M. 1994. Dynamics and distribution of cyanophages and their effect on marine *Synechococcus* spp. *Appl. Environ. Microbiol.* 60:3167–3174.

144. Suttle, C.A., Chen F. 1992. Mechanisms and rates of decay of marine viruses in seawater. *Appl. Environ. Microbiol*. 58:3721–3729.

145. 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. 2005. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* 309:1090-1092.

146. Mackinder, L.C., Worthy C.A., Biggi G., Hall M., Ryan K.P., Varsani A., Harper G.M., Wilson W.H., Brownlee C., Schroeder D.C. 2009. A unicellular algal virus, *Emiliania huxleyi* virus 86, exploits an animal-like infection strategy. *J Gen Virol.* 90:2306-2316.

147. Frada, M., Probert I., Allen M.J., Wilson W.H., de Vargas C. 2008. The "Cheshire Cat" escape strategy of the coccolithophore *Emiliania huxleyi* in response to viral infection. *Proc Natl Acad Sci USA* 105:15944-15949.

148. Balss, J., Papatheodorou P., Mehmel M., Baumeister D., Hertel B., Delaroque N., Chatelain F.C., Minor D.L. Jr., Van Etten J.L., Rassow J., Moroni A., Thiel G. 2008. Transmembrane domain length of viral K+ channels is a signal for mitochondria targeting. *Proc Natl Acad Sci USA* 105:12313-12318.

149. Weynberg, K.D., Allen M.J., Ashelford K., Scanlan D.J., Wilson W.H. 2009. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environ Microbiol.* 11:2821-2839.

150. Palenik, B., Grimwood J., Aerts A., Rouzé P., Salamov A., Putnam N., Dupont C., Jorgensen R., Derelle E., Rombauts S., Zhou K., Otillar R., Merchant S.S., Podell S., Gaasterland T., Napoli C., Gendler K., Manuell A., Tai V., Vallon O., Piganeau G., Jancek S., Heijde M., Jabbari K., Bowler C., Lohr M., Robbens S., Werner G., Dubchak I., Pazour G.J., Ren Q., Paulsen I., Delwiche C., Schmutz J., Rokhsar D., Van de Peer Y., Moreau H., Grigoriev I.V. 2007. The tiny eukaryote Ostreococcus provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci USA* 104:7705-7710.

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

152. Nagasaki, K., Tomaru Y., Katanozaka N., Shirai Y., Nishida K., Itakura S., Yamaguchi M. 2004. Isolation and characterization of a novel single-stranded RNA virus infecting the bloom-forming diatom *Rhizosolenia setigera*. *Appl. Environ. Microbiol.* 70:704-711.

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

154. Nagasaki, K., Tomaru Y., Takao Y., Nishida K., Shirai Y., Suzuki H., Nagumo T. 2005. Previously unknown virus infects marine diatom. *Appl. Environ. Microbiol.* 71:3528-3535.

155. Tomaru, Y., Shirai Y., Suzuki H., Nagumo T., Nagasaki K. 2008. Isolation and characterization of a novel single-stranded DNA virus infecting a cosmopolitan marine diatom *Chaetoceros debilis*. *Aquat. Microb. Ecol.* 50:103-112.

156. Nagasaki, K. 2008. Dinoflagellates, diatoms, and their viruses. *J Microbiol.* 46:235-243.

CHAPTER II

MICROARRAY ANALYSIS OF CHLORELLA VIRUS *PARAMECIUM BURSARIA* CHLORELLA VIRUS-1 TRANSCRIPTION

1. OVERVIEW

Paramecium bursaria chlorella virus 1 (PBCV-1), the prototype of the genus *Chlorovirus* (family Phycodnaviridae), is a large, icosahedral (190 nm in diameter), plaque-forming virus that infects the unicellular, eukaryotic green alga *Chlorella* NC64A. The PBCV-1 virion has a lipid membrane located inside an outer glycoprotein capsid. The 330-kb genome is a linear, non-permutated, dsDNA molecule with covalently closed hairpin ends that has approximately 365 protein encoding genes (CDS), as well as 11 tRNA encoding genes (reviews [1, 2, 3]). The CDSs are evenly distributed on both strands and intergenic space is minimal (typically less than 100 nucleotides); the exception is a 1788 bp sequence in the middle of the genome that encodes the tRNA genes. Approximately 35% of the 365 PBCV-1 gene products resemble proteins in the public databases.

PBCV-1 initiates infection by attaching rapidly and specifically to the cell wall of its host [4, 5], probably at a unique virus vertex [6, 7]. Attachment is immediately

followed by host cell wall degradation by a virus-packaged enzyme(s) at the point of contact. Following wall degradation, the viral internal membrane presumably fuses with the host membrane causing host membrane depolarization [8], potassium ion efflux [9], and increase in the cytoplasm pH [10]. These events are predicted to facilitate entry of the viral DNA and virion-associated proteins into the cell. PBCV-1 lacks a gene encoding a recognizable RNA polymerase or a subunit of it and RNA polymerase activity is not detected in PBCV-1 virions. Therefore, viral DNA and virion-associated proteins are predicted to migrate to the nucleus and early viral transcription is detected 5-10 min post-infection (p.i.), presumably by commandeering a host RNA polymerase(s) (possibly RNA polymerase II) [11, 12]. Virus DNA synthesis begins 60-90 min p.i., followed by virus assembly at 3-5 h p.i. in localized regions of the cytoplasm, called virus assembly centers [4]. At 6-8 h p.i., virus-induced host cell lysis occurs resulting in release of progeny virions (~1,000 viruses/cell of which ~25% are infectious). These events are depicted in Fig. 1.

To initiate PBCV-1 transcription, the host RNA polymerase(s), possibly in combination with a virus transcription factor(s), must recognize virus DNA promoter sequences. Recently, three short nucleotide sequences were identified in putative virus promoter regions (150 bp upstream and 50 bp downstream of the ATG translation site) that are conserved in PBCV-1 and other *Chlorovirus* members [13]. PBCV-1 CDSs are not spatially clustered on the genome by either temporal or functional class suggesting that transcription regulation must occur via *cis*- and possible *trans*-acting regulatory elements.

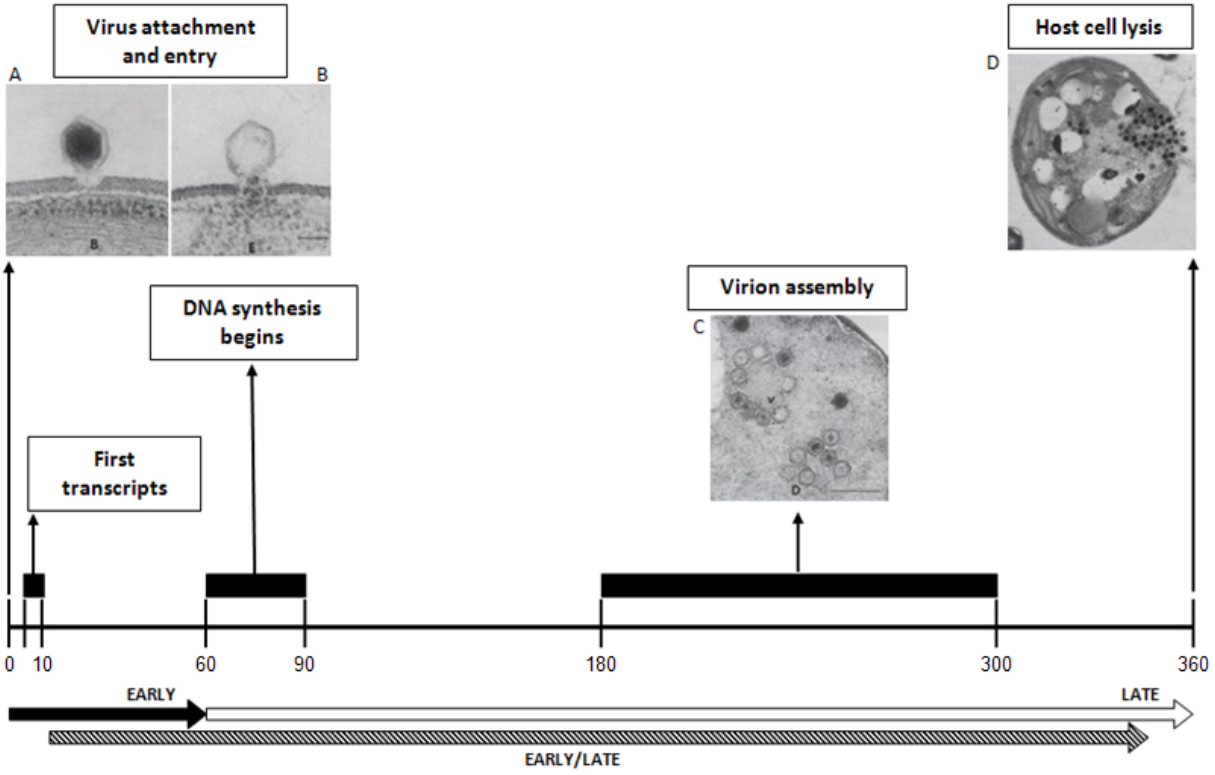


Figure 1. Timeline representing the PBCV-1 lifecycle in Chlorella NC64A*. Numbers represent minutes after infection. CDSs expressed before viral DNA synthesis begins were classified as early (black arrow), CDSs expressed after DNA synthesis begins were classified as late (white arrow), and CDSs expressed before and after DNA synthesis begins were classified as early/late (arrow with diagonal lines). Electron micrographs A and B were reproduced from [5] and micrographs C and D were reproduced from [4] with permission.*

To understand the dynamics of PBCV-1 global gene expression during virus replication, we constructed a microarray containing 50-mer probes to each of the 365 PBCV-1 CDSs. cDNAs from poly A-containing RNAs isolated from cells at seven times

after PBCV-1 infection were competitively hybridized against a reference sample on the microarray. To further delineate early and late gene expression, cells were treated with the DNA replication inhibitor, aphidicolin, prior to infection. The results provide the first comprehensive transcriptional map of the virus genome, conferring insights about the characterization of each PBCV-1 CDS, the majority of which have unknown functions. In addition, the microarray data suggest that viral DNA replication plays a significant role in the temporal regulation of gene expression.

2. MATERIALS AND METHODS

2.1. RNA isolation and drug treatment. *Chlorella* NC64A cells (1 x 10^8 cells/ml) were infected with PBCV-1 at a multiplicity of infection of 5 to ensure synchronous infection. Uninfected cells and cells at 20, 40, 60, 90, 120, 240 and 360 min p.i. were harvested by centrifugation (4,000 rpm) for 5 min at 4ºC and disrupted with glass beads (0.25-0.30 mm in diameter) using a bead beater (Disruptor Genie, Scientific Industries, Bohemia, NY) in the presence of Trizol (Invitrogen, Carlsbad, CA). RNAs were isolated using the Absolutely RNA Miniprep Kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. RNA integrity was verified in denaturing 1% agarose gels by monitoring host cytoplasmic and chloroplast ribosomal RNAs. Total RNA was quantified with a NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE).

To determine the effect of virus DNA synthesis on virus gene expression, aphidicolin (20 μg/ml) was added to the cells 15 min prior to infection and samples were

collected at the same times after infection as described above. Control samples were obtained from infected non-treated cells at the same times. Preliminary experiments to determine optimal drug dosage and time of application indicated that 20 μg/ml of aphidicolin completely inhibits DNA synthesis in 15 min.

2.2. Microarray construction and hybridization. The PBCV-1 genome is predicted to have 365 CDSs and 11 tRNA encoding genes. The tRNAs sequences were not included in the microarrays. Fifty-mer probes representing each CDS in the PBCV-1 genome were designed and synthesized by MWG Biotech (Ebersberg, Germany) (20-80% GC, 60-80º C melting temperature). A table with the probes' sequences is in Supplement 1. Probes were spotted onto CMT-GAPS silane-coated slides (Corning, Lowell, MA) using Omnigrid 100 (Genomic Solutions, Ann Arbor, MI) according to manufacturer's instructions. Probes were printed in quadruplicates on every slide. For each time point, 20 μg of total RNA was reverse-transcribed using oligo(dT) as a primer and cDNAs were labeled with either Cy3 or Cy5-dUTP (GE Healthcare, Piscataway, NJ) using the SuperScript Indirect cDNA Labeling System (Invitrogen) following the supplier's directions. The reference sample, for the time course experiments, consisted of a pool of transcripts obtained by mixing equal amounts of total RNA from each time point. Competitive hybridization experiments were conducted for each sample against the reference sample [14, 15, 16]. For the aphidicolin experiments, a direct comparison was carried out with each treated sample versus the corresponding non-treated infected control.

Labeled cDNAs were resuspended in pre-heated (68ºC) 40 µl of Ambion hybridization buffer #2 (Ambion, Austin, TX). Arrays were hybridized (42ºC) for 16 hours in a Corning hybridization chamber (Corning, Lowell, MA). Slides were washed twice (42ºC) in 2x SSC (1x SSC is 0.15 M sodium chloride, 0.015 M sodium citrate), 0.5% SDS for 15 min followed by two washes in 0.5x SSC, 0.5% SDS for 15 min. Slides were then dried with low-speed centrifugation and subjected to fluorescence detection with an Axon 4000B scanner (Molecular Devices, Sunnyvale, CA).

2.3. Microarray analysis. Results from three independent experiments were analyzed using the GenePix Pro v.6.0 software (Molecular Devices) and TIGR microarray software suite (TM4) [17]. Several transformations were performed to eliminate low quality data, to normalize the measured intensities using the Lowess algorithm, and to regulate the standard deviation of the intensity of the Cy5/Cy3 ratio across blocks. CDSs that displayed statistically significant modulation were identified by a one-way analysis of variance (ANOVA), using P values of <0.01 as a cutoff. For the aphidicolin experiments, Significant Analysis of Microarray (SAM) [18] was used to identify CDSs with statistically significant changes in expression in comparison to a non-treated infected sample (False discovery rate $\leq 5\%$). CDSs with similar expression profiles were grouped into different clusters with a K-means algorithm using Euclidean distance and 50 maximum iterations. PBCV-1 microarray data sets were deposited at NCBI's Gene Omnibus Express (GEO) under the accession number GSE18421.

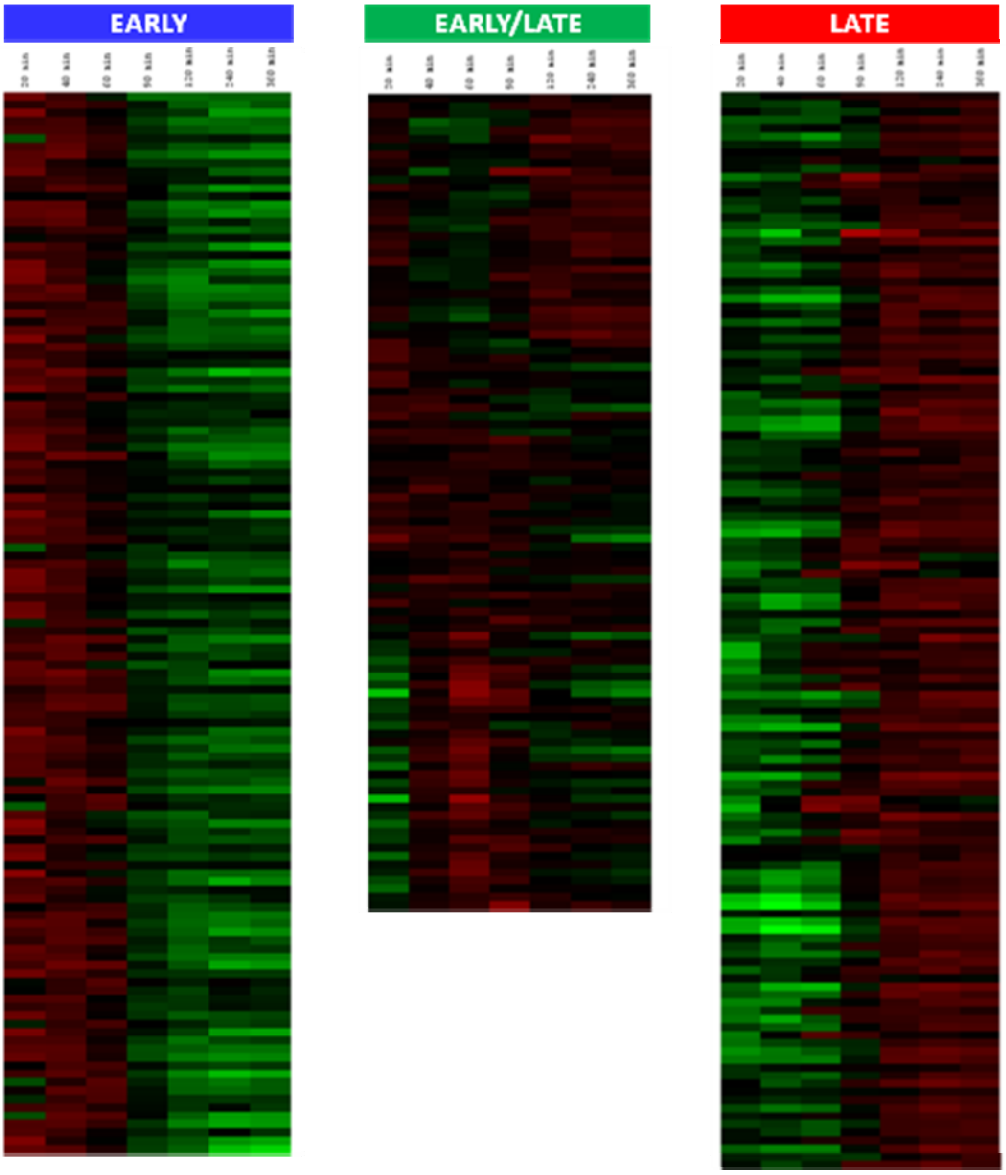


Figure 2. Heatmap illustrating the expression of 360 PBCV-1 CDSs during the infection cycle. *cDNAs from each time point were labeled with Cy5 and the reference control was labeled with Cy3.* ***Color code represents the log2 (Cy5/Cy3) ratio for each time point. CDSs with similar expression profiles were grouped into three classes representing early, early/late, and late using K-means algorithm. Each column corresponds to the time point when total RNA was collected (numbers represent minutes***

after infection). Each row represents a different CDS in PBCV-1. A list of all of the CDSs is available in Supplement 2.

3. RESULTS AND DISCUSSION

3.1. Microarray quality check. To evaluate our probes, equal amounts of PBCV-1 genomic DNA (2 μg) were labeled in two independent reactions with either Cy3-dCTP or Cy5-dCTP (GE Healthcare) using random primers (Invitrogen). The two reactions were then competitively hybridized with the probes in the microarray. No difference was detected in hybridization of the two DNA samples on all spots (results not shown) indicating the probes were specific and also excluding any preferential hybridization to one of the dyes. In addition, to check for host cross-hybridization, PBCV-1 DNA (2 μg) was labeled with Cy3-dCTP and host *Chlorella* NC64A DNA (2 μg) with Cy5-dCTP. Only two PBCV-1 CDSs (A260R and A625R) hybridized with host DNA; however, these two CDSs are false-positives since recently available *Chlorella* NC64A genome sequence did not have detectable homologous sequences.

3.2. PBCV-1 transcription program. RNAs were isolated from infected cells at 20, 40, 60, 90, 120, 240, and 360 min p.i. Competitive hybridization results from each time point against the RNA reference pool revealed that transcripts of 360 (99%) of the 365 PBCV-1 CDSs display statistically significant variation in at least one of the experimental time points (Fig. 2). CDSs A60L, A328L, A482R, and A646L did not pass

statistical tests in the time course experiment. CDS A689L was not spotted onto the array because it was accidentally omitted during the probe synthesis. The gene expression analysis was based on relative levels, rather than absolute levels of expression. Mapping the PBCV-1 transcription pattern to the genome revealed no large regions that were biased as to time of expression, indicating that gene expression in PBCV-1 is mostly controlled by multiple initiation sites (Fig. 3). However, a few early or late CDSs clustered in small regions of the genome as can be seen in shaded areas in Fig. 3. Interestingly, these small regions are also clustered and conserved in another sequenced chlorovirus (NY-2A) that infects the same host Chlorella (results not shown). Unlike the phage T4 genome, where most late CDSs are located in contiguous regions on the same DNA strand [19], PBCV-1 expression did not show a strong strand-specific bias. In addition, there is no relationship between time of expression and G+C content of the genome (Fig. 3).

Classification of PBCV-1 CDSs was based on when the transcript was detected by the microarray. Globally, the 360 statistically significant CDSs were grouped into three classes (Fig. 4): i) 227 (62%) of the CDSs were expressed before viral DNA synthesis begins at 60-90 min p.i.; ii) these 227 CDSs were divided into two classes: transcripts of 127 CDSs disappeared prior to initiation of virus DNA synthesis (considered early) while transcripts of 100 CDSs were still detected after virus DNA synthesis begins (considered early/late); iii) transcripts of 133 (37%) CDSs were detected after virus DNA synthesis begins (considered late). Functional categorization of PBCV-1 CDSs was reported elsewhere [20] and functional distribution compared to each transcriptional category is summarized in Fig. 4 and 5. Forty-four of the PBCV-1 encoded proteins have been

expressed and recombinant proteins shown to be functional enzymes. These are indicated with an asterisk in Fig. 5. The functions of the remaining CDSs are either putative or unknown.

A previously described putative promoter sequence (AATGACA) and a similar sequence (ATGACAA) [13, 11] were detected in 50 early or early/late PBCV-1 CDSs. However, promoter sequences for most early, early/late, and late CDSs remain unidentified.

3.3. Early CDSs. One hundred and twenty-seven (35%) of the 360 PBCV-1 CDSs were expressed early, 20-60 min p.i. (Supplement 2). Sixty-one percent of the early CDSs have no known function. Many of the early CDSs are predicted to encode the machinery for the virus to begin DNA replication. In fact, PBCV-1 encodes seven proteins involved in DNA replication, recombination and repair that were expressed early including: δ DNA polymerase (A185R), superfamily III helicase (A456L), DNA topoisomerase II (A583L), RNAse H (A399R), and PCNA (A574L). A pyrimidine dimer-specific glycosylase (A50L), a well-characterized DNA repair enzyme involved in pyrimidine photodimer excision [21], was also expressed early. Additional PBCV-1 encoded proteins involved in virus DNA synthesis and DNA recombination were in the early/late class including, DNA primase (A468R), a 5'-3' exonuclease (A166R) and a second PCNA (A193L). The PBCV-1 genome contains methylated nucleotides, both N6-methylated adenine and 5-methylcytosine [22]. Therefore, it is not surprising that the virus encodes three functional DNA methyltransferases that were transcribed early – two

enzymes that form N6-methyladenine (A251R and A581R) and one that forms 5-methylcytosine (A517L).

PBCV-1 DNA synthesis also requires large quantities of deoxynucleotide triphosphates (dNTPs) that cannot be accounted for simply by recycling deoxynucleotides from host DNA. By 4 h p.i., the total amount of DNA in the cell increases 4-fold due to viral DNA synthesis [23]. To guarantee a supply of dNTPs in non-proliferating host cells, large DNA viruses, including PBCV-1, encode proteins involved in dNTP biosynthesis: dUTP pyrophosphatase (A551L), thioredoxin (A427L), thymidylate synthase X (A674R), and cytosine deaminase (A200R) CDSs were transcribed early. Additional dNTP synthesizing CDSs in the early/late class included aspartate transcarbamylase (A169R), both subunits of ribonucleotide reductase (A476R and A629R), glutaredoxin (A438L), and dCMP deaminase (A596R).

Several PBCV-1 CDSs predicted to encode proteins involved in transcription were also expressed early. These proteins include three putative transcription factors, TFIIB (A107L), TFIID (A552R), TFIIS (A125L), two helicases, SWI/SNF helicase (A548L) and superfamily II helicase (A241R), and RNAse III (A464R). The genes for two enzymes involved in mRNA capping, an RNA triphosphatase (A449R) and a guanylyltransferase (A103R) were also transcribed early. The products of at least some of these early CDSs are undoubtedly involved in the switching of virus early gene transcription to late gene transcription.

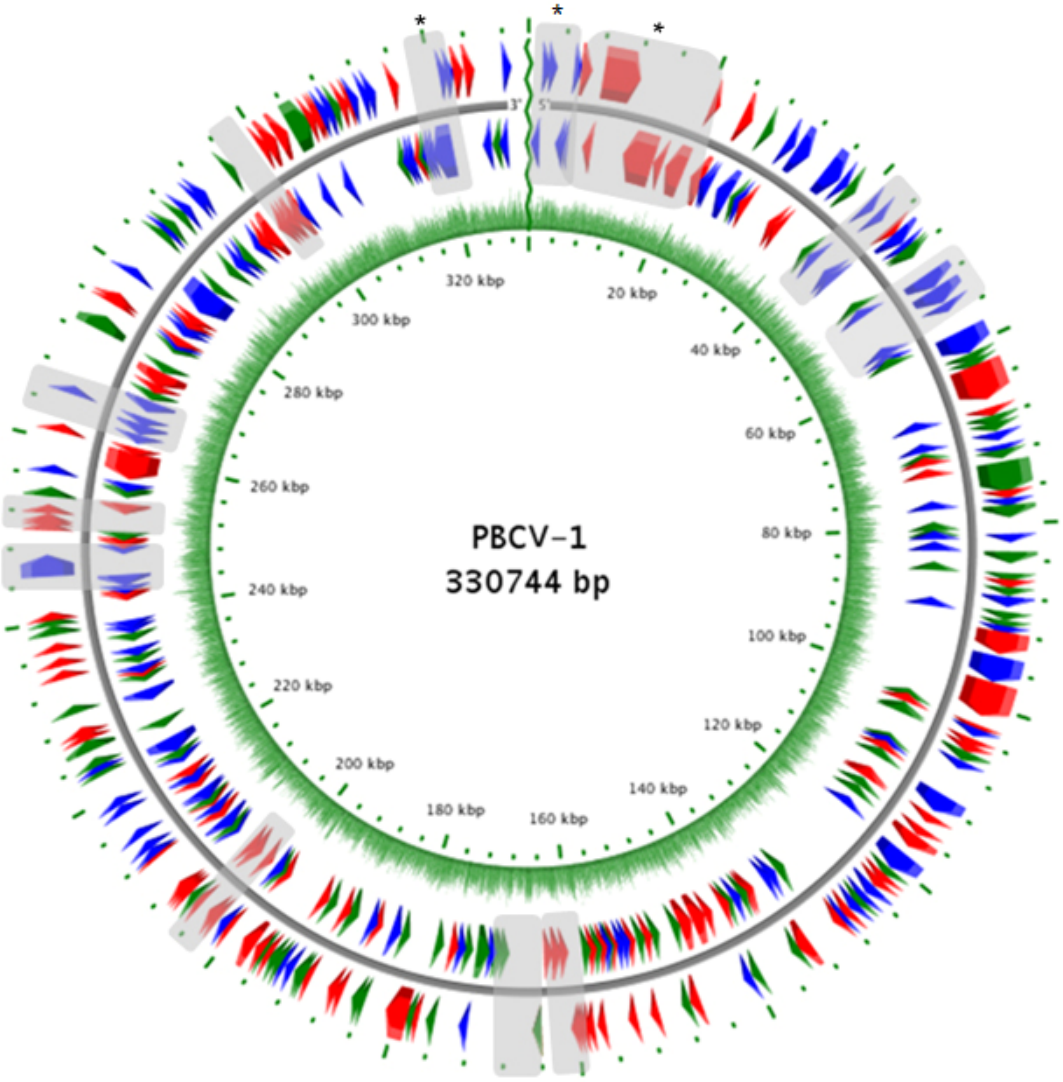


Figure 3. Mapping of the PBCV-1 transcriptome. *Blue arrows: early CDSs; green arrows: early/late CDSs; red arrows: late CDSs. Arrow points to transcription direction. Shaded areas indicate small transcription CDS clusters that are also conserved in another chlorovirus (NY-2A), except for the areas marked with asterisks. Middle circle shows G+C content of the genome. Note that the PBCV-1 genome is linear and a circular map was generated for illustration purposes only.*

A few PBCV-1 enzymes involved in protein synthesis and protein degradation were transcribed early including translation elongation factor-3 (A666L), ubiquitin C-terminal hydrolase (A105L), Skp1 protein (A39L), SCF-E3 ubiquitin ligase (A481L), a zinc metallopeptidase (A604L), and an ATPase (AAA + class) (A44L).

Unlike other viruses, PBCV-1 encodes at least part, if not all, of the machinery required to glycosylate its major capsid protein [19], including five glycosyltransferases [24, 25, 26]. Furthermore, glycosylation of the virus major capsid protein probably occurs independently of the host endoplasmic reticulum-Golgi system [19]. All of the glycosyltransferase CDSs were expressed early (A64R, A111/114R, A219/222/226R, A473L, A546L).

The chlorella viruses are also unusual because they encode enzymes involved in sugar metabolism. Two PBCV-1 encoded enzymes synthesize GDP-L-fucose from GDP-D-mannose, GDP-D-mannose dehydratase (A118R), and fucose synthase (A295L) [27, 28] and three enzymes, glucosamine synthetase (A100R), UDP-glucose dehydrogenase (A609L), and hyaluronan synthase (A98R), contribute to the synthesis of hyaluronan, a linear polysaccharide composed of alternating β–1,4-glucuronic acid and β-1,3-N-acetylglucosamine residues [29, 30]. The CDSs for these five enzymes were expressed early.

3.4. Early/Late CDSs. One hundred (27%) CDSs were classified as early/late, 67 (67%) of which have unknown function (Supplement 2). This class contains CDSs that were

expressed before 60 min p.i., but whose transcripts were also present after PBCV-1 DNA synthesis begins. At least three mechanisms can lead to classification of CDSs into the early/late class: (i) the CDSs were transcribed both before and after viral DNA replication begins; (ii) the CDSs were only transcribed prior to initiation of virus DNA synthesis but complete degradation of their transcripts only occurred after DNA synthesis begins; (iii) the CDSs encode polycistronic mRNAs, e.g., one could have a dicistronic mRNA in which one CDS is required for an early function and the other CDS is required for a late function. To add to the complexity, ~30 transcripts were detected early, disappeared and then reappeared as late transcripts (these CDSs are marked in Supplement 2). A similar phenomenon has been reported in transcriptional studies with the bacteriophage T4 [15] and Red Sea bream iridovirus [31].

In addition to the early/late CDSs described in the preceding section, the most striking feature of this early/late class was the presence of many genes encoding proteins associated with genome integration, including 5 GIY-YIG endonucleases (A134L, A287R, A315L, A495R, A651L) and 4 HNH endonucleases (A354R, A422R, A478L, A490L). Also, one of the two transposases (A366L) coded by PBCV-1, was expressed early/late. The functions of these proteins in the PBCV-1 life cycle are unknown.

3.5. Late CDSs. One hundred and thirty-three (37%) of the 360 CDSs were classified as late CDSs (Supplement 2), 74% of which have no match in the public databases. The expectation is that most late CDS products are involved in either virus capsid assembly, DNA packaging, virus release, or are packaged in the virus particles. Indeed, SDS-

PAGE mass spectrometry proteomic analysis of purified virions indicated that 118 PBCV-1 gene products were detected in the virions (Dunigan et al., unpublished results). Of these 118 proteins, 83 had their corresponding CDS transcribed late, 29 were transcribed early/late, while only 6 virion-associated proteins were expressed early.

At least 9 late CDSs are predicted to have their gene products serving a purely structural role in the virion including the major capsid protein A430L [32] and A140/145R, which is associated with the unique vertex of the virus [7]. Other putative structural proteins are A189/192R, A363R, A384L, A540L, A558L, A561L, A622L. Homologs of all these proteins are present in all of the chloroviruses (Dunigan et al., unpublished results). In addition to structural proteins, a DNA binding protein (A437L), predicted to aid in neutralization of the virus dsDNA, was in the virion and its gene transcribed late.

Several CDSs encoding proteins with putative enzyme functions were expressed late and their gene products were packaged in the PBCV-1 virion. Presumably these enzymes are released into the cell during virus infection and aid in establishing infection. These proteins include an ATPase (A561L), a superfamily II helicase (A363R), a SWI/SNF chromatin remodeling complex subunit (A189/192R), and a SET domain-containing histone 3, Lys27 methyltransferase (named vSET) (A612L). The vSET protein is predicted to aid in the rapid inhibition of host transcription during virus infection [33]. Two site-specific (restriction) endonucleases (A252 and A579L) were also expressed as late CDSs and their proteins are packaged in the virion. The restriction endonucleases are involved in host chromosome degradation, which begins within a few

minutes after virus infection [34]. Thus, PBCV-1 has at least two avenues to inhibit host transcription and subvert the host RNA polymerase (presumably RNA polymerase II) for virus transcription: virus packaged restriction endonucleases and a vSET histone 3, Lys27 methyltransferase.

Five of the 8 PBCV-1 encoded Ser/Thr protein kinase CDSs were expressed late (A34R, A277L, A278L, A282L and A614L). With the exception of A277L, the other four kinases were packaged in the virion (Dunigan et al., unpublished results). A dual specific phosphatase (A305L) was also expressed late and the protein is present in the virion. These results indicate that PBCV-1 has the potential to release several protein kinase/phosphatase proteins into the cell during infection; these enzymes are probably involved in regulatory mechanisms.

PBCV-1 encodes five proteins that degrade polysaccharides, and presumably some of these encoded proteins are involved in host cell wall degradation either during virus entry or in aiding lysis of the cell wall during virus release. With the exception of β 1,3-glucanase (A94L) that was expressed early, the remaining 4 CDSs were transcribed late. The late CDSs encode a β and α 1,4-linked glucuronic lyase (A215L), two chitinases (A260R and A181/182R) and one chitosanase (A292L).

Finally, 49 late gene products were not detected in the virion. Fifteen of these proteins have a putative function, for example a DNA packaging ATPase (A392R) that is predicted to be involved in packaging DNA into the virion. Interestingly, some CDSs that were expected to be transcribed early because of their putative involvement in the DNA replication, were transcribed late. These CDSs encode a replication factor C protein

(A417L) (RFC), which is similar to one of the two RFCs proteins in Archaea, an ATP-dependent DNA ligase (A544R), and a deoxynucleoside kinase (A416R). However, none of these proteins were detected in the virion (Dunigan et al., unpublished results).

3.6. Aphidicolin treatment. To confirm the early and late classification of PBCV-1 CDSs, aphidicolin was used to block viral DNA synthesis. The drug was added to the culture 15 min prior to the addition of PBCV-1 and cells were harvested at the same times after infection used in the previous experiments (20, 40, 60, 90, 120, 240, 360 min p.i.). Each sample was analyzed using the corresponding infected non-treated sample as a control.

Expression of 179 (49%) CDSs was inhibited by aphidicolin (Supplement 2). Of these 179, 14 were expressed early, 57 were expressed early/late, and 108 were expressed late. This experiment established that transcription of most late CDSs relies on the synthesis of viral DNA. In contrast, expression of 181 CDSs was not affected by aphidicolin. Of these 181, 113 were early, 43 were early/late, and 25 were late. Three CDSs (A328R, A482R, A464L) were not previously classified as early, early/late or late because they did not pass statistical tests. However, they were expressed in the presence of aphidicolin, which indicates that these CDSs might be early. Collectively, expression of most early CDSs was not affected by aphidicolin and expression of late CDSs was inhibited by the drug. We have no explanation why 14 early CDSs and 57 early/late CDSs were affected by aphidicolin nor do we know why 25 late CDSs were not affected by aphidicolin.

