Paramecium bursaria Chlorella Virus 1 Proteome Reveals Novel Architectural and Regulatory Features of a Giant Virus

David D. Dunigan, Ronald L. Cerny, Andrew T. Bauman, Jared C. Roach, Leslie C. Lane, Irina V. Agarkova, Kurt Wulser, Giane M. Yanai-Balser, James R. Gurnon, Jason C. Vitek, Bernard J. Kronschnabel, Adrien Jeanniard, Guillaume Blanc, Chris Upton, Garry A. Duncan, O. William McClung, Fangrui Ma, and James L. Van Etten

August 2012

This article was originally published at:

http://dx.doi.org/10.1128/JVI.00907-12

JVI Journals.ASM.org

Paramecium bursaria Chlorella Virus 1 Proteome Reveals Novel Architectural and Regulatory Features of a Giant Virus

David D. Dunigan,[a,b] Ronald L. Cerny,[c] Andrew T. Bauman,[d] Jared C. Roach,[e] Leslie C. Lane,[a] Irina V. Agarkova,[a,b] Kurt Wulser,[c] Giane M. Yanai-Balser,[a] James R. Gurnon,[a] Jason C. Vitek,[a] Bernard J. Kronschnabel,[a] Adrien Jeanniard,[f] Guillaume Blanc,[f] Chris Upton,[g] Garry A. Duncan,[h] O. William McClung,[h] Fangrui Ma,[b] and James L. Van Etten[a,b]

Department of Plant Pathology, University of Nebraska—Lincoln, Lincoln, Nebraska, USA[a]; Nebraska Center for Virology, University of Nebraska—Lincoln, Lincoln, Nebraska, USA[b]; Department of Chemistry, University of Nebraska—Lincoln, Lincoln, Nebraska, USA[c]; Ocean Biologics, Seattle, Washington, USA[d]; Institute of Systems Biology, Seattle, Washington, USA[e]; Structural and Genomic Information Laboratory, UMR7256 CNRS, Aix-Marseille University, Marseille, France[f]; Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada[g]; and Department of Biology, Nebraska Wesleyan University, Lincoln, Nebraska, USA[h]

The 331-kbp chlorovirus *Paramecium bursaria* chlorella virus 1 (PBCV-1) genome was resequenced and annotated to correct errors in the original 15-year-old sequence; 40 codons was considered the minimum protein size of an open reading frame. PBCV-1 has 416 predicted protein-encoding sequences and 11 tRNAs. A proteome analysis was also conducted on highly purified PBCV-1 virions using two mass spectrometry-based protocols. The mass spectrometry-derived data were compared to PBCV-1 and its host *Chlorella variabilis* NC64A predicted proteomes. Combined, these analyses revealed 148 unique virus-encoded proteins associated with the virion (about 35% of the coding capacity of the virus) and 1 host protein. Some of these proteins appear to be structural/architectural, whereas others have enzymatic, chromatin modification, and signal transduction functions. Most (106) of the proteins have no known function or homologs in the existing gene databases except as orthologs with proteins of other chloroviruses, phycodnaviruses, and nuclear-cytoplasmic large DNA viruses. The genes encoding these proteins are dispersed throughout the virus genome, and most are transcribed late or early-late in the infection cycle, which is consistent with virion morphogenesis.

Complex cellular and viral processes are modular and are accomplished by the concerted actions of functional modules. One of the important functional modules of a virus is the virion particle, which ranges in complexity from a single type of protein and a small nucleic acid (e.g., tomato bushy stunt virus) to having dozens of types of proteins and lipids, along with a large nucleic acid genome (e.g., poxviruses). Regardless, whether they are simple or complex in composition, all virions carry the legacy of their progenitors through encapsidation, release, and stabilization. Virions facilitate the propagation of progeny through a series of tightly regulated biochemical steps called the immediate-early phase of infection, which includes attachment, penetration, uncoating of the viral genome, intracellular trafficking of the viral genome to its replication center, and augmentation of cellular functions to "accept" the exotic nucleic acid/replicon. The architectural elements of virions tend to be prominent, but studies on the supergroup nucleocytoplasmic large DNA viruses (NCLDV) (7, 36, 42) indicate that, in addition to structural components, these virions perform multiple enzymatic and regulatory functions that are partitioned among several proteins. The purpose of this study was to determine the virion proteome of *Paramecium bursaria* chlorella virus 1 (PBCV-1), a member of the NCLDV (11, 53).

PBCV-1 is the type member of the genus *Chlorovirus* (family *Phycodnaviridae*) that infects certain chlorella-like green algae from freshwater sources; these viruses are found throughout the world (53, 55). The chlorovirus host algae are normally symbionts of aquatic protists and in that state are resistant to virus infection. Nevertheless, virus titers from natural sources as high as 10^5 PFU per ml have been measured; however, the titers fluctuate with the season (57, 60). Very little is known about the role chloroviruses play in freshwater ecology (40), but susceptible hosts lyse within 6 to 16 h in the laboratory, and burst sizes typically exceed 10^2 PFU per cell (53, 55). Thus, chloroviruses have the potential to alter microbial communities both quantitatively and qualitatively, as well as to act as a driving force for microbial evolution (11). Fortunately, some of the host algae can be grown in the laboratory independent of their cosymbiotic protists.

The 331-kbp PBCV-1 double-stranded DNA (dsDNA) genome was sequenced and annotated about 15 years ago (25) and was reported to have 689 open reading frames (ORFs) of at least 65 codons. Of these 689 ORFs, 377 were predicted to be coding DNA sequences (CDSs); PBCV-1 also encoded 11 tRNAs (reviewed in references 20, 54, and 56). The size of PBCV-1 extends beyond its coding capacity; the virion is a T=169d quasi-icosahedral particle with a diameter of 190 nm across the 5-fold axis (62, 63) and has an estimated molecular mass of greater than 1×10^9 Da (52). The virion is ~64% protein, consisting of at least 40 unique polypeptides, as seen on one-dimensional SDS-PAGE (41). The particle contains 5 to 10% lipid, which is associated with a bilayered membrane underneath an outer glycoprotein shell (5, 41, 63).

The capsid structure consists of the major capsid protein

Received 12 April 2012 **Accepted** 4 June 2012
Published ahead of print 13 June 2012
Address correspondence to David D. Dunigan, ddunigan2@unl.edu, or James L. Van Etten, jvanetten1@unl.edu.
Supplemental material for this article may be found at http://jvi.asm.org/.

doi:10.1128/JVI.00907-12

(MCP) Vp54, which is glycosylated at 6 sites (30) and is myristylated at least at 1 site (35). Vp54 complexes with itself, and perhaps other proteins, to form homotrimeric capsomers that are responsible for the planar features of the capsid. Initially, it was assumed that, except for the 12 vertices, Vp54 was the only protein contributing to the external capsid, and 5,040 copies of Vp54 were predicted per virion (63). However, recent studies indicate that the PBCV-1 virion is more complex than previously thought. (i) PBCV-1 contains a unique vertex with a 560-Å-long spike structure, which protrudes 340 Å from the surface of the virus. The part of the spike structure that is outside the capsid has an external diameter of 35 Å at the tip, expanding to 70 Å at the base. The spike structure widens to 160 Å inside the capsid and forms a closed cavity inside a large pocket between the capsid and the membrane, enclosing the virus DNA (5, 65). The related chlorovirus CVK1 has a virion-associated protein, Vp130 (a homolog of PBCV-1 A140/145R), that binds to algal cell walls and is located at a unique vertex (33, 34), suggesting the protein is associated with the spike structure. (ii) Regularly spaced appendages occurring on the surface of the virion are present at approximately 1 per trisymmetron (65). These appendages probably assist in attaching the virion to its host cell (55). (iii) The volumes of the capsomers at the common vertices and those surrounding the spike structure at the unique vertex differ significantly, suggesting they consist of different proteins (5, 65). (iv) At least one vertex region may have a retractable appendage, so that when probed with a scanning atomic-force stylet, the structure retracts but then resets, much like a plunger with a spring (22). It is not known if this plunger is at the unique spike structure vertex or one of the other 11 vertices. (v) Six minor capsid proteins of varying stoichiometries support the particle architecture and appear to interact with the internal membrane in both the tri- and pentasymmetron structures, as observed with an 8.5-Å-resolution map of the virion (65). Of these, a "long protein" (~32 kDa) with similarity to the PRD1 bacteriophage long glue proteins forms a hexagonal network over the internal surface of the trisymmetrons, and a "membrane protein" dimer (~28 kDa) is located at the edge of the trisymmetrons and is connected to the internal membrane (1, 8). (vi) PBCV-1 DNA binding proteins were evaluated by proteomic methods from isolated viral DNA of virions (58). Six proteins were identified that have high isoelectric points that are well suited for binding and neutralization of DNA. Thus, the PBCV-1 structure has both symmetric and asymmetric elements, adding to the complexity of the virus morphology. (vii) In addition to these structural features, PBCV-1 contains several functions that initiate infection. PBCV-1 attaches specifically to its host, *Chlorella variabilis* NC64A. Thus, we predict that one or more surface proteins of the virus, probably the spike structure, mediate attachment (65). Immediately upon PBCV-1 attachment, the cell wall is degraded at the site of attachment. (viii) Virions contain cell wall-degrading activity (27, 61). (ix) Within the first minutes of infection, the cell membrane depolarizes (12, 31), leaving the cell with significantly altered secondary transporter functions (2). This activity is hypothesized to be partially due to a PBCV-1-encoded K^+ channel, Kcv (A250R) (26); however, no direct evidence supports the presence of Kcv in the virion. (x) In the first 5 min of infection, host DNA begins to degrade, and this is likely due to the two virus-encoded DNA restriction endonucleases [R.CviAI (A579L) and R.CviAII (A252R)] packaged in PBCV-1 virions (3). Host chromatin degradation begins before viral transcripts appear. PBCV-1 DNA is resistant to the restriction enzymes because it is methylated. (xi) The next major intracellular event is the synthesis of early viral transcripts, observed 5 to 10 min postinfection (p.i.) (66; G. Blanc, J. Gurnon, D. Dunigan, Y. Xia, and J. Van Etten, unpublished data), which apparently occurs by pirating the cellular transcriptional machinery, because the virus does not carry a recognizable RNA polymerase gene and no polymerase activity was detected in virion-derived extracts (J. Rohozinski and J. Van Etten, unpublished results).

The purpose of the current study is to evaluate the total viral complement of proteins associated with the PBCV-1 virion using proteomic technologies and to reexamine the structural/architectural features of the virus, as well as the initial events of infection in the context of the protein complement. This evaluation led to the resequencing of the PBCV-1 genome after preliminary proteomic analyses suggested there were errors in the PBCV-1 genome sequence (25). This report presents the newly revised PBCV-1 genome and annotations and proteomic analyses of the infectious particles.

MATERIALS AND METHODS

Virus, cells, and culture conditions. Procedures for growing PBCV-1 in the alga *C. variabilis* have been described previously (3, 51, 52).

Virus purification scheme. The virus was purified essentially as described previously (51) with the following modifications. Prior to sucrose density gradient separation, the virus-cell lysate (2 liters) was clarified by incubation with 1% (vol/vol) NP-40 detergent at room temperature for 1 to 2 h with constant agitation, followed by centrifugation in a Beckman type 19 rotor at 53,000 × *g* for 50 min at 4°C. The pellet fraction was solubilized in virus storage buffer (VSB) (50 mM Tris-HCl, pH 7.8), layered onto a 10 to 40% (wt/vol) linear sucrose density gradient made up in VSB, and centrifuged in a Beckman SW28 rotor for 20 min at 72,000 × *g* at 4°C. The virus band was identified by light scattering, removed from the gradient, and concentrated by centrifugation. Resuspended virus was incubated with 50 μg/ml proteinase K in VSB for 4 h at 25°C to disassociate and degrade contaminating proteins (this treatment has no effect on virus infectivity). The proteinase K-treated virus was layered onto a 20 to 40% linear iodixanol (OptiPrep; Axis-Shield, Oslo, Norway) gradient in VSB and centrifuged at 72,000 × *g* in a Beckman SW28 rotor for 4 h at 25°C for isopycnic separation. The gradient produced a single major light-scattering band at ~32% iodixanol, corresponding to a density of 1.171 g/ml. The virus band was removed by side puncture of the centrifugation tube, diluted approximately 10-fold with VSB, and then concentrated by centrifugation in a Beckman Ti50.2 rotor at 80,000 × *g* for 3 h at 4°C. The pellet fraction was resuspended in VSB and then filter sterilized with a 0.45-μm-cutoff membrane and stored at 4°C. The virus was quantified by UV/visible scanning spectroscopy using an extinction coefficient ($A_{260/0.1\%}$) of 10.7 (51) and plaque assayed to determine the number of infectious particles. These preparations typically yielded several milliliters of stock virus at 1×10^{11} to 10×10^{11} PFU/ml. The infectious-particle/total-particle ratio is normally 0.25 to 0.5 for such preparations (52).

These preparations were used for both resequencing the PBCV-1 genome and determination of the proteome; the proteome was determined by two independent methods using mass spectrometry (MS) of trypsin-digested proteins.

Resequencing and annotation of the PBCV-1 genome. Preliminary proteomic analyses using the existing PBCV-1 gene annotations (National Center for Biotechnology Information [NCBI] Refseq, NC_000852) revealed possible errors in the genome sequence, which prompted us to resequence the PBCV-1 genome. PBCV-1 DNA was purified from virions treated with DNase I, sequenced using Roche 454 Life Sciences GS FLX Titanium chemistry, and assembled as described in the supplemental material. PBCV-1 contigs were identified and annotated as described in the supplemental material.