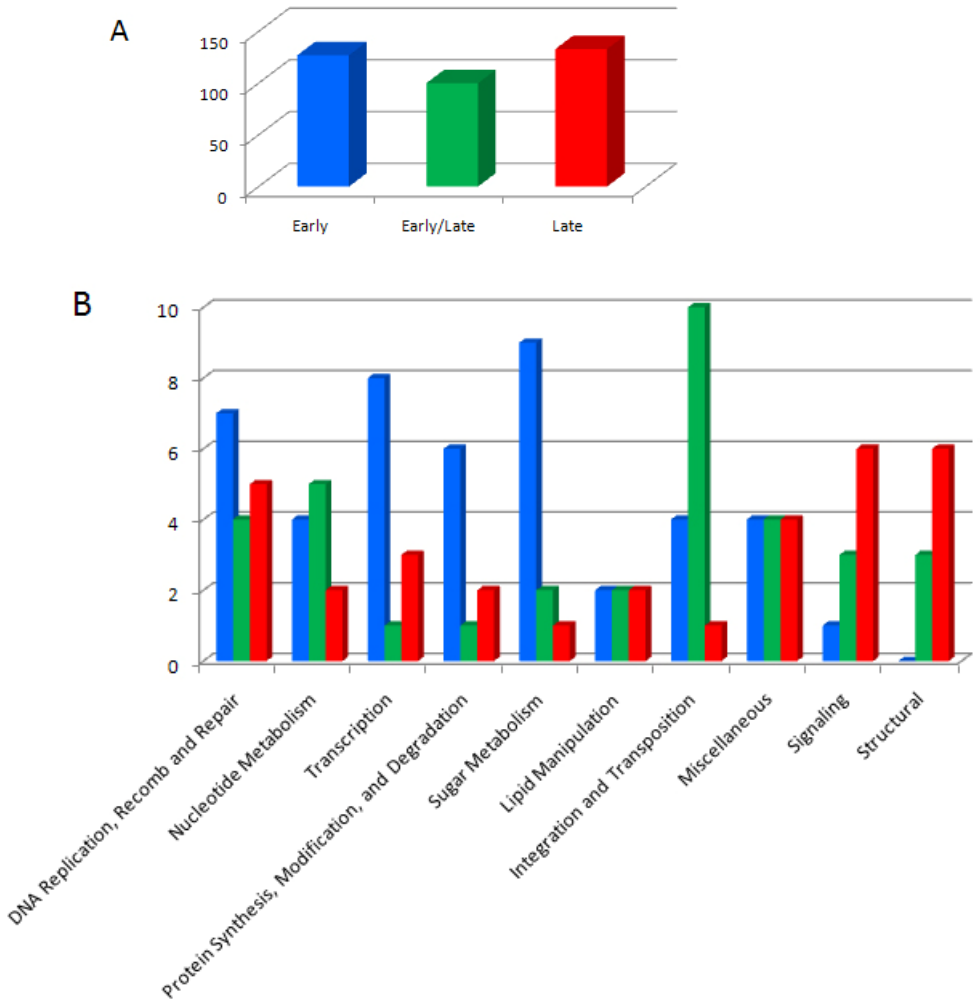


Figure 4. A. PBCV-1 global expression pattern. B. Distribution of CDSs according to their putative function and time of expression*. CDS products with unknown function are not listed in this graph.*

3.7. Verification of the microarray results. Three independent sets of experiments support the microarray results. i) Recombinant proteins from 19 PBCV-1 CDSs have

been biochemically characterized previously and their genes northern blotted to determine when they were expressed. Sixteen of these results completely agree with the microarray experiments. The other three northern analyses were similar, but not identical, with the microarray analyses. One is a glycosyltransferase (A64R), which northern analysis indicated is expressed from 45-360 min p.i. [25], whereas the microarray analysis indicated that the CDS was expressed only between 40-60 min p.i. The second CDS is a chitinase (A181/182R), which northern results indicate transcription occurs between 30-360 min p.i. [35], whereas, the microarray analysis showed the transcript was present from 60-360 min p.i. The third CDS is a dCMP deaminase (A596R), which northern analysis indicated expression occurs from 30-120 min p.i. [36]; the microarray experiments detected the transcript at 40-60 min p.i. and then again from 240-360 min p.i. This discontinuity in expression has also been observed with a few other virus CDSs and it has been observed in both northern blots and in the microarray results (e.g., homospermidine synthase (A207R) [37]). It is important to note that the microarrays measured relative levels of the transcript and this fact could explain the few differences mentioned above.

ii) The 46.2 Mb genome of the PBCV-1 host, *Chlorella* NC64A was recently sequenced to 9-fold coverage (http://genome.jgipsf.org/ChlNC64A_1/ChlNC64A_1.info.html). This prompted us to initiate a transcriptome analysis of *Chlorella* NC64A and PBCV-1 infected cells by shotgun sequencing cDNA derived from poly A+ RNA using a new high throughput sequencing instrument. To date, we have two sets of sequences, one from uninfected *Chlorella* NC64A and one from cells at 20 min p.i. We compared the 20 min p.i. viral microarray results with the infected 20 min p.i. viral transcriptome

sequencing results. Of the 172 CDSs detected by the microarray analysis at 20 min p.i., expression of 159 CDSs (92%) were detected in the transcriptome study using a 200-fold coverage cutoff (Yanai-Balser et al., unpublished results).

iii) Kawasaki et al. (2004) [11] identified 22 immediate-early PBCV-1 CDSs (expressed at 5-10 min p.i.). Our microarray results indicated that 20 of these 22 CDSs were expressed early. CDS A689L was not present in our array, and CDS A312L was classified as early/late and was not detected at 20 min p.i., which is the earliest time point in the present study. In the later case, the difference could be due to the relative measurement of the transcript.

3.8. Do late PBCV-1 mRNAs have a poly A tail? Both the microarray results and the preliminary transcriptome sequencing of PBCV-1 depended on cDNA synthesis using a oligo dT primer. We mention this issue for two reasons. First, about 20 years ago we conducted a set of pulse-labeling experiments where PBCV-1 infected cells were incubated with H^3-adenine for 30 min periods [38]. Poly A+ and poly A- RNAs were separated on an oligo dT-cellulose column and the radioactivity determined. The results indicated that 22-26% of the radioactivity eluted in the poly A+ fraction from cells infected from 0-30 min and from 30-60 min p.i. In contrast, 6.1%, 4.5% and 2.5% of the radioactivity eluted in the poly A+ fractions at 60-90, 90-120 and 120-240 min p.i., respectively. Therefore, we suggested that PBCV-1 early mRNAs probably contain poly A tails and late mRNAs might lack them. However, there are other explanations for these earlier results, e.g., the pool size of unlabeled adenine certainly increases

dramatically during infection, consequently, there could be a large dilution effect on the added H^3-adenine compared to the controls. In addition, we now know that PBCV-1 infection leads to rapid depolarization of the host plasma membrane, which causes an immediate decrease in adenine transport into the infected cell [10]. Both of these issues undoubtedly influenced our previous results.

Second, a report by Kawasaki et al. (2004) [11] also suggests that there may be a shift in poly A+ RNAs during chlorella virus replication. Twenty-two PBCV-1 immediate-early CDSs were analyzed and the transcripts gradually decreased in size after 20 min p.i., suggesting weakening or cessation of poly A polymerase activity.

The current manuscript provides evidence that most, if not all, PBCV-1 mRNAs have a poly A tail because expression of 99% of the PBCV-1 CDSs was detected in our microarray experiments. Furthermore, many of the CDSs we identified were clearly expressed late and packaged in the virion.

3.9. PBCV-1 transcription is complex. As mentioned above, we and others have characterized several individual PBCV-1 gene products. In most of these reports, a northern analysis was conducted when studying a specific CDS. The transcription patterns of ~50% of these CDSs are more complicated than just obtaining a single hybridizing RNA band of the predicted size at specific times. That is, full-length gene DNA probes often hybridize to mRNA transcripts that are 40-60% larger than the CDS itself, suggesting that PBCV-1 might have polycistronic mRNAs. In addition, some probes not only hybridize to mRNAs of the expected size, but they also hybridize to

larger mRNAs at other times in the virus replication cycle. These complex patterns occur even with ssDNA probes (e.g., the potassium ion channel CDS, *kcv*) [39]. One difference between the previous northern analyses and the microarray results is that hybridization to total RNA was used in the northern analyses and in the research described in this paper the cDNAs were synthesized from poly A+ RNAs.

We examined four of the CDSs that produced larger transcripts than expected to determine if expression of their flanking CDSs were identical in the microarray experiments to the target CDS, possibly suggesting a polycistronic mRNA. A common expression pattern was obtained for two of the four CDSs. The RNAse III (A464R) mRNA has a predicted size of 825 nucleotides (nts). However, in a northern blot the RNAse III hybridizing band is ~1,300 nts. Its two adjacent CDSs are 204 nts (A462R) and 354 nts (A465R), and like A464R, both of them are expressed early. The combined sizes of them with the RNAse III CDS are compatible with either a dicistronic or even a tricistronic mRNA. The other example is the potassium ion channel CDS (*kcv,* A250R) that northern analysis indicated has a complex expression pattern, at early times A250R is expressed as a large message and at late stages as a monocistronic mRNA [39]. However, when the early transcript for this CDS was mapped, the start and the stop sites were within the adjacent CDSs, so A250R is clearly not a polycistronic mRNA. The CDSs surrounding the other two CDSs with larger transcripts than expected, fucose synthase (A295L) and dCMP deaminase (A596R), were expressed at different times than the target CDS.

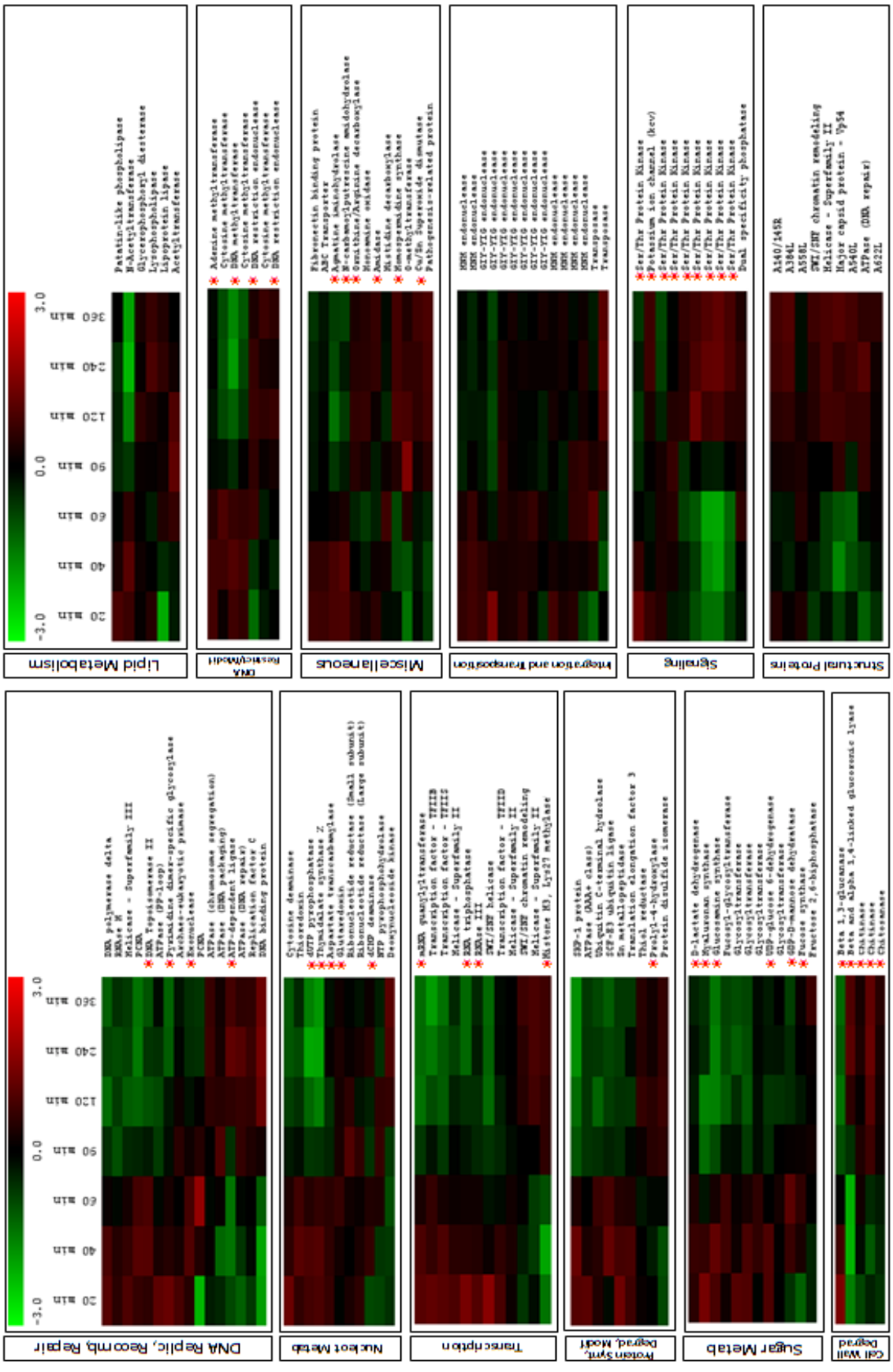

DNA Replic, Recomb, Repair
Nucleot Metab
Transcription
Protein Synt, Degrad, Modif
Sugar Metab
Cell Wall Degrad
Lipid Metabolism
DNA Restr/Modif
Miscellaneous
Integration and Transposition
Signaling
Structural Proteins
-3.0
0.0
3.0
20 min
40 min
60 min
90 min
120 min
240 min
360 min
DNA polymerase delta
RNAse M
Helicase - Superfamily III
PCNA
*DNA Topoisomerase II
ATPase (PP-loop)
*Pyrimidine dimer-specific glycosylase
Archaea-eukaryotic primase
*Exonuclease
PCNA
ATPase (chromosome segregation)
ATPase (DNA packaging)
*ATP-dependent ligase
ATPase (DNA repair)
Replication factor C
DNA binding protein
Cytosine deaminase
Thioredoxin
dUTP pyrophosphatase
*Thymidylate synthase X
*Aspartate transcarbamylase
Glutaredoxin
Ribonucleotide reductase (Small subunit)
Ribonucleotide reductase (Large subunit)
*dCTP deaminase
Pyrophosphohydrolase
Pseudouridylate kinase
*sigma guanylyltransferase
Transcription factor - TFIIB
Transcription factor - TFIIS
Helicase - Superfamily II
RNA triphosphatase
*RNAse III
SWI/SNF helicase
Transcription factor - TFIID
Helicase - Superfamily II
SWI/SNF chromatin remodeling
Helicase - Superfamily II
*Histone H3, Lys27 methylase
SEP-1 protein
ATPase (AAA+ class)
Ubiquitin C-terminal hydrolase
SCF-E3 ubiquitin ligase
Zn metallopeptidase
Translation elongation factor 3
Thiol reductase
*Prolyl-4-hydroxylase
Protein disulfide isomerase
*D-lactate dehydrogenase
Hyaluronan synthase
*Glucosamine synthase
Fucosyl-glycosyltransferase
Glycosyltransferase
Glycosyltransferase
*UDP-glucose 6-dehydrogenase
Glycosyltransferase
GDP-D-mannose dehydratase
*Fucose synthase
Fructose 2,6-bisphosphatase
Beta 1,3-glucanase
*Beta and Alpha 1,4-linked glucuronic lyase
Chitinase
Chitinase
Chitosomase
Patatin-like phospholipase
N-Acetyltransferase
Glycerophosphoryl diesterase
Lysophospholipase
Lipoprotein lipase
Acetyltransferase
*Adenine methyltransferase
Cytosine methyltransferase
DNA methyltransferase
Cytosine methyltransferase
DNA restriction endonuclease
Cytosine methyltransferase
DNA restriction endonuclease
Fibronectin binding protein
ABC Transporter
*Agmatine iminohydrolase
*N-carbamoylputrescine amidohydrolase
*Ornithine/Arginine decarboxylase
Monoamine oxidase
*Amidase
Histidine decarboxylase
Monoorganidine synthase
*O-methyltransferase
Cu/Zn Superoxide dismutase
*Pathogenesis-related protein
MGM endonuclease
MGM endonuclease
GIY-YIG endonuclease
GIY-YIG endonuclease
GIY-YIG endonuclease
GIY-YIG endonuclease
GIY-YIG endonuclease
GIY-YIG endonuclease
MGM endonuclease
MGM endonuclease
MGM endonuclease
MGM endonuclease
Transposase
Transposase
*Ser/Thr Protein Kinase
*Potassium ion channel (kcv)
Ser/Thr Protein Kinase
Ser/Thr Protein Kinase
Ser/Thr Protein Kinase
Ser/Thr Protein Kinase
Ser/Thr Protein Kinase
Ser/Thr Protein Kinase
Dual specificity phosphatase
A340/I45R
A384L
A550L
SWI/SNF chromatin remodeling
Helicase - Superfamily II
Major capsid protein - VP54
A540L
ATPase (DNA repair)
A622L

Figure 5. Heatmap showing distribution of selected CDSs according to their putative functional class. *cDNAs from each time point were labeled with Cy5 and the reference control was labeled with Cy3. Color code represents the log2 (Cy5/Cy3) ratio for each time point. Each column corresponds to the time point when total RNA was collected (numbers represent minutes after infection). Each row represents a different PBCV-1 CDS. Recombinant proteins have been characterized for the CDSs marked with an asterisk.*

It is clear that a detailed transcription analysis using a high throughput sequencing system of PBCV-1 infected cells will be required to precisely determine promoter and terminator sites, as well as splicing regions of PBCV-1 transcription. This analysis should include the use of random primers in addition to oligo dT primers, and also allow the detection of PBCV-1 encoded small RNAs.

3.10. Microarray results with other large dsDNA viruses. PBCV-1 expressed 99% of its CDSs at some point in the virus life cycle during infection in laboratory conditions. Similar results were obtained for other large DNA viruses such as vaccinia and monkeypox viruses in which ~95% of their CDSs are expressed in cell cultures [40, 41]. Transcription analysis using microarrays was also performed with two marine fish iridoviruses: (i) Singapore grouper iridovirus expressed 97% of its CDSs in cell culture [42] and (ii) Red Sea bream iridovirus expressed 96% of its CDSs during *in vitro*

infection [32]. Therefore, it appears that most CDSs are expressed in these large dsDNA viruses, even in laboratory conditions.

The temporal categorization of the transcripts in large viruses varies. Different studies classify virus transcripts in different ways, usually subdividing early CDSs into more specific categories. For example, Red Sea bream iridoviruses has ~9% of its CDSs classified as immediate-early, ~43% classified as early, and ~41% classified as late [32]. T4 bacteriophage transcripts are divided into immediate-early (42%), delayed early (~12%), middle early (21%), and late (~22%) [15]. Poxvirus transcription occurs in the cytoplasm of the infected cell and is programmed by a virus-encoded RNA polymerase and time-specific transcription factors to generate three transcription categories: early, intermediate, and late [43]. Using tiling array technology, a recent study reveals immediate-early transcripts in vaccinia [41]. An additional microarray-based classification of vaccinia and monkeypox virus gene expression divided the virus CDSs into early and late categories only (~50% of the CDSs in each class). The method used in this later study categorized CDSs according to the time the transcript is first detected, so the intermediate class (expressed early and late) was not distinguished, implying that this class was a subgroup of the early CDSs [42]. For PBCV-1, we classified a portion of the early CDSs as early/late when the transcripts were detected before and after the onset of viral DNA synthesis.

4. CONCLUSION

For the first time, a global mRNA transcription profile was conducted for a chlorella virus. The PBCV-1 life cycle is temporally programmed. This regulation is controlled by a precise gene expression pattern, where the time of transcription is dictated by initiation of viral DNA replication, which begins 60-90 min p.i. Early CDSs were transcribed between 20-60 min p.i, and products of the majority of the early CDSs are responsible for providing the machinery for viral DNA replication. Late CDSs were transcribed after 60 min p.i., and most of them were dependent on initiation of viral DNA synthesis, since aphidicolin blocked the expression of most late CDSs. Products of many late CDSs serve either a structural role or were packaged in the virion, presumably to aid in virus infection. Forty-six percent of the early transcripts were still detected after DNA synthesis begins and were called early/late CDSs, indicating a complex mechanism for mRNA maturation and degradation. Some of these early/late CDSs may only be synthesized early, but their transcripts were not degraded until after DNA synthesis begins. However, this scenario probably does not apply to all early/late CDSs because several early-late CDS products were packaged in the virion. These CDSs may also be transcribed as late CDSs.

This study reveals that PBCV-1 gene expression time is independent of GC content, as well as the transcription direction and CDS location in the genome. It is anticipated that the temporal transcription map will provide clues to the functions of the large category of viral proteins whose functions are not yet known and to the events that must be orchestrated for a successful viral infection.

5. SUPPLEMENT

5.1. Supplement 1. PBCV-1 50-mer probes' sequences

CDS	Sequence
A002L	TAGATGATTTGAGAAAGTTTGCCGAAGAGTATGCGAATACTGAAATTATA
A003R	CTGGTGAAAGATACACTCATCTACAACGGAACCGTATTTATACCTACGGC
A005R	GCGTAAAGATGCTCATAGAGGCAGGTGCCGGTCTTAACATCGACGGTGAT
A007L	CATCAACAATGAAGGAGCGACACCATTACACTTTGCGTCTCGATATGCAC
A008L	CAATCACTAAAGGGTGTACTGATTATGTTATGTTATTAATCAACGCAGGG
A009R	CGTGATTAAGATGTGTGCTTCGGGAGATCATAGAATGTATAAACTCCCAC
A010R	TCACGTTGTTTTCGACTAACTGGAATGTCCTACAAGTGCGGAGTGGTCTT
A011L	TACAGCAGCAACTATCTCAGATGTGTATAACCAATCGACTACATTAGCAA
A014R	TTTCCTAAATACACTATACATTGGTGCCCAATATAACACGAATTTTAGTA
A018L	CATCAACTTCATTTTTCGGCCTGATGGTGACGGCGGTACAAACAGTGTAA
A025L	CGTTGGAAACTTATTCGCAAACGGAAGCACCATACAACTTTCAAATCTTG
A029L	TAATAACGAACTTGATACCTACCGTTGATATAACGGCTAATGTTCTGCGT
A034R	TCCATGCTCTCAGAGTGTTTTACGATAAGATGAGTGCTGAAGATAAAAAA
A035L	CGCAGTAATAACTTCAAAAGTATCTGGAAAGGTTACCAAGGTTGTATCCC
A037L	AAACTCTTAACTAGTGTGATTGACCAACGAGGTATTTTTTATTTCTTATA
A039L	TGAAAACGATCTTACACCCGAAGAAGAAGCTGCTGCTCTTGCAGAACATT
A041R	AACATTCTAAAAGATCCGAAGGATAAATCTTGGTACAATGGGTTGAAAAA
A044L	TCGGATTACTACACACCAGCAGAAGTGTCTCAGATCATCCTCAAAAATAT
A048R	GCGTCTTGAATAGGATAAATATTCTCATGGATACATATGATATTACGCAT
A049L	ATACAAAAATCGAATATCCCATACGAAAATACCCTGGTGACTTCATTTAA
A050L	AAAAATCGCTATGAAGCCAAGTTTTTACAGGTTCACGAAAGCTAAAACCA
A051L	ATGTGATAAATATTCACGATGATAAAGTTACTTTTTACCACGATCACCAA
A053R	CTTAAAAATATTGCAGATACAACCATCGAAAACCTAACTTCAGCATTCAA
A057R	GACCATCCATACTTAGATTATACTCAGGATAATGCGAATGTGAAACCTTC
A060L	GTCTTCCGATGGATAGGAGTTCAATCATTGATACAGTACCATTTCCATAG
A061L	ATTTTTTCAACATCGGAAACGGAAAGACTGCCGCAATCATAAAAAAATAA
A063L	TGGTTCGTTTCTAAACTCAATGTGGATGTGCCGAAAATCTTCCACTCGTA
A064R	GTAGAGGATATACCTCATTTAGAATGGTTAGAAATTCTGGCAAATGAAAT
A067R	GCTGACATCCGACGGACAGCAGTTCAATGTGACTCTGCCACATGCTCTCA
A071R	CGATTGTGTACTTTGGGATAGCCGAGGGAAAAACTTGGAATGTAATTTGA
A075L	AAGAATTTCCCGGAAGAACGGCAAAGGATCGTTGATTTATACAGTGATAA
A077L	TTATGGAAACATTCCGTTATCAAGGTATTGTAACAACGGGTTTTATGTTC
A078R	TGGTGGTGTTGATCCTGAACCTGTTGATCTTAAAGGTTATACGAAATATA

A079R	AAAGATGATATCCACACTCCTGCTCCCATTGGTAATATATCAGTCGTTGG
A081L	GTTGAAGCGGCGAGCGTAATAATAGAAGACAACGACGAAACTATCATTTT
A084L	ATGACACGAGCACTCGATTTCATAGATGCAGCAGAAGACAGGTTCCCGTT
A085R	AGAAACTTTGCACGCAGGACTTCCCGTGAAAAGTGGAGAAAAAATCATCG
A087R	TAAACACTCCGAACAAGGTGTATGGTCGCACATGGATTCGAGCACATTTC
A088R	ATTACAAGTTGGAGGTTGCAAATGCTTCTGGTGACTATGATAATGAATAT
A089R	GAGGTGAGAAGATTGGGTTTACTGCGAAGAGTAGCTTGAAGTCAATGGGT
A090R	TCCTTGTATACCTAAAAGGAAAACACCCAACTTTCATATTAGAACATATT
A092/93L	AAACGTGCCTCGAATGAAGCCCTCAAACTCGCGTTACAGATGTATGTTAA
A094L	CTGATGCACCTTACAACCGACCTTTTTATATTATTCTGAATACATCTATC
A098R	TTCTGTATGATTCCGGCCAGGATTACTGCAATGATGACGCTTTGGGACAT
A100R	AATTTTCTCCGATAATTCACACTATCCCGATGCAACTGCTTTCGTATTAC
A103R	TCCAACTATCATCAAGGACCATCTCAAAAAGGCCAATGCTATTTATCATA
A105L	AATAAATGGTTCAGATATGACGATGAAGTTGTTACACAAATTTATGATAT
A107L	TCTCAAACGAAGACATCAAGAAGCTACGCCGGGATAGCAACAAAATGATT
A109L	CAAATGAAACATAAATTCCTACAGAGAGATGAAATTCGTCAAATTGAACT
A111/114R	CCGTGGATAATCAACGCGGATTATATCTCATATCTAGACGATGATAATAT
A118R	CCAGATGACTACGTTATTGCGACTGGACAGACGACGAGTGTCCGTGAATT
A121R	GGTCAAAACCATCATTTACAGACGCTCTTAAAGTAACTGAAAAAAAAATA
A122/123R	TGCACTAATTTTAAGAGGTATTGGAGGTACAAACACTGTTAATCTGTTCA
A125L	TATAAAAAACTTCGTAAATTCAAAACCGTGGGAAGTATGTCCCGAAAAAT
A127R	AGAACCCATTCTCATAAAGAACGCTGGACTATTCGCACATAAACGAGGCT
A129R	ACGATACTCTCAGGCTGTTTCGTCGATACGAGAACTCCTTGTGAATCCAA
A130R	CACCTCCGTTTTTATTTTCGATACTCGTGGCAATTGGATGCGCTGCAATA
A131L	TTTACGAACAACTGGAAGCTCTGAAGACTATTTCAACGCTCCAGATTAAC
A133R	TATGCTCGTTAAACAACACATACACAGGTATAATAAGAACTATACATATT
A134L	CTATTTACAAAGATTTCCCGATCATCAAGAAGACTTCGACTTCCTCTTAA
A135L	CGTCACTCTTTTCTTTACTACTTTTTTCACAGGAGCCATTGCACACACCG
A137R	TTCGCGCCAATAGATATAAGAGAGCATTCATCTCTCCCGTAATTCGACCA
A138R	TTCTTATAACATTAATAGTCCAGTAGAAATGGTATGTAAACCTCGTAAAC
A139L	CAGTTGGTAGAACTCACACCTAAACTCAAGGAACAAATAGAATGTAAATT
A140/145R	GTGAAACCCACGGCACCCGTTGGAACCATCATTATCGCAGCGAAGGTGAC
A148R	TGGTGCTAAAGAATCTACCCGGACGTTTAACCAAGAAGGTCATTGTTTGA
A150L	CGACACACGGTGTCAACTGATGTAATCTCTAAATACAAAAAGAAAGTTAA
A151R	ACCATAACGCCAAGTATTGTTGGAGTTATCTCTGGACCTTTTTTGCTCAA
A153R	AAGAAAAAATTGAAGAGAAAATGGAATCAATGTTCATAGATGAGGACTAA
A154L	TGGATATGGCAACATTCCGTGCCCCGTAAGAACTCAACTCAAATACGAGA
A157L	AGGGACAAGGAGTCCGAAGTTGTGAAGGGCGTTGAGAATTCCCAAGTGTA
A158L	CTACAAGCGGTTTTGACAACGGAGAACTTCCTGATGTGGTAGATGATGTT
A161R	TCCCGATATTATTAAGAAGAAGAAGAGGGCAAAGCGCAATAGCAAACTGA

A162L	ACCAAAAACACATTCACCAGATGATCTAGAGAAATATGAAGCCGAGGGAT
A163R	TGAACTTGTAGAATTTATGCCTACCAATAATAGTAGTATACGAGACAATT
A165L	GAAGTGGAATCTTTGAAAATGGTGATAGATACCAACGGAAAGAATATTAA
A166R	CGACGTGCCAAGAGAAGAATACGTCAGAGAACTTGAAGACACGGACGATT
A168R	TGGTCAATGACACCCGCGCAGAAAATGATTACACCTACAATTCCAGATGT
A169R	ATGTGGATACAAACCATCGTGCCGTGTATTTTGACCAAGTGGAACGTGGG
A171R	CTTGGTATTTTACACTCGGAAAAACAGTACCTATTCCAGAAACAACCATA
A173L	AATTTTTCCAGAGTCTCATCGTCACAGCGACGAAAGACAAATTTTCAAAA
A175R	GATGCGCGACGCTTAGTGTTTTTAGTCGGAGCCATATATTTTTACCAAAT
A177R	AAGGTTGTTGATTCCGCAGTTGGGTATGTGAAAGACGCAATAATAGATGC
A180R	GCGAGTATGTCGAAAGGTAAAGGAAATGTCAACGTGATTTACACAAATAT
A181/182R	AGCAGCGATCAGTGCCGCTACTGCAACGAAGAAGCAGTGTGATGACTTGG
A185R	GGATACAAGAACGAGTGCCAACCGCATCTTCACGTAGCGAACAAAATTTA
A189/192R	AAATCACCTTCACCGTCTCGTTCGGCTATAAGAACACCAACTGGGAAAAC
A193L	TGTCGTCTGCTGTAGAAATTTACTTGAAGGAAAGTTACCCACTTATTTTG
A196L	TTGTTTTTGATTTCTTGGATCGTGTTTTTCTCCATATTGTATGCCCTTAA
A199R	TCTAACAGCGGAAACCGTGATGATCACAGCTTCTCTAGCAGGTCCTGGGT
A200R	AATAATTTCAGAAGGGCATAATCACACGAGGAAGCAAATAGGAACATTTT
A201L	GACGGAGGTGTGTTTTATAAGTTTGCCAAAGGAGGAGTTTGGGTAACAGC
A202L	ATTGTTTTAAGTGGCTTCACGAAAAATATGGAAATGTAGAACCTGATGGG
A203R	TAGAAAATATTGATGAAACACCGCTCGCGGAATTCATTCCAGATTTTGTA
A205R	TCGCAAAATTTATAGAGGAAAACGCGGTGGTGAGTTTGTAATTTCTGATA
A207R	CCAGCTTCTCAGAGATGTCCCGGATGATGAGGAATATGTACCATCTGTAT
A208R	GATGGTGTCAGGGTAGACCTAACAAAAGTGGAATTTGAAGTTGCTGATGA
A213L	ATTCCTACGGATATTTATCAGAACTCTGGTCGTGTTTCTCTGAAAGAATT
A214L	TTGCGGATAGTCTATCTTCTGGCGAGACCAAGGTGTCTCTGTCTATCAAG
A215L	AGAATCAGGTAGCATACTTCACGAATTTCCAAATGAAGAAATACGAATGA
A217L	GATAATAATAGGAGAAATGCTGTCAAAACGTATAGGTTATGTTGAGGGTG
A219/222/226R	CAGACATTGTCTTCCATTTGTGGTCTATTACCGTCAATCATAAGATTTCT
A227L	TTTGGTTCCAAACTTCTATCAAGATTTACCACTGGTACACAGAAAGTTAC
A229L	CTTCGTATGTTCTCAGGGATAATATTGCTTTTTGTTCTCTCGTAATTTAA
A230R	AAATTGTTCGCATATGGTATAGACATCGAAAACAAAATCATCATTCTATT
A231L	AAAGAAAGAAACTCGAAGCCTTAAAAAAGAAAATACAAACTCGGAAATAA
A233R	CCAACTTCCAAAAAATTAGTGACTACGAAATGTGGGAGCATTATACTTAT
A234L	AAGGGATTCAAACAATATTTCCTTCGCCAGGACAACGGAGACGTTCAATT
A237R	GGAATGGAGGTCTGCGAAGATAGTTTGGGAGCCGTCTCTAAGCAATGTCC
A239L	GGACTGCCAAGTTGAAGGCCGTAGGAGCTGCCATCAATGATTTTGCATAT
A241R	CAAAAACAACATTTTAGAGGAATTTTGAAAACTTCATCGCTTTACATTTA
A243R	AAAGGGTTTGCAGCTTTGAGAAATCTAGAGATCCATGTTCATCAGATATA
A245R	AGGAAAACGAATCGGATGCGCTGTTATAGGATATGCCAAAGAAAACTTTT

A246R	AGGTGGAATAAGTAAAAAGGATGTGAGAATGACATATCATTATTATGCAC
A247R	ACAACGATGGAAATTTGCCGGATGACCTAACGTCGGACAAACAAATTATT
A248R	GAGACGGATGAAAATTGGAGAACTTTTTGAGCACCCATTCATGATTCTTA
A250R	AACTTCTCTGTTGCAAACCCGGACAAAAAGGCATCATGGATAGATTGTAT
A251R	CCGCAGTGGGAAATGATGTTGAAGAATATCTTTTCTTTGTATATATTGAA
A252R	AAAAGATTAGGAGAGTTCCAGATTCACACATCCAGAAGTTGTATTAAATT
A253R	AAAAATATGTTGGCGAGAATTGCAAAGGAACGCAAAACTTCTAAAGAAAA
A254R	GTACATACTATTGATGATGCCGTATATATTGTATCAGGAGGATTTGGTGA
A256L	GAATCTCTAAAAGAAGTAAAACCTATTGCTCCAAGTGAATATGAAGAATT
A257L	GGCAGTACAACTTCAAAAACAACTAAAATAAATTCATCTGTATTTACATT
A259L	TTCGAAACAAAATATTCGGCTCGGGAAATGAGAGACGCAGGAGGGTCGTG
A260R	ACTCGTTCCCTACTCGGTAACCGACGCCGCCGAAACGATGCTCAAACTCG
A261R	CCTCCAATATACAAACCATTGTTAGAACGAGACATAAAAAATGTAGATAT
A262L	GATGGAGAAAACGGTGGAAAACAGTGCAAGTACTATCCATGATTTTTATA
A263L	AGTTGTGAACGTGGTAAAACCACAGGGATACAAACCAGAATTTGTAAACC
A265L	TCATTGGCTCAGATGAAATCCGGGAATTACACGCAAGGTGGTTTGAAATT
A267L	CTCGAACTTACAACTGATGACTGGGAACAACTTATCAAAGGAGAATGTGC
A271L	TTGACGAGGAATTCAACACTCATTGGTTTAACGACCCGCTGACACGCCAT
A273L	CGTTAAAAATAATAGCGTCTCAACAACAAGAGATAAAATTGTTAGAGAAA
A274R	TATCAACATACACACTGTATGATGGAGAGTTAAATTATGATAGTTTTCAC
A275R	TCGTTTCAAGAGCTGTAGATGCCGTCAAAAAAGAGATACGACATGCGAAT
A277L	TCCTCAACCACGCGGTGTTTCAACCATTCGCATGTCGTATCTCTTTTTTG
A278L	ATAGTAAGAAACTTACTCCACCAACTGAAGTAGATGCTAAAAAACGCAAA
A282L	CAATTACACGAATAGCGTAGCCCAAGCCAGAGATTATGGATGGTACAACC
A284L	GTTGTTCCATGTTTTAAATAATTTTGATATTCCAAAAGGAGTAGTGGCGT
A286R	TAAACTCTGTGAAAACGGTTCCTCATGGTAACGATTATATGTATCTTTAT
A287R	AGAAACGAAGAAGAAAATTAGTGAAAATAATCCTAATTCAAAGATAGTGT
A289L	CTCTCACCGACCGAGTTGATGAAAAATACATCCCAACACAAACCGACATT
A292L	GCTGTACACTTTGGGCAAACATCTTCAAAGAAGGAAACGTTGGGCTGAAA
A295L	ATCTACGATACTTATATGCCCGATGGAACTCTGAAAAAACTCATCGACTC
A296R	ATTAAGGACCAGAGAAGCTAAAGTAGAAAGGCGTCGGTTGCAAGCAGAAG
A297L	TGGCTATTATTTTCATCCGTTACGAGAGAAATACTGCAAAGTTTATTTAA
A298L	TGAGAGATGACCTGGTAAAGAATCTACATATCAAAGGATTTCACAGGAGG
A301L	CGCTCTAAATGGTAATGTTGAATATATCAACCAACTACTTGCGGATCTAT
A304R	ACATACACGAAATAACCCGACATATACAATATTCCGAAAATTATCTATTA
A305L	AAAAAATACAAAAAACACAAAAAACACGAAGAAAAGCAAGAATCTTAAAT
A306L	GTTACACGAAGATGATCCAGAAGTATGGTTTGACTACATGAGCGAGGAGC
A308L	TCGTAGTAGAAGTCGAATGCCTTGATGACGAGTGTGTAGTCGTAAATTAT
A310L	GTGGCAGTACAATTACGCAGAATGGTCGTGTTCTCGGTGGAACGAGATAA
A312L	TCTCGAAGTGAGCGACGAAGACGATGCCCGCGTAGCTCAAAACTTCCTTT