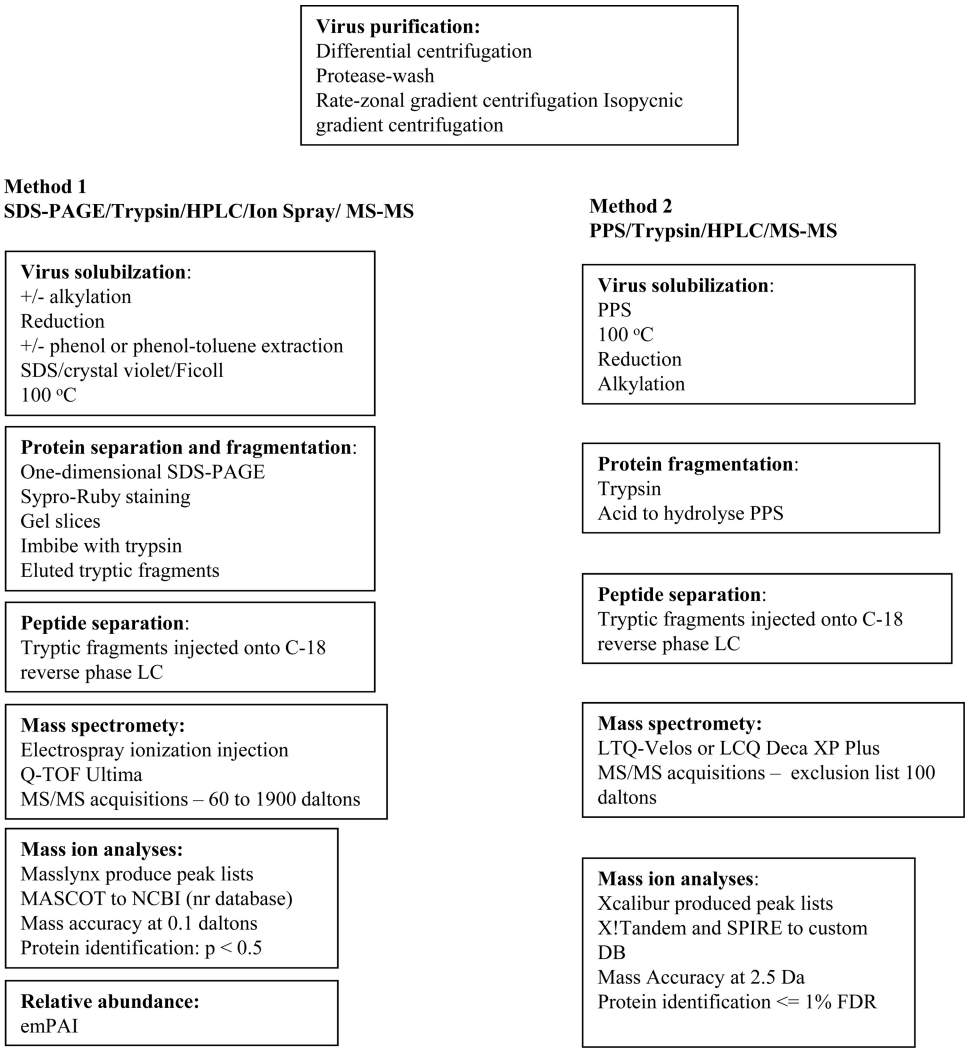

FIG 1 Proteomic methodologies for PBCV-1 virions.

Proteomics method 1: SDS-PAGE/trypsin/high-performance liquid chromatography (HPLC)/ion spray/tandem MS (MS-MS). (i) Particle disruption and protein extraction. The PBCV-1 virion proteome was evaluated by two independent methodologies (Fig. 1). In the first method, virion proteins were solubilized essentially as described previously (24), with reduction of the proteins by adjusting 50 μg of virions in 50 μl. An equal volume of cracking buffer (50 mM Tris, pH 8.5, 5 mM reducing agent dithiothreitol [DTT] [freshly reduced with tributylphosphine; in some experiments, beta-mercaptoethanol was substituted for dithiothreitol], 1% SDS, 0.1% crystal violet, and 1% Ficoll 400) was added. The sample was heated to 100°C for 3 min. The reduced proteins were subsequently alkylated by adjusting the solution to 12.5 mM iodoacetamide with a 0.25 M stock and then heating to 100°C for 1 min. These samples were immediately subjected to SDS-PAGE. Alternatively, the proteins were alkylated without previous reduction by the same procedure.

Phenolic extractions were also used to isolate virion proteins. Reduced and alkylated proteins were adjusted to 40% sucrose to increase the density of the solution. These preparations were then extracted with an equal volume of water-saturated phenol or water-saturated phenol with toluene added to increase the hydrophobicity of the phenol. The protein-containing phenolic phase was removed, and the protein was precipitated with 10 volumes of methanol and then dissolved and heated in cracking buffer.

(ii) One-dimensional SDS-PAGE. Proteins were separated on 32-cm linear-gradient (4 to 20%) polyacrylamide gels with 0.1% SDS and 375 mM Tris, pH 8.7, and tank buffer of 25 mM Tris-190 mM glycine. The samples were electrophoresed at room temperature till the crystal violet tracking dye reached the bottom of the gel.

The gel was fixed and stained with Sypro-Ruby according to the manufacturer's recommendations (Life Technologies Corporation). The stained gel was imaged using a blue-box transluminator. Once imaged, the gel was cut into 32 1-cm pieces, being careful to clean the scalpel between samples. These gel pieces were then processed for trypsin digestion and mass spectrometry analyses.

(iii) MS-based microsequencing. The excised gel pieces were digested for peptide sequencing using a slightly modified version of a method described previously (39). Briefly, the samples were washed with 100 mM ammonium bicarbonate, reduced with 10 mM DTT, alkylated with 55 mM iodoacetamide, washed twice with 100 mM ammonium bicarbonate, and digested *in situ* with 10 ng/μl trypsin. Peptides were extracted with two 60-μl aliquots of 1:1 acetonitrile-water containing 1% formic acid. The extracts were reduced in volume to approximately 25 μl using vacuum centrifugation.

Ten microliters of the extract solution was injected onto a trapping column (300 μm by 1 mm) in line with a 75-μm by 15-cm C_{18} reversed-

phase LC column (LC- Packings). Peptides were eluted from the column using a water plus 0.1% formic acid (A)/95% acetonitrile-5% water plus 0.1% formic acid (B) gradient with a flow rate of 270 μl/min. The gradient was developed with the following time profile: 0 min 5% B, 5 min 5% B, 35 min 35% B, 40 min 45% B, 42 min 60% B, 45 min 90% B, 48 min 90% B, and 50 min 5% B.

The eluting peptides were analyzed using a Q-TOF Ultima tandem mass spectrometer (Micromass/Waters, Milford, MA) with electrospray ionization. Analyses were performed using data-dependent acquisition (DDA) with the following parameters: a 1-s survey scan (380 to 1,900 Da), followed by up to three 2.4-s MS-MS acquisitions (60 to 1,900 Da). The instrument was operated at a mass resolution of 8,000. The instrument was calibrated using fragment ion masses of doubly protonated Glu-fibrinopeptide.

(iv) Mass ion analyses. The MS-MS data were processed using Masslynx software (Micromass) to produce peak lists for database searching. MASCOT (Matrix Science, Boston, MA) was used as the search engine. The data were searched against the NCBI nonredundant database. The following search parameters were used: mass accuracy, 0.1 Da; enzyme specificity, trypsin; fixed modification, CAM; and variable modification, oxidized methionine. Protein identifications were based on random probability scores with a minimum value of 25. Although this number varied from experiment to experiment, typically it was 25 or less for confidence at a P value of <0.05.

(v) Relative abundances. Approximate relative quantitation of the proteins was done using the exponentially modified protein abundance index (emPAI) (17). This method uses the number of observed peptides compared to the number of observable peptides, giving a ratio that is directly proportional to the relative abundance of the protein in the mixture when adjusted exponentially ($\text{emPAI} = 10^{\text{PAI}} - 1$, where PAI is the number of observed peptides per protein divided by the number of observable peptides per protein).

Proteomics method 2: PPS/trypsin/HPLC/MS-MS. (i) Protein extraction and trypsin digest. One hundred micrograms of PBCV-1 was mixed 1:1 with 100 mM ammonium bicarbonate buffer, pH 8.3, containing 0.2% PPS (Protein Discovery Laboratories, San Diego, CA; final concentration, 50 mM ammonium bicarbonate, 0.1% PPS), boiled for 5 min, cooled to room temperature, reduced, alkylated with 5 mM dithiothreitol and 15 mM iodoacetamide, and then digested with sequencing grade trypsin at a 1:50 trypsin/protein ratio for 4 h at 37°C with shaking. The digested samples were acidified with HCl (200 mM), incubated at 37°C, and centrifuged at 4°C to remove the PPS prior to LC-MS application.

(ii) LC methods. Buffer solutions were made with LC-MS grade water, acetonitrile, and formic acid and consisted of 5% acetonitrile-0.1% formic acid in water (buffer A) and 100% acetonitrile-0.1% formic acid (buffer B). Two or 4 μg total protein from each sample was loaded onto a reverse-phase (RP) trap (Magic; 5 μm, 200 Å; Michrom Bioresources, Auburn, CA) with 100% buffer A and washed for 10 min prior to separation on a microcapillary column. The microcapillary column was constructed by slurry packing 18 cm of C_{18} material (HALO; 2.7 μm, 100 Å; Michrom Bioresources) into a 75-μm (inside diameter [i.d.]) fused silica capillary, which was previously pulled to a tip diameter of 5 μm using a Sutter Instruments laser puller (Sutter Manufacturing, Novato, CA). Separations were performed on an Eksigent (Dublin, CA) 1D+ nano-LC (LCQ-Deca XP Plus, 0 to 30% B over 240 min, 30 to 70% B over 10 min at 300 μl/min; LTQ-Velos, 0 to 30% B over 80 min, 35 to 70% B over 10 min at 300 μl/min).

(iii) Mass spectrometry methods. Data-dependent MS-MS analysis was performed using an LTQ-Velos or LCQ Deca XP Plus mass spectrometer (ThermoFisher, San Jose, CA). Full MS spectra were acquired in centroid mode, with a mass range of 400 to 2,000 Da. To prevent repetitive analysis, dynamic exclusion was enabled with a LTQ-Velos: repeat count, 1; repeat duration, 30 s; exclusion list size, 500; and exclusion duration, 90 s. Tandem mass spectra were collected using a normalized collision energy of 35% and an isolation window of 3 Da.

For the LTQ, one full scan was followed by 6 MS-MS scans of the 6 most intense precursor ions not on the dynamic-exclusion list. LCQ-Deca XP Plus: repeat count, 1; repeat duration, 30 s; exclusion list size, 100; and exclusion duration, 20 s. Tandem mass spectra were collected using a normalized collision energy of 35% and an isolation window of 4 Da.

For the LCQ, one full scan was followed by 3 MS-MS scans of the 3 most intense precursor ions not on the dynamic-exclusion list.

(iv) Mass ion analyses. Processing and searching of MS-MS spectra and analyzing peptide and protein identification data were performed using the SPIRE (Systematic Protein Investigative Research Environment [http://www.proteinspire.org]) system with default parameters. Searches were conducted using the X!Tandem search engine (9) with a 2.5-Da mass error, a variable modification for methionine oxidation (16@M), and a fixed modification for iodoacetamide (57@C), along with the default search parameters. The sequence file for the searches of the modules contained PBCV-1 appended to a decoy database of *Ostreococcus tauri*. In addition, a randomly reshuffled version of each database was appended for error estimation. The search results were processed with the LIPS (logistic identification of peptide sequences) model (15) to generate peptide spectrum scores. Peptide identification probabilities and false-discovery rates (FDR) were calculated based on the reshuffled matches using an isotonic-regression model (16). A 90% certainty was used as the basis for spectrum identifications. A recently introduced approach was used to estimate the protein identification FDR from individual peptide identification probabilities (16).

Nucleotide sequence accession number. The genome sequence and annotation are deposited at the NCBI as reference sequence NC_000852.5.

RESULTS AND DISCUSSION

Resequenced and reannotated PBCV-1 genome. The original sequence and annotation of PBCV-1 were completed over 15 years ago using primitive procedures compared to current technology. During the past 15 years, we have corrected the sequences of individual genes as mistakes were detected. Those mistakes and preliminary results from the current proteomic analyses that indicated sequencing errors prompted us to resequence PBCV-1. The revised PBCV-1 genome contains 330,805 nucleotide pairs compared to 330,743 nucleotide pairs from the earlier sequencing effort. The two genome versions differed by 458 indel positions (mostly single-nucleotide indels) and 188 substitutions. The genome annotation is listed in Table S1 in the supplemental material. The resequenced genome submitted to NCBI includes the 2,222-bp terminal-inverted-repeat ends, but not the incompletely base-paired covalently closed hairpin 35-nucleotide loops at each end of the genome. Thus, the genome is a linear double-stranded DNA of 330,805 bp with two 35-nucleotide partially paired terminal loops. Sequencing reads were obtained through the hairpin loops (data not shown). The terminal repeats and hairpin loops are identical to the published results of Zhang et al. (67). Nucleotide 1 refers to the first paired nucleotide following the hairpin loop.

One significant change in the new annotation is that ORFs of 40 codons or more were classified as potential CDSs; the previous annotation used 65 codons as the minimum size. This resulted in 802 ORFs, of which 416 ORFs were classified as "major" CDSs (designated with an upper case A) based on the following supporting evidence: these ORFs did not have larger overlapping ORFs and/or were expressed transcriptionally (64) and/or the protein was identified in the proteomic analyses. The major ORFs cover 92.8% of the genome sequence and have an average protein prod-