A313L	AAATTTACGAGAGTATCACCACTGGAAAGTGCATGATGTGTGGGACAATA
A314R	ACTGGTATTATATTGATTGCTACAAACGAATTAACATATAATCGCCCTCG
A315L	TTCATGGTGGTATAGCGAGTTGTGCTCGTGGTGATAGAAATTCTGCTTAT
A316R	GAAAGAAGATTTATTATTCCCGGAATTACCAACTACAAATATAGCACCTA
A318R	ATTGTAACGCTGACATAGAACCGATGGCCGGTGGTTATGGAGCTAAATAT
A320R	TCGAAACCCTCAAGAGAAAATACACAGCGGAGATGAATAAAATTAAGTAA
A321R	GGCGTCCTTGGGTATTTACTCGAAGCCGCTGGAAAACCAGAACTTGCGAA
A322L	AACAATTTTACACACCTGTTATTCCACAAAACCCTATACAACTTCAGTGA
A324L	GAGTTTTCATCACCGGAGGATACAACTGTAAAGGTTGTTAAAGGAAGTGC
A326L	TATGAAATATAAAAACATAGGGACGAACCTCATAGAAAAATCGGAGCTTG
A328L	ATGATTACTACATGTTATACCCGGTTAGGATGTATAAACGATACAAAATA
A329R	GATGGTTCTATCTGTAGCCATTAACGCGATAATAGATATCATAGACGAAG
A330R	TCTTTAATAATGTCAGGAGGAAAACAACTCTTTTGTGTCTAAATAAAATA
A333L	AGATATGAGAGCGTGTGGAGAACCTGGTGCTGTACCCGAAGAATTTTGGA
A337L	TAAGATGGTGGCAGAAAGCATCAAGGAAGAATTCAACTTTGCTCATGTCC
A339L	TCGTCATCAGATAGGTTGAGTATTATGAAGTCGTGTCTTGGGTGGGTAAG
A341L	CACGAGGTTCGCAGGTCTATTTTCGACCTAATCGGCTCGTTGTTCAAGAA
A342L	GGAAACGGAAGCAACTACAAGCGCAAAAAAACAAGCAACAGAGGAAGTGA
A348R	CTACCGGCAAGAGGATGTTCAGAAATCCTAAAATTCACAATGATTTTAAT
A349L	GATTACCGCGCATTCCTCCAAAAGAACGCGAAGGAGGTCGAAAAAACTGT
A351L	ACGAGACGCAACAAGAAGATAGAGGAATATGCGTTGGTAGGAATTGATAT
A352L	CACACAGGGCAGCTCCCTGAAGAACGCGAACTACGATCTCCGCGCTGACC
A354R	AAATGTGTATCTTATATTGATGGCGTAGAAGAAAAAGAACACAATAGTCT
A357L	TGTGATAAATGCGTCTAGGGAAAAACGCACGAGTGGAACGAATTTCTCGG
A360R	TCGTAGATAACGGTAAAACATCCTTAAAAAGAATGAGCGAAGTTCGGTAA
A361R	CTTTGTCAAAGGGATGGAAGAACGACTCAAACCAATCATACAGCTCAAAA
A363R	TGACAAGTTCGAGAAAAATAAGATCATCAGGAAGCCCGAGGGATGTCGTG
A366L	AAAATTGTAGGAAAATTCGAAGAACACACTGAGACAGAATTAAACAATAT
A368L	TGATGACGATGGTTGGGACTCCGAAGACTATGATCCCGAAGAATGGTATT
A373R	ATTCGCAAAAGATTCAACGTAAAGAAACCACGAAAATTTCTAAATCATTT
A375R	GTAAAAAATACTATGGAACTACGAAGGATAGGTTTTATTCGTTTAGTCAA
A378L	ATCTTGTCCTCGTTTCCAAAAATGGAATGGAAGTAAATGCGAATGTGATT
A379L	AATAATACTCCTCGCAATATGAATAATACACGGACCGCAAAGAAATCAAC
A382R	TGTTTATTTGACGTACAACCCGCAGATTTCATTGTTCAAGAGAAGTATAA
A383R	TGGGAACTAATAATACCATTGAAATTTTTTTGCTGCAAACGGAACACAGG
A384L	GCGCCGAAACTGCCGGAAAATGTAGCGGTAAAGAAGGAGGTGAACCATAT
A390L	ACGAATGTGAGTACACAACTCGACAAAGATGGAAAGGAACTTGGGTATGT
A392R	TTCCAAAGAAAGGAAAAGGGATGATGACTCTGATGATGAAGCCGATAACG
A394R	ATAGATGACACGGAGCAATTTCCCGACATGGATTGTGTATCTACGAAAGA
A395R	GATTTTCAGGACATATAGACCTTGGTGATTTCGGACGCACCGCGATTTGT

A396L	CAAATTCTGATAACACATATTTTTACTACCTTGTACTTCTTTCCGTAATC
A397R	ATAATGAGGAATATCCATTATTCATAGTAGTACACGAACTACATGAAGAA
A398L	ACTCTCAGGTCTCGGACTTAAAGCAGCGACACTTTTAGCACCGCGCATCG
A399R	AGTTAAGGCGCACAGTGGAAATATTGGAAACACATTTGCTGATAAACTTG
A400R	AAGTGGTCGTAAAACCAACATTGTCTTGCCGTGTTCAAAAAACGAAATAA
A401R	GTAAAACAAAAAATCGCTGCTGGAAGTCAAGAAGATATTCGTCAAATGGG
A402R	TATGGAAAACCTATTTACGCAGATGCCAACCTACAATGTTCTAAAAGTGC
A403R	CGTTCGTTTTGGTCCTGTCCTTGATGAAGCAACGTTCAAAGAATATCAAC
A404R	GTAGAAGAAGTCGTAGAAGATTATGAGGATGACGAGGTCGTAGAAGAAGT
A405R	GGAGTTGTATCTGTTGTTTTCATTATAGCACTAAAAAACGCCAAAAATAT
A407L	CTGTAACAACTGCCGCTCCAGATCTCACGGCGTTTGGTCAGTTTTTGTAT
A408L	AGAAGACCAAGGATTCAAAATCCGGGAAGAAGCACCGCAAACATAAATAA
A410L	GAATCTATGATTCCCTATGCGGAATATGTAGCCAACCGTGACGAGACTAT
A411R	GTTGTCATGGATTTGATGAAAAATATAATACCTCCAAATACATCTCAGTT
A412R	ACGAAGTAACACCAAATAATGAAAATAAATTTCTATTCCTTGACGACGAA
A413L	CACCTTTAAAAGAACGTCGTGTAAAACCTATTGTTAAAATTGAGGAACCG
A414R	GCCTATTCGTACATTCGATTGTTTTCGGTCTTCTGACGTACATATACCTG
A416R	CAATAAAATTTATAGACGCTAGTAAAAACGAAGATGAAGTCGAGGAAGAT
A417L	AGAGCCGAGAACGGGCTGCCATTTCTAGATGTTTTAGAGTTGTCGTATGT
A420L	TCAATGTCTTGCAACCATATTCTTGATAATAATGGAGATTAAAAGAAATT
A421R	AACTTGGTTCAGCCGCGATCTTAATTATGGAAAAGCGAATAGTAAGATAT
A422R	ACATATAGGACCGAATAATATAGGTTGCACCGCTTACGCAAACTACACTC
A423R	ATGTTCAGGAAATAGGGAAAGAAGTTATTGAAAAATACAAAACATTAGAT
A424R	AATAACGGGAATAATAAAGGGAATAATAGCGTCAAGAAAAACAACGTCAA
A426R	CTTTACAAAATCATCTCGAAAAGTGTTCTACTCCAGAACTCGTTCATCTT
A427L	GTATGAATGGTCTAAAGGAATATGACGAATTTATTGAACTCGTAAAACCT
A428L	TAGCAGAACCAGAGTACCACCCAGAATCTTCTCTAAATTATTTTAACAGA
A429L	CTAATGTCATATTCAGCTCAATCTAAGGTTTACACTATTTTGATGTCTTT
A430L	GCCAAAAATTACAACGTCCTTCGCATAATGAGTGGCATGGGAGGCTTAGC
A431L	TGAACACTACGATGGAAGGAAAAATGTCAAGGTGGAAGTTATCTAGGAAA
A432R	CAAAGTATTGGTTCTCACATCACACCCGTAAACGAGTATATGGTATCTAA
A435R	GGGCGAGCTTAGATCCGTTACGGCCGCGTACGAATTCGCCTCCACGAGGG
A437L	TGATCGCCGGTAAACCCGTGAAGATCGAGAAAGAATCCAAGGCGAAGTTC
A438L	GTTTTAACATTTCCTCGTGTGTATAATGGTTCGCGTTTAATTGGTGGTTG
A439R	TTCTCCTTATGCCTTTCATAATGATTCACTGGATTCTGTTGGATGATACG
A440L	GGGAGAGAGAAGACGCCTCAAAATACACTTAAAACTCTCGACAATGCCCA
A441L	AGTTCCGCGAAAAGATTGCCGACTACGGTTTCGAGCCACTTATGGATAAA
A443R	TGATCAGAAATCTAGGAACTCTGTTGTCATGGTCTGTGAGTAAAAAGTAA
A444L	TAAAGCTCGTGGTCAGCAAGCATCCGATAGTCTAAAGGGTTTCTTCGAGA
A445L	ATCGCAGATGCAAAAAAGATTGCGCGTTATGGGTTTGCAATGTTTCTCCT

A448L	CCGTCCTGAAAAACGAGATGCCTGTTTACGTTGTTGATGCCGAAAAGAAC
A449R	GCGACCGACAGAGACTGTGAATCCACGTATGAGATCGAAGTTGAATTGTA
A450R	AGGATTTGTTGACCGTGCTGTTGACGCTGTGAAGACTGAGATTCGAAATG
A452L	AGATGATACTGAAATGGCGATAGAGACCATTACAAAAGCTGGAGAAAAGT
A454L	ACAGCAAACACCACACTCAACTACCCATGCGTTCTCTAAACGAACCACCC
A456L	AGTTCATCCAAGGGTTGTCTCTAAAATCTGCAAAGGCAAAGGATGCCAAT
A462R	AATATTTGTATTCGTTTCATAATGGACACTCCTCAAACACGCCTCAGCCC
A464R	GAAAATTTCAACAGGAACTGGAAAATCCCGCAAAGATGCAGAGCAAAATG
A465R	CGTCAGTATGGTTGATGATCCATTTATCCGCATTACGTTATCCTAAAAAC
A467L	AACTCGAAGGATTTGTTGACTTCATAACGAGGATTGTGTATGACGAGAAA
A468R	CCTAAGAAACTCATAGAAGAATTATTTCCAGAAGAGGTTGATGATAAATC
A470R	ATACCGAGGAGGCCGGGACTTCCGCTTCATGCAGTTGAAAAAATCTTTGA
A471R	GAACAGGGAGATAAAAGAAGAGAATTGTCAAAAGCTGCGATCAAACGTTG
A473L	TTTTTTGACAACCGTCAACACATTCATCTCTCTACCTTTTATTTGTAACA
A476R	AGCTCATTTCGCTTCAAGGGAAGACCAACTTCTTCGAAAAGCGTGTCAGT
A478L	ATATTAGAGACGGTGCTCTACGGCGAAATATTCAATTCGACCTCACGAAA
A480L	AATGCTGTTCAGTTATTTTGCCGCATACCAACTAGGGAAAATGCAAGAAA
A481L	ATATTGCACAAGAAGTCGAACGCAGAATTAACAATACCGAACACGAATAA
A482R	CTGAAACTCAGACGACCCAAACCAATGACCAAGGGAAAAGCAACTCTTGA
A484L	TATTTACCCTGGGACGATGAGTTCCGCATTCGAAGCTAGAAATTATAGAG
A485R	AGGAGACGCTGTAAGAAGAATGGAAATCATCAAATTGGCATTACAAAATT
A486L	TTGGTAAAGTCATTTCCCACTGAGTTTGACGATAGTGAGAAAATTATGTG
A488R	CCGTTAGAGAACTTCTGCCCATAGAGCAAATTCTCAGGGCATATCTTGGA
A490L	CAGAAGATTTTGTGAGAATTATAACCAATGAAGAATGCTTTTACTGTGGG
A491R	CAACAAGTACGCAGAACCTGGTGAAAAGAATGTTGGTAGAAACTTTTTGA
A492L	TTTCCGTTCGACGACCACGAGGACAATGCAGAAAGATTTTGTGATTTATA
A493L	TGTTCCCGGTTGACAACCATAAGGAAAACGCAAAGAAGTTTTCAGCATTG
A494R	AAAAATTTACACAACCAAGACGTAATATGGAAGAAAATATGTAATGACTT
A495R	AAAAAATGAGCGACTCAAGAAAAGGTGGAAATAATCCGAGAGCAAGACCG
A497R	AGCATCTTGAAGATGGTGTTTGGAAACTGGAAAATTCTGTTGTGCGTTGC
A501L	TGACAGCCTCAAGGAATATCTTGTATCTGTTGCAAAGTTTGTCGCCGTAG
A502L	GATGATTTGGCATCATTGTTTTCAAAGAGACTCTTACCAGATATAATCCC
A503L	GTCGTAGGTCTCAAATTTCATCTAAAGGTAAAGAAAAACATTCGCTTTCC
A505L	GGTAATGGTCTATATCGGAAACTCAATGATTTTAAATTTCAAAATACTCT
A512R	GTTGACAATGTGATTGATTTTGATGTGTCTATGATGCCTGGTCGTACTAA
A517L	CATATAAAACATCCTACACGTTCTCCTGAAAATATCACTTTAACACCGAG
A519L	CTCAAATGGTTGGTATTCGCGTTCGTTTTAGGATTTGTGTTACATTATAT
A520L	TATTACGACAGAAATGGTCTCAAGAAGTTTATCGCGGAACAAGGTTCTAA
A521L	GTCCCGATCTATCATCAGTGACGATGTTCAACCCATTTGCCAAGAAATCT
A523R	CAGATGATTGATGTTCCTCTCGCAGCGGCAAAGACATCTTTCATTGCTGG

A526R	CACCTTCTCGGTTGTTCGATAGGTCCGATGGTGTAAAATATGTTAGATAA
A527R	ATATCAACGTCAATCGCCTCGTGCTTAATTATATTCGTTCATTCCAATAA
A530R	GTGAAAATGAGGCATCCTTCGTGGTTTTGTGTGATCCCAAAAAGGTGCTC
A531L	AAGAGAAGTATTCAGGTCCTCAAACCGTATTCACGGAAACTCCTAAGCTC
A532L	GGAAAAAAGGCCGAAAAGTTCAACGACGCCTATTACGGTGTTCCATCTAC
A533R	CGTATTACCGCGTCAGGACCTTTTACTCTTCTTGGTAGCACAGGTACTAT
A535L	TCGGATGACGAAGAGGACATCATTCCAAAGAACATAAACATAGAAAAAAT
A536L	TGAAACTATATCGCCCAGATATCACCAGAGTACCAGACGATGAGTGGGAC
A537L	TCTATACTGAACGACGACCCGGAAGAAATCTTAGACCCAATTGTAGAATT
A539R	GATACTATGCTATCCGAAGAAGAATTTCAAGCCATCAGACGGCTTTTGTA
A540L	AGTCTCTTGAAACTCTAGAAGAACGTGTGAAATTATTAGAATCCCGGTGA
A543L	AATTCTACACTTGTAGGGCATACGCCGGGTACACAATGCAAAACACGGAA
A544R	TAAAATGGTGAAATTCAAATACTTCGAAATGGGTTCCAAAGATTGCCCAA
A546L	CCGAGGAAAAGGTTCGCTTTTTGTGTAATAGTCATCTTGATTCAAAATTC
A548L	GAACCTACAGGTGGCAACTCGTGTGTTTATCAATAGTTTGTCGTGGAATG
A551L	TTCTACAGGTATTTCTATTCGCGTTCCAAATGGGACATATGGTCGAATTG
A552R	TGGAAATAGCACATGGTTACCTAACAAATAGCTGGAGATTGTGTATTTGA
A554/556/557L	GGTCATAATATTCCTTACCTTCCTAATAGTACACCCGCGCATTTCAAAAG
A558L	AAACGAATCTCCACGGGCAGTATACCACATCAAATAGCACATTCGTGACG
A559L	CAAAGTAGACTCTGTTAAGAAGGTGTGGACAAAAAGGATGGCAGAATATA
A561L	AAGAGAAATGCGTGGAATAACATCAAAATTGGTCTTAAATTAAATACGAT
A564L	TGTGGGACGTGTAAACCAGAAGACGCAGTTAATGTCGTATCAAAGAAATG
A565R	AAGAAGAATGGAGGTCCTTTGAATTACGGAGACACATTATACTTGAAGTT
A567L	AAACAAGATGTAAACCGCCGCGATTTTGCTACTGAACTTAATAATTCATT
A568L	TGGTTTTTTCATGTTATTATTACAGGAGTATTGACAACATTTATCGTTAT
A570L	AAGTTGTAGATACGCCAGAAGAAGAAGTTATAGAACATGAAGAAGCTTAA
A571R	CTTGCAAAAAATAAAGCATCTGTCGTGAAAAATATGACCAAAAAAATGTA
A572R	CGAGGTCCCATGTTCTTCTGTGAGCTTTTATGGCCTTACAAAAGACGATT
A574L	TAAGCGAGTCGTAATAAAACTACTCGATGATCAACCGGGTATTTTTGAAT
A575L	GTGTACCCTTATAATTTCATTTGCGCTAACGAAAATATGTTGAGGACGAG
A577L	ATGATACAACAAAAGCAGAAAGCCAAAAGGAAAACTGTTCATTATAAATA
A579L	TAAAACCAAGTGGTCAGTGGACGACGAAATTAGGAGAACATATAGCAGAG
A581R	AATTCAAAGAAACCAGGTTCTACAACAACTGAAATTATTGTTATCAAAAT
A583L	AATATGCGGATTAGCAGTCTCACACTGGAACGTGCGACAGAATTGGAACG
A590L	AATATTTATTGGATGGTTTTGCAATTTTGGAGGCAATAGAAACGATATTA
A592R	CGTATATCGATTTTGATTACGATAACTATGACGAAACTCTAGTTTTTTAA
A594R	CGATTCAACTTCAACGCATCTCCTTCTTCCGGGCCATTCAAGTCTATTTT
A596R	TTTCAAATGAAATGTTTGATGAATGTGGAATAGAAGTCGAATACCTATGA
A598L	TGTTCAAAAAACCGAGTGACGAAATAATCTTCAAATGGTCTTTGGCAACG
A601R	GACAAAAAGAGGTGACCCGAGTGTTTATCTACCGTTGATCACGAGCATAT

A602L	GCGTCACTGATGCAGTCCGCAAAATCCCTCAAGTTTCTTCAAAAATCCTC
A603R	ATGTCCGTAGGTTCATTTACTCTCTCATTGGGAAAAAAGGTTCATTCGTT
A604L	ACATATTGTGAAAATTACAAGAGAATATACTCAAAATCAGGCGATTTTAA
A605L	CATTTCCAAAGCTCACGCCAACCAAGATGAGGTACGTTAATCTAATGTAG
A607R	ATACTCCAATTCATTTATTTCTACGCTTTATACAACGCCCTTCCCTGACA
A608R	CACGTTTTTTACACACGAAAAGGTATTCATTTAGAAAGAGAAATCATTAA
A609L	AGAGTTCCCGTTGAACATCGCATTCTCTTTGGTAAAAAATTAATCACACG
A612L	CAATGGCTCTTGGTTTTGGTGCAATTTTTAACCATAGCAAAGACCCTAAC
A614L	TGTGCTCGTATTGGAAAAAAATGCGGTCCTAAAGGAAGATGTGTCAAACT
A617R	CCAAACATTAAGAGACTCGAAAATAAAAGAGTGACGAGTGGAAACTTTAG
A618L	ATATTCGTCAACAGCGAAGGCAGCACTGTGTCCGACATTCTCACACGTAT
A619L	GAAAACCATTATAAAAGAACAGAAGGAAAAACCAAAACGAGTGTTCATTG
A620L	TTGTTTGAAAACAAGAAGGAAAATTTCACTACCGGAAAAAGTCTGACTAA
A621L	ATAAAAACATAGAAAACTCTATCATTGCCTCTATAATGCTTCAATCAAAT
A622L	TCAGATTTATGCTCCCAACTTTAACATCCTGCGTATCGCTGCTGGAATGG
A623L	AACATCGTTTGATGGAATCTCATTCATGTCCCACTTTACAGGCAAAAGAA
A624R	TCTGAAACGACGACAAACTTCAAAGCGTTCCTATTGGGTTTTCTGGCATA
A625R	TAATGTATGCGACCGCGATATTCACGCCGCCCGCAACATCCTCTTGAAAT
A627R	AGTTATAGGAAGACAAGTGCGTCGTCCAGTCAGTCGTCAAGTTAGAAAAC
A628L	GTTTGGCTGTTTTACTACCCAGAATACATTAACTTCTCTCAGTACTCACA
A629R	AAACAGCCTGGGACCTTTCCATGAAAAGTGTGATTGACATGGCGGCAGAC
A633R	CTGTATTGTTCACAAAAGGTTCCCGAAAGATGGAAAAATAGTGACGGGTT
A634L	GCTAACTGGGAACCATACGAAAAATATCTTAATAAGGGTGATAAAGTTGA
A635R	ATTCATCGTGGGAATTTACGCAGTGTGGCTGAGCTGGAGCTGCTCAAAAG
A636R	AAGTTTTTTAAGACAGGAGAAAGCACCAAGTGTCACGTCAATACGTCAAT
A637R	AACTCATTATGAAAGAATTCGGTGTTTTCTTCGACAAGGCAAAACGATTT
A638R	GTGATATGTTCATCACTGAAGAGGAAGCCACGGGAGTCGTAAATTCAGGT
A642R	AAGAACTTTATGGAATCAAAAACCTTAACGAATTTCCACCGCAGGTTATA
A643R	AGGTAGCTTGAGGGTGCCATATAGCCGTAAGTCAATGAGAAGACCAGGGT
A644R	ATGCTTGAAGCAAATTCTTCCCGGCGTTCGCGCATACGCACTATACGTCA
A645R	AAGGCCGAAAAAAACGGGCTGCTTTCAATGGATAAAAATATACCACTGAT
A646L	TAGAAAAGTATCCCTTTCTCCGAGAAGTTAATACACTCGTATACAAAGGA
A647R	AATTTGGGAGGACGACAAATACCGACATCAGGAAACTAAAAACAAGATAT
A649R	GATGGAGATACTCGCACGGTTATTTATAGAGCACTACAATCGCCACAGTT
A651L	AAATGTGCAAGAGGTAAAAGAAAAACCGCGTATAAGTTTCGTTGGTCATA
A654L	ACTTCTCAAAAAAACTATTGAAGAATCATCTTCACATGGGTTCAAGTATA
A656L	CCAGATTATTCGAAAATATTACCAAAGAGAACTAGTGGTAAAGTGATCCC
A658R	GAGTATATTGTGCAGACGTGGTTCTTGGGTTGTTGTAAACGAAGATATCA
A659L	TTCCGAATTAGACCATGTTGATGAAAATGAGGTGGAAGAGGAAGAAAATT
A662L	TAAACAAGACCATTGTGAATCAGATTTTTTTTGCACCCATCAACATTTCT

A664L	AGGCCATTAGAACTCGTGTCAATTGCTCGTGTACAAGAAGAAGAACTTAT
A665L	TGTAGGAGAAATTTTATCGCCTATGATGAAGGACTCGGAACTTACAGTAT
A666L	TGAGGAAGGCGCGTGGTGAAGAAGTATCTGACAGCGATGAAGACTACTAA
A672R	TGAACGGAAAACTGGGAATATACCCTGAAGCAATTATAAATGACACATAA
A674R	CAAGGTTTATTCTTCCTCTGAGTACACCAACTACTATTTACATGTCGGGT
A676R	TAACTGAACTTTATGGAATTCGCTACGGTTATACGTTCGGAAATGATTAA
A678R	TTGGTGGGACATACTTGATAGGAACTACTATTTTCCTTGCTCCTGATTTC
A682L	ATAAAAACACTCATCGAAGCAGGTGGTAACATCAACGCCGTCAACAATTC
A683L	CACACCATGTGCTAGTTATGGCACTGCTACTGCTGGATGCAATGTTCTGA
A686L	TTTGTTATCATTCCCAAGGGTACAGAAGTCCATATGAAATATGTTAATCT
A687R	CCCGTTCAAGATCGAAGAGAGTGACTTGAGACGAGCTTTGGAATCCGTGT

5.2. Supplement 2. List of PBCV-1 CDSs with expression profile

CDS	RefSeq (NCBI)	Expression time[c]	Expressed with aphidicolin	Putative function	COG[d]
A002L	**NP_048350**	**E**	**Yes**	Unknown function	Uncategorized
A003R	**NP_048351**	**E**	**Yes**	Unknown function	Uncategorized
A005R	**NP_048353**	**E**	**Yes**	Unknown function	Uncategorized
A007L	**NP_048355**	**E**	**Yes**	Unknown function	Uncategorized
A008L	**NP_048356**	**E**	**Yes**	Unknown function	Uncategorized
A009R	**NP_048357**	**E**	**Yes**	Unknown function	Uncategorized
A010R	**NP_048358**	**L**	**No**	Unknown function	Uncategorized
A011L	**NP_048359**	**L**	**No**	Unknown function	Uncategorized
A014R	**NP_048362**	**L**	**Yes**	Unknown function	Uncategorized
A018L	**NP_048366**	**L**	**No**	Unknown function	Uncategorized
A025L	**NP_048373**	**L**	**No**	Unknown function	Uncategorized
A029L	**NP_048377**	**L**	**No**	Unknown function	Uncategorized
A034R[a]	**NP_048382**	**L**	**Yes**	Ser/Thr protein kinase	Signaling
A035L	**NP_048383**	**L**	**No**	Unknown function	Uncategorized
A037L	**NP_048385**	**E**	**Yes**	Unknown function	Uncategorized
A039L	**NP_048387**	**E**	**Yes**	SKP-1 Protein	Protein Synthesis, Modification, and Degradation
A041R	**NP_048389**	**L**	**No**	Unknown function	Uncategorized
A044L	**NP_048392**	**E**	**Yes**	ATPase (AAA + class)	Protein Synthesis, Modification, and Degradation
A048R	**NP_048396**	**EL**	**Yes**	Unknown function	Uncategorized
A049L[b]	**NP_048397**	**EL**	**Yes**	Glycerophosphoryl diesterase	Lipid Metabolism
A050L[a]	**NP_048398**	**E**	**Yes**	Pyrimidine dimer-specific glycosylase	DNA Replication, Recombination, and Repair
A051L	**NP_048399**	**L**	**No**	Unknown function	Uncategorized
A053R[a]	**NP_048401**	**E**	**Yes**	D-lactate dehydrogenase	Sugar Metabolism
A057R	**NP_048405**	**E**	**Yes**	Unknown function	Uncategorized
A060L	**NP_048408**	**N/A**	**N/A**	Unknown function	Uncategorized
A061L	**NP_048409**	**L**	**No**	O-methyltransferase	Miscellaneous
A063L	**NP_048411**	**L**	**No**	Unknown function	Uncategorized
A064R	**NP_048412**	**E**	**Yes**	Glycosyltransferase	Sugar Metabolism
A067R	**NP_048415**	**E**	**Yes**	Unknown function	Uncategorized
A071R	**NP_048419**	**EL**	**No**	Unknown function	Uncategorized
A075L	**NP_048423**	**EL**	**Yes**	Unknown function	Uncategorized
A077L	**NP_048425**	**E**	**Yes**	Unknown function	Uncategorized
A078R[a]	**NP_048426**	**E**	**Yes**	N-carbomoylputrescine amidohydrolase	Miscellaneous
A079R	**NP_048427**	**E**	**Yes**	Unknown function	Uncategorized
A081L	**NP_048429**	**E**	**No**	Unknown function	Uncategorized
A084L	**NP_048432**	**E**	**Yes**	Unknown function	Uncategorized

A085R[a]	**NP_048433**	**L**	**No**	Prolyl-4-hydroxylase	Protein Synthesis, Modification, and Degradation
A087R	**NP_048435**	**E**	**Yes**	HNH endonuclease	Integration and Transposition
A088R	**NP_048436**	**E**	**Yes**	Unknown function	Uncategorized
A089R[b]	**NP_048437**	**EL**	**Yes**	Unknown function	Uncategorized
A090R	**NP_048438**	**EL**	**No**	Unknown function	Uncategorized
A092/93L[b]	**NP_048441**	**EL**	**No**	Unknown function	Uncategorized
A094L[a]	**NP_048442**	**E**	**Yes**	Beta-1,3-glucanase	Cell Wall Degradation
A098R[a]	**NP_048446**	**E**	**Yes**	Hyaluronan synthase	Sugar Manipulation
A100R[a]	**NP_048448**	**E**	**Yes**	Glucosamine synthetase	Sugar Manipulation
A103R[a]	**NP_048451**	**E**	**Yes**	mRNA guanylyltransferase	Transcription
A105L	**NP_048453**	**E**	**Yes**	Ubiquitin C-terminal hydrolase	Protein Synthesis, Modification, and Degradation
A107L	**NP_048455**	**E**	**Yes**	Trancription factor TFIIB	Transcription
A109L	**NP_048457**	**EL**	**Yes**	Unknown function	Uncategorized
A111/114R	**NP_048459**	**E**	**Yes**	Fucosyl-Glycosyltransferase	Sugar Manipulation
A118R[a]	**NP_048466**	**EL**	**No**	GDP-D-mannose dehydratase	Sugar Manipulation
A121R	**NP_048469**	**EL**	**No**	Unknown function	Uncategorized
A122/123R	**NP_048470** **NP_048471**	**L**	**Yes**	Unknown function	Uncategorized
A125L	**NP_048472**	**E**	**Yes**	Transcription factor TFIIS	Transcription
A127R	**NP_048475**	**L**	**No**	Unknown function	Uncategorized
A129R	**NP_048477**	**EL**	**Yes**	Unknown function	Uncategorized
A130R[b]	**NP_048478**	**EL**	**Yes**	Unknown function	Uncategorized
A131L	**NP_048479**	**E**	**Yes**	Unknown function	Uncategorized
A133R	**NP_048481**	**E**	**Yes**	Unknown function	Uncategorized
A134L	**NP_048482**	**EL**	**No**	GIY-YIG endonuclease	Integration and Transposition
A135L	**NP_048483**	**L**	**Yes**	Unknown function	Uncategorized
A137R	**NP_048485**	**E**	**Yes**	Unknown function	Uncategorized
A138R	**NP_048486**	**EL**	**Yes**	Unknown function	Uncategorized
A139L	**NP_048487**	**L**	**No**	Unknown function	Uncategorized
A140/145R[b]	**NP_048488**	**EL**	**No**	Unknown function	Uncategorized
A148R	**NP_048496**	**L**	**No**	Unknown function	Uncategorized
A150L	**NP_048498**	**E**	**No**	Unknown function	Uncategorized
A151R	**NP_048499**	**E**	**Yes**	Unknown function	Uncategorized
A153R	**NP_048501**	**EL**	**No**	Superfamily II helicase	Transcription
A154L	**NP_048502**	**EL**	**No**	Unknown function	Uncategorized
A157L[b]	**NP_048505**	**EL**	**No**	Unknown function	Uncategorized
A158L	**NP_048506**	**E**	**No**	Unknown function	Uncategorized
A161R	**NP_048509**	**E**	**Yes**	Unknown function	Uncategorized
A162L	**NP_048510**	**E**	**No**	Unknown function	Uncategorized
A163R	**NP_048511**	**EL**	**No**	Unknown function	Uncategorized
A165L	**NP_048513**	**EL**	**No**	Unknown function	Uncategorized

A166R[a]	**NP_048514**	**EL**	**No**	Exonuclease	DNA Replication, Recombination, and Repair
A168R	**NP_048516**	**L**	**No**	Unknown function	Uncategorized
A169R[a]	**NP_048517**	**EL**	**No**	Aspartate transcarbamylase	Nucleotide metabolism
A171R	**NP_048519**	**E**	**Yes**	Unknown function	Uncategorized
A173L	**NP_048521**	**E**	**Yes**	Patatin-like phospholipase	Lipid Manipulation
A175R	**NP_048523**	**EL**	**Yes**	Unknown function	Uncategorized
A177R	**NP_048525**	**EL**	**No**	Unknown function	Uncategorized
A180R	**NP_048528**	**E**	**Yes**	Fibronectin binding protein	Miscellaneous
A181/182R[a]	**NP_048529**	**L**	**No**	Chitinase	Cell Wall Degradation
A185R	**NP_048532**	**E**	**Yes**	DNA Pol	DNA Replication, Recombination, and Repair
A189/192R	**NP_048536**	**L**	**Yes**	SWI/SNF chromatin remodeling complex	Transcription
A193L	**NP_048540**	**EL**	**No**	PCNA	DNA replication, Recombination, and Repair
A196L	**NP_048543**	**L**	**No**	Unknown function	Uncategorized
A199R	**NP_048546**	**L**	**No**	Unknown function	Uncategorized
A200R	**NP_048547**	**E**	**Yes**	Cytosine deaminase	Nucleotide Metabolism
A201L	**NP_048548**	**EL**	**No**	Unknown function	Uncategorized
A202L	**NP_048549**	**EL**	**Yes**	Unknown function	Uncategorized
A203R	**NP_048550**	**L**	**No**	Unknown function	Uncategorized
A205R	**NP_048552**	**L**	**No**	Unknown function	Uncategorized
A207R[a,b]	**NP_048554**	**EL**	**No**	Ornithine/Arginine decarboxylase	Miscellaneous
A208R	**NP_048555**	**E**	**Yes**	Unknown function	Uncategorized
A213L	**NP_048560**	**EL**	**Yes**	Unknown function	Uncategorized
A214L	**NP_048561**	**E**	**Yes**	Unknown function	Uncategorized
A215L[a]	**NP_048562**	**L**	**No**	Beta and Alpha 1,4 linked glucoronic lyase	Cell Wall Degradation
A217L[b]	**NP_048564**	**EL**	**No**	Monoamine oxidase	Miscellaneous
A219/222/226R	**NP_048569**	**E**	**Yes**	Glycosyltransferase	Sugar Manipulation
A227L	**NP_048575**	**L**	**No**	Unknown function	Uncategorized
A229L	**NP_048577**	**L**	**Yes**	Unknown function	Uncategorized
A230R	**NP_048578**	**L**	**No**	Unknown function	Uncategorized
A231L[b]	**NP_048579**	**EL**	**No**	Unknown function	Uncategorized
A233R	**NP_048581**	**L**	**No**	Unknown function	Uncategorized
A234L	**NP_048582**	**EL**	**No**	Unknown function	Uncategorized
A237R[a]	**NP_048585**	**L**	**No**	Homospermidine synthase	Miscellaneous
A239L	**NP_048587**	**E**	**Yes**	Unknown function	Uncategorized
A241R	**NP_048589**	**E**	**Yes**	Superfamily II helicase	Transcription
A243R	**NP_048591**	**EL**	**No**	Unknown function	Uncategorized
A245R[a]	**NP_048593**	**L**	**No**	Cu/Zn-Superoxide Dismutase	Miscellaneous
A246R	**NP_048594**	**L**	**No**	Pathogenesis-related protein	Miscellaneous
A247R	**NP_048596**	**E**	**No**	Unknown function	Uncategorized
A248R[a]	**NP_048597**	**E**	**Yes**	Ser/Thr protein kinase	Signaling
A250R[a,b]	**NP_048599**	**EL**	**Yes**	Potassium channel protein	Signaling

A251R[a]	**NP_048600**	**E**	**Yes**	Adenine methyltransferase	DNA Restriction/Modification
A252R[a]	**NP_048602**	**L**	**No**	DNA restriction endonuclease	DNA Restriction/Modification
A253R	**NP_048604**	**E**	**Yes**	Unknown function	Uncategorized
A254R	**NP_048605**	**L**	**No**	Acetyltransferase	Lipid Manipulation
A256L	**NP_048607**	**EL**	**Yes**	Unknown function	Uncategorized
A257L	**NP_048608**	**E**	**Yes**	Unknown function	Uncategorized
A259L	**NP_048610**	**E**	**Yes**	Unknown function	Uncategorized
A260R[a]	**NP_048613**	**L**	**Yes**	Chitinase	Cell Wall Degradation
A261R	**NP_048615**	**EL**	**Yes**	Unknown function	Uncategorized
A262L	**NP_048616**	**L**	**Yes**	Unknown function	Uncategorized
A263L	**NP_048617**	**L**	**No**	Unknown function	Uncategorized
A265L	**NP_048619**	**EL**	**Yes**	Unknown function	Uncategorized
A267L	**NP_048621**	**E**	**No**	HNH endonuclease	Integration and Transposition
A271L[b]	**NP_048625**	**EL**	**Yes**	Lysophospholipase	Lipid Metabolism
A273L	**NP_048627**	**L**	**No**	Unknown function	Uncategorized
A274R	**NP_048628**	**EL**	**No**	Unknown function	Uncategorized
A275R	**NP_048629**	**E**	**Yes**	Unknown function	Uncategorized
A277L[a]	**NP_048631**	**L**	**No**	Ser/Thr protein kinase	Signaling
A278L[a]	**NP_048632**	**L**	**No**	Ser/Thr protein kinase	Signaling
A282L[a]	**NP_048636**	**L**	**No**	Ser/Thr protein kinase	Signaling
A284L[a,b]	**NP_048638**	**EL**	**Yes**	Amidase	Miscellaneous
A286R	**NP_048640**	**L**	**No**	Unknown function	Uncategorized
A287R	**NP_048641**	**EL**	**No**	GIY-YIG endonuclease	Integration and Transposition
A289L[a]	**NP_048643**	**EL**	**Yes**	Ser/Thr protein kinase	Signaling
A292L[a]	**NP_048646**	**L**	**No**	Chitosanase	Cell Wall Degradation
A295L[a]	**NP_048649**	**EL**	**No**	Fucose synthase	Sugar Manipulation
A296R	**NP_048650**	**L**	**Yes**	Unknown function	Uncategorized
A297L	**NP_048651**	**L**	**Yes**	Fructose-2,6 biphosphatase	Sugar Metabolism
A298L	**NP_048652**	**E**	**No**	Unknown function	Uncategorized
A301L	**NP_048655**	**E**	**Yes**	Unknown function	Uncategorized
A304R	**NP_048658**	**L**	**Yes**	Unknown function	Uncategorized
A305L	**NP_048659**	**L**	**No**	Dual specificity phosphatase	Signaling
A306L	**NP_048660**	**EL**	**Yes**	Unknown function	Uncategorized
A308L	**NP_048663**	**E**	**Yes**	Unknown function	Uncategorized
A310L	**NP_048665**	**L**	**No**	Unknown function	Uncategorized
A312L	**NP_048667**	**EL**	**Yes**	Unknown function	Uncategorized
A313L	**NP_048669**	**L**	**No**	Unknown function	Uncategorized
A314R	**NP_048670**	**L**	**No**	Unknown function	Uncategorized
A315L	**NP_048671**	**EL**	**Yes**	GIY-YIG endonuclease	Integration and Transposition
A316R	**NP_048672**	**L**	**Yes**	Unknown function	Uncategorized