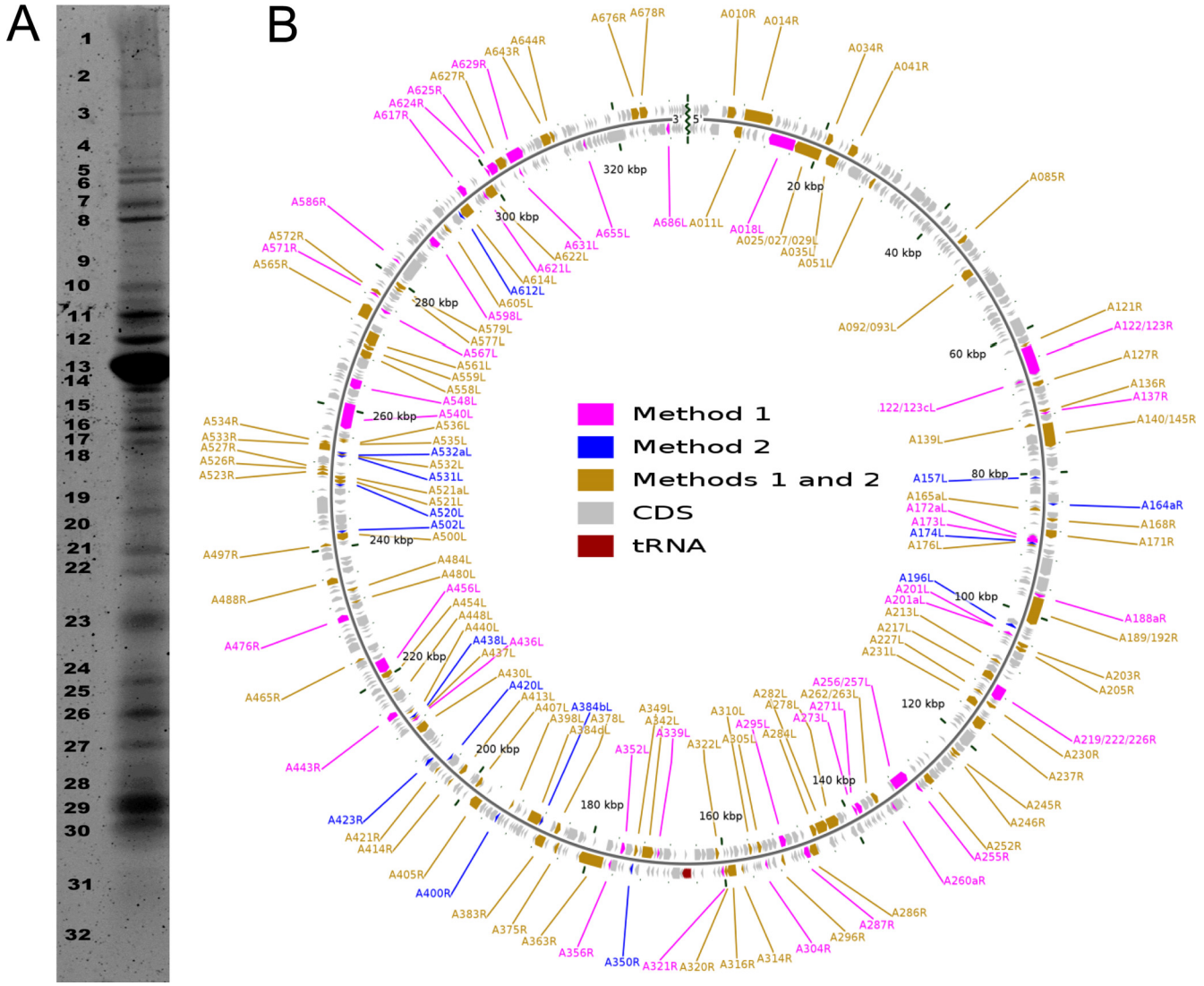

FIG 2 SDS-PAGE protein separation and virion proteome mapping onto the PBCV-1 genome. (A) Distribution of virion proteins with SDS polyacrylamide gel separation. The numbers on the left indicate the gel fragment that was analyzed. (B) The PBCV-1 genome was resequenced, assembled, and annotated to correct existing sequence errors. The 416 predicted CDSs are represented as gray arrows running both clockwise and counterclockwise along the genome. Note that the diagram is circular, but there is a break at the 12 o'clock position because the viral genome is a linear molecule with terminal inverted repeats and closed hairpin ends. The terminal sequences (inverted repeats and hairpin ends) were found to be identical to those reported previously (67). The polycistronic gene encoding 11 tRNAs is presented in red (at 6 o'clock). The 148 proteins of the virion proteome were determined using two independent mass spectrometry-based methods (see Materials and Methods). The results of each method are shown. The map was developed with CGView software (43).

uct size of 249 amino acids. In addition, 11 tRNA genes were identified, as reported previously. The remaining 386 ORFs were labeled "minor" ORFs (designated with a lower case a), and most of them are probably not CDSs. They encode putative proteins with an average size of 86 amino acids. The gene annotations, along with functional assignments, are listed in Table S1 in the supplemental material.

To avoid confusion in the literature, we kept the same gene numbering system used previously, i.e., a gene labeled *a250r* is still labeled *a250r*. When two adjacent ORFs were found to be a single ORF, e.g., A189R and A192R, we named it A189/192R. Finally, where smaller ORFs that were not considered previously were identified, we labeled them with a lowercase letter, e.g., A254aR. These new gene annotations were used for the proteomic analyses of the virion proteins.

PBCV-1 virion proteome. Highly purified virions were used for the proteome analyses, including a "protease treatment" step in which the particles were incubated with proteinase K to degrade proteins nonspecifically associated with the particle surface. Proteinase K treatment does not affect PBCV-1 infectivity (3). Using a combination of sample treatment, separation, and mass spectrometry methods, 148 virus-encoded proteins were detected in the PBCV-1 virion (Fig. 2B). For abundant proteins, any method was sufficient to detect mass ions, allowing identification with high confidence. However, some of the low-abundance and small proteins were identified by only one of the two methods, primarily due to differential separation, where the protein of interest was separated from an abundant, and consequently masking, protein. The dynamic range of these analyses was $\sim 10^4$, with the MCP present at approximately 10^3 copies per virion relative to a hypo-

thetical protein present at 1 copy per virion. Thus, the sample treatment and separation method selected were important elements in the proteome determination. The proteins were identified by two independent methods, and 62% of the proteins were detected by both methods. Twenty-six percent were uniquely identified by the SDS-PAGE method (method 1), and 11% were uniquely identified by the PPS solubilization method (method 2). It is important to note that some proteins are not readily detected using mass spectrometric methods, e.g., small proteins associated with membranes (38). Thus, the proteome reported here may increase with additional data in the future. However, the results presented are a compilation of many experiments under varying conditions for protein extraction and isolation, giving us high confidence in the compiled list of proteins, including several proteins with predicted transmembrane domains, as well as many small proteins, i.e., less than 10 kDa (Table 1).

Method 1: SDS-PAGE/trypsin/HPLC/ion Spray/MS-MS. Method 1 identified 132 virus-encoded proteins in the virion. Virion proteins were either (i) extracted directly into the gel sample buffer, (ii) first extracted into a phenolic phase to remove nucleic acids, or (iii) extracted into a hypopolarized phenolic phase supplemented with toluene to further extract highly polar proteins, such as glycosylated proteins. The extracted proteins either were alkylated with iodoacetamide and then reduced or were left alkylated. While these methods helped extract certain proteins, others were excluded, and no additional proteins were detected beyond the standard method of extraction into the gel sample buffer.

Protein separation using one-dimensional gel electrophoresis resolved ~30 distinct Sypro-Ruby-stained bands. The dynamic range of observed polypeptides is large. For example, the MCP migrates at approximately 54 kDa and is the most abundant protein in the virion, migrating near the midpoint of the gel (Fig. 2A, gel position 13). The MCP has a nominal mass of 48 kDa and is posttranslationally modified with sugars at 6 positions (30) and with at least one myristyl group (35), as well as having the amino-terminal methionine removed (13). This very abundant protein contrasts with proteins detected in regions of the gel where little or no staining was observed, e.g., gel positions 1, 8, 9, 31, and 32 in Fig. 2A. Although very little staining was observed in these regions, several proteins were detected by the mass spectrometry analyses. Indeed, proteins were detected in all regions of the gel.

Qualitative changes in protein mobility were observed with different sample treatments (see Fig. S1 in the supplemental material). Samples that were alkylated with iodoacetamide gave nearly the same number of bands as those that were reduced with dithiothreitol (or beta-mercaptoethanol) and alkylated. However, the mobilities of a few proteins were altered by this differential treatment, as visualized by Sypro-Ruby staining. For example, a protein band(s) migrating at gel position 5 in the alkylated sample is absent in the sample that was both reduced and alkylated. Conversely, proteins observed at gel positions 7 and 8 for the reduced and alkylated sample are not visible in samples that were only alkylated. Several other differentials occurred between these two treatments; nevertheless, the protein profiles determined for the treatments were similar for the prominent proteins. The use of multiple treatment and separation methods was most useful for low-abundance polypeptides, as indicated by the MASCOT score.

Method 2: trypsin/HPLC/MS-MS. The trypsin/HPLC/MS-MS method identified 109 virus-encoded proteins, 16 of which were unique to the method. All tryptic or semitryptic peptide matches were analyzed using the SPIRE analysis suite (14–16) against PBCV-1 and *C. variabilis* genome databases. Restricting the matches to tryptic peptides did not decrease the false-positive rate, so full semitryptic searching was employed. The false-positive rate was estimated from searches of a decoy database of the *O. tauri* proteome. The false-positive rate was computed to be 0.42%, so one of the 109 proteins identified in this group of experiments might be a false positive. All the proteins identified had a confidence level of high or very high in at least one of the 10 analyses in the group and were considered to be in the virion.

Of the 10 analyses performed by this method, 6 proteins were detected in only one analysis. One of the proteins was found in 2 analyses, one in 3 analyses, 4 in four analyses, 21 in 5 analyses, 2 in 6 analyses, 2 in 9 analyses, and 89 in all 10 analyses. The number of analyses in which a protein is observed can be influenced by either variability inherent in mass spectrometry-based proteomics experiments, variability in expression, stability of the proteins, or false-positive results.

The proteome is L strand and R biased. The genes predicted to encode proteins in the PBCV-1 genome are biased to the right side (R) (262 of 416) relative to the midpoint of the genome; this is also reflected in the number of gene products in the proteome (81 CDSs from the right side and 67 CDSs from the left side) (Fig. 3). In addition, there is a bias for the reverse (L) strand for the right half of the genome in both the total predicted proteins (159 of 416) (Fig. 3A) and the virion proteome (48 of 148) (Fig. 3B). This bias is consistent with certain viable PBCV-1 spontaneous large deletion mutants, where up to 40 kbp of the left side of the genome can be deleted (23, 53), and they are recapitulated in the chlorovirus CVK2 (6). The right side L strand virion-coding genes have a mean G+C content of 22%, whereas the overall G+C content of the genome is 40% and the mean G+C content of all the coding genes is 31%. These observations suggest the left side of the genome has less selection pressure than the right side for the essential functions of virion assembly and maturation. Virion-associated genes (38% of the total) with atypical nucleotide compositions are relatively dense in the right-side L strand, whereas virion proteins are relatively sparse (14%) in the corresponding left side of the genome.

The proteome is skewed toward small basic proteins. The PBCV-1 proteome has proteins ranging in mass from 4.9 to 143 kDa and in isoelectric points from 3.6 to 13.0, assuming there are no posttranslational modifications (Fig. 4). Quantitatively, the proteome is dominated by the MCP, centrally located in these distributions. Qualitatively, the proteome is skewed toward basic (~75%) and relatively small proteins: approximately 50% are less than 20 kDa, and 63% of the proteins have molecular masses of less than 50 kDa and pI values greater than 7.0. This skewing toward the more basic side is interesting because the electrostatic charges of the 6×10^5 phosphate moieties in the virus genome are probably neutralized by basic proteins (58). However, this prediction must be evaluated further, because the stoichiometry of the virion proteins is uncertain. Additionally, how they relate to the chlorovirus CVK2 proteins with DNA binding and protein kinase activities needs to be clarified (59).

Two-dimensional (2-D) gel analyses using isoelectric focusing versus mass separations support the skewing toward basic and small proteins, suggesting that the majority of these proteins are not posttranslationally modified in a way that causes significant deviations of the predicted charge-mass migration (results not