A318R	**NP_048674**	**L**	**No**	Unknown function	Uncategorized
A320R	**NP_048676**	**L**	**No**	Unknown function	Uncategorized
A321R	**NP_048677**	**L**	**No**	Unknown function	Uncategorized
A322L	**NP_048678**	**L**	**Yes**	Unknown function	Uncategorized
A324L	**NP_048680**	**L**	**No**	Unknown function	Uncategorized
A326L	**NP_048682**	**L**	**No**	NTP pyrophosphohydrolase	Nucleotide Metabolism
A328L	**NP_048684**	**N/A**	**Yes**	Unknown function	Uncategorized
A329R	**NP_048685**	**L**	**No**	Unknown function	Uncategorized
A330R	**NP_048686**	**EL**	**No**	Unknown function	Uncategorized
A333L	**NP_048689**	**EL**	**No**	Unknown function	Uncategorized
A337L[b]	**NP_048693**	**EL**	**No**	Unknown function	Uncategorized
A339L[b]	**NP_048695**	**EL**	**No**	Unknown function	Uncategorized
A341L	**NP_048697**	**E**	**Yes**	Unknown function	Uncategorized
A342L[b]	**NP_048699**	**EL**	**No**	Unknown function	Uncategorized
A348R	**NP_048705**	**E**	**Yes**	Unknown function	Uncategorized
A349L[b]	**NP_048706**	**EL**	**No**	Unknown function	Uncategorized
A351L	**NP_048708**	**E**	**Yes**	GIY-YIG endonuclease	Integration and Transposition
A352L	**NP_048709**	**L**	**No**	Unknown function	Uncategorized
A354R	**NP_048711**	**EL**	**No**	HNH endonuclease	Integration and Transposition
A357L	**NP_048714**	**EL**	**Yes**	Unknown function	Uncategorized
A360R	**NP_048717**	**L**	**No**	Unknown function	Uncategorized
A361R[b]	**NP_048718**	**EL**	**No**	Unknown function	Uncategorized
A363R	**NP_048720**	**L**	**No**	Superfamily II helicase	Transcription
A366L	**NP_048723**	**EL**	**No**	Transposase	Integration and Transposition
A368L	**NP_048725**	**E**	**Yes**	Unknown function	Uncategorized
A373R	**NP_048730**	**EL**	**Yes**	Unknown function	Uncategorized
A375R[b]	**NP_048732**	**EL**	**Yes**	Unknown function	Uncategorized
A378L	**NP_048735**	**L**	**No**	Unknown function	Uncategorized
A379L	**NP_048736**	**E**	**Yes**	Unknown function	Uncategorized
A382R	**NP_048739**	**L**	**No**	Unknown function	Uncategorized
A383R	**NP_048740**	**L**	**Yes**	Unknown function	Uncategorized
A384L[b]	**NP_048741**	**EL**	**Yes**	Unknown function	Uncategorized
A390L	**NP_048747**	**L**	**No**	Unknown function	Uncategorized
A392R	**NP_048749**	**L**	**No**	ATPase (DNA packaging)	DNA Replication, Recombination, and Repair
A394R[b]	**NP_048751**	**EL**	**No**	Unknown function	Uncategorized
A395R	**NP_048752**	**EL**	**Yes**	Unknown function	Uncategorized
A396L	**NP_048753**	**EL**	**No**	Unknown function	Uncategorized
A397R	**NP_048754**	**E**	**Yes**	Unknown function	Uncategorized
A398L	**NP_048755**	**L**	**No**	Unknown function	Uncategorized
A399R	**NP_048756**	**E**	**No**	Rnase H	DNA Replication, Recombination, and Repair

A400R	**NP_048757**	**EL**	**Yes**	Unknown function	Uncategorized
A401R	**NP_048758**	**EL**	**No**	Unknown function	Uncategorized
A402R	**NP_048759**	**L**	**No**	Lipoprotein lipase	Lipid Manipulation
A403R	**NP_048760**	**EL**	**Yes**	Unknown function	Uncategorized
A404R	**NP_048761**	**L**	**No**	Unknown function	Uncategorized
A405R	**NP_048762**	**L**	**No**	Unknown function	Uncategorized
A407L	**NP_048764**	**L**	**No**	Unknown function	Uncategorized
A408L	**NP_048765**	**EL**	**No**	Unknown function	Uncategorized
A410L	**NP_048767**	**E**	**Yes**	Unknown function	Uncategorized
A411R	**NP_048768**	**L**	**No**	Unknown function	Uncategorized
A412R	**NP_048769**	**E**	**Yes**	Unknown function	Uncategorized
A413L	**NP_048770**	**L**	**No**	Unknown function	Uncategorized
A414R	**NP_048771**	**L**	**No**	Unknown function	Uncategorized
A416R	**NP_048773**	**L**	**No**	Deoxynucleoside kinase	Nucleotide Metabolism
A417L	**NP_048774**	**L**	**No**	Replication factor C	DNA Replication, Recombination, and Repair
A420L	**NP_048777**	**L**	**No**	Unknown function	Uncategorized
A421R	**NP_048778**	**L**	**No**	Unknown function	Uncategorized
A422R[b]	**NP_048779**	**EL**	**No**	HNH endonuclease	Integration and Transposition
A423R	**NP_048780**	**L**	**No**	Unknown function	Uncategorized
A424R	**NP_048781**	**L**	**Yes**	Unknown function	Uncategorized
A426R	**NP_048783**	**L**	**No**	Unknown function	Uncategorized
A427L	**NP_048784**	**E**	**Yes**	Thioredoxin	Nucleotide Metabolism
A428L	**NP_048785**	**EL**	**Yes**	Unknown function	Uncategorized
A429L	**NP_048786**	**E**	**Yes**	Unknown function	Uncategorized
A430L	**NP_048787**	**L**	**Yes**	Major capsid	Structural Protein
A431L	**NP_048788**	**E**	**Yes**	Unknown function	Uncategorized
A432R	**NP_048789**	**L**	**No**	Unknown function	Uncategorized
A435R	**NP_048793**	**E**	**Yes**	Unknown function	Uncategorized
A437L	**NP_048794**	**L**	**No**	DNA binding protein	DNA Replication, Recombination, and Repair
A438L[a]	**NP_048795**	**EL**	**No**	Glutaredoxin	Nucleotide Metabolism
A439R	**NP_048796**	**E**	**Yes**	Unknown function	Uncategorized
A440L	**NP_048797**	**E**	**Yes**	Unknown function	Uncategorized
A441L	**NP_048798**	**E**	**Yes**	Unknown function	Uncategorized
A443R	**NP_048800**	**E**	**No**	Unknown function	Uncategorized
A444L	**NP_048801**	**L**	**No**	Unknown function	Uncategorized
A445L	**NP_048802**	**E**	**Yes**	ABC Transporter	Miscellaneous
A448L	**NP_048805**	**L**	**No**	Protein disulfide isomerase	Protein Synthesis, Modification, and Degradation
A449R[a]	**NP_048806**	**E**	**Yes**	RNA triphosphatase	Transcription
A450R	**NP_048807**	**E**	**Yes**	Unknown function	Uncategorized
A452L	**NP_048809**	**E**	**Yes**	Unknown function	Uncategorized

A454L	**NP_048811**	**EL**	**Yes**	Unknown function	Uncategorized
A456L	**NP_048813**	**E**	**Yes**	Helicase -Superfamily III	DNA Replication, Recombination, and Repair
A462R[b]	**NP_048818**	**EL**	**No**	Unknown function	Uncategorized
A464R[a]	**NP_048820**	**E**	**Yes**	RNAse III	Transcription
A465R[b]	**NP_048821**	**EL**	**No**	Thiol oxidoreductase	Protein Synthesis, Modification, and Degradation
A467L	**NP_048823**	**EL**	**Yes**	Unknown function	Uncategorized
A468R	**NP_048824**	**EL**	**No**	Archaeo-eukaryotic primase	DNA Replication, Recombination, and Repair
A470R	**NP_048826**	**L**	**No**	Unknown function	Uncategorized
A471R	**NP_048827**	**L**	**No**	Unknown function	Uncategorized
A473L	**NP_048829**	**E**	**Yes**	Glycosyltransferase	Sugar Metabolism
A476R	**NP_048832**	**EL**	**No**	Ribo. Reductase	Nucleotide Metabolism
A478L	**NP_048834**	**EL**	**Yes**	HNH endonuclease	Integration and Transposition
A480L	**NP_048836**	**L**	**No**	Unknown function	Uncategorized
A481L	**NP_048837**	**E**	**Yes**	SCF-E3 ubiquitin ligase	Protein Synthesis, Modification, and Degradation
A482R	**NP_048838**	**N/A**	**Yes**	Unknown function	Uncategorized
A484L[b]	**NP_048840**	**EL**	**No**	Unknown function	Uncategorized
A485R	**NP_048841**	**L**	**No**	Unknown function	Uncategorized
A486L	**NP_048842**	**E**	**Yes**	Unknown function	Uncategorized
A488R	**NP_048844**	**L**	**No**	Unknown function	Uncategorized
A490L	**NP_048846**	**EL**	**No**	HNH endonuclease	Integration and Transposition
A491R	**NP_048847**	**L**	**No**	Unknown function	Uncategorized
A492L	**NP_048848**	**E**	**Yes**	Unknown function	Uncategorized
A493L	**NP_048849**	**E**	**No**	Unknown function	Uncategorized
A494R	**NP_048850**	**EL**	**No**	Unknown function	Uncategorized
A495R	**NP_048851**	**EL**	**Yes**	GIY-YIG endonuclease	Integration and Transposition
A497R	**NP_048853**	**L**	**No**	Unknown function	Uncategorized
A501L	**NP_048857**	**L**	**Yes**	Unknown function	Uncategorized
A502L	**NP_048858**	**L**	**No**	Unknown function	Uncategorized
A503L	**NP_048859**	**E**	**No**	Unknown function	Uncategorized
A505L	**NP_048861**	**E**	**Yes**	Unknown function	Uncategorized
A512R	**NP_048868**	**E**	**Yes**	Unknown function	Uncategorized
A517L	**NP_048873**	**E**	**Yes**	Cytosine methyltransferase	DNA Restriction/Modification
A519L	**NP_048875**	**L**	**No**	Unknown function	Uncategorized
A520L	**NP_048876**	**L**	**No**	Unknown function	Uncategorized
A521L[b]	**NP_048877**	**EL**	**Yes**	Unknown function	Uncategorized
A523R	**NP_048879**	**L**	**No**	Unknown function	Uncategorized
A526R	**NP_048882**	**L**	**No**	Unknown function	Uncategorized
A527R	**NP_048883**	**L**	**No**	Unknown function	Uncategorized
A530R	**NP_048886**	**L**	**No**	Cytosine methyltransferase	DNA Restriction/Modification
A531L	**NP_048887**	**L**	**No**	Unknown function	Uncategorized

A532L	**NP_048888**	**L**	**No**	Unknown function	Uncategorized
A533R	**NP_048889**	**EL**	**No**	Unknown function	Uncategorized
A535L[b]	**NP_048891**	**EL**	**No**	Unknown function	Uncategorized
A536L[b]	**NP_048892**	**EL**	**No**	Unknown function	Uncategorized
A537L	**NP_048893**	**E**	**No**	Unknown function	Uncategorized
A539R	**NP_048895**	**E**	**Yes**	GIY-YIG endonuclease	Integration and Transposition
A540L	**NP_048896**	**L**	**No**	Unknown function	Uncategorized
A543L	**NP_048899**	**L**	**No**	Unknown function	Uncategorized
A544R	**NP_048900**	**L**	**No**	ATP-dependent ligase	DNA Replication, Recombination, and Repair
A546L	**NP_048902**	**E**	**No**	Glycosyltransferase	Sugar Metabolism
A548L	**NP_048904**	**E**	**Yes**	SWI/SNF helicase	Transcription
A551L[a]	**NP_048907**	**E**	**Yes**	dUTP pyrophosphatase	Nucleotide Metabolism
A552R	**NP_048908**	**E**	**Yes**	Transcription factor TFIID	Transcription
A554/556/557L	**NP_048910**	**E**	**Yes**	ATPase (PP-loop)	DNA Replication, Recombination, and Repair
A558L	**NP_048914**	**EL**	**No**	Unknown function	Uncategorized
A559L	**NP_048915**	**L**	**No**	Unknown function	Uncategorized
A561L	**NP_048917**	**L**	**Yes**	ATPase (DNA repair)	DNA Replication, Recombination, and Repair
A564L	**NP_048920**	**EL**	**Yes**	Unknown function	Uncategorized
A565R[b]	**NP_048921**	**EL**	**Yes**	ATPase (Chromosome segregation)	DNA Replication, Recombination, and Repair
A567L	**NP_048923**	**EL**	**No**	Unknown function	Uncategorized
A568L	**NP_048924**	**E**	**Yes**	Unknown function	Uncategorized
A570L	**NP_048926**	**L**	**Yes**	Unknown function	Uncategorized
A571R	**NP_048927**	**L**	**No**	Unknown function	Uncategorized
A572R	**NP_048928**	**L**	**No**	Unknown function	Uncategorized
A574L	**NP_048930**	**E**	**Yes**	PCNA	DNA Replication, Recombination, and Repair
A575L	**NP_048931**	**E**	**Yes**	Unknown function	Uncategorized
A577L	**NP_048933**	**L**	**No**	Unknown function	Uncategorized
A579L[a]	**NP_048935**	**L**	**No**	DNA restriction endonuclease	DNA Restriction/Modification
A581R[a]	**NP_048937**	**E**	**Yes**	DNA methyltransferase	DNA Restriction/Modification
A583L[a]	**NP_048939**	**E**	**Yes**	DNA Topoisomerase II	DNA Replication, Recombination, and Repair
A590L[b]	**NP_048946**	**EL**	**No**	Unknown function	Uncategorized
A592R	**NP_048948**	**EL**	**No**	Unknown function	Uncategorized
A594R	**NP_048950**	**E**	**Yes**	Unknown function	Uncategorized
A596R[a]	**NP_048952**	**EL**	**Yes**	dCMP deaminase	Nucleotide Metabolism
A598L	**NP_048954**	**EL**	**Yes**	Histidine decarboxylase	Miscellaneous
A601R	**NP_048957**	**EL**	**No**	Unknown function	Uncategorized
A602L	**NP_048958**	**E**	**Yes**	Unknown function	Uncategorized
A603R	**NP_048959**	**E**	**Yes**	Unknown function	Uncategorized
A604L	**NP_048960**	**E**	**Yes**	Zn metallopeptidase	Protein Synthesis, Modification, and Degradation
A605L[b]	**NP_048961**	**EL**	**No**	Unknown function	Uncategorized

A607R	**NP_048963**	**E**	**Yes**	Unknown function	Uncategorized
A608R	**NP_048964**	**E**	**Yes**	Unknown function	Uncategorized
A609L[a]	**NP_048965**	**E**	**Yes**	UDP-glucose 6-dehydrogenase	Sugar Metabolism
A612L[a]	**NP_048968**	**L**	**No**	Histone H3, Lys27 methylase	Transcription
A614L[a]	**NP_048970**	**L**	**No**	Ser/Thr protein kinase	Signaling
A617R[a]	**NP_048973**	**EL**	**Yes**	Ser/Thr protein kinase	Signaling
A618L	**NP_048974**	**L**	**Yes**	Unknown function	Uncategorized
A619L	**NP_048975**	**L**	**No**	Unknown function	Uncategorized
A620L	**NP_048976**	**L**	**No**	Unknown function	Uncategorized
A621L	**NP_048977**	**L**	**No**	Unknown function	Uncategorized
A622L	**NP_048978**	**L**	**No**	Unknown function	Uncategorized
A623L	**NP_048979**	**E**	**No**	Unknown function	Uncategorized
A624R	**NP_048980**	**L**	**No**	Unknown function	Uncategorized
A625R	**NP_048981**	**L**	**Yes**	Transposase	Integration and Transposition
A627R	**NP_048983**	**L**	**Yes**	Unknown function	Uncategorized
A628L	**NP_048984**	**E**	**Yes**	Unknown function	Uncategorized
A629R	**NP_048985**	**EL**	**Yes**	Ribo. Reductase (Large subunit)	Nucleotide Metabolism
A633R	**NP_048989**	**L**	**Yes**	Unknown function	Uncategorized
A634L	**NP_048990**	**E**	**Yes**	Unknown function	Uncategorized
A635R	**NP_048991**	**L**	**Yes**	Unknown function	Uncategorized
A636R	**NP_048992**	**L**	**No**	Unknown function	Uncategorized
A637R	**NP_048993**	**E**	**Yes**	Unknown function	Uncategorized
A638R[a]	**NP_048994**	**E**	**Yes**	Agmatine iminohydrolase	Miscellaneous
A642R[b]	**NP_048998**	**EL**	**No**	Unknown function	Uncategorized
A643R	**NP_048999**	**L**	**No**	Unknown function	Uncategorized
A644R	**NP_049000**	**L**	**No**	Unknown function	Uncategorized
A645R	**NP_049001**	**E**	**Yes**	Unknown function	Uncategorized
A646L	**NP_049002**	**N/A**	**Yes**	Unknown function	Uncategorized
A647R	**NP_049003**	**E**	**Yes**	Unknown function	Uncategorized
A649R	**NP_049005**	**E**	**Yes**	Unknown function	Uncategorized
A651L	**NP_049007**	**EL**	**Yes**	GIY-YIG endonuclease	Integration and Transposition
A654L	**NP_049010**	**E**	**Yes**	N-Acetyltransferase	Lipid Metabolism
A656L	**NP_049012**	**E**	**Yes**	Unknown function	Uncategorized
A658R	**NP_049014**	**L**	**No**	Unknown function	Uncategorized
A659L	**NP_049015**	**L**	**No**	Unknown function	Uncategorized
A662L	**NP_049018**	**EL**	**Yes**	Unknown function	Uncategorized
A664L	**NP_049020**	**E**	**Yes**	Unknown function	Uncategorized
A665L	**NP_049021**	**E**	**Yes**	Unknown function	Uncategorized
A666L	**NP_049022**	**E**	**Yes**	Translation elongation factor -3	Protein Synthesis, Modification, and Degradation
A672R	**NP_049028**	**E**	**Yes**	Unknown function	Uncategorized

A674R[a]	**NP_049030**	**E**	**Yes**	Thymidylate synthase X	Nucleotide Metabolism
A676R	**NP_049032**	**L**	**No**	Unknown function	Uncategorized
A678R	**NP_049034**	**L**	**Yes**	Unknown function	Uncategorized
A682L	**NP_049038**	**E**	**Yes**	Unknown function	Uncategorized
A683L	**NP_049039**	**EL**	**Yes**	Cytosine methyltransferase	DNA Restriction/Modification
A686L	**NP_049042**	**E**	**Yes**	Unknown function	Uncategorized
A687R	**NP_049043**	**E**	**Yes**	Unknown function	Uncategorized
A689L	**NP_049045**	**N/A**	**N/A**	Unknown function	Uncategorized

[a] Recombinant proteins have been characterized for these CDSs

[b] Non-continuous transcription was detected for these CDSs

[c] E = Early genes; EL = Early/Late genes; L = Late genes; N/A = Not Available

[d] COG = Cluster of Orthologus Group

6. REFERENCES

1. Van Etten, J. L. 2003. Unusual life style of giant chlorella viruses. *Annu. Rev. Genetics* 37:153-195.

2. Wilson, W. H., Van Etten J. L., Allen M. J. 2008. The Phycodnaviridae: the story of how tiny giants rule the world, p. 1-42. *In* J. L. Van Etten (ed.), Lesser known large dsDNA viruses. Springer-Verlag, Berlin-Heidelberg, Germany.

3. Yamada, T., Onimatsu H., Van Etten J. L. 2006. Chlorella viruses, p. 293-336. *In* K. Maramorosch, and A. J. Shatkin (ed.), Advances in virus research. Elsevier Academic Press, San Diego, CA.

4. 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.

5. 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.

6. Cherrier, M. V., Kostyuchenko V. A., Xiao C., Bowman V. D., Battisti A. J., Yan X., Chipman P. R., Baker T. S., Van Etten J. L., Rossmann M. G. 2009. An icosahedral algal virus has a complex unique vertex decorated by a spike. *Proc. Natl. Acad. Sci. USA* 106:11085-11090.

7. Onimatsu, H., Suganuma K., Uenoyama S., Yamada T. 2006. C-terminal repetitive motifs in Vp130 present at the unique vertex of the Chlorovirus capsid are essential for binding to the host Chlorella cell wall. *Virology* 353:433-442.

8. Frohns, F., Käsmann A., Kramer D., Schäfer B., Mehmel M., Kang M., Van Etten J. L., Gazzarrini S., Moroni A., Thiel G. 2006. Potassium ion channels of Chlorella viruses cause rapid depolarization of host cells during infection. *J. Virol*. 80:2437-2444.

9. Neupärtl, M., Meyer C., Woll I., Frohns F., Kang M., Van Etten J. L., Kramer D., Hertel B., Moroni A., Thiel G. 2008. Chlorella viruses evoke a rapid release of K^+ from host cells during the early phase of infection. *Virology* 372:340-348.

10. Agarkova, I. V., Dunigan D., Gurnon J., Greiner T., Barres J., Thiel G., Van Etten J. L. 2008. Chlorovirus-mediated membrane depolarization of Chlorella alters secondary active transport of solutes. *J. Virol*. 82:12181-12190.

11. Kawasaki, T., Tanaka M., Fujie M., Usami S., Yamada T. 2004. Immediate early genes expressed in chlorovirus infections. *Virology* 318:214-223.

12. 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.

13. Fitzgerald, L. A., Boucher P. T., Yanai-Balser G. M., Suhre K., Graves M. V., Van Etten J. L. 2008. Putative gene promoter sequences in the chlorella viruses. *Virology* 380:388-393.

14. Knapen, D., Vergauwen L., Laukens K., Blust R. 2009. Best practices for hybridization design in two-colour microarray analysis. *Trends in Biotechnology* 27:406-414.

15. Luke, K., Radek A., Liu X., Campbell J., Uzan M., Haselkorn R., Kogan Y. 2002. Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* 299:182-191.

16. Yang, Y. H., Speed T. 2002. Design issues for cDNA microarray experiments. *Nature Reviews* 3:579-588.

17. Saeed, A. I., Sharov V., White J., Li J., Liang W., Bhagabati N., Braisted J., Klapa M., Currier T., Thiagarajan M., Sturn A., Snuffin M., Rezantsev A., Popov D., Ryltsov A., Kostukovich E., Borisovsky I., Liu Z., Vinsavich A., Trush V., Quackenbush J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374-378.

18. Tusher, V. G., Tibshirani R., Chu G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* 98:5116-21.

19. 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.

20. Fitzgerald, L. A., Graves M. V., Li X., Feldblyum T., Nierman W. C., Van Etten J. L. 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A. *Virology* 358:472-484.

21. McCullough, A. K., Romberg M. T., Nyaga S., Wei Y., Wood T. G., Taylor J. S., Van Etten J. L., Dodson M. L., Lloyd R. S. 1998. Characterization of a novel cis-syn and trans-syn II pyrimidine dimer glycosylase/AP lyase from a eukaryotic algal virus, PBCV-1. *J. Biol. Chem*. 273:13136-13142.

22. Van Etten, J. L., Schuster A. M., Girton L., Burbank D. E., Swinton D., Hattman S. 1985. DNA methylation of viruses infecting a eukaryotic Chlorella-like green alga. *Nucleic Acids Res*. 13:3471-3478.

23. 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.

24. Graves, M. V., Bernardt C. T., Cerny R., Van Etten J. L. 2001. Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. *Virology* 285:332-345.

25. Wang, I.-N., Li Y., Que Q., Bhattacharya M., Lane L. C., Chaney W. G., Van Etten J. L. 1993. Evidence for virus-encoded glycosylation specificity. *Proc. Natl. Acad. Sci. USA* 90:3840-3844.

26. Zhang, Y., Xiang Y., Van Etten J. L., Rossmann M. G. 2007. Structure and function of a chlorella virus PBCV-1 encoded glycosyltransferase. *Structure* 15:1031-1039.

27. Fruscione, F., Sturla L., Duncan G. A., Van Etten J. L., Valbuzzi P., De Flora A., Zanni E. D., Tonetti M. 2008. Differential role of $NADP^+$ and NADPH in the

activity and structure of GDP-D-mannose 4,6-dehydratase from two chlorella viruses. *J. Biol. Chem*. 283:184-193.

28. Tonetti, M., Zanardi D., Gurnon J. R., Fruscione F., Armirotti A., Damonte G., Sturla L., De Flora A., Van Etten J. L. 2003. Chlorella virus PBCV-1 encodes two enzymes involved in the biosynthesis of GDP-L-fucose and GDP-D-rhamnose. *J. Biol. Chem*. 278:21672-21677.

29. DeAngelis, P., Jing W., Graves M. V., Burbank D. E., Van Etten J. L. 1997. Hyaluronan synthase of chlorella virus PBCV-1. *Science* 278:1800-1803.

30. 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.

31. Lua, D. T., Yasuike M., Hirono I., Aoki T. 2005. Transcription program of Red Sea bream iridovirus as revealed by DNA microarrays. *J. Virol*. 79:15151-15164.

32. Graves, M. V., Meints R. H. 1992. Characterization of the major capsid protein and cloning of its gene from algal virus PBCV-1. *Virology* 188:198-207.

33. Mujtaba, S., Manzur K. L., Gurnon J. R., Kang M., Van Etten J. L., Zhou M-M. 2008. Epigenetic transcriptional repression of cellular genes by a viral SET protein. *Nature Cell Biology* 10:1114 - 1122

34. Agarkova, I. V., Dunigan D. D., Van Etten J. L. 2006. Virion-associated restriction endonucleases of chloroviruses. *J. Virol*. 80:8114-8123.

35. Sun, L., Adams B., Gurnon J. R., Ye Y., Van Etten J. L. 1999. Characterization of two chitinase genes and one chitosanase gene encoded by chlorella virus PBCV-1. *Virology* 263:376-387.

36. Zhang, Y., Maley F., Maley G. F., Duncan G., Dunigan D. D., Van Etten J. L. 2007. Chloroviruses encode a bifunctional dCMP-dCTP deaminase that produces two key intermediates in dTTP formation. *J. Virol*. 81:7662-7671.

37. 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.

38. Van Etten, J. L., Schuster A. M., Meints R. H. 1988. Viruses of eukaryotic chlorella-like algae, p. 411-428. *In* Y. Koltin, and M. J. Leibowitz (ed.), Viruses of fungi and simple eukaryotes. Marcel Dekker, Inc., New York, NY.

39. Kang, M., Graves M., Mehmel M., Moroni A., Gazzarrini S., Thiel G., Gurnon J. R., Van Etten J. L. 2004. Genetic diversity in chlorella viruses flanking *kcv*, a gene that encodes a potassium ion channel protein. *Virology* 326:150-159.

40. Assarsson, E., Greenbaum J. A., Sundström M., Schaffer L., Hammond J. A., Pasquetto V., Oseroff C., Hendrickson R. C., Lefkowitz E. J., Tscharke D. C., Sidney J., Grey H. M., Head S. R., Peters B., Sette A. 2008. Kinetic analysis of a

complete poxvirus transcriptome reveals an immediate-early class of genes. *Proc. Natl. Acad. Sci. USA* 105:2140-2145.

41. Rubins, K. H., Hensley L. E., Bell G. W., Wang C., Lefkowitz E. J., Brown P. O., Relman D. A. 2008. Comparative analysis of viral gene expression programs during poxvirus infection: a transcriptional map of the vaccinia and monkeypox genomes. *PLoS ONE* 3:e2628.

42. Teng, Y., Hou Z., Gong J., Liu H., Xie X., Zhang L., Chen X., Qin Q. W. 2008. Whole-genome transcriptional profiles of a novel marine fish iridovirus, Singapore grouper iridovirus (SGIV) in virus-infected grouper spleen cell cultures and in orange-spotted grouper, *Epinephulus coioides*. *Virology* 377:39-48.

43. Condit, R. C. 2007. Vaccinia, Inc. - probing the functional substrate of poxviral replication factories. *Cell Host Microbe* 2:205-207.

CHAPTER III

PUTATIVE GENE PROMOTER SEQUENCES IN THE CHLORELLA VIRUSES

1. OVERVIEW

Chlorella viruses (Family *Phycodnaviridae*, genus *Chlorovirus*) are large, icosahedral, plaque-forming, dsDNA-containing viruses that infect and replicate in certain isolates of chlorella-like green algae. The 330-kb genome of the prototype virus, *Paramecium bursaria* chlorella virus 1 (PBCV-1), was sequenced and annotated more than 10 years ago [1]. The virus contains 366 putative protein-encoding genes and a polycistronic gene that encodes 11 tRNAs [1, 2]. Approximately 40% of its predicted gene products resemble proteins of known function and many are unexpected for a virus. Currently, three species are included in the genus *Chlorovirus*: i) Viruses that infect *Chlorella* NC64A (NC64A viruses), ii) viruses that infect *Chlorella* Pbi (Pbi viruses) and iii) viruses that infect symbiotic zoochlorella in the coelenterate *Hydra viridis* [3]. *Chlorella* NC64A and *Chlorella* Pbi are normally endosymbionts of the protozoan *Paramecium bursaria*, but they can be cultured independently of the protozoan. The current study involves three chlorella viruses whose genomes have been sequenced: viruses PBCV-1 and NY-2A [4] are NC64A viruses with 330-kb and 369-kb genomes, respectively, and virus MT325, a Pbi virus with a 314-kb genome [5].

PBCV-1 infects its host by attaching rapidly to the external surface of the algal cell wall [6]. Attachment occurs at a unique virus vertex [7] and is followed by digestion of the cell wall at the attachment point. Following host wall degradation, the PBCV-1 internal membrane presumably fuses with the host membrane, which leads to the entry of virus DNA and probably associated proteins into the host. An empty capsid remains on the host surface. Circumstantial evidence suggests that the infecting PBCV-1 DNA, and DNA associated proteins, rapidly move to the nucleus to initiate virus transcription [2]. Support for this hypothesis includes the fact that neither PBCV-1 nor any of the chlorella viruses encode a recognizable RNA polymerase or RNA polymerase subunit. Furthermore, RNA polymerase activity was not detected in PBCV-1 virions (Rohozinski and Van Etten, unpublished results). Consequently, assuming that the infecting viral DNA moves to the nucleus, it must commandeer one of the host's RNA polymerases (probably RNA polymerase II) to initiate viral transcription [2]. Therefore, the host polymerase(s), possibly in combination with a virus protein(s), must recognize some virus DNA promoter sequence(s) to initiate transcription. This process occurs rapidly because early PBCV-1 transcripts can be detected within 5 to 10 min post-infection (p.i.) [8]. Virus DNA replication starts 60 to 90 min p.i., followed by transcription of late virus genes. Nascent virus capsids begin to assemble in localized regions of the cytoplasm, called virus assembly centers, at 2 to 4 hr p.i. By 5 to 6 hr p.i. the cytoplasm contains many progeny viruses and by 6 to 8 hr p.i. the cell lyses and releases progeny viruses (~1,000 particles/cell).

Thus, PBCV-1 transcription is temporally programmed. Genes defined as "early" are transcribed within 5-60 min p.i.; some of the earliest transcripts form in the absence

of *de novo* protein synthesis [8] (Yanai-Balser et al., unpublished results). Transcripts of genes defined as "late" begin to appear 60-90 min p.i.; their appearance probably requires translation of early viral genes. However, some early gene transcripts can also be detected in later stages of infection. The PBCV-1 genes are not spatially clustered on the genome by either temporal or functional class. Therefore, temporal regulation of transcription must occur via *cis*- and possibly *trans*-acting regulatory elements.

The purpose of the current study is to identify conserved DNA sequences that might be involved in activation and regulation of viral transcription by using bioinformatic procedures. We identified three conserved nucleotide sequences that appear within 150 nucleotides of the ATG translation start codon of many virus open reading frames (ORFs). One of these motifs is associated predominately with early viral gene transcription and is likely to serve as a promoter for early genes.

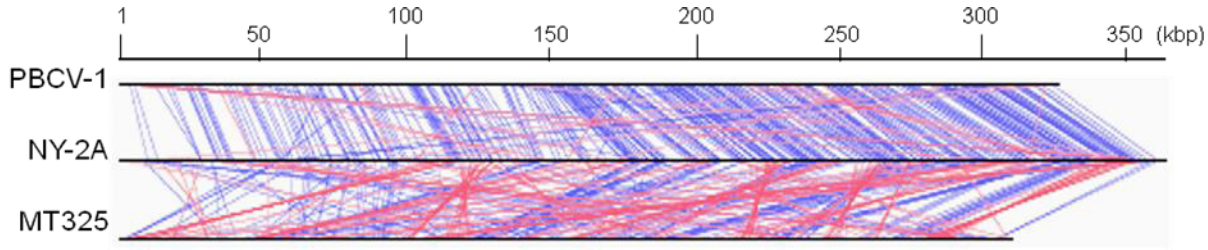


Figure 1. Genomic locations of homologous genes between NY-2A and either PBCV-1 or MT325. *When a homologous gene is detected between NY-2A and another genome a line is drawn. If the gene is transcribed in the same direction the line is blue. If the gene is transcribed in the opposite direction the line is red.*

2. MATERIALS AND METHODS

2.1. Bioinformatics. The genome sequences and annotations for viruses PBCV-1, NY-2A, and MT325 are available from GenBank under accession numbers U42580, DQ491002, and DQ491001, respectively. The same material is also located at http://greengene.uml.edu. For this study, the promoter region was defined as the region encompassing 150 nucleotides upstream of the ATG translation initiation codon and 50 nucleotides downstream of the ATG translation start codon.

AlignAce software [9] was used to identify conserved motifs in the promoter regions of the 366 PBCV-1 genes. Three conserved sequences were initially identified (Fig. 2). Two of the sequences were 10-mers and one was a 12-mer. These sequences were optimized and shortened (from each end) one base at a time by trial and error to generate the highest ratio of sequence hits in the promoter region relative to total sequence hits in the PBCV-1 genome (Table 1). The PBCV-1, NY-2A, and MT325 genomes were then searched for the occurrence of the three optimized sequences under complete stringency; the locations of the sequences within the promoter region were identified for each gene. The position of each sequence was then plotted with respect to the ATG translation initiation codon (Fig. 3). In addition, the PBCV-1 genome was searched for the three conserved sequences allowing one nucleotide mismatch and plotted.

2.2 RNA isolation. Infected chlorella cells (m.o.i. of 5) were collected at 20, 40, 60, 90, 120, 240, and 360 min p.i. Cells were disrupted with glass beads in the presence of Trizol (Invitrogen, Carlsbad, CA) and RNA was isolated using the Absolutely RNA

Miniprep kit (Stratagene, LaJolla, CA), according to the manufacturer's instructions. RNA integrity was verified in denaturing 1% agarose gels where intact host cytoplasmic and chloroplast rRNAs were visualized.

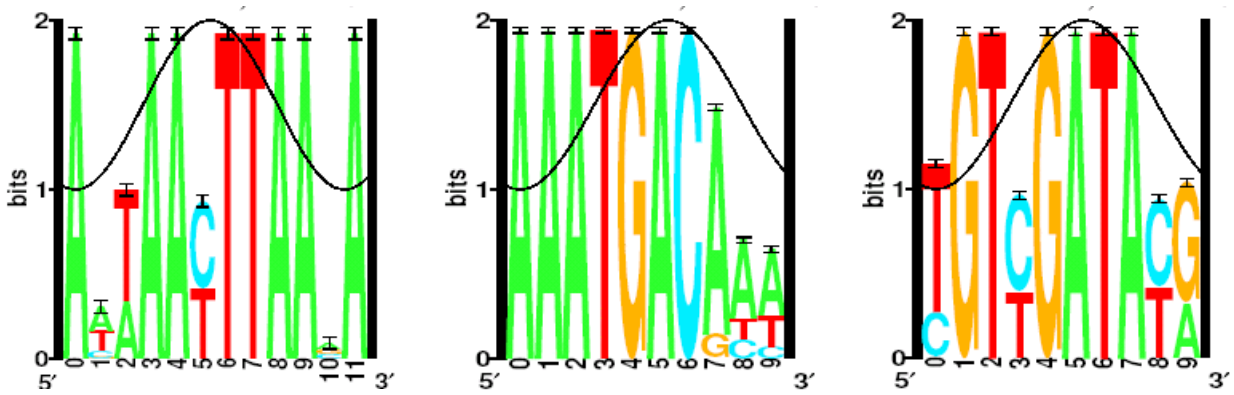


Figure 2. AlignAce results for the three conserved nucleotide sequences that frequently occur in the virus PBCV-1 promoter regions. *The black line indicates a potential groove of DNA.*

2.3 Microarrays fabrication and hybridization. A microarray containing 50-mer oligonucleotide probes representing each ORF in the PBCV-1 genome was constructed by MWG Biotech (Ebersberg, Germany) and the Microarray Core Facility (University of Nebraska Medical Center). For each time point, 20 mg of total RNA was reverse-transcribed using oligo(dT) as primers and cDNA was labeled with Cy3 or Cy5-dUTP (GE Healthcare, Piscataway, NJ) with the aid of a SuperScript Indirect cDNA Labeling System (Invitrogen, Carlsbad, CA) following the supplier's directions. Competitive

hybridization experiments were conducted for each time point against a pool of transcripts representing every gene isolated in the time course.

2.4 Microarrays analysis. Results from three independent biological hybridizations were analyzed using the GenePix Pro v.6.0 software (Molecular Devices, Sunnyvale, CA) and TIGR microarray software suite (TM4) [10]. Many transformations were performed to eliminate low quality data, to normalize the measured intensities using Lowess algorithm, and to regularize the standard deviation of the intensity of the Cy3/Cy5 ratio across the blocks. Genes that displayed statistically significant modulation were identified by a one-way analysis of variance, using P values of <0.01 as a cutoff. Genes with similar expression profiles were grouped into 10 different clusters using a K-means algorithm.

3. RESULTS and DISCUSSION

3.1. Viruses analyzed and criteria used to define genes and promoter regions in this study. The three chlorella viruses chosen for this study are PBCV-1, NY-2A, and MT325 that have 366, 404, and 329 putative protein-encoding genes, respectively. Approximately 80% of the genes are present in all three viruses. PBCV-1 and NY-2A infect the same host, *Chlorella* NC64A, and presumably are more closely related, in terms of evolutionary distance, to each other than to MT325, which infects *Chlorella* Pbi. However, the two NC64A viruses are among the most diverse of the NC64A viruses. The average amino acid identity between PBCV-1 and NY-2A homologs is ~75% [4], whereas the average amino acid identity between PBCV-1 and MT325 is ~50% [5].

Most PBCV-1 and NY-2A gene homologs are located co-linearly; in contrast, homologous genes in PBCV-1 and MT325 have almost no co-linearity with each other (Fig. 1). Thus, the promoter elements of the two NC64A viruses might be expected to be more similar to each other than between NC64A and Pbi viruses.

Table 1. Promoter motif optimization based on generating the motif that has the most occurrences in the promoter region relative to the total number of occurrences in the PBCV-1 genome.

Sequence	# Promoter region	Total # in genome	# Promoter region/ # genome
ARNTTAANA			
ANNAANYYAANA	78	208	0.38
ANNRRNYYAANA	99	344	0.29
NNRRNTTAANA	108	305	0.35
ARNTTAANA	98	222	0.44
RNTTAANA	165	563	0.29
NTTAANA	323	1278	0.25
AATGACA			
AAATGACRHH	89	182	0.49
ATGACRHH	107	240	0.45
AATGACRH	112	251	0.45
AATGACR	116	269	0.43
AATGACG	12	78	0.15
AATGACA	104	191	0.54
AATGAC	128	337	0.38
ATGACA	127	403	0.32
GTNGATAYR			
NGTNGATANN	69	329	0.21
GTNGATAYR	51	182	0.28
TYGATAYR	59	262	0.23
TNGATAYR	82	418	0.20
YGATAYR	103	604	0.17
YGTYGATAY	42	154	0.27
YGTNGATAY	48	194	0.25
YGTYGATA	45	189	0.24
YGTNGAT	84	513	0.16
TNGATAY	111	706	0.16

**Shown is the total number of times the motif occurred in the promoter regions. Note the same motif can occur multiple times within the same promoter region. For example, there are 91 unique genes that contain the promoter sequence ARNTTAANA; 84 of these genes contain the sequence one time and 7 genes contain the sequence twice.*

The following criteria were originally used to define genes in the three viruses: i) a minimal size of 65 codons initiated by an ATG codon, ii) when genes overlapped, the largest gene was chosen and iii) genes typically contain A+T-rich (>70%) regions in the 50 nucleotides upstream of the ATG translation start codon [1]. For this study, promoter regions were defined as encompassing a 200 bp region (150 bp upstream and 50 bp downstream of the ATG translation start site) of each viral encoded gene. However, the intergenic regions between PBCV-1 genes have an average size of 81 nucleotides with a standard deviation of 83 nucleotides (excluding the two-tailed 5% most extreme data points). In fact, 260 of the 366 PBCV-1 genes have less than 100 nucleotides between them. Using this definition, many of the putative viral promoter regions are located in an adjacent gene.

3.2. Three conserved sequences occur in the chlorella virus promoter regions. Using AlignAce software, three highly conserved nucleotide sequences were identified in the PBCV-1 promoter regions (Fig. 2). These sequences were optimized as described in the Materials and Methods section to generate three sequences that range in size from 7 to 9 nucleotides (Table 1); one or more degenerate positions occur in two of the three sequences. Some promoter regions contain more than one copy of either the same or different conserved sequences. As reported in Fig. 3, most of the sequences occurred in the -150 to 0 nucleotide region.

*Sequence **ARNTTAANA.*** The sequence ARNTTAANA occurs in the promoter region in 91 of the 366 PBCV-1 genes (25%), in 90 of the 404 virus NY-2A genes (22%), and in 40 of the 329 MT325 genes (12%) (Table 2). Relative to the entire genome, this sequence is present within the 200-nucleotide promoter region 44% of the time in PBCV-1, 49% of the time in NY-2A, and 37% of the time in MT325. Furthermore, the location of the sequence is biased to nucleotide position -15 to -45, relative to the ATG translation start codon (64% in PBCV-1, 66% in NY-2A, and 65% in MT325) (Fig. 3A). Thus the region between nucleotides -15 and -45 is a hotspot for the ARNTTAANA sequence.

*Sequence **AATGACA.*** The sequence AATGACA occurs in the promoter region in 60 of the 366 PBCV-1 genes (16%), in 74 of the 404 NY-2A genes (18%), and 25 of the 329 MT325 genes (8%) (Table 2). Relative to the entire genome, this sequence is present within the 200-nucleotide promoter region in 54% of the PBCV-1 genes, 53% of the NY-2A genes, and 25% of the MT325 genes. Furthermore, the AATGACA sequence is biased to nucleotide position -60 to -90, relative to the ATG initiation codon (44% in PBCV-1, 37% in NY-2A, and 33% in MT325) (Fig. 3B). These results indicate that the region between nucleotides -60 and -90 is a hotspot for the AATGACA sequence. This sequence resembles the consensus -35 element (TTGACA) in *E. coli* promoters.

*Sequence **GTNGATAYR.*** The sequence GTNGATAYR occurs in the promoter region in 49 of the 366 PBCV-1 genes (13%), 58 of the 404 NY-2A genes (14%), and 36 of the 329 MT325 genes (11%) (Table 2). Relative to the entire genome, this sequence is found

specifically within the 200-nucleotide promoter region in 28% of the PBCV-1 genes, 22% of the NY-2A genes, and 21% of the MT325 genes. The location of the sequence is biased to nucleotide positions -50 to -80, relative to the ATG initiation codon (39% in PBCV-1, 38% in NY-2A, and 70% in MT325) (Fig. 3C). These results indicate that the region between nucleotides -50 and -80 is a hotspot for the GTNGATAYR sequence.

Table 2. General characteristics of three conserved, putative promoter elements in three chlorella viruses.

	Percentage of genes with motif in promoter region			Ratio of motif hits in the promoter region to total hits in the genome			Predicted promoter location	Percentage of promoter motifs found within predicted promoter location		
	PBCV-1	NY-2A	MT325	PBCV-1	NY-2A	MT325	(nt from ATG start)	PBCV-1	NY-2A	MT325
ARNTTAANA	25%	22%	12%	0.44	0.49	0.37	−15 to −45	64%	66%	65%
AATGACA	16%	18%	8%	0.54	0.53	0.25	−60 to −90	44%	37%	33%
GTNGATAYR	13%	14%	11%	0.28	0.22	0.21	−50 to −80	39%	38%	70%

3.3. Occurrence of conserved sequences in PBCV-1 allowing a one base mismatch. The presence of these three conserved sequences in the promoter regions was also determined with one base mismatch in PBCV-1. Under complete stringency, 48% of the 366 PBCV-1 gene promoter regions contain at least one of the three conserved sequences. With one base-pair mismatch, ARNTTAANA occurs in 306 (84%) of the gene promoter regions, AATGACA occurs in 204 (56%) of the gene promoter regions, and GTNGATAYR occurs in 155 (42%) of the PBCV-1 366 promoter regions. [Note: some of the genes have two sequences located one or more times in the same promoter region (supplement 1).] Allowing a one base mismatch, one of these three motifs is present in all but 15 of the 366 PBCV-1 promoter regions. The locations of the

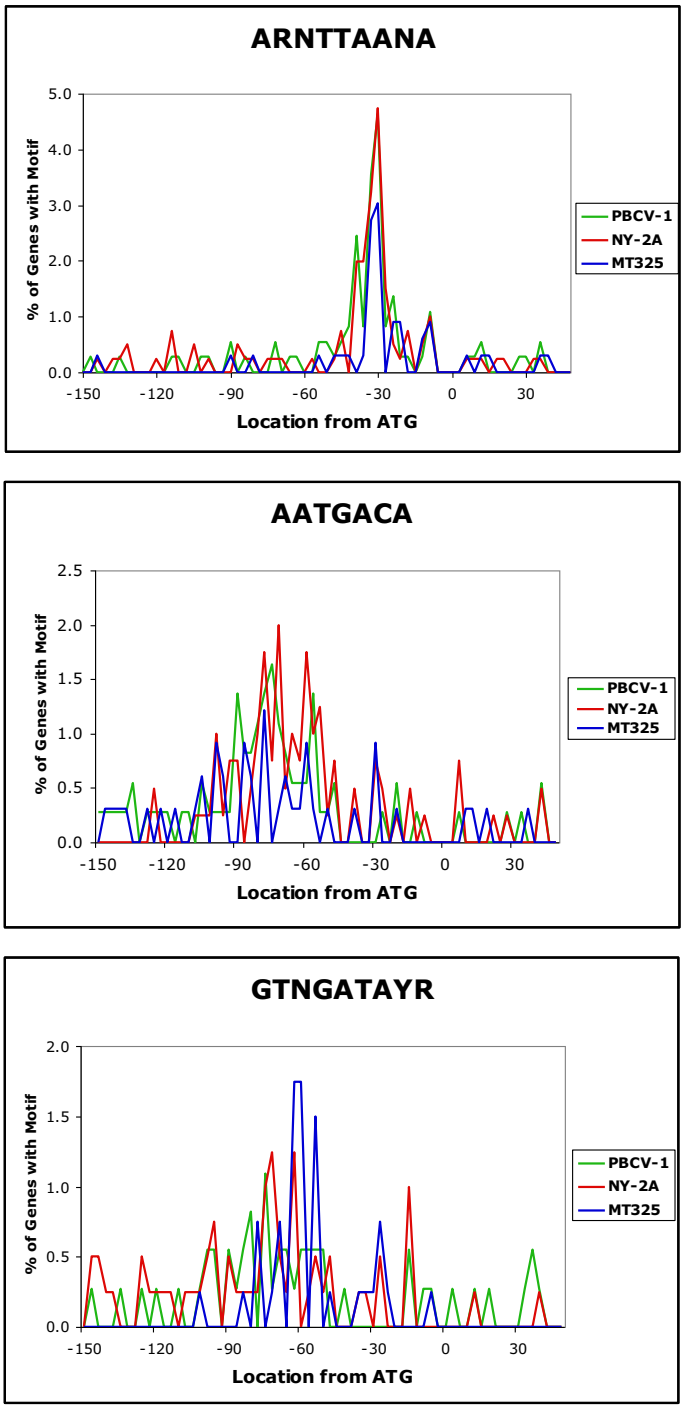


Figure 3. The positional distributions of three conserved nucleotide sequences in the gene promoter region of chlorella viruses PBCV-1, NY-2A, and MT325 with respect to the ATG translation start codon.

sequences with one base pair mismatch relative to the ATG translation start codon are similar to the locations under complete stringency (results not shown).

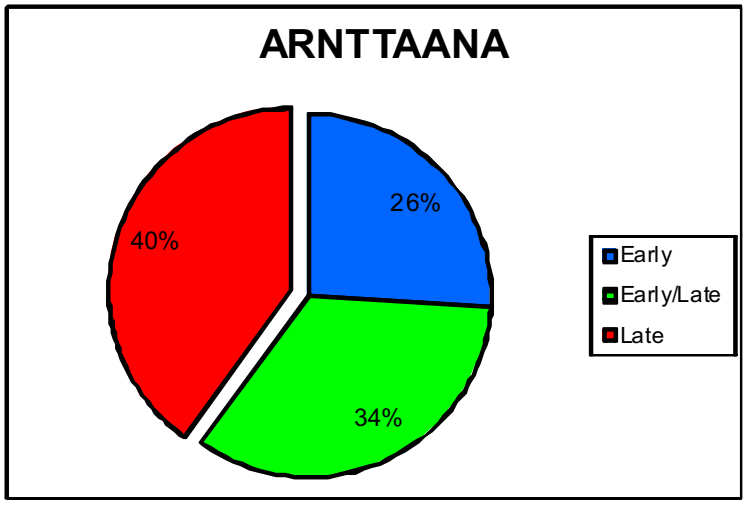
3.4. The conserved motifs are not specific to direction or location within the genome. None of the three conserved sequences exhibit a preference for direction or location within the three viral genomes. This finding is not surprising because genes classified as early or late occur throughout the PBCV-1 genome (Yanai-Balser et al., unpublished results) and they are approximately equally positioned in both orientations.

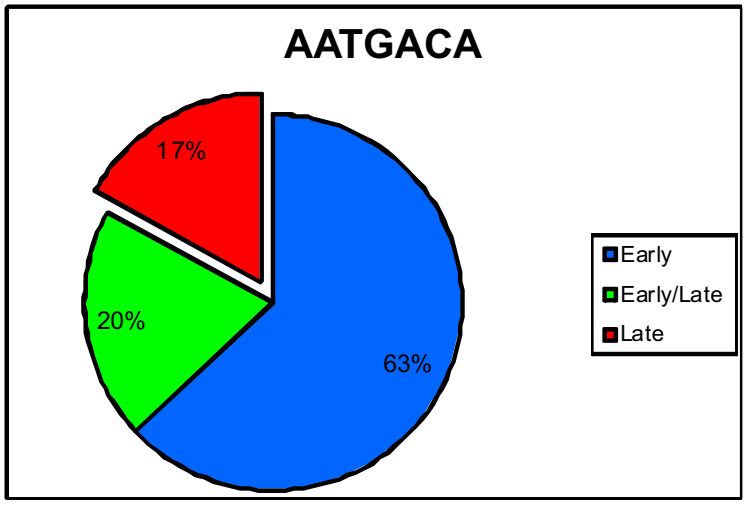
3.5. Homologous virus genes often share similar motif patterns at conserved locations. Viruses PBCV-1, NY-2A, and MT325 share many homologs (~80% of the genes are conserved among the three viruses). Therefore, we examined the occurrence of the conserved sequences among homologs in the three viruses. Homologous gene products often share similar motif patterns at conserved locations relative to the ATG translation start site (supplement 1). For example, a putative VLTF2-type transcription factor is a gene product encoded by all three viruses. Homologs in each of the viruses have the same motif (ARNTTAANA) in a similar location (-32, -31, and -33 nucleotides from the ATG translation start codon in viruses PBCV-1, NY-2A and MT325, respectively). Furthermore, if a specific motif occurs outside of the expected promoter region (*e.g.* outside of the -50 and -80 region for GTNGATAYR), homologous genes contain the motif in a similar location. For example, a gene encoding a putative PBCV-1 replication factor C protein subunit (*a417l*) contains the GTNGATAYR sequence

beginning at -146. The homologous gene in NY-2A contains the same sequence beginning at nucleotide –142. This sequence is not present in the MT325 replication factor C gene homolog.

3.6. Conserved sequences in the promoter region of homologous genes often contain identical nucleotides at degenerate positions. In addition to conserved sequences in their promoter regions, homologous genes often have similar nucleotide preferences at degenerate nucleotide positions within those sequences. Two of the three conserved nucleotide sequences (ARNTTAANA and GTNGATAYR) have degenerate nucleotide positions, and conserved nucleotide preferences occur among homologs for each of these two sequences. For example, the promoter region of the ribonucleotide reductase large subunit gene (*a629r* in PBCV-1 and *b832r* in NY-2A) contains the sequence GTNGATAYR, a sequence with three degenerate nucleotide positions. At the 'N' position, the PBCV-1 and NY-2A homologs contain a cytidine residue. At the 'Y' position, both viral genomes contain a cytidine residue and at the 'R' position, both viral genomes contain an adenine residue (Table 3). Not surprisingly, the nucleotide conservation at degenerate positions is more frequent between PBCV-1 and NY-2A than between either of these viruses and MT325.

3.7. Motif AATGACA is strongly associated with PBCV-1 early gene expression. To determine if there is a correlation between time of expression and the presence of a putative promoter sequence in the PBCV-1 genes, we constructed a microarray


ARNTTAANA
26%
40%
34%
Early
Early/Late
Late


AATGACA
17%
20%
63%
Early
Early/Late
Late

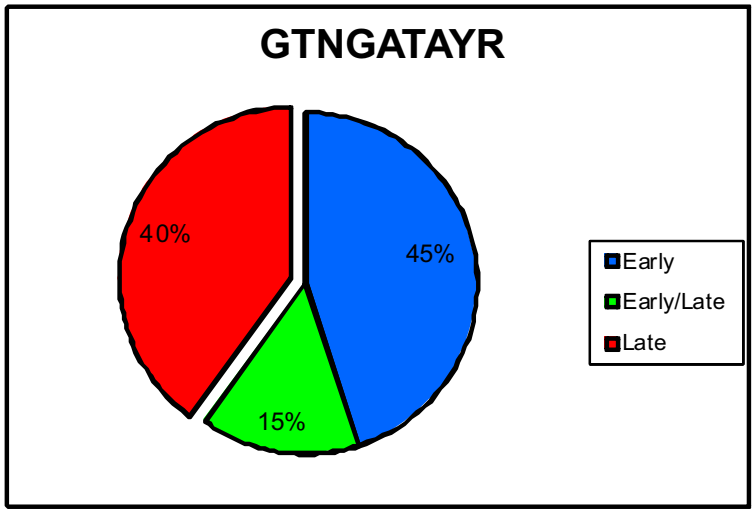

GTNGATAYR
40%
45%
15%
Early
Early/Late
Late

Figure 4. Distribution of three putative promoter motifs relative to when PBCV-1 genes are expressed. *Transcripts of genes classified as early/late are detected prior to the beginning of DNA replication and are present after DNA synthesis begins.*

containing probes from each gene in the genome. Competitive hybridization experiments were conducted employing cDNA from poly A-containing viral RNAs obtained from cells at 20, 40, 60, 90, 120, 240, and 360 min p.i., which allows us to follow global transcription of PBCV-1 replication.

The microarray results established that PBCV-1 transcripts fall into two groups: (i) early genes (59%), expressed before 60 min p.i. (the beginning of DNA synthesis) and (ii) late genes (41%), expressed after 60 min p.i. However, transcripts of 42% of the early genes are also present at late times after infection, referred to as early/late genes in Fig. 4 (Yanai-Balser et al., unpublished results).

Most of the genes with the AATGACA sequence are expressed early during infection (83%); transcripts from 24% of these early genes are also present after virus DNA synthesis begins (Fig. 4). The remaining 17% of the genes containing the AATGACA sequence are expressed late.

The other two sequences, ARNTTAANA and GTNGATAYR, have a no correlation with expression time. Sixty percent of the genes with the sequence ARNTTAANA are transcribed early; transcripts from 56% of these early genes are also present after virus DNA synthesis begins. The remaining 40% of the genes with the ARNTTAANA sequence are expressed late. Likewise, 60% of the genes with the

sequence GTNGATAYR in the promoter region are transcribed early and 25% of these genes produce transcripts that are also detected late during infection. The remaining 40% of the genes containing the GTNGATAYR sequence are expressed late. However, since 60% of the total genes are expressed early and 40% of the total genes are expressed late, there is no correlation with time of expression and these two sequences.

Table 3. *Examples of homologous proteins of known function containing promoter motifs with conserved nucleotides at degenerate positions.*

	PBCV-1	NY-2A	MT-325
100% similar promoter motifs in at least two genomes			
Ribo. reductase (large)	GTCGATACA	GTCGATACA	–
6-Phosphofructokinase	AACTTAAGA	AACTTAAGA	–
GDP-D-mannose dehydrogenase	AACTTAACA	AACTTAACA	–
Glycosyltransferase	AACTTAAGA	AACTTAAGA	–
PCNA	AACTTAAGA	AACTTAAGA	–
RNA triphosphatase	AACTTAACA	AACTTAACA	AGCTTAACA
RNase III	AATTTAAGA	AATTTAAGA	AACTTAATA
TFIID	AATTTAAAA	AATTTAAAA	–
VLTF2-type transcription factor	AATTTAAGA	AATTTAAGA	AACTTAACA
Only one difference in a degenerate position in at least two genomes			
Replication factor C	GTCGATACG	GTCGATATG	–
dUTP pyrophosphatase	GTTGATACG	GTTGATATA	GTCGATATA
Coat protein-like	AACTTAAAA	AAATTAATA	AACTTAAGA
Fructose-2,6 bisphosphatase	AATTTAAAA	–	AACTTAAAA
Fucose synthase	AATTTAAGA	AACTTAAGA	–
Ubiquitin C-terminal hydrolase	AGCTTAACA	AGTTTAACA	–
UDP-glucose dehydrogenase	AGATTAACA	AATTTAACA	–
Two or more differences in degenerate positions			
Adenine DNA methylase	AAGTTAATA	AATTTAAAA	–
ATPase (AAA+ Class)	AAATTAATA	AATTTAAGA	–
ATPase (DNA repair)	AAATTAATA	–	AATTTAACA
Histidine decarboxylase	AAATTAATA	AATTTAAGA	–
Transposase	AACTTAAGA	AATTTAATA	AATTTAACA

Bold nucleotides represent non-degenerate positions. A (-) denotes either a homolog does not exist or a homologous protein does not contain the motif. Degenerate positions are as follows: N = A/C/G/T, R = A/G, and Y = C/T

3.8. Promoter elements in related viruses. This is the first attempt to identify promoter elements by bioinformatic procedures in the phycodnaviruses. However, two previous reports described conserved nucleotide sequences in promoter regions that are associated either with a single chlorella virus gene, a gene encoding a potassium ion channel protein [11], or with 23 immediate early expressed genes in chlorella virus CVK2 [12]. The motif identified in the immediate early genes by Kawasaki et al. (ATGACAA) is similar to a motif identified in this manuscript (AATGACA), which also correlated with early transcripts.

The phycodnaviruses probably share a common evolutionary ancestry with the poxviruses, iridoviruses, asfarviruses, and the mimivirus [13, 14, 15]. All of these viruses have nine gene products in common and at least two of these viral families have an additional 41 homologous ORFs [14]. Collectively, these large dsDNA viruses are referred to as nucleocytoplasmic large DNA viruses [13].

A bioinformatics study on mimivirus identified an eight-nucleotide sequence, AAAATTGA, which occurs in the putative promoter regions (-150 to 0) of 403 of the 911 (45%) mimivirus ORFs [16]. This element is specific to the mimivirus lineage and the authors suggest that the element may correspond to an ancestral promoter structure predating the radiation of the eukaryotic kingdom.

In the iridovirus, Chilo iridescent virus (CIV), 5 nucleotides (AAAAT) located between -19 and -15 have been described as essential for promoter activity [17]. Interestingly, this promoter sequence is not only in the putative promoter regions of other CIV genes but also in other iridoviruses. Conserved nucleotide sequences in the

promoter regions of the poxviruses [18] and the asfarvirus, African swine fever virus [19], have also been reported.

4. CONCLUSIONS

This study identified three conserved 7 to 9 nucleotide sequences that probably function as promoter elements in the chlorella viruses. One of these sequences is associated primarily with early viral gene transcription and is likely to serve as a promoter for early genes. One way to test these predictions is to place one or more of these suspected early gene promoter regions in front of a late virus gene and determine if the "late" gene is now expressed early. Unfortunately, these experiments are not possible at the present time because procedures for manipulating the chlorella virus genomes are lacking.

5. SUPPLEMENT

5.1. Supplement Table 1. Putative PBCV-1, NY-2A and MT325 genes are grouped by their functional categories. If a conserved nucleotide motif has been identified within the promoter region of the gene, the sequence and location are noted.

DNA Replication, Recombination and Repair									
	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
δ DNA polymerase	*a185r*			*b249r*			*m019l*		
Archaeo-eukaryotic primase	*a468r*			*b633r*	GTCGATATG	-85	*m664l*		
				b633r	AATGACA	-71			
PCNA	*a193l*	AACTTAAGA	-33	*b261l*	AACTTAAGA	-29	*m014r*		
	a574l			*b767l*			*m697l*		
Replication factor C	*a417l*	AGATTAAAA	-49	*b571l*	GTCGATATG	-142	*m430l*		
	a417l	GTCGATACG	-146						
RNase H	*a399r*			*b547r*			*m570l*	GTCGATATA	-61
Helicase-Superfamily III	*a456l*			*b623l*	GTCGATATG	-83	*m674r*		
DNA Topoisomerase II	*a583l*			*b781l*			*m546r*		
ATP-dependent DNA ligase	*a544r*	AACTTAATA	-28	*b734r*	GTCGATACA	-13			
				b734r	AATGACA	0			
				b734r	AATGACA	7			
ATPase (PP-loop)	*a554/556/557l*			*b744l*			*m389r*	AATGACA	37
ATPase (DNA packaging)	*a392r*			*b536r*			*m586l*		
Pyrimidine dimer-specific glycosylase	*a050l*			*b076l*			*m627l*		
Exonuclease	*a166r*	AACTTAACA	-29	*b214r*			*m215r*		

Transcription									
	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Transcription factor TFIIB	*a107l*	AATGACA	-86	*b154l*	AATGACA	-25	*m139l*		
Transcription factor TFIID	*a552r*	AATTTAAAA	-147	*b743r*	AATTTAAAA	-7	*m266r*	GTCGATATA	-78
				b743r	GTCGATATG	-73			
Transcription factor TFIIS	*a125l*			*b175l*			*m163l*	AGATTAATA	-33
							m163l	GTCGATATA	-62
VLTF2-type transcription factor	*a482r*	AATTTAAGA	-32	*b647r*	AATTTAAGA	-31	*m635r*	AACTTAACA	-33
Superfamily II helicase	*a153r*	AAATTAAGA	38	*b203r*	AACTTAATA	-9	*m201r*		
	a241r	AATGACA	-61	*b203r*	AAATTAAGA	38	*m225l*	GTCGATATA	-60
	a241r	AAATTAAAA	-8	*b316r*			*m372r*		
	a241r	GTCGATACG	-75	*b508r*					
	a363r								
mRNA guanylyltransferase	*a103r*	AATGACA	-56	*b148r*	AATGACA	-84	*m133r*	AGCTTAAAA	-44
RNA triphosphatase	*a449r*	AACTTAACA	-30	*b612r*	AACTTAACA	-36	*m399l*	AGCTTAACA	-41
Histone H3, Lys 27 methylase	*a612l*	AATGACA	7	*b268l*			*m727l*	AATGACA	10
	a612l	GTGGATATG	41	*b813l*					
SWI/SNF chromatin remodeling complex	*a189/192r*			*b258r*			*m015l*		
SWI/SNF helicase	*a548l*	AATGACA	-147	*b738l*	AATGACA	-70	*m272l*		
	a548l	AATGACA	-140	*b738l*	AATGACA	-63			
	a548l	AATGACA	-133						

	a548l	AATGACA	-126						
	a548l	AATGACA	-119						
	a548l	AATGACA	-112						
RNase III	*a464r*	AATTTAAGA	-7	*b628r*	AATTTAAGA	-27	*m672l*	AACTTAATA	13
							m672l	GTCGATATG	-27
Cytosine deaminase	*a200r*	AATGACA	-75	*b271r*	AATGACA	-20	*m010l*		
	a200r	AATGACA	0	*b271r*	AATGACA	0			

Sugar Manipulation

	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
D-lactate dehydrogenase	*a053r*						*m026l*		
GDP-D-mannose dehydratase	*a118r*	AACTTAACA	-28	*b163r*	AACTTAACA	-28			
				b163r	AATGACA	0			
Fucose synthase	*a295l*	AATTTAAGA	-22	*b395l*	AACTTAAGA	-28			
UDP-glucose 6-dehydrogenase	*a609l*	AGATTAACA	-101	*b465r*	AATTTAACA	-131	*m719l*		
Glucosamine synthetase	*a100r*			*b143r*			*m037r*		
Hyaluronan synthase	*a098r*						*m128r*		
Chitin synthase				*b139r*					
				b472r	AAATTAAAA	-16			
Cellulase precursor							*m354r*		
Polysaccaride deacetylase				*b469l*					
Glycosyltransferase	*a064r*			*b159r*			*m186r*	AATGACA	-28
	a111/114r	AAGTTAATA	-83	*b618r*			*m467r*	GTCGATATA	-78
	a219/222/226r			*b736l*	AACTTAAGA	-31	*m467r*	GTTGATACA	-52
	a473l	GTCGATATG	-88				*m491r*		
	a546l	AACTTAAGA	-90				*m721l*		
dTDP glucose pyrophosphorylase							*m174l*		

Lipid Manipulation

	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Acetyltransferase	*a254r*	GTTGATACG	-60						
N-acetyltransferase	*a654l*			*b853l*			*m758r*		
Glycerophosphoryl diesterase	*a049l*			*b075l*					
Lipoprotein lipase	*a402r*	AGCTTAACA	-32	*b550r*	AGCTTAACA	-16	*m564l*		
Lysophospholipase	*a271l*			*b354l*					
Patatin-like phospholipase	*a173l*			*b226l*			*m219l*		

Cell Wall Degradation

	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Chitinase	*a181/182r*	AATTTAAAA	-28	*b239r*	GTCGATACG	-56	*m085r*		
	a260r	GTCGATATG	-68				*m791r*		
	a260r	GTTGATATG	-57						
Chitosanase	*a292l*			*b393l*			*m091r*		
β & α 1,4 linked glucuronic lyase	*a215l*			*b288l*			*m289r*	GTCGATATA	-101
				b468r					
β-1,3-glucanase	*a094l*	AATGACA	-90	*b137l*	AATGACA	-14	*m124l*		
	a094l	AATGACA	-83	*b137l*	AATGACA	0			
	a094l	AATGACA	-76						
	a094l	AATGACA	-69						
	a094l	AATGACA	-62						
	a094l	AATGACA	-55						

Signaling

	PBCV-1	NY-2A	MT325

	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Aquaglyceroporin							*m030r*		
Potassium channel protein	*a250r*			*b336r*			*m183r*		
Ligand-gated channel protein	*a163r*	AGTTTAAGA	-36						
Glutamate receptor	*a162l*								
Ca2+ transporting ATPase							*m535l*		
Dual specificity phosphatase	*a305l*			*b430l*			*m313l*	AAATTAAAA	-28
Serine/Threonine protein kinase	*a034r*			*b331r*	GTCGATATG	-13	*m143l*	AATTTAATA	-21
	a248r	AATGACA	-80	*b331r*	AATGACA	0	*m221l*		
	a277l			*b331r*	AATGACA	7	*m543r*	AATGACA	-76
	a278l			*b365l*	AAATTAACA	-55	*m729l*	AAATTAAAA	-23
	a282l			*b368l*			*m729l*	AATTTAATA	-9
	a289l			*b388l*	GTCGATACG	-74	*m729l*	AATGACA	-139
	a614l			*b816l*			*m733r*		
	a617r			*b818r*			*m794r*		

Integration and Tranposition

	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Transposase	*a625r*			*b080l*	AATTTAACA	-37			
				b080l	AATGACA	-92			
				b334r	AATGACA	-73			
				b334r	AATGACA	0			
				b378l	AATTTAACA	-37			
				b378l	AATGACA	-92			
				b702l	AATGACA	0			
				b711l					
				b829r	AATTTAAAA	-144			
Resolvase				*b083l*	AATTTAACA	-42			
				b083l	AATGACA	-97			
				b381l	AATTTAACA	-42			
				b381l	AATGACA	-97			
				b715l	AATTTAACA	-42			
				b715l	AATGACA	-97			
Tlr 6Fp DNA mobile protein	*a121r*			*b168r*	AAATTAATA	-25	*m150r*	AACTTAACA	-11
GIY-YIG endonuclease	*a134l*	AATTTAAGA	-37	*b039l*			*m032l*		
	a287r	AATGACA	-80	*b185l*	AACTTAAGA	-36	*m111r*		
	a315l			*b185l*	GTCGATACA	-47			
	a351l	AATGACA	-80	*b206l*			*m180l*		
	a351l	AATGACA	-73	*b246r*			*m288l*	GTCGATATA	-60
	a351l	GTTGATATA	-100	*b286r*	AAATTAATA	-85	*m370r*	AATGACA	-143
	a495r			*b286r*	GTCGATACG	-71	*m370r*	AATGACA	-123
	a539r			*b286r*	AATGACA	-58	*m370r*	AATGACA	-103
	a651l	AATGACA	-64	*b346l*	GTCGATACA	-145	*m370r*	AATGACA	-83
	a651l	AACTTAAAA	-38	*b346l*	AATGACA	-77	*m460r*	AGGTTAACA	41
				b389r	AATGACA	0	*m465l*	AATTTAACA	-90
				b389r	AATGACA	7	*m465l*	GTCGATATA	-61
				b433l			*m502l*		
				b440l			*m517r*	AATGACA	-62
				b497l	AATTTAAAA	-132	*m683l*		
				b497l	GTCGATATA	-142			
				b499l	GTCGATATA	-99			
				b602l					
				b629r	AATGACA	-46			
				b724r	AATTTAATA	-47			
				b724r	AATGACA	-53			
				b850l	GTCGATATG	-99			
HNH endonuclease	*a087r*	AATGACA	-26	*b022r*			*m069r*		

	a267l			*b061r*	AATTTAATA	-29	*m093l*		
	a354r			*b133r*	AATGACA	-77	*m578r*	GTCGATATA	-60
	a422r			*b165r*			*m622l*	GTCGATATA	-59
	a478l			*b173l*	AATGACA	-37			
	a490l			*b173l*	AATGACA	-25			
				b199r	GTCGATACA	-71			
				b199r	AATGACA	-58			
				b218r	AATGACA	-55			
				b324l					
				b370l	AATGACA	-70			
				b370l	AATGACA	-63			
				b424l					
				b446r	AATGACA	-55			
				b598l					
				b718l					
				b747l					
				b753l	AATGACA	-81			
				b798r	AAATTAATA	-82			
				b798r	AATGACA	-62			
				b805r					
				b878l					

Protein Synthesis, Modification and Degradation

	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Translation elongation factor-3	*a646l* *a666l*	AAGTTAATA	29				*m375r* *m742r*	GTTGATATA	-27
Prolyl-4-hydroxylase	*a085r*	AATGACA	0	*b126r*	AATGACA	-29	*m401r*	AATGACA	-29
Thiol oxidoreductase	*a465r*			*b630r*			*m670r*		
Protein disulfide isomerase	*a448l*	AATGACA	-55	*b611l* *b611l*	GTTGATATA AATGACA	-88 -66	*m403r*		
SKP-1 protein	*a039l*	AATGACA	-76	*b068l* *b068l*	AATGACA AATGACA	-77 -64	*m807r*		
Ubiquitin				*b699l*					
Ring finger ubiquitin ligase	*a481l* *a481l*	AAATTAAAA AACTTAAGA	-88 -29	*b645l* *b645l*	AAATTAAAA AACTTAAGA	-45 -28	*m629l*	GTCGATATA	-59
Ubiquitin C-terminal hydrolase	*a105l*	AGCTTAACA	-31	*b150l*	AGTTTAACA	-29	*m137l*		
Zn metallopeptidase	*a521l* *a604l* *a604l* *a604l*	AACTTAATA AATGACA AATGACA GTTGATATA	-44 -96 -89 -111	*b685l* *b803l* *b803l* *b803l*	 AAATTAAGA GTCGATATA AATGACA	 -86 -73 -59	*m496l*	AATTTAATA	-20
Initiation factor 2							*m488l* *m489r*	AATGACA	12

Nucleotide Metabolism

	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Aspartate transcarbamylase	*a169r*	GTCGATACG	-80	*b222r*	AATTTAACA	-28			
Ribo. Reductase (small subunit)	*a476r*			*b641r*			*m653l*	AATGACA	-97
Ribo. Reductase (large subunit)	*a629r* *a629r*	AATTTAAAA GTCGATACA	-133 -96	*b832r*	GTGGATACA	-105	*m777l*	AATGACA	-66
Deoxynucleoside kinase	*a416r*						*m425r*		
dCMP deaminase	*a596r*	AATTTAAAA	-29	*b795r*	AATTTAACA	-33	*m530l*		
dUTP pyrophosphatase	*a551l*	GTTGATACG	-95	*b741l* *b741l*	AAATTAAAA GTTGATATA	-114 -75	*m264l*	GTCGATATA	-78
Thymidylate synthase X	*a674r*	AATGACA	-84	*b865r*	AATGACA	-78	*m034l*	GTCGATATA	-84

Glutaredoxin	*a438l*			*b592l*			*m241l* *m423r*	AGCTTAAAA	-29
Thioredoxin	*a427l* *a427l*	AATGACA AAATTAACA	-47 32	*b581l*	AATGACA	-47	*m445l* *m448l* *m449l*	GTCGATATG AATTTAAGA	-61 -30

DNA Restriction/Modification									
	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Adenine methyltransferase	*a251r* *a581r*	AAGTTAATA	-55	*b010r* *b016l* *b230l* *b236l* *b359r* *b399r* *b416r* *b418r* *b543l* *b567l* *b774r*	AAGTTAAAA AATTTAAAA	34 -34			
Cytosine methyltransferase	*a517l* *a530r* *a683l*	AGTTTAAGA	-29	*b008r* *b088r* *b411l* *b566r* *b681l* *b681l* *b697r* *b769r*	AATTTAAGA AACTTAACA AACTTAACA AATTTAAGA AATGACA GTCGATATG	7 -27 19 -44 -60 -54	*m359l*		
DNA restriction endonuclease	*a252r* *a579l* *a579l* *a579l*	AATGACA AATGACA AACTTAAAA GTTGATATA	0 -17 -29 -40	*b361r* *b542r*					

Miscellaneous									
	PBCV-1			NY-2A			MT325		
	Gene	Sequence	Location	Gene	Sequence	Location	Gene	Sequence	Location
Ornithine/Arginine decarboxylase	*a207r*			*b278r*			*m307l*		
Agmatine iminohydrolase	*a638r*			*b844r*			*m766l*		
N-carbamoylput. amidohydrolase	*a078r*	GTTGATATG	-75	*b116r*			*m103l*		
Homospermidine synthase	*a237r*			*b305r*			*m233l*		
Histidine decarboxylase	*a598l*	AAATTAATA	-51	*b796l* *b796l*	AATTTAAGA GTCGATATA	-33 -70	*m601l*		
Monoamine oxidase	*a217l* *a217l* *a217l* *a217l*	AATGACA AATGACA AATGACA AATGACA	-20 -10 0 33	*b289l* *b289l*	AATGACA AATGACA	-29 0	*m283r*		
Amidase	*a284l*			*b371l*			*m101r*		
Cu/Zn-superoxide dismutase	*a245r*	AAATTAACA	-114				*m099r*		
O-methyltransferase	*a061l*								
FkbM Methyltransferase				*b183l*			*m177l*		
ABC transporter protein	*a445l*			*b606l*	GTCGATACG	-147	*m404r*	AATTTAATA	-33
ATPase (SequenceA+ class)	*a044l* *a044l*	AAATTAATA AACTTAAAA	-40 -32	*b073l* *b073l*	AATTTAAGA GTTGATACA	-29 -126			
Fibronectin binding protein	*a180r*	AAATTAATA	-8				*m789l*		

6. REFERENCES

1. Li, Y., Lu Z., Sun L., Ropp S., Kutish G.F. 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.
2. Van Etten, J. L. 2003. Unusual life style of giant chlorella viruses. *Ann. Rev. Genetics* 37:153-195.
3. Yamada, T., Onimatsu H., Van Etten J.L. 2006. Chlorella viruses. *Adv Virus Res* 66:293-336.
4. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Nierman W.C., Van Etten J.L. 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A. *Virology* 358:472-84.
5. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Hartigan J., Van Etten J.L. 2007. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect *Chlorella* Pbi. *Virology* 358:459-471.
6. 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.
7. Onimatsu, H., Suganuma K., Uenoyama S., Yamada T. 2006. C-terminal repetitive motifs in Vp130 present at the unique vertex of the chlorovirus capsid are essential for binding to the host chlorella cell wall. *Virology* 353:433-442.
8. 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.