TABLE 1 PBCV-1 virion proteome

Protein (CDS)	Mass (Da)	pI	Expression stage	Function or putative function	Proteomic method(s)	TM prediction[a]		
						T	H	P
A010R	44,998	5.2	Late	Capsid protein; PfamA, PF4451.5 [1.9e−50]	1, 2	0	0	0
A011L	45,076	5.4	Late	Capsid protein; PfamA, PF4451.5 [2.9e−61]	1, 2	0	0	0
A014R	141,382	6.3	Late	Unknown protein	1, 2	0	0	0
A018L	137,639	4.9	Late	Unknown protein; PfamA, PF06598.4 (chlorovirus glycoprotein repeat) [1.2e−11]	1	0	0	0
A025/027/029L	140,095	4.4	Late	Unknown protein	1, 2	0	0	0
A034R	35,163	10.4	Late	Protein kinase; PfamA, PF00069.18 (protein kinase domain) [1.4e−07]	1, 2	0	1	0
A035L	65,606	8.9	Late	Unknown protein	1, 2	0	1	0
A041R	44,315	10.8	Late	Unknown protein	1, 2	0	1	0
A051L	22,804	8.6	Late	Unknown protein	1, 2	1	2	1
A085R	27,812	7.8	Late	Prolyl 4-hydroxylase; PfamA, PF03171.13 [2OG-Fe(II) oxygenase superfamily] [3.5e−11]	1, 2	1	1	1
A092/093L	49,577	10.7	Early-late	Unknown protein; PfamA, PF08789.3 (PBCV-specific basic adaptor domain) [1.2e−15]	1, 2	0	0	0
A121R	12,486	10.8	Early-late	Unknown protein	1, 2	0	0	0
A122/123cL	4,912	10.1	NA[b]	Unknown protein	1	0	0	0
A122/123R	137,880	5.0	Late	COG5295 (autotransporter adhesin) [4e−12]; PfamA, PF06598.4 (chlorovirus glycoprotein repeat) [3.6e−11]/PF11962.1 (domain of unknown function [DUF3476]) [8.2e−66]	1	0	31	0
A127R	27,126	10.1	Late	Unknown protein	1, 2	0	0	0
A136R	16,367	11.5	NA	Unknown protein	1, 2	0	0	0
A137R	8,777	10.9	Early	Unknown protein	1	0	0	0
A139L	17,701	8.4	Late	Unknown protein	1, 2	2	2	2
A140/145R	120,898	11.0	Early-late	Unknown protein	1, 2	0	1	0
A157L	12,328	3.9	Early-late	Unknown protein	2	1	1	1
A164aR	7,094	5.8	NA	Unknown protein	2	1	0	0
A165aL	19,024	10.1	NA	Unknown protein	1, 2	0	0	0
A168R	18,317	4.6	Late	Unknown protein	1, 2	1	1	1
A171R	42,413	10.2	Early	Unknown protein	1, 2	0	0	0
A172aL	6,053	9.8	NA	Unknown protein	1	1	1	0
A173L	31,933	8.2	Early	COG1752 (predicted esterase of the alpha-beta hydrolase superfamily) [2e−06]; PfamA, PF01734.15 (patatin-like phospholipase) [4.2e−27]	1	0	2	0
A174L	7,453	12.2	NA	Unknown protein	2	0	0	0
A176L	9,167	11.3	NA	Unknown protein; PfamA, PF08789.3 (PBCV-specific basic adaptor domain) [9e−12]	1, 2	0	0	0
A188aR	17,326	10.0	NA	COG0417 (DNA polymerase elongation subunit [family B]) [3e−07]; PfamA, PF00136.14 (DNA polymerase family B) [6.5e−17]	1	0	0	0
A189/192R	143,575	11.4	Late	Unknown protein	1, 2	0	0	0
A196L	17,456	8.4	Late	Unknown protein	2	3	3	1
A201aL	6,787	8.8	NA	Unknown protein	1	0	0	0
A201L	10,005	10.7	Early-late	Unknown protein	1	2	2	2
A202L	12,232	5.0	Early-late	Unknown protein	2	0	0	0
A203R	24,011	6.0	Late	Unknown protein	1, 2	1	2	0
A205R	22,452	12.1	Late	Unknown protein; PfamA, PF08789.3 (PBCV-specific basic adaptor domain) [4.2e−16]	1, 2	0	0	0
A213L	16,483	4.5	Early-late	Unknown protein	1, 2	1	1	1
A217L	45,248	9.9	Early-late	Unknown protein	1, 2	0	0	1
A219/222/226R	77,797	7.0	Early	COG1215 (glycosyltransferases probably involved in cell wall biogenesis) [4e−06]; Swissprot, P58932 (RecName, FullCellulose synthase catalytic subunit [UDP forming]) [6e−07]	1	9	8	10
A227L	15,689	10.0	Late	Unknown protein	1, 2	0	0	0
A230R	22,055	8.4	Late	Unknown protein	1, 2	4	4	4
A231L	43,644	9.9	Early-late	Unknown protein	1, 2	1	0	0
A237R	58,565	9.5	Late	Homospermidine synthase	1, 2	0	0	0
A245R	19,748	9.3	Late	Cu/Zn superoxide dismutase	1, 2	1	1	0
A246R	12,017	11.5	Late	Unknown protein	1, 2	0	0	0
A252R	39,856	10.3	Early	R.CviAII restriction endonuclease	1, 2	0	0	0
A255R	17,300	5.1	NA	Unknown protein	1	0	0	0
A256/257L	96,729	7.2	Early-late	Unknown protein	1	0	0	0

(Continued on following page)

TABLE 1 (Continued)

Protein (CDS)	Mass (Da)	pI	Expression stage	Function or putative function	Proteomic method(s)	TM prediction[a]		
						T	H	P
A260aR	7,742	11.9	NA	Unknown protein	1	0	0	0
A262/263L	29,470	9.6	NA	Unknown protein	1, 2	2	3	2
A271L	31,114	7.1	Early-late	COG2267 (lysophospholipase) [1e−07]	1	0	3	0
A273L	15,713	9.9	Late	PF03713.6 (domain of unknown function [DUF305]) [6.8e−13]	1	3	3	3
A278L	69,231	10.8	Late	Protein kinase; PfamA, PF00069.18 (protein kinase domain) [1.2e−07]/PF08789.3 (PBCV-specific basic adaptor domain) [7.5e−10]	1, 2	0	1	0
A282L	63,371	10.8	Late	Protein kinase; PfamA, PF00069.18 (protein kinase domain) [1.2e−07]/PF08789.3 (PBCV-specific basic adaptor domain) [1.3e−17]	1, 2	0	1	0
A284L	30,766	9.2	Early-late	Aminidase	1, 2	0	0	0
A286R	43,042	9.6	Late	Unknown protein	1, 2	0	0	0
A287R	31,349	9.4	Early-late	PfamA, PF01541.17 (GIY-YIG catalytic domain) [4.2e−11]/PF07453.6 (NUMOD1 domain) [8.6e−11]	1	0	0	0
A295L	35,626	7.9	Early-late	Fucose synthetase; Swissprot, Q9LMU0 (RecName, FullPutative GDP-L-fucose synthase 2; AltName, FullGDP-4-keto-6-deoxy-D-mannose-3 5-epimerase-4-reductase 2 ShortAtGER2) [1e−100]	1	0	0	0
A296R	17,393	12.2	Late	Unknown protein	1, 2	0	1	1
A304R	9,490	5.8	Late	Unknown protein	1	0	0	0
A305L	22,910	10.7	Late	Protein phosphatase; Swissprot, Q9BY84 (RecName, FullDual specificity protein phosphatase 16; AltName, FullMitogen-activated protein kinase phosphatase 7 ShortMAP kinase phosphatase 7 ShortMKP-7) [7e−12]	1, 2	0	0	0
A310L	18,268	8.5	Late	Unknown protein	1, 2	0	0	0
A314R	9,114	6.7	Late	Unknown protein	1, 2	1	1	1
A316R	48,779	10.7	Late	Unknown protein	1, 2	0	1	0
A320R	15,685	10.5	Late	Unknown protein	1, 2	1	1	1
A321R	12,830	8.8	Late	Unknown protein	1	2	2	2
A322L	20,039	5.0	Late	Unknown protein	1, 2	1	1	1
A339L	7,372	11.1	Early-late	Unknown protein	1	0	0	0
A342L	63,813	9.2	Early-late	Unknown protein	1, 2	1	1	1
A349L	21,077	10.0	Early-late	Unknown protein	1, 2	0	1	0
A350R	14,676	9.7	NA	PfamA, PF12239.1 (protein of unknown function [DUF3605]) [4.4e−23]	2	0	0	0
A352L	23,310	3.6	Late	Swissprot, Q5UQF7 (RecName, FullUncharacterized protein R489; Flags, Precursor) [1e−05]	1, 2	0	1	1
A356R	12,512	10.5	NA	Unknown protein	1	0	0	0
A363R	128,448	10.9	Early	Swissprot, P0C9B2 (RecName, FullPutative ATP-dependent RNA helicase Q706L) [2e−06]	1, 2	0	2	0
A375R	19,085	9.4	Early-late	Unknown protein	1, 2	2	2	2
A378L	29,219	9.4	Late	Unknown protein	1, 2	1	1	0
A383R	52,511	5.2	Late	Capsid protein; Pfam, PF04451.5 (large eukaryotic DNA virus major capsid protein) [1.6e−25]	1, 2	0	0	0
A384bL	6,809	9.0	NA	Unknown protein	2	1	1	1
A384dL	69,009	8.0	Early-late	Capsid protein; PfamA, PF01607.17 (chitin binding peritrophin-A domain) [2.4e−07]/PF04451.5 (large eukaryotic DNA virus major capsid protein) [2e−11]	1, 2	1	2	1
A398L	12,987	9.9	Late	Unknown protein	1, 2	2	3	3
A400R	13,634	9.5	Early-late	Unknown protein	2	0	0	0
A405R	53,502	10.3	Late	Unknown protein	1, 2	1	2	1
A407L	23,382	8.9	Late	Unknown protein	1, 2	1	2	2
A413L	26,998	9.5	Late	Unknown protein	1, 2	2	2	2
A414R	10,612	10.8	Late	Unknown protein	1, 2	2	2	2
A420L	7,918	6.4	Late	Unknown protein	2	1	1	1
A421R	11,056	10.1	Late	Unknown protein	1, 2	1	1	1
A423R	18,458	6.5	Late	Unknown protein	2	0	1	0
A430L	48,165	7.5	Late	Major capsid protein	1, 2	0	0	0
A436L	6,932	13.0	NA	Unknown protein; Pfam, PF08789.3 (PBCV-specific basic adaptor domain) [1.5e−16]	1	0	0	0
A437L	10,876	11.0	Late	PfamA, PF05854.4 (nonhistone chromosomal protein MC1) [5.9e−07]	1, 2	0	1	0
A438L	8,988	10.7	Early-late	Glutaredoxin	2	0	0	0
A440L	10,112	11.1	Early	Unknown protein	1, 2	0	0	0
A443R	34,961	5.3	Early	Unknown protein	1	0	0	0