9. Roth, F.P., Hughes J.D., Estep P.W., Church G.M., 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat. Biotechnol.* 16:939-945.

10. Saeed, A.I., Sharov V., White J., Li J., Liang W., Bhagabati N., Braisted J., Klapa M., Currier T., Thiagarajan M., Sturn A., Snuffin M., Rezantsev A., Popov D., Rylstov A., Kostukovich E., Borisovsky I., Liu Z., Vinsavich A., Trush V., Quackenbush J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374-378.

11. Kang, M., Graves M., Mehmel M., Moroni A., Gazzarrini S., Thiel G., Gurnon J.R., Van Etten J.L. 2004. Genetic diversity in chlorella viruses flanking kcv, a gene that encodes a potassium ion channel protein. *Virology* 326:150-159.

12. Kawasaki, T., Tanaka M., Fujie M., Usami S., Yamada T. 2004. Immediate early genes expressed in chlorovirus infections. *Virology* 318:214-223.

13. 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.

14. Iyer, L.M., Balaj S., Koonin E.V., Aravind L. 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res.* 117:156-184.

15. 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.

16. Suhre, K., Audic S., Claverie J.M. 2005. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. *Proc. Natl. Acad. Sci. USA* 102:14689-14693.

17. Nalcacioglu, R., Ince I.A., Vlak J.M., Demirbag Z., van Oers M.M. 2007. The Chilo iridescent virus DNA polymerase promoter contains an essential AAAAT motif. *J. Gen. Virol*. 88:2488-2494.

18. Moss, B. 2007. Poxviridae: the viruses and their replication, p. 2905-2946. *In* P. M. H. D.M. Knipe, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, S.E. Straus (eds.), Fields Virology, 5th Ed., WoltersKluwer/Lippincott Williams & Wilkins, Philadelphia.

19. Garcia-Escudero, R., Vinuela E. 2000. Structure of African swine fever virus late promoters: requirement of a TATA sequence at the initiation region. *J. Virol* 74:8176-8182.

CHAPTER IV

CONCLUSIONS AND FUTURE PROSPECTS

1. OVERVIEW

Only recently, the role of viruses in the ecology of aqueous environments has begun to be appreciated [1, 2, 3, 4]. Viruses are major players in the mortality of aquatic microorganisms and, consequently, contribute to geochemical cycles as well as structure of microbial communities [4]. Research focusing on individual viruses infecting unicellular and multicellular algae has made progress in recent years, but the field of algal virology is still in its infancy [5, 6, 7, 8, 9, 10, 11]. Questions about the long-term effects of viral-mediated processes and their true biological potential in maintaining aquatic communities remain unanswered.

To help recognize the importance of viruses in freshwater, we investigate the virus PBCV-1 that infects the eukaryotic unicellular green alga, *Chlorella* NC64A. Since its discovery in the early 1980s [12], many fascinating characteristics about this virus have been found. For example, this virus has a large genome and it is predicted to encode ~365 proteins. PBCV-1 CDSs resemble genes from all domains of life. Some of the PBCV-1 encoded proteins have been independently characterized by producing recombinant proteins. There was, however, a big gap in the understanding of how, where, and when viral gene expression occurs and if the 365 CDSs were transcribed. It was also not clear how the expression of each CDS is regulated, and ultimately, how virus proteins

take control of the host and drive the successful infection resulting in host cell lysis and spread of progeny virion. In an attempt to find answers as to which and when PBCV-1 CDSs are expressed, we monitored virus transcription during the replication cycle in *Chlorella* NC64A by microarray analysis. This allowed us to group each CDS with respect to time of expression. The results from this study led to the following conclusions: (i) most PBCV-1 CDSs (99%) are expressed during the virus life cycle in laboratory conditions. However, many PBCV-1 CDSs are not essential for replication *in vitro* because PBCV-1 deletion mutants (with deletions up to ~37 kb) produce viable progeny [13]. Microarray analysis of other large DNA viruses also reveals expression of the majority of genes (e.g., poxviruses, iridoviruses). (ii) PBCV-1 temporally regulates the expression of each CDS and most late CDS expression depends on the onset of viral DNA synthesis. This pattern is common for other viral systems (e.g., poxviruses and iridoviruses). (iii) By classifying PBCV-1 CDSs according to time of expression and putative function, it was determined that many of the early CDSs were involved with virus DNA replication and many late CDSs were packaged in the virion.

This transcriptional study contributes to our understanding of PBCV-1 infection, but it also raises many questions that are mentioned below.

- **Do PBCV-1 CDSs have additional promoters?**

Although three sequences were found in many putative promoter regions of the PBCV-1 genome, only one of them (5'-AATGACA-3') correlates with early expression of PBCV-1 CDSs (Chapter III). However, many early CDSs do not possess this sequence and more importantly, no conserved putative promoter sequences were found for CDSs

that are expressed early/late and late stages of infection. In Chapter III, the search for putative promoter sequences is described and it primarily focused on a specific region (-150 to +50 from the translation start site), assuming promoter sequences are found within this range and also assuming they are ubiquitous. Rare promoter sequences are not easily identified since they cannot always be distinguished from random background noise using bioinformatics tools. Microarray results, as described in Chapter II, revealed that 35% of the PBCV-1 CDSs are transcribed early, 27% are transcribed early/late, and 37% are transcribed late. The earliest time point analyzed was 20 min p.i., however, not all early CDSs were detected at 20 min p.i. The same phenomenon was observed for late CDS, not all late viral transcripts were detected at the same time post-infection. The pattern is even more complicated for early/late CDSs, as their transcripts are detected at different times, and sometimes, the transcript is present at an early time, then disappears, and reappears at a late time. This suggests that there are additional promoters elements in each transcriptional class that have not been identified.

Because ~80% of the PBCV-1 CDSs are conserved in other chloroviruses, it may be possible to compare, for example, the promoter regions of all early PBCV-1 CDSs to the promoter regions of their orthologs. Assuming CDSs with the same function are transcribed at the same time during infection, and also assuming the promoter sequence is conserved, it may be possible to find putative signals in this comparison

- **Assuming PBCV-1 is dependent on host transcription machinery, do viral CDSs, promoters, and terminators resemble host genes?**

Now that the genome sequence of the host *Chlorella* NC64A is available and assuming virus gene expression depends on host transcriptional machinery, it will be interesting to search for the putative promoter sequences identified in PBCV-1 genome and determine whether they are also found in the *Chlorella* NC64A genome. Also, if putative promoter sequences are found in *Chlorella,* they should also be compared to the putative promoter regions in the PBCV-1 genome. An experiment using the yeast-two-hybrid system could be carried out to test if transcription factors from the virus or host bind to the putative promoter sequences.

- **Are "minor" CDSs important for PBCV-1 replication?**

All of the PBCV-1 CDSs described in this work were considered "major" CDSs; i.e., protein encoding genes. "Minor" CDSs are also present in the virus genome and they are mostly ignored. A CDS is classified as minor in PBCV-1 if: (i) it resides inside a major CDS, regardless of orientation; (ii) if it overlaps within a major CDS, regardless of orientation. However, mass-spectrometry analysis of virion-associated proteins (D. Dunigan, unpublished results) has revealed several (~12) minor CDSs expressed and found to be virion-associated. This indicates that some minor CDSs might have important roles in PBCV-1 replication. In addition, preliminary analysis of the transcriptome sequence of uninfected cells and PBCV-1 infected cells at 20 min p.i. reveals that some of the minor CDSs are expressed.

Minor CDS microarray probes were not used in this study, thus it would be interesting to check if they are expressed and when they are expressed. However, it is not

trivial to distinguish a minor from a major CDS especially when they are co-localized in the genome. Thus, the results from this type of experiment must be carefully interpreted.

- **Are predicted initiation and termination sites correct for PBCV-1 CDSs? Are there additional introns in PBCV-1 CDSs?**

In the microarray analysis, PBCV-1 CDSs were classified according to their expression time, i.e., when hybridization occurred with the corresponding 50 nucleotide-long probe. The probes were designed to be specific and they correspond to various sites within each known CDS. Thus, microarrays do not provide any information regarding initiation and termination sites or intron regions.

As mentioned above, transcriptome mapping of PBCV-1 infected cells at 20 min p.i. shows transcription of a few CDSs starting before the predicted start codon and bases after the termination codon. The mapping also reveals the presence of introns that were not previously identified. It will be important to repeat this experiment in order to confirm these results.

- **Are some PBCV-1 messages are polycistronic?**

We suspect that some of PBCV-1 messages are polycistronic because Northern blot analyses revealed RNA bands running larger than their expected sizes from denaturing agarose gels electrophoresis. However, a comprehensive characterization of PBCV-1 transcripts and processing is lacking. We expect most CDSs to be monocistronic because microarray analysis shows very few clusters with nearby genes showing the same expression pattern.

- **Where does PBCV-1 transcription occur?**

It has been assumed that PBCV-1 DNA and some virion-associated proteins are directed to the nucleus upon infection. This is in part because the virus lacks a recognizable RNA polymerase gene and the assumption is that the virus uses the host transcription machinery for expression of viral messages. Although a few viral proteins (e.g. restriction endonucleases) likely go to the nucleus, there is no direct evidence that the virus DNA goes to the nucleus.

The results obtained by the microarray analysis do not provide any evidence as to where transcription occurs. However, if it is shown that virus promoters are similar to host promoters, then this will probably indicate that the virus uses transcription machinery from the host and consequently goes to the cell nucleus.

- **How does PBCV-1 transcription time correlate to translation time?**

The presence of a transcript as indicated by the microarray may or may not directly correlate with the presence of translation products. Western blots of a few PBCV-1 proteins have shown a ~30 min lag after expression of the mRNA. However, translation of the majority of PBCV-1 CDSs have not been analyzed.

- **Are mRNAs packaged in the virion?**

Since virus transcription begins almost immediately after infection (within 2-5 min p.i.), it is possible that some messages required very early are packaged in the virion. It is possible to test this hypothesis using the PBCV-1 microarray. Disrupted virus

particles can be used for cDNA synthesis, and the product can later be fluorescently labeled and hybridized with the probes in the microarray. It will not be surprising if some mRNAs are packaged in the virion since viral messages can be detected as early as 5 min p.i. [14]. However, recent calculations of transcription rates *in vivo* reveal that transcription can occur as fast as 4.4 kb/min and the average PBCV-1 average CDS size is 0.9 kb, which means it could be synthesized in ~ 15 seconds.

- **How is viral mRNA degradation carried out during infection?**

PBCV-1 microarray analysis classified each viral CDSs according to the presence of the transcript in the time points investigated. The presence of specific transcription classes also indicates that the infected cell has a very efficient process for degrading virus mRNAs. It is not known how this is regulated by the virus.

2. REFERENCES

1. Dunigan D.D., Fitzgerald L.A., Van Etten J.L. 2006. Phycodnaviruses: a peek at genetic diversity. *Virus Res*. 117:119-132.

2. Allen M.J., Wilson W.H. 2008. Aquatic virus diversity accessed through omic techniques: a route map to function. *Curr Opin Microbiol*. 11:226-232.

3. Wilson WH, Van Etten JL, Allen MJ. 2009. The Phycodnaviridae: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol*. 328:1-42.

4. Suttle C.A. Marine viruses--major players in the global ecosystem. 2005. *Nat Rev Microbiol.* 5:801-812

5. Van Etten J.L. 2003. Unusual life style of giant chlorella viruses. *Annu Rev Genet.* 37:153-195.

6. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Nierman W.C., Van Etten J.L. 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect *Chlorella* NC64A. *Virology* 358:472-484.

7. Fitzgerald L.A., Graves M.V., Li X., Feldblyum T., Hartigan J., Van Etten J.L. 2007. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect Chlorella Pbi. *Virology* 358:459-471.

8. Fitzgerald L.A., Graves M.V., Li X., Hartigan J., Pfitzner A.J., Hoffart E., Van Etten J.L. 2007. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. *Virology* 362:350-361.

9. 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. 2005. Complete genome sequence and lytic phase transcription profile of a Coccolithovirus. *Science* 309:1090-1092.

10. Derelle E., Ferraz C., Escande M.L., Eychenié S., Cooke R., Piganeau G., Desdevises Y., Bellec L., Moreau H., Grimsley N. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri. PLoS One.* 3(5):e2250.

11. Nagasaki, K. 2008. Dinoflagellates, diatoms, and their viruses. *J Microbiol.* 46:235-243.

12. 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. USA* 79:3867-3871.

13. 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* 214:413–20.

14. 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.

APPENDIX I

CHLORELLA VIRUSES CONTAIN GENES ENCODING A COMPLETE POLYAMINE BIOSYNTHETIC PATHWAY

1. OVERVIEW

Polyamines are small cationic organic molecules of vital importance for all living organisms. The primary polyamines in eukaryotes are putrescine and the putrescine-derived spermidine and spermine (Fig. 1A). Putrescine is formed either directly by the decarboxylation of ornithine by ornithine decarboxylase (ODC) or indirectly from arginine by the sequential action of arginine decarboxylase (ADC), agmatine iminohydrolase (AIH, also known as agmatine deiminase), and N-carbamoylputrescine amidohydrolase (CPA). Both pathways occur in plants although the model plant Arabidopsis thaliana lacks an ODC gene [1] and is therefore completely dependent on the ADC pathway. Putrescine is converted to spermidine by spermidine synthase (SPDS) and S-adenosylmethionine decarboxylase; the latter enzyme produces decarboxylated S-adenosylmethionine (dcSAM) which serves as the aminopropyl donor for SPDS. Spermine is synthesized in a similar way from spermidine and dcSAM by spermine synthase. Although putrescine, spermidine,and spermine are the main polyamines in eukaryotes, many other polyamines have been identified in nature that are either derived from the putrescine pathway, such as 1,3-diaminopropane and homospermidine, or from other precursors,e.g. cadaverine is derived from lysine [2].

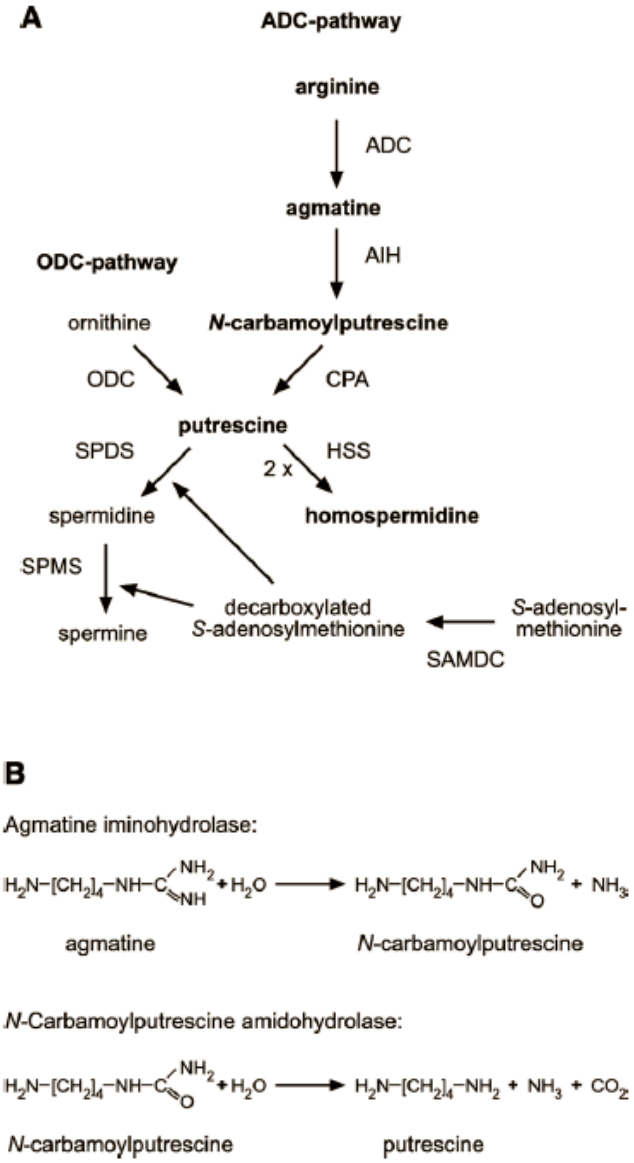


Figure 1. Polyamine biosynthesis pathways*. (A) Overview. (B) Details of the reactions catalyzed by agmatine iminohydrolase and N-carbamoylputrescine amidohydrolase. ADC, arginine decarboxylase; AIH, agmatine iminohydrolase; CPA, N-carbamoylputrescine amidohydrolase; HSS, homospermidine synthase, ODC, ornithine decarboxylase; SAMDC, S-adenosylmethionine decarboxylase; SPDS, spermidine synthase; SPMS, spermine synthase.*

Polyamines are not only essential for living cells butthey also occur in the capsids of many viruses and, in some cases, are important for virus multiplication (reviewed in [3, 4]). In other systems, host polyamines are important for resistance against viruses (reviewed in [4]).

The relationship between viruses and polyamines took a new turn with the discovery that chlorella virus *Paramecium bursaria* chlorella virus (PBCV-1) (family Phycodnaviridae) has genes encoding the two polyamine biosynthetic enzymes, ODC [5] and homospermidine synthase (HSS) [7]. HSS synthesizes the rare polyamine homospermidine from two molecules of putrescine (Fig. 1A). Although PBCV-1 ODC is able to directly convert ornithine to putrescine, its preferred substrate is arginine [8]; consequently the PBCV-1 enzyme is referred to as ODC/ADC.

PBCV-1, the prototype chlorella virus, has a 331-kb dsDNA genome that contains 366 putative protein-encoding genes and a polycistronic gene that encodes 11 tRNAs [9, 10]. PBCV-1 and another virus used in this study, NY-2A (370 kb genome), infect Chlorella NC64A (NC64A viruses), which is an endosymbiont of the protozoan P. bursaria that was originally isolated in North America. Other viruses, such as MT325 (314 kb genome) infect Chlorella Pbi (Pbi viruses), an endosymbiont of *P. bursaria* that was isolated in Europe [11, 12].

The genes for the polyamine biosynthetic enzymes AIH and CPA from A. thaliana were recently identified [12, 13, 14]. This identification was due to their similarity to the corresponding enzymes from *Pseudomonas aeruginosa* PAO1, which were identified previously [15, 16]. Database searches revealed that homologs of both enzymes are also encoded by PBCV-1. Here we report the characterization of these genes

and their products from PBCV-1, as wellas from chlorella viruses NY-2A and MT325. Our results establish that chlorella viruses encode a complete polyamine biosynthetic pathway that allows the formation of homospermidine from arginine via agmatine, N-carbamoylputrescine, and putrescine.

2. MATERIALS AND METHODS

2.1 Cloning of viral genes for AIH and CPA. Isolation of viral DNAs has been described [17]. The ORFs for the AIH and CPA proteins were amplified by PCR using the DyNAazyme EXT DNA polymerase (Finnzymes, Espoo, Finland). Each forward primer begins with an NdeI restriction site including an in frame start codon (CATATG), the reverse primers omit the stop codon and end with XhoI restriction sites for in frame cloning with the (His)6-tag of the pET-21b(+) vector (Novagen, Nottingham, UK). If XhoI sites were present in the ORF, XhoI-compatible SalI restriction sites were used instead. PCR products were T/A cloned into pGEM-T (Promega, Madison, WI). After establishing that the PCR products were free of error by sequencing (MWG Biotech, Ebersberg, Germany), the ORFs were cloned into the NdeI and XhoI sites of pET-21b(+).

The accession numbers for the virus genomes are U42580, DQ491002, DQ491003, DQ491001, and DQ890022 for viruses PBCV-1, NY-2A, AR158, MT325, and FR483, respectively. The DNA sequences can also be found at http://greengene.uml.edu.

2.1 Expression of viral proteins in *E. coli* and purification of the (His)6-tagged proteins. The proteins were expressed in E. coli strain CodonPlus (DE3) RIL

(Stratagene, La Jolla, CA). Pre-cultures were grown overnight in 2YT medium [18] containing ampicillin (100 μgml− 1)and chloramphenicol (30 μgml− 1)in an orbital shaker (220 rpm) at 37 °C. The expression cultures were inoculated 1:10 in 2YT medium (including ampicillin and chloramphenicol) with the pre-cultures. Optimized expression conditions to produce the recombinant proteins are described in the Supplementary Table S3.

Table1. Characteristics of AIH and CPA proteins from chlorella viruses in comparison to the enzymes from Arabidopsis thaliana and Pseudomonas aeruginosa PAO1

Enzyme	Length (amino acids)	Calculated MW (kDa)	Calculated IEP	Locus tag	Identity to *Arabidopsis*/ *Pseudomonas* enzymes (%)
AIH					
*At*AIH	383	43.1	4.92	At5g08170	100/60
*Pa*AIH	368	41.2	4.60	PA0292	60/100
PBCV-1 AIH	359	40.8	5.83	A638R	48/51
NY-2A AIH	359	40.9	5.46	B844R	47/50
MT325 AIH	363	41.1	5.42	M766L	45/51
CPA					
*At*CPA	299	33.5	5.93	At2g27450	100/63
*Pa*CPA	292	32.7	5.92	PA0293	63/100
PBCV-1 CPA	298	33.2	5.70	A78R	49/50
NY-2A CPA	298	33.2	5.42	B116R	49/51
MT325 CPA	296	32.9	7.04	M103L	47/47

Bacteria were harvested by centrifugation and stored frozen at −80 °C. For protein purification the cell pellets were resuspended in 1/10 of the culture volume of lysis buffer (50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 5 mM 2mercaptoethanol, 1 mg ml− 1 lysozyme) and incubated for 15 min on ice. Cell lysis was completed by sonification with an ultrasound tip for 4×1 min at 50 W (Sonifier B17, Branson, USA). After removing the insoluble material by centrifugation (Sorvall GSA rotor, 10,000 rpm, 10 min), the

supernatant was loaded onto a Ni2+-NTA agarose (Qiagen, Hilden, Germany) column (bed volume 2 ml), which was equilibrated in 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 10 mM imidazole. The column was washed with 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, 30 mM imidazole, and bound proteins were eluted in a total volume of 2.5 ml with 50 mM sodium phosphate, pH 8.0, 300 mM NaCl, and 250 mM imidazole. After desalting on PD-10 columns (GE Healthcare, Freiburg, Germany) into 50 mM potassium phosphate, pH 8.0 and 1 mM DTT (for the CPA proteins) or 50 mM potassium phosphate, pH 8.0 and 0.5 mM DTT (for AIH proteins), the purified recombinant proteins were frozen in 0.5 ml aliquots in liquid nitrogen and stored at −80 °C. Under these conditions the enzymes were stable for at least 2 months.

2.3 Enzyme assays. Enzymes assays were conducted in triplicates with a background control in which the protein was inactivated by boiling for 10 min. The typical CPA assay was performed in 50 mM MES, pH 6.0, 1 mM DTT, 1 mM N-carbamoylpu-trescine and 1µg ofprotein in a totalvolume of0.5 ml at 37 °C. The AIH assays consisted of 50 mM potassium phosphate, pH 7.5, 0.5 mM DTT, 1mM agmatine and1µg protein in a total volume of 1 ml and were incubated at 30 °C.

Activity was measured as the release of ammonia from the substrates (Fig. 1B). Ammonia was detected with the indophenol blue method: a 100 µl aliquot of sample was mixed with 100 µlof 0.33 Msodium phenolate, 0.02 Msodium hypochlorite (prepared fresh daily), and 0.01% (w/v) sodium nitroprusside. The reaction was incubated for 3 min in a boiling water bath and then 600 µlof water was added. The absorbance at 640 nm was determined and compared to a calibration curve made with NH4Cl.

Thin-layer chromatography was conducted on silica plates (PolygramSIL 201 G/UV254, Macherey-Nagel,Düren, Germany) that were developed in ethanol/NH4OH (25%) 1:1. About 5 to 10 µlof sample was applied to the plates. Amines were detected by spraying with ninhydrin (0.2% in ethanol).

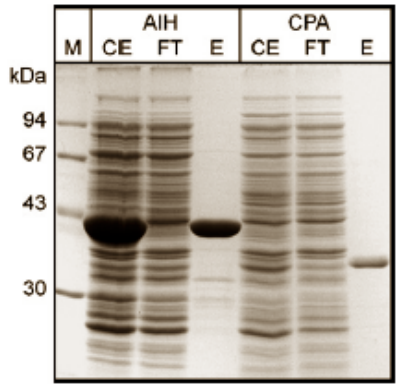


Figure 2. Expression and purification of (His)6-tagged PBCV-1 AIH and CPA from E. coli. *About 10 µl each of bacterial crude extracts (CE) and the column flowthrough (FT) were loaded on the gel. 2and 10 µlofthe AIH and CPAeluates (E) were loaded, respectively. M: low-molecular weight markers.*

2.4 RNA isolation. Infected chlorella cells (m.o.i. of 5) were collected at 20, 40, 60, 90, 120, 240, 360 min p.i. Cells were disrupted with glass beads in the presence of Trizol (Invitrogen, Carlsbad, CA) and RNA was isolated using an Absolutely RNA Miniprep Kit (Stratagene, LaJolla, CA) according to the manufacturer's instructions. RNA integrity was verified in denaturing 1% agarose gels where intact host cytoplasmic and chloroplast rRNAs were visualized.

2.5 Microarray and dot blot analyses. A microarray containing 50-mer probes representing each ORF in the PBCV-1 genome was constructed by MWG Biotech (Ebersberg, Germany) and the Microarray Core Facility (University of Nebraska Medical Center). Four replicas of the entire genome are present in each array. For each time point, 20 μgof total RNA was reverse-tr anscribed using oligo(dT) as primers and cDNA was labeled with Cy3 or Cy5-dUTP (GE Healthcare, Piscataway, NJ) with the aid of a SuperScript Indirect cDNA Labeling System (Invitrogen, Carlsbad, CA) following the supplier's directions. Competitive hybridization experiments were conducted for each time point against a pool of transcripts representing every gene present in the time course.

Results from three independent biological hybridizations were analyzed using the GenePix Pro v.6.0 software (Molecular Devices, Sunnyvale, CA) and TIGR microarray software suite (TM4) [19]. A number of transformations were performed in order to eliminate low quality data, to normalize the measured intensities using the Lowess algorithm, and to regularize the standard deviation of the intensity of the Cy3/Cy5 ratio across the blocks. Genes that displayed statistically significant modulation were identified by a one-way analysis of variance, using P values of <0.01 as a cutoff. Genes with similar expression profiles were grouped into 10 different clusters using a K-means algorithm.

Viral DNAs used for dotblots were denatured, applied to nylon membranes fixed by UV cross-linking, and hybridized with 32Plabelled gene probes as described previously [20].

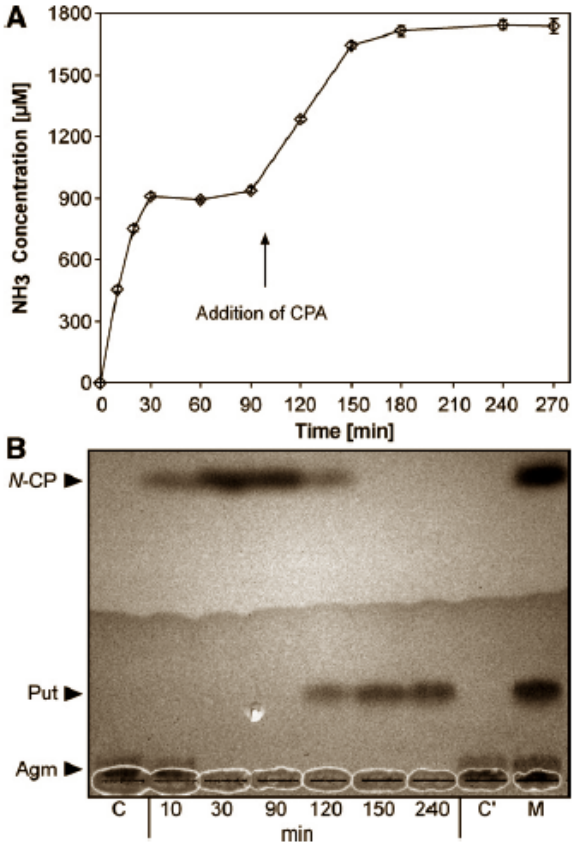


Figure 3. Enzymatic activities of PBCV-1 AIH and CPA. *The release of ammonia (A) as well as the turnover of the substrates to the products (B) were analyzed. Agmatine (1 mM) was incubated with PBCV-1 AIH (20 µg) for 90 min, then 15 µgof PBCV-1 CPAwas added. C, C': controls with heat denatured enzyme after 10 and 240 min incubation, respectively; M: standard mixture of 2mMeach of agmatine (Agm), N-carbamoylputrescine (N-CP) and putrescine (Put).*

2.6 Phylogenetic analysis. The amino acid sequences of PBCV-1 polyamine biosynthesis enzymes were used to search the "non-redundant" dataset of GenBank using

BLASTP [21]. All sequences showing similarities greater than 50% over the full sequence were collected (between 30 and 130 entries) and multiple entries from the same species were discarded. The remaining 30–70 sequences were aligned using the DNAMAN program (version 5.2.2, Lynnon Corp., Vaudreuil-Dorion, Canada) and sequences which obviously altered the alignment were discarded. A new alignment was produced with the remaining sequences, extremely diverse N-and C-terminal sequences were deleted, and neighbor-joining trees using Kimura distances and 1000 bootstrap replicates were constructed with the DNAMAN software. These trees were used to discard additional sequences in overrepresented monophyletic clades with high bootstrap support. To root the trees paralogous sequences were included (arginine deiminase for the AIH sequences and N-carbamoyl-β-alanine amidohydrolase for the CPA sequences). Again a new alignment was created with the remaining sequences (43 and 57 for the AIHs and CPAs, respectively). From these alignments phylogenetic inference was performed by Bayesian analysis using the MrBayes software (version 3.1.2; [22]) (gamma rates, model: mixed, 250,000 generations, burn-in: 25% of saved trees).

4. RESULTS

In the following sections we only show results obtained with recombinant AIH and CPA proteins from PBCV-1 genes; results obtained with recombinant AIH and CPA proteins from viruses NY-2A and MT325 will be mentioned in the text or, where appropriate, listed in the tables.

Table 2. Enzymatic properties of AIH and CPA proteins from chlorella viruses, Arabidopsis thaliana, and Pseudomonas aeruginosa PAO1

Enzyme	pH optimum	Temperature optimum (°C)	K_m (mmol L^{-1})	k_{cat} (s^{-1})
AIH				
PBCV-1 AIH	7.5	25	0.043	10.2
NY-2A AIH	7.5	30	0.054	9.7
MT325 AIH	7.5	20	0.054	–[a]
*At*AIH[b]	7	35–40	0.112	18.0
*Pa*AIH[c]	8.0	45	0.6	4.2
CPA				
PBCV-1 CPA	5.5–6.0	40	0.250	3.1
NY-2A CPA	5.5–6.0	45	0.216	2.8
MT325 CPA	6.5	30	0.214	–[a]
*At*CPA[d]	8–9	40	0.135	2.9
*Pa*CPA[c]	8.0	40	0.5	3.3

[a] Not determined (see text).

[b] From [13].

[c] From [16].

[d] From [14].

3.1 Identification of *aih* and *cpa* genes in chlorovirus genomes. BLAST [21] searches using the AIH and CPA sequences from A. thaliana revealed homologous genes encoding both enzymes in the virus PBCV-1 genome. The predicted gene products from PBCV-1 open reading frames (ORF) A638R and A78R have 48% and 49% amino acid identity to AtAIH and AtCPA, respectively (Table 1). Two other chlorella viruses have been sequenced recently [23, 24], virus NY-2A that infects the same host as PBCV-1 (Chlorella NC64A) and virus MT325 that infects Chlorella Pbi. Both of these viruses also contain putative AIH and CPA encoding genes (Table 1) as well as PBCV-1 ODC/ ADC and HSS homologs (Table 3). The PBCV-1 AIH has 97% and 63% amino acid identity to its NY-2A and MT325 homologs, respectively. Similar amino acid identities occur between PBCV-1 CPA and its homologs from NY-2A and MT325 (97% and 68%amino

acid identity, respectively). These values are similar to the amino acid identities of PBCV-1 ODC/ ADC (86% and 63%) and HSS (93% and 69%) to their respective NY-2A and MT325 homologs (see also Supplementary Fig. S1). The viral AIH proteins are 5 to 24 amino acids shorter than the enzymes from P. aeruginosa and A. thaliana, while all of the CPA proteins are of similar size.

3.2 Chlorella virus AIH and CPA proteins are functional enzymes. We amplified the aih and cpa genes from PBCV-1, NY-2A and MT325 genomes by PCR. Recombinant proteins were expressed in Escherichia coli with C-terminal (His)6-tags and were subsequently purified by metal affinity chromatography (Fig. 2). AIH proteins from PBCV-1 and NY-2A were obtained as soluble proteins in large amounts (up to 20 mg from 100 ml bacterial cultures), whereas the CPA proteins were expressed in lower amounts (up to 1 mg from 300 ml cultures). In contrast, only small amounts of soluble proteins were obtained from both MT325 genes (about 200 μg per 300 ml culture) because most of the expressed proteins appeared in inclusion bodies.

A colorimetric ammonia assay was used to test the recombinant proteins for activity because AIH and CPA each release ammonia from their presumed substrates in stoichiometric amounts with their products (Fig. 1B). In addition, thin-layer chromatography was used to verify the products of the two reactions as N-carbamoylputrescine (for AIH) and putrescine (for CPA). PBCV-1 AIH produced ammonia from agmatine, while at the same time N-carbamoylputrescine was formed (Figs. 3A, B, 0–90 min). This N-carbamoylputrescine was subsequently converted to putrescine with a concomitant release of more ammonia after addition of PBCV-1 CPA

(Figs. 3A, B, 90–240 min). Thus, PBCV-1 AIH and CPAwere both functional enzymes displaying their predicted activities. Similar results were obtained with recombinant proteins from

NY-2A and MT325 (results not shown). When PBCV-1 AIH was heat-denatured before assaying, no release of ammonia and no conversion of agmatine occurred, even after prolonged incubation (Fig. 3B, lane C′). This experiment also establishes that agmatine is not a substrate for PBCV-1 CPA.

3.3 Enzyme characterization. The properties of both enzymes (pH and temperature optima, m and kcat)were determined and the results are summarized in Table 2. The pH optima of the viral AIH enzymes (pH 7.5)were similar to those from A. thaliana and P. aeruginosa. In contrast, the viral CPA enzymes had highest activities in slightly acidic conditions (pH 5.5–6.5), when compared to the CPA enzymes from A. thaliana and P. aeruginosa (pH 8.0–9.0). The temperature optima for both enzymes from the Pbi virus MT325 (20 °C for AIH and 30 °C for CPA) were slightly lower than the two enzymes from the NC64A viruses and from A. thaliana and P. aeruginosa (Table 2).

The Km values were similar within each group of enzymes with the exception of the P. aeruginosa enzymes, which had higher Km values for both enzymes. The protein concentrations of the purified MT325 enzymes were so low that they could not be reliably determined; therefore, we did not calculate kcat values for these two enzymes. The kcat values for the other viral enzymes were similar to those obtained from A. thaliana and P. aeruginosa enzymes.

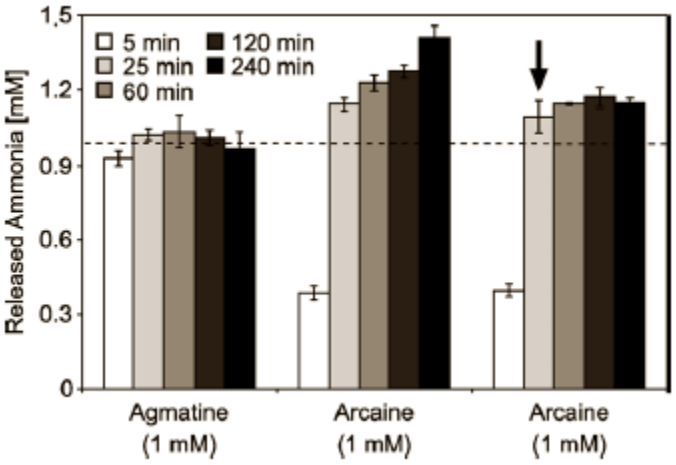


Figure 4. Release of ammonia from agmatine and arcaine by PBCV-1 AIH*. 40 μg of protein was incubated with 1mM of the substrates and the released ammonia was quantified at the indicated time points. In the second arcaine sample (third set of columns) the protein was heat inactivated after 30 min (indicated with an arrow). The dotted line shows the 1 mM level.*

3.4 Substrate specificities. We tested the structurally similar compounds arginine, arginine methyl ester, argininamide, and arcaine (1,4-diguani-dinobutane), which is a derivative of agmatine (1-amino-4-guandinobutane), as possible substrates for PBCV-1 AIH. Under standard conditions (0.5 μg protein/ml, 1mM substrate, 30 °C, 30 min), no ammonia was released from arginine, arginine methyl ester or argininamide. However, increased amounts of enzyme (80 μg/ml) and prolonged incubation (1 and 2 h), resulted in very weak activity with arginine methyl ester (0.2 nkat/mg protein). In contrast, arcaine was a relatively good substrate for the viral AIH enzymes (Fig. 4); the specific activity of PBCV-1 AIH with arcaine was only 4-fold less than with agmatine (results not shown). Arcaine has been reported to inhibit plant AIH enzymes [25, 26]. These authors

reported about 28% to 50% inhibition of maize AIH with 1 or 2 mM arcaine; 2 mM arcaine inhibited A. thaliana recombinant AIH by 39% (Janowitz and Piotrowski, unpublished results). Therefore, we were surprised that arcaine was a substrate for the viral AIHs. We tested the plant AtAIH and also observed release of ammonia from arcaine by this enzyme (results not shown). Although arcaine has two guanidino groups, only one molecule of ammonia per molecule of arcaine is released initially by the viral enzymes. However, prolonged incubation with the enzyme results in the release of ammonia that clearly exceeds the molar amount of arcaine (Fig. 4). This result can be explained if the product of the first reaction, 1-carbamoyl-4-guanidinobutane, is also a substrate for the viral AIH but the specific activity of the enzyme for this substrate is lower.

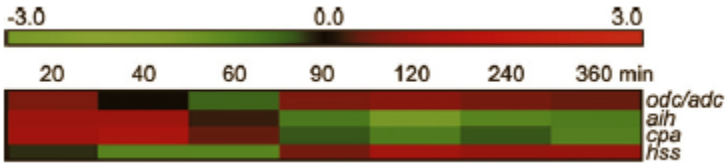


Figure 5. Expression of PBCV-1 polyamine biosynthesis genes during the infection cycle as revealed by microarray analyses*. Color code represents the log2 (Cy5/ Cy3) ratio for each time point and has Cy3 as a reference.*

We tested N-carbamoyl-D, L-aspartic acid, N-carbamoyl-β-alanine, and L-citrulline as substrates for the viral CPA enzymes. However, no ammonia was released from any of these compounds indicating that they are not substrates for the viral CPA enzymes.

3.5 Expression of PBCV-1 polyamine biosynthesis genes during infection. The expression of the PBCV-1 aih and cpa genes during virus replication was monitored by Northern blot (results not shown) and by microarray analyses (Fig. 5). Both genes are expressed only within the first 20 to 40 min of infection. Since PBCV-1 DNA replication begins between 60 and 90 min p.i. [27], aih and cpa are early expressed genes. Their expression overlaps with expression of the PBCV-1 odc/ adc gene (ORF A207R) that is also expressed early (Fig. 5 and [5]). However, the PBCV-1 odc/adc gene is also expressed late (from 90 to 360 min) and, therefore, also overlaps with expression of the hss gene (ORF A237R) that is a typical late gene (Fig. 5). The aih and cpa transcripts are each about 1600 nt in size as revealed by Northern blot analyses (results not shown). These are appropriate sizes for the AIH and CPA proteins, which are 359 and 298 amino acids, respectively.

Immediate early genes (detectable expression within the first 5 to 10min after infection) often have ATGACAA sequences in their predicted promoter regions [28]. However, this sequence was absent in the putative promoter regions of aih and cpa genes from all three viruses (data not shown).

3.6 Occurrence of aih and cpa genes in other chlorella viruses. To determine if the aih and cpa genes are common among the chlorella viruses, a638r and a78r were hybridized to DNA isolated from 37 additional viruses that infect Chlorella NC64A and 5 viruses that infect Chlorella Pbi. The a638r (aih) probe hybridized to some extent with all 37 NC64A viruses (Fig. 6A). The a78r (cpa) probe hybridized to 33 of the NC64A viruses (Fig. 6B). Neither probe hybridized to the host DNA or to the 5 Pbi virus DNAs. These

results suggest that the aih gene is common in all of the NC64A viruses and that the cpa gene might be absent in a few NC64A viruses, i.e. NYs-1, IL-5-2s1, MA-1D, and NY-2B, or that the nucleotide sequence is sufficiently different from the PBCV-1 cpa gene so that hybridization is not detected. The lack of hybridization to the 5 Pbi viruses by both genes is very likely due to differences in their nucleotide sequences.

Table 3. Polyamine biosynthesis ORFs in several chlorella viruses

Enzyme	Chlorella virus					
	NC64A viruses			Pbi viruses		ATCV-1
	PBCV-1	NY-2A	AR158	FR483	MT325	
ODC/ADC	A207R	B278R	C256R	N312L	M307L	Z760R
AIH	A638R	B844R	C759R	–	M766L	Z806R
CPA	A78R	B116R	C104R	N095L	M103L	Z169R
HSS	A237R	B305R	C286R	N232L	M233L	Z590L

New sequencing projects have identified aih and cpa genes, as well as odc/adc and hss genes, in the NC64A virus AR158 [24] and in a new chlorella virus, ATCV-1, that infects endosymbiontic chlorella from the heliozoon *Acanthocystis turfacea* [29] (Table 3, unpublished results). However, a newly sequenced Pbi virus, FR483, lacks the aih gene, but contains the other three polyamine biosynthetic genes [23].

3.7 Phylogenetic analysis of viral polyamine biosynthesis genes. Neighbor-joining phylogenetic analyses of the polyamine biosynthetic gene products (ADC/ODC, AIH, CPA, and HSS) of PBCV-1, NY-2A, AR158, MT325, and FR483 and of 30–60 homologs from other species resulted in four trees in which the virus sequences formed

monophyletic clades with bootstrap support of 100%, indicating that the last common ancestor of the NC64A and Pbi viruses already contained these genes. Additionally, all four trees supported the monophylly of the NC64A and the Pbi viruses with 100% bootstrap support (data not shown and Supplementary Fig. S2).

Previous studies indicate that the PBCV-1 HSS enzyme is bacterial-like [6] and phylogenetic analyses depict the PBCV-1 ODC/ADC protein arising near the ancestral origin of the ODC clade [5, 7]. Aih and cpa genes are common in bacteria and plants; in *Pseudomonas aeruginosa* PAO1 both genes are organized in an operon [15] and they are also clustered in many other bacteria (Supplementary Fig. S2). However, determining the ancestors of the viral aih and cpa genes was difficult because of apparent horizontal gene transfer events, indicated by the unusual groupings of sequences from distant bacterial species (e.g. Firmicutes together with γ-Proteobacteria), and by the absence of a clear plant–bacteria split (Supplementary Fig. S2). Therefore, we cannot unequivocally determine if the viral aih and cpa genes are of bacterial or plant origin.

Polyamine biosynthesis genes have only been found in chlorella viruses and are absent in viruses from the other five Phycodnaviridae genera. Therefore, the chlorella viruses either acquired these genes after separation from the other phycodnaviruses or the genes were lost by the other phycodnaviruses during evolution. It is interesting that the G+C content of the four genes encoding the polyamine biosynthetic enzymes is always higher than the mean G+C content of their respective viral genomes (Table 4). This higher G+C content is especially striking for the cpa gene. Deviation of the G+C content of a gene from the mean G+C is often interpreted as a sign of a recent horizontal gene

transfer [30]. Therefore, maybe the chlorovirus polyamine biosynthesis genes were acquired after separation from the other phycodnaviruses.

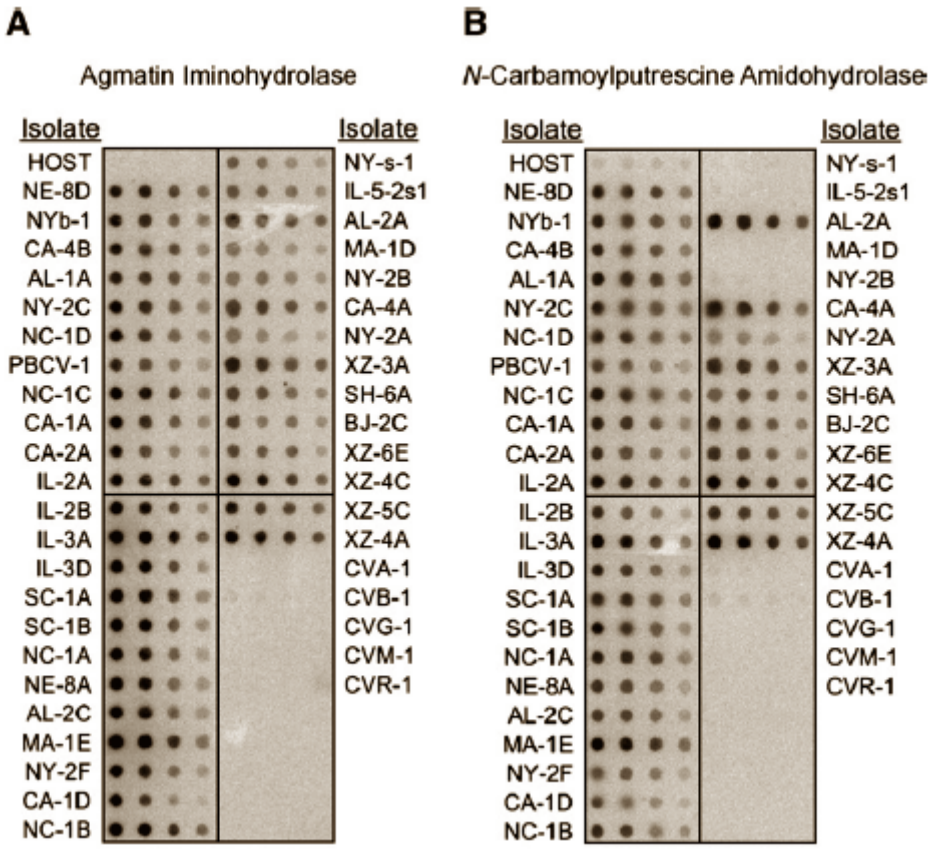


Figure 6. Hybridization of the PBCV-1 aih gene (A) and cpa gene (B) to DNAs isolated from the host Chlorella NC64A, 37 NC64A viruses and 5 Pbi viruses (CVA-1, CVB-1, CVG-1, CVM-1, and CVR-1)*. The DNAs were hybridized with a* ^{32}*P-labelled a638r (aih)gene probe (A) or a a78r (cpa)gene probe (B). The blots in each row contain 1, 0.5, 0.25, and 0.12 µgof DNA from left to right, respectively.*

4. DISCUSSION AND CONCLUSIONS

The importance of polyamines in virus replication has often been discussed in the literature (reviewed in [3, 31]) and these discussions can be summarized as follows: i) Polyamines, primarily putrescine, spermidine and spermine, are present in large amounts in the capsids of some viruses where they can neutralize up to 50% of the viral nucleic acid (e.g. [3, 32, 33, 34]). In contrast, other viruses are devoid of polyamines (e.g. [35]). ii) Virus infection often leads to an increase in polyamine biosynthesis and consequently, in the polyamine content of infected cells [36, 37]. iii) Depletion of intracellular polyamine pools (e.g. by using inhibitors of polyamine biosynthesis) inhibits the replication of some viruses but not others (e.g. [38, 39, 40, 41]).

These previous studies assumed that the polyamines associated with virus replication are synthesized by host enzymes. However, this assumption is no longer valid for all viruses because two genes encoding functional polyamine biosynthetic enzymes were recently discovered in the chlorella virus PBCV-1, an ODC/ADC [5, 7] and a HSS [6]. As reported here, PBCV-1 and at least two other chlorella viruses, NY-2A and MT325, have two additional genes that encode functional AIH and CPA enzymes. Thus, PBCV-1, and most likely NY-2A and MT325, encode a complete ADC pathway that converts arginine to putrescine (ODC/ADC proteins from NY-2A and MT325 have not been tested for activity). Consequently, these viruses are not dependent on their hosts for polyamine biosynthesis.

Table 4. G+C content of the chlorella virus genes encoding polyamine biosynthetic enzymes.

	Virus					
	PBCV-1	NY-2A	AR158	MT325	FR483	ATCV-1
Gene						
odc/adc	41%	45%	41%	46%	46%	51%
aih	43%	42%	43%	47%	–	51%
cpa	45%	46%	48%	51%	50%	55%
hss	44%	44%	44%	47%	46%	54%
Complete genome	40%	40.7%	40.7%	45.3%	44.6%	49%

The fact that all four polyamine biosynthetic encoding genes are present in most chloroviruses, as shown by Southern blot analyses (Fig. 6; [5, 6]) and genome-sequencing projects (Table 3), suggests that these proteins are important for chlorella virus replication. However, at least one chlorella virus, FR483, which infects Chlorella Pbi lacks an aih gene [23]. The other three genes encoding polyamine biosynthetic enzymes are present in all chlorella viruses sequenced to date. Chlorovirus is the only genus within the family Phycodnaviridae, currently consisting of 6 genera [42], with genes encoding polyamine biosynthetic enzymes; thus, either polyamines are especially important for chlorella virus replication or the chlorella host, in contrast to other algal hosts, is unable to provide the necessary amounts of polyamines for virus replication. The results also indicate that the chlorella viruses either acquired these enzymes after separation from the other phycodnaviruses or that all of the phycodnaviruses originally had the polyamine biosynthetic genes and that they have been selectively lost by the other phycodnaviruses during evolution. However, one should keep in mind that only a few other phycodnavirus genomes have been sequenced [43] and so the concept that only

chlorella viruses encode these enzymes could change as more algal virus genomes are sequenced. It may not be relevant, but one distinction between the chlorella viruses and members of the five other phycodnavirus genera is that the chlorella viruses exist in fresh water and the other viruses occur in marine environments.

Unfortunately, not much is known about polyamine metabolism in chlorellae. ADC and ODC activities were reported in Chlorella vulgaris Beijerinck [44, 45]. The related green alga Chlamydomonas reinhardtii only uses the ODC pathway to synthesize putrescine [46]. However, although Chlamydomonas lacks an adc gene it has aih and cpa genes (available at: http:// genome.jgi-psf.org/Chlre3/Chlre3.home.html), but, according to EST data, they are not expressed. It is unknown if Chlorella NC64A and Chlorella Pbi possess a complete ADC pathway consisting of ADC, AIH and CPA; if they do the host chlorella AIH may functionally replace the absent viral AIH in FR483 during infection. The genome of Chlorella NC64A, the host for PBCV-1 and NY-2A, is currently being sequenced and annotation of its sequence will provide information on its polyamine biosynthetic capability. However, it is unlikely that the four PBCV-1 polyamine biosynthetic genes are acquired from Chlorella NC64A because none of the four virus genes hybridize with Chlorella NC64A DNA (Fig. 6A, B; [5, 6]).

PBCV-1 aih and cpa genes are transcribed as early genes, appearing within the first 20 min p.i. Their expression overlaps with transcription of the PBCV-1 odc/adc gene so that all three putrescine biosynthetic enzymes are presumably present early in infection. Therefore, one might expect the polyamine levels to increase during the first 60 min of PBCV-1 infection. However, little change occurs in either the polyamine concentration or its composition (putrescine, cadaverine, spermidine, and

homospermidine are present) during the first 60 min of virus infection [6]. By 240 min p.i., the concentration of putrescine increases about 3.5 times, whereas the other polyamines decrease duringthis time. The net result is that the total polyamine concentration decreases slightly during virus replication [6]. It is unlikely that polyamines play a role in the neutralization of PBCV-1 DNA in the capsid because the number of polyamine molecules per PBCV-1 virion is so low that they can only neutralize ~0.2% of the virus phosphate residues [6]. Furthermore, the physiological significance of the polyamines in the PBCV-1 particles must be limited because they are only loosely associated with the virions, i.e. they can be replaced by Tris or displaced by washing the particles in a polyamine-free buffer without affecting virus infectivity [6].

The dual expression pattern (early and late) of the viral genes encoding the polyamine biosynthetic enzymes may indicate that putrescine (early in viral replication) and homospermidine (late in viral replication) have different functions in virus replication. It should be mentioned that the HSS protein, but not the other three polyamine biosynthetic enzymes, is a major component of PBCV-1 virions (Dunigan et al., manuscript in preparation). Therefore, the HSS protein may also serve a structural role in the virus life cycle.

In summary, most of the chlorella viruses have four genes that encode functional enzymes involved in polyamine biosynthesis. However, the biological function(s) of these enzymes and their products is currently an enigma.

5. REFERENCES

1. Hanfrey, C., Sommer S., Mayer M.J., Burtin D., Michael A.J. 2001. *Arabidopsis* polyamine biosynthesis: absence of ornithine decarboxylase and the mechanism of arginine decarboxylase activity. *Plant J* 27: 551–560.
2. Cohen, S.S. 1998. A Guide to the Polyamines. Oxford Univ. Press Inc., New York.
3. Cohen, S.S. 1998. Viruses, p. 366–395. *In* A Guide to the Polyamines. Oxford Univ. Press Inc., New York.
4. Walters, D.R. 2003. Polyamines and plant disease. *Phytochemistry* 64:97–107.
5. 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.
6. 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.
7. 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.
8. Van Etten, J.L., 2003. Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.* 37:153–195.
9. Yamada, T., Onimatsu H., Van Etten J.L. 2006. Chlorella viruses. *Adv. Virus Res.* 66:293–336.

10. Reisser, W., Becker B., Klein T. 1986. Studies on ultrastructure and host range of a Chlorella attacking virus. *Protoplasma* 135:162–165.

11. Reisser, W., Burbank D.E., Meints S.M., Meints R.H., Becker B., Van Etten J.L. 1988. A comparison of viruses infecting two different Chlorella-like green algae. *Virology* 167:143–149.

12. Illingworth, C., Mayer M.J., Elliott K., Hanfrey C., Walton N.J., Michael A.J. 2003. The diverse bacterial origins of the *Arabidopsis* polyamine biosynthetic pathway. *FEBS Lett.* 549:26–30.

13. Janowitz, T., Kneifel H. Piotrowski M. 2003. Identification and characterization of plant agmatine iminohydrolase, the last missing link in polyamine biosynthesis of plants. *FEBS Lett.* 544:258–261.

14. Piotrowski, M., Janowitz T., Kneifel H. 2003. Plant C–N hydrolases and the identification of a plant N-carbamoylputrescine amidohydrolase involved in polyamine biosynthesis. *J. Biol. Chem.* 278:1708–1712.

15. 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.

16. Nakada, Y., Itoh Y. 2003. Identification of the putrescine biosynthetic genes in *Pseudomonas aeruginosa* and characterization of agmatine deiminase and N-carbamoylputrescine amidohydrolase of the arginine decarboxylase pathway. *Microbiology* 149:707–714.

17. 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.

18. Ausubel, F.M., Brent R., Kingston R.E., Moore D.D., Seidman J.G., Struhl K. 2004. Current Protocols in Molecular Biology. John Wiley and Sons, New York.

19. Saeed, A.I., Sharov V., White J., Li J., Liang W., Bhagabati N., Braisted J., Klap, M., Currier T., Thiagarajan M., Sturn A., Snuffin M., Rezantsev A., Popov D., Ryltsov A., Kostukovich E., Borisovsky I., Liu Z., Vinsavich A., Trush V., Quackenbush J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 34:374–378.

20. Sun, L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L. 2000. Characterization of a β-1,3-glucanase encoded by chlorella virus PBCV-1. *Virology* 276:27–36.

21. Altschul, S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

22. Ronquist, F., Huelsenbeck J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

23. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Hartigan J., Van Etten J.L. 2007. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect Chlorella Pbi. *Virology* 358:459-471.

24. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Nierman W.C., Van Etten J.L. 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect Chlorella NC64A. *Virology* 358:472-484.

25. Smith, T.A. 1969. Agmatine iminohydrolase in maize. *Phytochemistry* 8:2111–2117.

26. Yanagisawa, H., Suzuki Y. 1981. Corn agmatine iminohydrolase. Purification and properties. *Plant Physiol.* 67:697–700.

27. 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.

28. Kawasaki, T., Tanaka M., Fujie M., Usami S., Yamada T. 2004. Immediate early genes expressed in chlorovirus infections. *Virology* 318:214–223.

29. Bubeck, J.A., Pfitzner A.J. 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.

30. Lawrence, J.G., Ochman H., 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44:383–397.

31. Cohen, S.S., McCormick F.P. 1979. Polyamines and virus multiplication. *Adv. Virus Res.* 24:331–387.

32. Ames, B.N., Dubin D.T. 1960. The role of polyamines in the neutralization of bacteriophage deoxyribonucleic acid. *J. Biol. Chem.* 235:769–775.

33. Gibson, W., Roizman B. 1971. Compartmentalization of spermine and spermidine in the herpes simplex virion. *Proc. Natl. Acad. Sci. USA* 68:2818–2821.

34. Tyms, A.S., Clarke J.R., Ford L., Eloranta T. 1990. Polyamines and the growth of pathogenic viruses, p.122-133. *In* S. H. Goldemberg, I. D. Algranati (eds.), The Biology and Chemistry of Polyamines. IRL Press, Oxford.

35. Nickerson, K.W., Lane L.C. 1977. Polyamine content of several RNA plant viruses. *Virology* 81:455–459.

36. Goldstein, D.A., Heby O., Marton L.J. 1976. Biphasic stimulation of polyamine biosynthesis in primary mouse kidney cells by infection with polyoma virus: uncoupling from DNA and rRNA synthesis. *Proc. Natl. Acad. Sci. USA* 73:4022–4026.

37. Torget, R., Lapi L., Cohen S.S. 1979. Synthesis and accumulation of polyamines and S-adenosylmethionine in Chinese cabbage infected by turnip yellow mosaic virus. *Biochem. Biophys. Res. Commun.* 87:1132–1139.

38. Gibson, W., van Breemen R., Fields A., LaFemina R., Irmiere A. 1984. D,L-alpha-Difluoromethylornithine inhibits human cytomegalovirus replication. *J. Virol.* 50:145–154.

39. Greco, A., Calle A., Morfin F., Thouvenot D., Cayre M., Kindbeiter K., Martin L., Levillain O., Diaz J.J. 2005. S-Adenosyl methionine decarboxylase activity is required for the outcome of herpes simplex virus type 1 infection and represents a new potential therapeutic target. *FASEB J.* 19:1128–1130.

40. Pohjanpelto, P., Sekki A., Hukkanen V., von Bonsdorff C.H. 1988. Polyamine depletion of cells reduces the infectivity of herpes simplex virus but not the infectivity of Sindbis virus. *Life Sci.* 42:2011–2018.

41. Raina, A., Tuomi K., Mantyjarvi R. 1981. Roles of polyamines in the replication of animal viruses. *Med. Biol.* 59:428–432.

42. Wilson, W.H., Van Etten, J.L., Schroeder, D.S., Nagasaki, K., Brussaard, C., Delaroque, N., Bratbak, G., and Suttle, C. 2005. Phycodnaviridae, p.163-175. *In* Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., and Ball, L. A.

(eds.), Virus Taxonomy: Classification and Nomenclature of Viruses. Elsevier Academic Press, San Diego, CA.

43. Dunigan, D.D., Fitzgerald L.A., Van Etten J.L. 2006. Phycodnaviruses: a peek at genetic diversity. *Virus Res.* 117:119–132.

44. Cohen E., Arad S.M., Heimer Y.M., Mizrahi Y. 1983. Polyamine biosynthetic enzymes in Chlorella: characterization of ornithine and arginine decarboxylase. *Plant Cell Physiol.* 24:1003–1010.

45. Cohen, E., Arad S.M., Heimer Y.H., Mizrahi Y. 1984. Polyamine biosynthetic enzymes in the cell cycle of Chlorella: correlation between ornithine decarboxylase and DNA synthesis at different light intensities. *Plant Physiol.* 74:385–388.

46. Theiss, C., Bohley P., Voigt J. 2002. Regulation by polyamines of ornithine decarboxylase activity and cell division in the unicellular green alga *Chlamydomonas reinhardtii*. *Plant Physiol*. 128:1470–1479.

APPENDIX II

CHLORELLA VIRUSES ENCODE MOST, IF NOT ALL, OF THE MACHINERY TO GLYCOSYLATE THEIR GLYCOPROTEINS INDEPENDENT OF THE ENDOPLASMIC RETICULUM AND GOLGI

1. VIRUS ENCODED GLYCOSYLTRANSFERASES

Structural proteins of many viruses, such as rhabdoviruses, herpesviruses, poxviruses, and paramyxoviruses are glycosylated. Most virus glycoproteins, which can have very complex glycans, are linked to Asn via N-acetylglucosamine, although some viruses have O-linked sugars attached to Ser or Thr residues via an amino sugar, usually acetylglucosamine or acetylgalactosamine. Typically, viruses use host-encoded glycosyltransferases and glycosidases located in the endoplasmic reticulum (ER) and Golgi apparatus to add and remove N-linked sugar residues from virus glycoproteins either co-translationally or shortly after translation of the protein. This post-translational processing aids in protein folding and involves other host-encoded enzymes such as protein disulfide isomerase. After folding and assembly, virus glycoproteins are transported by host sorting and membrane transport functions to virus specified regions in host membranes where they displace host glycoproteins. Progeny viruses then bud through these virus-specific target membranes, which is usually the final step in the assembly of infectious virions [1-4]. Thus, nascent viruses only become infectious by budding through the membrane, usually the plasma membrane, as they are released from

the cell. Consequently, the glycan portion of virus glycoproteins is host-specific. The theme that emerges from these studies is that virus glycoproteins are synthesized and glycosylated by the same mechanisms as host glycoproteins. Therefore, the only way to alter glycosylation of virus proteins is to either grow the virus in a different host or have a mutation in the virus protein that alters the protein glycosylation site.

One outcome of this scenario is that, in general, viruses lack genes encoding glycosyltransferases. However, a few virus-encoded glycosyltransferases have been reported in the last few years (see review [5]). Sometimes these virus-encoded glycosyltransferases add sugars to compounds other than proteins. For instance, some phage-encoded glycosyltransferases modify virus DNA to protect it from host restriction endonucleases [e.g., 6-8], and a glycosyltransferase encoded by baculoviruses modifies a host insect ecdysteroid hormone leading to its inactivation [9]. Bovine herpesvirus-4 encodes a β-1,6-N-acetyl-glucosaminyltransferase that is localized in the Golgi and is probably involved in post-translational modification of virus structural proteins [10-12]. Additional putative virus encoded glycosyltransferases of unknown function include certain pox viruses [13, 14], *Ectocarpus siliculosus* virus 1 [15], *Ostreococus* virus OsV5 [16], *Acanthamoeba polyphaga* Mimivirus [17], and two Archaea viruses, *Acidianus filamentous* virus 3 [18], and *Sulfolobus* virus STSV1 [19]. With the rapid increase in sequencing virus genomes, additional virus encoded glycosyltransferases will undoubtedly be discovered.

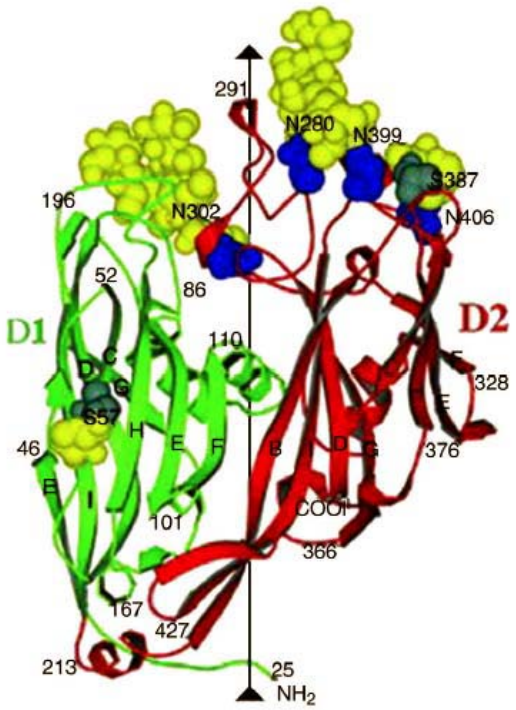


Figure 1. A ribbon diagram of the crystal structure of the virus PBCV-1 major capsid protein, Vp54, monomer with strategic amino acids labeled. *The carbohydrate moieties (yellow) and glycosylated Asn and Ser residues (blue and gray, respectively) are shown as space-filling atoms. The green and red regions indicate the consecutive "jelly roll" domains of the protein. The arrow shows the orientation of the protein from the inside to the outside of the virus. Reproduced from [62].*

One group of viruses that differs from the scenario that viruses use the host machinery located in the ER and Golgi to glycosylate their glycoproteins is the chloroviruses (family Phycodnaviridae) that infects eukaryotic algae. Phycodnaviruses are large (150 to 190 nm in diameter) polyhedral, dsDNA-containing viruses with an internal lipid membrane (genomes of 170 to 560 kb) [20-23]. They are present in aqueous

environments throughout the world and play dynamic, albeit largely unknown, roles in regulating algal communities in aqueous environments, such as the termination of massive algal blooms commonly referred to as red and brown tides [e.g., 24-27]. Some phycodnaviruses encode as many as 600 proteins, which is more protein-encoding genes than in the smallest bacteria.

The phycodnaviruses probably have a common ancestor with the poxviruses, iridoviruses, African swine fever virus, and Mimivirus [e.g., 28-30], and accumulating evidence indicates that the phycodnaviruses have a long evolutionary history, possibly dating from the time eukaryotes arose from prokaryotes (2-3 billion years ago) [e.g., 30-32]. Some algal virus genes encode commercially important enzymes such as DNA restriction endonucleases [e.g., 33, 34], whereas other viral genes encode proteins that are the smallest in their class and may represent the minimal catalytic unit. Consequently, these 'small' proteins serve as models for mechanistic and structural studies (e.g., DNA topoisomerase type II [35] and a K^+ ion channel protein [36]).

Phycodnavirus members in the genus *Chlorovirus* are plaque-forming viruses that infect certain isolates of unicellular, chlorella-like green algae. The prototype chlorella virus *Paramecium bursaria* chlorella virus (PBCV-1) infects a host, *Chlorella* NC64A, that is normally a symbiont in the protozoan *P. bursaria*. The 330 kb PBCV-1 genome has ~690 potential open reading frames (ORFs) of 65 codons or larger [20]. About 365 of these ORFs are predicted to encode proteins (CDSs) and the functions of 35-40% of these CDSs have been tentatively identified. Pertinent to this review is that 17 of the PBCV-1 genes encode enzymes that manipulate sugars, many of which have never been found previously in a virus (Table 1). At least five of the enzymes are putative

glycosyltransferases, presumably involved in glycosylation of the PBCV-1 major capsid protein Vp54. [Note: in our original report on the annotation of the PBCV-1 genome we identified 7 PBCV-1 encoded glycosyltransferases [37]. However, two of these CDSs, which were co-linear with an adjacent glycosyltransferase, were the result of sequencing errors and are part of the adjacent glycosyltransferase.]

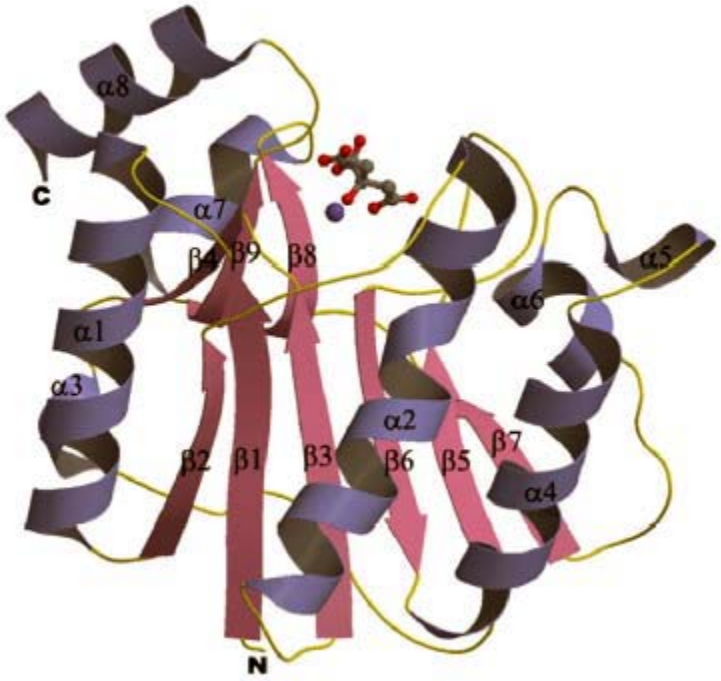


Figure 2. A ribbon diagram of the crystal structure of the glycosyltransferase domain from virus PBCV-1 CDS A64R. *Bound Mn^{2+} and citrate ions are shown in a ball-and-stick representation. Alpha (purple) and beta sheets (pink) are color coded. Reproduced from reference [70] with permission.*

2. CHLOROVIRUS ENCODED SUGAR ENZYMES

Three PBCV-1 encoded enzymes are involved in the synthesis of the extracellular matrix polysaccharide hyaluronan (also referred to as hyaluronic acid), including the glycosyltransferase hyaluronan synthase (HAS) [38, 39]. Hyaluronan, a ubiquitous constituent of the extracellular matrix in vertebrates, consists of ~20,000 alternating β-1,4-glucuronic acid and β-1,3-N-acetylglucosamine residues [40]. Hyaluronan synthesis occurs at the plasma membrane and is simultaneously extruded through the membrane to the exterior of the cell. Until the *has* gene was discovered in PBCV-1, hyaluronan was thought to only occur in vertebrates and a few pathogenic bacteria, where it forms an extracellular capsule, presumably to avoid the immune system [40, 41].

PBCV-1 also encodes two enzymes involved in the biosynthesis of hyaluronan precursors, glutamine:fructose-6-phosphate amidotransferase and UDP-glucose dehydrogenase [42]. All three genes are expressed early during PBCV-1 infection. These results led to the discovery that hyaluronan lyase-sensitive hair-like fibers begin to accumulate on the surface of PBCV-1 infected host cells by 15 min post infection (p.i.). By 4 hr p.i., the infected cells are covered with a dense fibrous hyaluronan network [39].

The *has* gene is present in many, but not all, chloroviruses isolated from diverse geographical regions [39]. Surprisingly, many chloroviruses that lack a *has* gene have a gene encoding a functional chitin synthase (CHS). Furthermore, cells infected with these viruses produce chitin fibers on their external surface [43]. Chitin, an insoluble linear homopolymer of □-1,4-linked N-acetyl-glucosamine residues, is a common component of insect exoskeletons, shells of crustaceans and fungal cell walls [44].

A few chloroviruses contain both *has* and *chs* genes, and form both hyaluronan and chitin on the surface of their infected cells [43, 45]. Finally, a few chloroviruses probably lack both genes because no extracellular polysaccharides are formed on the surface of cells infected with these viruses [39]. The fact that many chloroviruses encode enzymes involved in extracellular polysaccharide biosynthesis suggests that the polysaccharides, which require a huge expenditure of ATP for their synthesis, are important in the virus life cycle. However, at present this function(s) is unknown.

Table 1. Virus PBCV-1 encoded enzymes involved in sugar manipulation.

Activity	CDS	Size (aa)	Expressed[1] (min p.i.)	Reference
Glycosyltransferases				
	A064R	638	20–90	[64]
	A111/114R	860	40–60	–
	A219/222/226R	677	20–60	–
	A473L	517	20–60	–
	A546L	396	40–60	–
Hyaluronan biosynthesis				
UDP-glucose dehydrogenase	A609L	389	20–40	[37]
Glutamine:fructose-6-PO_4 amidotransferase	A100R	595	20–60	[37]
Hyaluronan synthase	A098R	568	20–40	[33]
Sugar nucleotide biosynthesis				
GDP-D-mannose 4,6 dehydratase	A118R	345	40–90	[41]
Fucose synthase[2]	A295L	317	40–90	[41]
Polysaccharide degrading enzymes				
Chitinase	A181/182R	830	60–360	[44]
Chitinase	A260R	484	90–360	[44]
Chitosanase	A292L	328	90–360	[44]
β-1,3-glucanase	A094L	364	20–40	[45]
β and α 1,4-linked glucuronic lyase	A215L	321	120–360	[46]
Others				
D-lactate dehydrogenase	A053R	363	20–60	–
Pyrimidine dimer-specific glycosylase	A050L	141	20–60	[48]

[1] *The time the genes were expressed was determined by microarray analysis (Yanai-Balser, G.M. et al., manuscript in preparation).*

[2] *Another name is GDP-4-keto-6-deoxy-D-mannose epimerase reductase.*

The chloroviruses encode additional enzymes involved in nucleotide sugar metabolism. Two enzymes encoded by virus PBCV-1, GDP-D-mannose 4,6 dehydratase (GMD) and a GDP-4-keto-6-deoxy-D-mannose epimerase reductase (GMER) comprise the highly conserved pathway that converts GDP-D-mannose to GDP-L-fucose. In vitro reconstruction of the biosynthetic pathway using recombinant PBCV-1 GMD and GMER synthesized GDP-L-fucose [46, 47]. Unexpectedly, however, the PBCV-1 GMD also catalyzes the NADPH-dependent reduction of the intermediate GDP-4-keto-6-deoxy-D-mannose, forming GDP-D-rhamnose. Both fucose and rhamnose are constituents of the glycans attached to the PBCV-1 major capsid protein Vp54 (see below). Therefore, the virus might encode the enzymes to meet this need. Other chloroviruses encode additional sugar metabolizing enzymes including a putative mannose-6-phosphate isomerase and a UDP-glucose 4,6 dehydratase [48] (Tonetti, M. et al., unpublished data).

At least five PBCV-1-encoded enzymes are involved in polysaccharide degradation, including 2 chitinases, a chitosanase, a □-1,3-glucanase, and a □- and □-1,4-linked glucuronic acid lyase (summarized in [21], [49-51]). Finally, one enzyme is a glycosylase that initiates pyrimidine photodimer excision [52, 53].

3. GLYCOSYLATION OF VIRUS PBCV-1 MAJOR CAPSID PROTEIN VP54

The major capsid protein of PBCV-1, Vp54 (comprises ~40% of total virus protein), is a glycoprotein with a molecular weight of 53,790 (Table 2). Glycosylation of Vp54 and two minor structural proteins Vp280 (~1% of the total virus protein) and Vp260 (~0.1% of total virus protein) differs from other viruses. CDS A430L encodes

Vp54 with a predicted molecular weight of 48,165 Da. This protein undergoes additional post-translational modifications in addition to being glycosylated. The amino terminal Met is removed [54] and the protein is myristylated in the carboxyl-terminal portion of the protein, probably at an internal Lys or Arg [55]. Vp260 is coded by CDS A122/123R, a protein of 138 kDa. Since the protein migrates on SDS-PAGE with a weight of ~260 kDa, it probably consists of more than 40% carbohydrate. The CDS encoding Vp280 is unknown.

Cryo-electron microscopy and three-dimensional (3-D) icosahedral image reconstruction of the PBCV-1 virion indicate that the outer capsid is icosahedral and covers a lipid bilayered membrane [56, 57]. The capsid shell consists of 1,680 donut-shaped trimeric capsomers plus 12 pentameric capsomers at each icosahedral vertex. The trimeric capsomers are arranged into 20 triangular facets (trisymmerons, each containing 66 trimers) and 12 pentagonal facets (pentasymmetrons, each containing 30 trimers and one pentamer at the icosahedral vertices). Trisymmetrons and pentasymmetrons are made up of trimeric capsomers, with each doughnut shaped capsomer composed of 3 monomers of Vp54. The trimeric capsomers are easy to isolate because heating PBCV-1 to 65^{o}C for a few minutes leads to solubilization of the trimers. However, the monomer form of Vp54 is insoluble.

Recent five-fold symmetry, rather than icosahedral, averaging of cryoelectron microscopy data has produced some exciting results on the structure of PBCV-1 [58]. PBCV-1 has a unique pentameric vertex with a 250 Å long and 50 Å wide cylindrical spike that connects to a large pocket between the capsid and the enveloped nucleocapsid. Circumstantial evidence suggests that Vp260 may be a component of the spike structure,

which presumably is involved in virus attachment.

The conclusion that the PBCV-1 proteins are glycosylated by a different mechanism than those used by other viruses originally arose from antibody studies. Polyclonal antiserum prepared against intact PBCV-1 virions inhibits virus plaque formation by agglutinating particles. Spontaneously derived, antiserum-resistant, plaque-forming variants of PBCV-1 occur at a frequency of 10^{-5} to 10^{-6}. These antiserum-resistant variants fall into four serologically distinct classes (Table 2) [59]. Polyclonal antisera prepared against members of each of these antigenic classes react exclusively or predominately with the Vp54 (and Vp280 and Vp260) equivalents from the viruses in the class used for the immunization. While each of the three glycoproteins from the antigenic variants migrate faster on SDS-PAGE than those of the strains from which they are derived, all of the major capsid proteins co-migrate after deglycosylation. Western blot analyses of Vp54 proteins isolated from the variants, before and after removing the glycans with trifluoromethane-sulfonic acid or altering the glycan with periodic acid, established that the antigenic variants reflect differences in the Vp54 glycans. In addition, the ratio of the seven neutral sugars [glucose, fucose, galactose, mannose, xylose, rhamnose and arabinose [59] attached to PBCV-1 Vp54 and the variants change in a manner that correlates with antigenicity and Vp54 migration on SDS-PAGE (Table 3). Variants from different classes can complement and recombine in dual infection experiments to produce wild-type progeny, indicating that the enzymes involved in glycosylation reside in different virus-encoded complementation groups [60].

Additional observations indicate that Vp54 glycosylation is unusual: i) Unlike viruses that acquire their glycoprotein(s) by budding through a plasma membrane, intact

infectious PBCV-1 particles accumulate inside the host 30-40 min before virus release [61]. ii) The nucleotide sequence of the Vp54 gene (*a430l*) in each of the variants is identical to wild type PBCV-1, thus the peptide portion of Vp54 is not altered in the mutants [59]. iii) All of the antigenic variants are grown in the same host so the differences cannot be attributed to the host. iv) Polyclonal antibodies to Vp54 do not react with host glycoproteins. v) Neither N-acetylglucosamine (GlcNAc) nor N-acetylgalactosamine were detected in Vp54 glycans, sugars commonly found in Asn-linked (N-linked) and many Ser/Thr-linked (O-linked) glycoproteins produced via the cellular ER-Golgi pathway (Table 3) [59]. vi) Unlike most glycoproteins that exhibit size microheterogeneity, Vp54 appears homogeneous on SDS-PAGE; in addition, mass spectrometry analysis only reveals one satellite peak that differs in molecular weight by 140 Da, the approximate weight of either one arabinose, xylose, fucose or rhamnose sugar (Cerny, R. and Van Etten, J.L., unpublished data). vii) The ability to easily crystallize Vp54 as a homotrimer provides additional evidence that the protein is essentially homogeneous (Fig. 1) [62].

The glycans are not linked to the major capsid proteins by a traditional N-linkage because: i) treatment of Vp54 with the enzymes endo-β-N-acetylglucosaminidase F or N-glucosidase F, which cleave glycans from N-linked glycoproteins [63], does not alter Vp54 migration on SDS-PAGE (Chaney, W.G. and Van Etten, J.L., unpublished data). ii) The drug tunicamycin, which inhibits synthesis of N-linked glycoproteins [64], has no effect on PBCV-1 replication nor Vp54 electrophoretic mobility, even at concentrations three times that required to inhibit host chlorella growth [55]. Finally, an antibody (provided by Felix Wieland) to a β–glucose Asn linkage present in bacterial

glycoproteins [65] reacts strongly with Vp54 (Gurnon, J.R. and Van Etten, J.L., unpublished data), suggesting that one or more of the Asn linked glycans in Vp54 have this linkage. This same linkage exists in the B2 chain of mammalian laminin, a major basement membrane glycoprotein [66].

Table 2. Properties of antigenic classes of viruses derived from PBCV-1.

Antigenic classes			Properties of members of each antigenic class		
Class	Proposed genotype	M_r Vp54 SDS-PAGE	Virus	MW Vp54, mass spectrometry	Total glycan mass[a]
Wild-type	wt	54 kDa	PBCV-1	53,790	5545
P91	*mp91*	53 kDa	P91	52,780	4535
EPA-1	*mepa1*	52 kDa	EPA-1	51,560	3315
			EPA-2	nd[b]	nd
			EPA-3	nd	nd
			P31	51,593	3348
			P1050	nd	nd
			P1056	nd	nd
E11	*mel 1*	51 kDa	E11	nd	nd
			P41	nd	nd
			P1210	nd	nd
			P1219	nd	nd
P100	*mp100*	50.5 kDa	P100	51,038	2793

[a] *Calculated by subtracting the deduced peptide mass minus the N-terminal initiating Met residue (48,017) and the mass of one myristate (228.4) from the MW determined by mass spectrometry. All the protein backbones are identical.*

[b] *nd, not determined.*

Modified from [60].

We originally suggested that Vp54 had a single ~30mer glycan or two-15mer glycans attached to it [60]. However, the Vp54 crystal structure revealed sugars attached to six amino acids in the protein (Fig. 1) [62]. Comparison of the molecular mass of Vp54 (53,790 Da) with its predicted molecular weight from the amino acid sequence (48,165 Da) indicates that the protein contains 30-35 sugar moieties, of which 20 could

be detected, but not identified, in the crystal density map. At a minimum, single sugars were found attached to Ser^{57} and Ser^{387}; six and seven-branched chain sugar moieties were found attached to Asn^{280} and Asn^{302}, respectively, as well as three and two sugar residues were located at Asn^{399} and Asn^{406}, respectively. As predicted from the antibody results, the four N-linked glycans are on the external surface of the virus. GlcNAc fits the carbohydrate electron density at residues Asn^{280}, Asn^{302}, and Asn^{399}. However, the inability to detect GlcNAc in Vp54 suggests that another modified sugar might be in this position, e.g., glucose with an O-linked acetate.

The identification of the glycan-linked Asn residues in Vp54 provided additional evidence that Vp54 glycosylation does not involve host glycosyltransferases. Asn^{302}, Asn^{399}, and Asn^{406} occur in the amino acid sequence (A/G)NTXT, and Asn^{280} occurs in an ANIPG sequence. None of these Asn residues reside in a NX(T/S) sequence commonly recognized by ER located glycosyltransferases [67]. These findings also explain why previous enzymatic tests for N-glycosylation were negative. Finally, the major capsid protein, like the glycosyltransferases, lacks an ER or Golgi signal peptide. Therefore, taken together, the results suggest that PBCV-1 encodes some, if not all, of the enzymes involved in constructing the glycans attached to Vp54.

4. PBCV-1 ENCODED GLYCOSYLTRANSFERASES

As mentioned above, five putative glycosyltransferase-encoding genes have been found in PBCV-1, which are scattered throughout the PBCV-1 genome (Table 1). None of these PBCV-1 encoded glycosyltransferases have an identifiable signal peptide that

would target them to the ER. Furthermore, the cellular protein localization program PSORT predicts that all of these proteins, with the exception of A473L, are located in the cytoplasm. A473L is predicted to contain one transmembrane domain.

The genes for all five PBCV-1 encoded glycosyltransferases are expressed early during PBCV-1 infection (Yanai-Balser, G.M. et al., manuscript in preparation). Thus, assuming the proteins are stable, the enzymes should be available for either adding sugars to the Vp54 glycans or transferring the glycans. The *a430l* gene is expressed late.

PBCV-1 *a64r* gene. The PBCV-1 *a64r* gene encodes a 638 amino acid protein with 4 motifs that are conserved in the "fringe-class" of glycosyltransferases [68, 69]. Five conserved domains, designated 1-5, have been identified in this class of enzymes [69]. Domains 3 and 5 contain the proposed catalytic amino acids, the "DXD" sequence in domain 3 and the first "D" residue in domain 5. The A64R protein contains domains 2-5 although the spacing between some of the domains differs from that of fringe-glycosyltransferases. As mentioned above, the A64R protein lacks both an identifiable signal peptide that would target the protein to the ER and a membrane-spanning domain in contrast to fringe glycosyltransferases.

Analysis of 13 PBCV-1 antigenic variants revealed mutations in *a64r* that correlated with a specific antigenic class, EPA-1 (Table 2). The *a64r* gene in six of these antigenic variants was sequenced to determine if mutations in *a64r* correlated with the EPA-1 antigenic variation [60]. The *a64r* sequences from mutants EPA-1, EPA-3, and P31 have single nucleotide substitutions, which produce a single amino acid substitution in the A64R protein. Two of the amino acid substitutions occur in the DXD motif

(domain 3) and the other one is in domain 4. A fourth variant has an extra base in the coding sequence, which creates a frame shift mutation in the gene. Finally, the entire gene is deleted in the other two antigenic variants.

Dual infection experiments with the different antigenic variants indicate that viruses containing wild-type *a64r* complement and recombine with viruses that contain variant *a64r* to form wild-type virus. Therefore, we concluded that *a64r* encodes a glycosyltransferase involved in the synthesis of the Vp54 glycan [60].

5. STRUCTURE OF PBCV-1 ENCODED A64R GLYCOSYLTRANSFERASE

Blast searches with the 638 amino acid A64R protein established that the glycosyltransferase domains are located in the amino-terminal 211 amino acids. The A64R carboxyl terminus resembles several bacterial proteins of unknown function.

The amino-terminal A64R glycosyltransferase domain (residues 1-211) was expressed in a bacterial system for crystallization studies. The crystal structure of the peptide was solved at 1.6 Å resolution (Fig. 2) [70]. The peptide has a mixed □/□ fold containing a central, six-stranded □ sheet flanked by □helices. The overall fold is similar to catalytic domains of glycosyltransferases in the GT-A group, retaining glycosyltransferase family 34, although the amino acid similarity between them is low (less than 14% for equivalent C□ atoms). Crystal structures of A64R complexed with UDP, CMP, or GDP, established that only UDP bound to A64R in the presence of Mn^{2+}, consistent with its high structural similarity to glycosyltransferases that use UDP as the sugar carrier. The structure of the A64R, UDP-glucose, and Mn^{2+} complex showed that

the largest conformational change occurred when hydrogen bonds were formed with the ligands. Unlike UDP-glucose, UDP-galactose and UDP-GlcNAc did not bind to A64R, suggesting a selective binding of UDP-glucose. Thus UDP-glucose is most likely the sugar donor for A64R, consistent with glucose occurring in the virus major capsid protein glycans (Table 3).

However, no enzymology has been conducted on the A64R peptide, or any of the other virus encoded glycosyltransferases because the structure of the Vp54 glycans are unknown. Consequently, the substrate(s) for A64R or any of the putative viral glycosyltransferases are unknown.

Table 3. Sugar composition of the glycoproteins from PBCV-1 serotypes.

Sugar	Viruses — relative sugar content						
	PBCV-1	P91	EPA-1	P31	P41	E11	P100
Fucose	0.69	0.73	0.88	0.81	0.78	0.83	0.61
Ara/Rham[a]	1.57	1.24	0.64	0.65	0.65	0.65	0.19
Galactose	1.02	0.97	1.02	1.03	0.99	0.97	0.91
Glucose	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Xylose	1.58	1.47	1.59	1.57	0.88	0.77	0.60
Mannose	1.02	0.70	0.14	0.17	0.24	0.15	0.12

Values are normalized to that of glucose. The mutants are arranged so that the major capsid protein decreases in size from PBCV-1 to P100. Viruses EPA-1 and P31 and viruses P41 and E11 are serologically similar even though they are separate isolates. Viruses P91, EPA-1, P31 and P41 are derived from PBCV-1; virus E11 is derived from EPA-1; and virus P100 is from P31. Ara/Rham, arabinose/rhamnose. (Reproduced from [59]).

**Subsequent experiments established that both Ara and Rham are present in the PBCV-1 major capsid protein.*

6. WHERE DOES VP54 ACQUIRE ITS GLYCANS?

A brief description of the PBCV-1 life cycle follows. PBCV-1 infects its host *Chlorella* NC64A by attaching rapidly, probably at a unique virus vertex [71, 58] via one of its vertices, to the algal cell wall, followed by degradation of the wall at the point of attachment [72]. After digestion of the wall, the PBCV-1 internal membrane presumably fuses with the host membrane to translocate virus DNA and probably some virus associated proteins to the inside of the host, leaving an empty capsid on the surface.

Two observations suggest that upon infection PBCV-1 DNA and probably DNA-associated proteins quickly move to the nucleus and commandeer at least some of the host transcription machinery to initiate early viral RNA synthesis; early virus transcripts can be detected within 5-10 min p.i. [73, 74]. i) PBCV-1 does not encode a recognizable RNA polymerase gene(s) nor is RNA polymerase activity detected in isolated virions (Rohozinski, J. and Van Etten, J.L., unpublished data). ii) A small intron with splice-site sequences characteristic of nuclear-spliced mRNAs exists in the PBCV-1 DNA polymerase gene [75]. Presumably, this intron is excised in the cell nucleus.

Viral DNA synthesis begins between 60 and 90 min p.i., followed by late viral transcription [76, 73]. Virus assembly begins ~3 hr p.i. in localized regions of the cytoplasm called virus assembly centers. By 6-8 hr p.i., virus induced host cell lysis occurs and results in the release of progeny viruses (~1000 virus particles per cell of which ~25% are infectious) [72].

The intracellular location of virus protein glycosylation is unknown, although gold-labelled Vp54 antibody only reacts with proteins in the virus assembly centers

(Graves, M.V., Gurnon, J.R. and Van Etten, J.L., unpublished data). This finding suggests that Vp54 might acquire its glycans in the virus assembly centers.

Table 4. Glycosyltransferase-like genes encoded by six viruses that infect three Chlorella species*.

PBCV-1		NY-2A		AR158		MT325		FR483		ATCV-1	
CDS	aa	CDS	aa	CDS	aa	CDS	aa	CDS	aa	CDS	aa
A064R	632	–	–	–	–	–	–	–	–	–	–
A111/114R	860	B159R	860	C150R	860	M467R	849	N472R	848	Z120R	846
A219/222/226R	677	–	–	C265R	634	M721L	604	N715L	629	Z425R	612
A473L	517	–	–	–	–	M186R	520	N191R	520	Z178L	533
A546L	321	B763L	405	C661L	396	–	–	–	–	–	–
–	–	B618R	270	C559R	270	–	–	–	–	–	–
–	–	–	–	–	–	–	–	–	–	Z417L	553
–	–	–	–	–	–	–	–	–	–	Z667L	546
–	–	–	–	–	–	–	–	–	–	Z823R	553

**CDS's listed in the source line across are considered to be orthologs. The amino acid sequences of each of them are available at http://GreenGene.uml.edu. Viruses PBCV-1, NY-2A and AR158 infect Chlorella NC64A, viruses MT325 and FR483 infect Chlorella Pbi and virus ATCV-1 infects Chlorella SAG 3.83.*

7. HOW ARE SUGARS ADDED TO VP54?

The structures of the glycans attached to Vp54 are unknown and we are happy to collaborate with anyone who wants to solve their structures. The prediction is that the glycan structures differ from typical glycans because it is unlikely that the host ER and Golgi apparatus are involved in adding sugars to PBCV-1 Vp54. Vp54 lacks a recognizable ER signal peptide and so presumably, it does not enter the ER. However, one can ask the question: Are the sugars added to Vp54 either sequentially or are they synthesized independently of Vp54, possibly on a lipid carrier, and then attached to the protein in a single step? A slight variation of these two possibilities is to synthesize a

core glycan(s) independently of the protein and attach it to Vp54. This core glycan(s), which could be either larger or smaller than the final product, is processed further after attachment to the protein.

To distinguish between these possibilities [2,6-^{3}H]-galactose was added to cells at the time of PBCV-1 infection and cells collected at various times p.i. [We are aware that galactose can be converted to other products.] After disrupting the cells, SDS-soluble materials were separated on SDS-PAGE, and labeled material viewed via fluorography (Fig. 3) (Graves, M.V. and Van Etten, J.L., unpublished data). Vp54 first appears between 60 and 90 min p.i. and is a tight homogeneous band suggesting that the fully synthesized glycans are transferred to the Vp54 peptide "en mass" [Note: the Vp54 equivalent from antigenic variant EPA-1 is easily distinguished from wild type Vp54 even though it is only 2,230 Da smaller; the slightly smaller variant Vp54 from mutant E11 can be distinguished as well (Fig. 3)] It is interesting that a slowly migrating radioactive smear is present at the top of the gel first appearing between 15 and 30 min p.i., prior to the appearance of Vp54; the smear increases in intensity to 90 min p.i. and then becomes less intense as Vp54 intensifies. Could this "smear" be an intermediate(s) in the synthesis of the Vp54 glycan? Interestingly, the appearance of this smear correlates with the expression of A64R, as well as the other putative glycosyltransferases; each of the glycosyltransferase gene transcripts begin to be detected at 20 to 40 min p.i. (Table 3; [Yanai-Balser, G.M. et al., manuscript in preparation]). Note: the smear is absent in normal cells even after 5 hr of continuous labeling (Fig. 3, lane U).

Assuming the Vp54 glycan precursors are transferred "en mass", one predicts that they are attached to a lipid carrier such as either undecaprenol-phosphate, which is the

carrier for bacterial peptidoglycans and cell surface polysaccharides [77], or dolicholdiphosphate, which is the carrier in eukaryotic cells [67]. It is unlikely that dolicholphosphate is the carrier because tunicamycin, which interferes with the transfer of UDP-N-acetylglucoseamine to dolicholphosphate [64], has no effect on either PBCV-1 replication or Vp54 synthesis, even at concentrations three times that required to inhibit host chlorella growth [55].

8. GLYCOSYLTRANSFERASES CODED BY OTHER CHLORELLA VIRUSES

Polyconal antibodies to wild type PBCV-1 and its antigenic variants do not react with many other independent virus isolates that infect *Chlorella* NC64A (NC64A viruses). This lack of cross-reactivity suggests that the major capsid proteins from these non-reacting viruses have different glycans and as such these viruses code unique glycosyltransferases. In addition, two other related virus/chlorella systems are available for study, viruses that infect *Chlorella* Pbi (Pbi viruses) and *Chlorella* SAG 3.83 (SAG viruses). None of the Pbi viruses or SAG viruses react with PBCV-1 antiserum.

We recently sequenced and annotated five additional chlorella viruses, two more NC64A viruses [78], two Pbi viruses [79], and one SAG virus [48]. These newly sequenced viruses encode from 3 to 6 putative glycosyltransferases (Table 4). The PBCV-1 glycosyltransferase A111/114R is the only protein with an ortholog in all of the viruses. The orthologs B618R and C559R from NC64A viruses NY-2A and AR158, respectively, produce a high Blast score with N-acetyglucosaminyltransferases, suggesting that N-acetylglucosamine might be present in their major capsid proteins. The

major capsid proteins from NY-2A and AR158 clearly differ from PBCV-1 Vp54 because neither virus reacts with antibodies to PBCV-1 or any of its antigenic variants. The sugar components of the major capsid proteins have not been determined for any of these 5 newly sequenced viruses. The prediction is that the sugars will probably differ both qualitatively and quantitatively from PBCV-1.

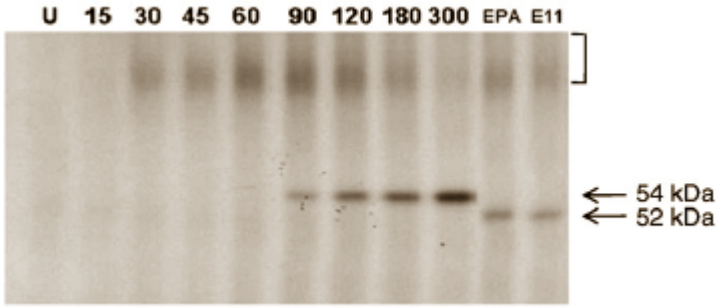


Figure 3. Total proteins isolated from virus PBCV-1 infected Chlorella NC64A at the indicated times (min) p.i. *The infection was done in the presence of [^{3}H]-galactose. Uninfected (lane U) cells were labeled for 300 min. Lanes labeled EPA and E11 contain total proteins from cells at 120 min after infection with antigenic variant viruses EPA-1 and E11. The slower migrating "smear" described in the text is indicated by the bracket. Proteins were separated on a 7.5% SDS-PAGE, and labeled proteins visualized by autoradiography (Graves, M.V. and Van Etten, J.L., unpublished data).*

9. IMPLICATIONS

The chlorella viruses contain genes encoding a variety of glycosyltransferases,

most of which, if not all, lack a recognizable ER signal peptide. Furthermore, hundreds of these viruses can be isolated easily from natural sources. We assume that different glycosyltransferases use different sugars and form distinct linkages. Because all of the virus-encoded glycosyltransferases lack ER signaling domains and most of them lack transmembrane domains, the virus encoded recombinant glycosyltransferases are likely to be soluble, which makes them easier to purify and use as reagents for adding sugars to proteins.

10. CONCLUSION

Collectively, the results indicate that glycosylation of the PBCV-1 major capsid protein differs from that of other viruses and that glycosylation probably occurs independently of the ER and Golgi apparatus. Furthermore, most, if not all, of the machinery to carry out this glycosylation is encoded by the virus. These conclusions lead to many questions including: i) Could Vp54 glycosylation reflect an ancestral pathway that existed prior to ER and Golgi formation in eukaryotic cells? ii) Is PBCV-1 glycosylation similar to that of O-antigen-containing bacteria, which add a precise number of sugars to the basal core polysaccharide [80]. iii) Are the Vp54 glycan precursors synthesized and assembled on a lipid carrier? iv) How much diversity in the number of glycosyltransferases and glycan structures exists among chlorella viruses? v) What is the ultimate function of the capsid protein glycan?

11. REFERENCES

1. Doms, R.W., Lamb R.A., Rose J.K., Helenius A. 1993. Folding and assembly of viral membrane proteins. *Virology* 193:545-562.
2. Knipe, D.M. 1996. Virus-host cell interactions, p. 273-299. *In* B.N. Fields, D.M. Knipe, P.M. Howley, R.M. Chanock, J.L. Melnick, T.P. Monath, B. Roizman, S.E. Straus (eds.), Fields Virology, 3rd ed., Lippincott-Raven Publ., Philadelphia.
3. Olofsson, S., Hansen J.E.S. 1998. Host cell glycosylation of viral glycoproteins: a battlefield for host defense and viral resistance. *Scand. J. Infect Dis.* 30:435-440.
4. Vigerust, D.J., Shepherd V.L. 2007. Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol*. 15:211-218.
5. 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.
6. Gram, H., Ruger W. 1986. The α-glucosyltransferases of bacteriophages T2, T4, and T6. A comparison of their primary structures. Mol. Gen. Genet. 202:467-470.
7. Freemont, P.S., Ruger W. 1988. Crystallization and preliminare X-ray studies of T4 phage β-glucosyltransferase. *J. Mol. Biol.* 203:525-626.
8. Winkler, M., Ruger W. 1993. Cloning and sequencing of the genes of β-glucosyl-HMC-α-glucosyl-transferases of bacteriophages T2 and T6. *Nucleic Acids Res.* 21:1500.
9. O'Reilly, D.R. 1995. Baculovirus-encoded ecdysteroid UDP-glucosyltransferases. *Insect Biochem. Mol. Biol*. 25:541-550.

10. Vanderplasschen, A., Markine-Goriaynoff N., Lomonte P. et al. 2000. A multipotential β-1,6-N-acetylglucosaminyl-transferase is encoded by bovine herpes virus type 4. *Proc. Natl. Acad. Sci. USA* 97:5756-5761.

11. Markine-Goriaynoff, N., Georgin J.P., Goltz M. et al. 2003. The core 2 b-1,6-N-acetylglucosaminyltransferase-mucin encoded by bovine herpesvirus 4 was acquired from an ancestor of the African buffalo. *J. Virol.* 77:1784-1792.

12. Markine-Goriaynoff, N., Gillet L., Karlsen O.A. et al. 2004. The core 2 b-1,6-N-acetylglucosaminyl-transferase-M encoded by bovine herpesvirus 4 is not essential for virus replication despite contributing to post-translational modifications of structural proteins. *J. Gen. Virol*. 85:355-367.

13. Alfonso, C.L., Tulman E.R., Lu Z., Oma E., Kutish G.F, Rock D.L. 1999. The genome of *Melanoplus sanguinipes* entomopoxvirus. *J. Virol*. 73:533-552.

14. Bawden, A.L., Glassberg K.J., Diggans J., Shaw R., Farmerie W., Moyer R.W. 2000. Complete genomic sequence of the *Amsacta moorei* entomopoxvirus: analysis and comparison with other poxviruses. *Virology* 274:120-139.

15. 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.

16. Derelle E., Ferraz C., Escande M.L., Eychenié S., Cooke R., Piganeau G., Desdevises Y., Bellec L., Moreau H., Grimsley N. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri. PLoS One.* 3(5):e2250.

17. 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.

18. Vestergaard, G., Aramayo R., Basta T., Haring M., Peng X., Brugger K., Chen L., Rachel R., Boisset N., Garrett R. A., Prangishvili D. 2008. Structure of the acidianus filamentous virus 3 and genomics of related archaeal lipothrixviruses. *J. Virol*. 82:371-381.

19. Xiang, X., Chen L., Huang X., Luo Y., She Q., Huang L. 2005. *Sulfolobus tengchongensis* spindle-shaped virus STSV1: interactions and genomic features. *J. Virol*. 79:8677-8686.

20. Van Etten, J.L. 2003. Unusual life style of giant chlorella viruses. *Annu Rev Genet.* 37:153-95.

21. Yamada, T., H. Onimatsu, J.L. Van Etten. 2006. Chlorella viruses, p. 293-366. *In* K. Maramorosch, A.J. Shatkin (eds.), Advances in Virus Research, Vol. 66, Elsevier Inc.

22. Dunigan D.D., Fitzgerald L.A., Van Etten J.L. 2006. Phycodnaviruses: a peek at genetic diversity. *Virus Res*. 117:119-132.

23. Wilson WH, Van Etten JL, Allen MJ. 2009. The Phycodnaviridae: the story of how tiny giants rule the world. *Curr Top Microbiol Immunol*. 328:1-42.

24. Fuhrman, J.A. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399:541-548.

25. Wilhelm, S.W., Suttle C.A. 1999. Viruses and nutrient cycles in the sea - viruses play critical roles in the structure and function of aquatic food webs. *Bioscience* 49:781-784.

26. Suttle, C.A. 2000. Marine viruses - major players in the global ecosystem. *Nat. Rev. Microbiol.* 5:801-812.

27. Wommack, K.E., Colwell R.R. 2000. Viroplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* 64:69-114.

28. Iyer, L.M., Aravind L., Koonin E.V. 2001. Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virology* 75:11720-11734.

29. Iyer L. M., Balaji S., Koonin E. V., Aravind L. 2006. Evolutionary genomics of Nucleo-Cytoplasmic Large DNA Viruses. *Virus Research* 117:156-84.

30. 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* 306:1344-50.

31. Villarreal, L.P., DeFilippis V.R. 2000. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *J. Virol*. 74:7079-7084.

32. Villarreal, L.P. 2005. Viruses and the Evolution of Life, p. 395, American Society Microbiology Press, Washington, D.C.

33. Xia, Y., Van Etten J.L. 1986. DNA methyltransferase induced by PBCV-1 virus infection of a chlorella-like green alga. *Mol. Cell. Biol*. 6:1440–445.

34. Xia, Y., Burbank D.E., Uher L., Rabussay D., Van Etten J.L. 1987. IL-3A virus infection of a chlorella-like green alga induces a DNA restriction endonuclease with novel sequence specificity. *Nucleic Acids Res*. 15:6075-6090.

35. 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*. 275:6915–21.

36. Plugge B., Gazzarrini S., Nelson M., Cerana R., Van Etten J.L., Derst C, DiFrancesco D., Moroni A., Thiel G. 2000. A potassium channel protein encoded by chlorella virus PBCV-1. *Science* 287: 1641–1644.

37. 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* 237:360–77.

38. DeAngelis, P.L., Jing W., Graves M.V., Burbank D.E., Van Etten J.L. 1997. Hyaluronan synthase of chlorella virus PBCV-1. *Science* 278:1800–1803.

39. 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.

40. DeAngelis, P.L. 1999. Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cell Mol. Life Sci*. 56: 670–682.

41. DeAngelis, P.L. 2002. Evolution of glycosaminoglycans and their glycosyltransferases: Implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. *Anat. Rec.* 268:317–326.

42. Landstein, D., Graves M.V., Burbank D.E., DeAngelis P.L., 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.

43. Kawasaki, T., Tanaka M., Fujie M., Usami S., Sakai K., Yamada T. 2002. Chitin synthesis in chlorovirus CVK2–infected chlorella cells. *Virology* 302:123–131.

44. Gooday G.W., Humphreys A.M., McIntosh W.H. 1986. *In* R. Muzzarelli, C. Jeuniaus, G.W Gooday (eds.), p 83-91, Role of chitinases in fungal growth, in Chitin in Nature and Technology, Plenum Press, New York.

45. Yamada, T., Kawasaki T. 2005. Microbial synthesis of hyaluronan and chitin: New approaches. *J Biosci Bioeng*. 99:521-528.

46. 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.

47. Fruscione, F., Sturla L., Duncan G., Van Etten J.L., Valbuzzi P., De Flora A., Di Zanni E., Tonetti M. 2008. Differential role of NADP+ and NADPH in the activity and structure of GDP-D-mannose 4,6-dehydratase from two chlorella viruses. *J Biol Chem.* 283:184-193.

48. Fitzgerald L.A., Graves M.V., Li X., Hartigan J., Pfitzner A.J., Hoffart E., Van Etten J.L. 2007. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic chlorella strain of the heliozoon *Acanthocystis turfacea*. *Virology* 362:350-361.

49. Sun, L., Adams B., Gurnon J.R., Ye Y., Van Etten J.L. 1999. Characterization of two chitinase genes and one chitosanase gene encoded by chlorella virus PBCV-1. *Virology* 263:376-387.

50. Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L. 2000. Characterization of a β-1,3-glucanase encoded by chlorella virus PBCV-1. *Virology* 276:27-36.

51. Sugimoto, I., Onimatsu H., Fujie M., Usami S., Yamada T. 2004. vAL-1, a novel polysaccharide lyase encoded by chlorovirus CVK2. *FEBS Lett*. 559:51-56.

52. Furuta, M., Schrader J.O., Schrader H.S. et al. 1997. Chlorella virus PBCV-1 encodes a homolog of the bacteriophage T4 UV damage repair gene, DenV. *Appl. Environ. Microbiol.* 63:1551-1556.

53. McCullough, A.K., Romberg M.T., Nyaga S. et al. 1998. Characterization of a novel *cis-syn* and *trans-syn II* pyrimidine dimer glycosylase/AP lyase from a eukaryotic algal virus. PBCV-1. *J. Biol. Chem.* 273:13136-13142.

54. Graves, M.V., Meints R.H. 1992. Characterization of the major capsid protein and cloning of its gene from algal virus PBCV-1. *Virology* 188:198-207.

55. Que, Q., Li Y., Wang I.N., Lane L.C., Chaney W.G., Van Etten J.L. 1994. Protein glycosylation and myristylation in chlorella virus PBCV-1 and its antigenic variants. *Virology* 203:320-327.

56. 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. *Nature Struct. Biol.* 7:101-103.

57. Simpson, A.A., Nandhagopal N., Van Etten J.L., Rossmann M.G. 2003. Structural analyses of Phycodnaviridae and Iridoviridae requires extension of the rules governing quasi-symmetry. *Acta Cryst*. D59:2053-2059.

58. Cherrier, M. V., V. A. Kostyuchenko, C. Xiao, V. D. Bowman, A. J. Battisti, X. Yan, P. R. Chipman, T. S. Baker, J. L. Van Etten, and M. G. Rossmann. 2009. An icosahedral algal virus has a complex unique vertex decorated by a spike. *Proc. Natl. Acad. Sci. USA* 106:11085-11090.

59. Wang, I-.N., Li Y., Que Q., Bhattacharya M., Lane L.C., Chaney W.G., Van Etten J.L. 1993. Evidence for virus-encoded glycosylation specificity. *Proc. Natl. Acad. Sci. USA* 90:3840-3844.

60. Graves, M.V., Bernadt C.T., Cerny R., Van Etten J.L. 2001. Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. *Virology* 285:332-345.

61. Van Etten, J.L., Burbank D.E., Kuczmarski D., Meints R.H. 1983. Virus infection of culturable chlorella-like algae and development of a plaque assay. *Science* 219:994-996.

62. Nandhagopal, N., Simpson A., Gurnon J.R. et al. 2002. The structure and evolution of the major capsid protein of a large, lipid-containing, DNA virus. *Proc. Natl. Acad. Sci. USA* 99:14758-14763.

63. Maley, F., Trimble R.B., Tarentino A.L., Plummer T.H. 1989. Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal. Biochem*. 180:195-204.

64. Elbein, A.D. 1987. Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains. *Annu. Rev. Biochem*. 56:497-534.

65. Wieland, F., Heitzer R., Schaefer W. 1983. Asparaginylglucose: novel type of carbohydrate linkage. *Proc. Natl. Acad. Sci. USA* 80:5470-5474.

66. Schreiner, R., Schnabel E., Wieland F. 1994. Novel glycosylation in eukaryotes: laminin contains the linkage unit beta-glucosylasparagine. *J. Cell Biol*. 124:1071-1081.

67. Reuter, G., Gabius H.J. 1999. Eukaryotic glycosylation: whim of nature or multipurpose tool. *Cell. Mol. Life Sci.* 55:368-422.

68. Bruckner K., Perez L., Clausen H., Cohen S. 2000. Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. *Nature* 406:411-415.

69. Yan Y.P., Schultz J., Mlodzik M., Bork P. 1997. Secreted fringe-like signaling molecules may be glycosyltrsferases. *Cell* 88:9-11.

70. Zhang Y., Xiang Y., Van Etten J.L., Rossmann M.G. 2007. Structure and function of a chlorella virus PBCV-1 encoded glycosyltransferase. *Structure* 15:1031-1039.

71. Onimatsu H., Suganuma K., Uenoyama S., Yamada T. 2006. C-terminal repetitive motifs in Vp130 present at the unique vertex of the chlorovirus capsid are essential for binding to the host chlorella cell wall. *Virology* 353:432-442.

72. 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.

73. 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.

74. Kawasaki, T., Tanaka M., Fujie M., Usami S., Yamada T. 2004. Immediate early genes expressed in chlorovirus infections. *Virology* 318:214-223.

75. 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.

76. 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.

77. Raetz, C.R., Whitfield C. 2002. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* 71:635-700.

78. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Nierman W.C., Van Etten J.L. 2007. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect Chlorella NC64A. *Virology* 358:472-484.

79. Fitzgerald, L.A., Graves M.V., Li X., Feldblyum T., Hartigan J., Van Etten J.L. 2007. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect Chlorella Pbi. *Virology* 358:459-471.

80. Bastin, D.A., Stevenson G., Brown P.K., Haase A., Reeves P.R. 1993. Repeat unit polysaccharides of bacteria: A model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol. Microbiol.* 7:725-734.