(Continued on following page)

TABLE 1 (Continued)

Protein (CDS)	Mass (Da)	pI	Expression stage	Function or putative function	Proteomic method(s)	TM prediction[a]		
						T	H	P
A448L	12,369	10.4	Late	Protein disulfide isomerase with heme binding site	1, 2	0	0	0
A454L	31,194	4.7	Early-late	Unknown protein	1, 2	1	1	0
A456L	75,235	5.5	Early	COG3378 (predicted ATPase) [3e−06]; PfamA, PF08706.4 (D5 N terminal like) [3.9e−09]	1	0	0	0
A465R	13,528	10.2	Early-late	COG5054 (mitochondrial sulfhydryl oxidase involved in the biogenesis of cytosolic Fe/S proteins) [4e−06]; PfamA, PF04777.6 (Erv1/Alr family) [3.5e−22]	1, 2	0	0	0
A476R	37,393	4.4	Early-late	Swissprot, Q6Y657 (RecName, FullPutative ribonucleoside-diphosphate reductase small chain B; AltName, FullRibonucleotide reductase small subunit B; AltName, FullRibonucleoside-diphosphate reductase R2B subunit) [1e−113]	1	0	0	1
A480L	9,838	10.0	Late	Unknown protein	1, 2	2	2	2
A484L	18,604	9.6	Early-late	Unknown protein	1, 2	0	0	0
A488R	34,631	5.0	Late	Swissprot, Q5UQL4 (RecName, FullUncharacterized protein L417) [2e−09]	1, 2	0	3	0
A497R	15,378	10.4	Late	Unknown protein	1, 2	2	2	1
A500L	38,463	5.0	NA	Unknown protein	1, 2	1	2	1
A502L	11,069	9.4	Late	Unknown protein	2	1	1	1
A520L	11,674	10.7	Late	Unknown protein	2	0	0	0
A521aL	22,578	6.3	NA	Swissprot, O55742 (RecName, FullUncharacterized protein 136R) [2e−07]	1, 2	0	0	0
A521L	23,738	11.4	Early-late	Unknown protein	1, 2	0	0	0
A523R	19,096	9.6	Late	Unknown protein	1, 2	0	0	0
A526R	16,434	9.3	Late	Unknown protein	1, 2	0	1	0
A527R	11,605	10.7	Late	Unknown protein	1, 2	0	0	0
A531L	7,670	7.5	Late	Unknown protein	2	1	1	1
A532aL	5,479	4.5	NA	Unknown protein	2	1	1	1
A532L	8,698	9.7	Late	Unknown protein	1, 2	1	1	1
A533R	40,132	3.8	Early-late	Unknown protein	1, 2	0	0	0
A534R	11,783	9.7	NA	Unknown protein	1, 2	0	0	0
A535L	8,210	4.7	Early-late	Unknown protein	1, 2	0	0	0
A536L	8,485	10.0	Early-late	Unknown protein	1, 2	1	1	0
A540L	127,197	6.2	Late	Unknown protein	1	0	0	0
A548L	57,432	9.5	Early	PfamA, PF00176.16 (SNF2 family N-terminal domain) [6.7e−34]/PF00271.24 (helicase conserved C-terminal domain) [1.5e−10]	1	0	0	0
A558L	45,547	5.1	Early-late	Capsid protein; PfamA, PF04451.5 (large eukaryotic DNA virus major capsid protein) [6.6e−60]	1, 2	0	0	0
A559L	24,034	10.2	Late	Unknown protein	1, 2	1	1	0
A561L	71,004	9.9	Late	Unknown protein	1, 2	1	2	1
A565R	73,169	7.3	Early-late	Unknown protein	1, 2	1	1	1
A567L	17,418	10.1	Early-late	Unknown protein	1	0	0	0
A571R	12,972	12.0	Late	Pfam hit, PF08789.3 (PBCV-specific basic adaptor domain) [5.7e−17]; Refseq best hit, YP_001426112 (hypothetical protein FR483_N480R [*Paramecium bursaria* chlorella virus FR483]) [3e−39]	1	0	0	0
A572R	20,606	7.1	Late	Unknown protein	1, 2	0	0	0
A577L	15,442	11.0	Late	Unknown protein	1, 2	0	0	0
A579L	27,445	10.1	Late	R.CviAI restriction endonuclease	1, 2	0	0	0
A586R	8,567	11.8	NA	Unknown protein	1	0	0	0
A598L	41,558	6.9	Early-late	COG0076 (glutamate decarboxylase and related PLP-dependent proteins) [5e−06]; PfamA, PF00282.12 (pyridoxal-dependent decarboxylase conserved domain) [1.1e−17]	1	0	0	0
A605L	17,769	10.9	Early-late	Unknown protein	1, 2	1	1	1
A612L	13,587	8.7	Late	Histone H3K27 methylase	2	0	0	0
A614L	64,733	11.2	Late	Protein kinase; PfamA, PF00069.18 (protein kinase domain) [5.6e−11]	1, 2	0	0	0
A617R	37,586	9.9	Early-late	Swissprot, Q5UQJ6 (RecName, FullPutative serine/threonine-protein kinase R400) [7e−12]	1	0	0	0
A621L	12,935	9.5	Late	Unknown protein	1	2	2	2
A622L	58,097	5.7	Late	Capsid protein; PfamA, PF04451.5 (large eukaryotic DNA virus major capsid protein) [1.7e−66]	1, 2	0	0	0
A624R	13,570	9.3	Late	Unknown protein; PfamA, PF09945.2 (Predicted membrane protein [DUF2177]) [3.4e−26]	1	3	4	3

(Continued on following page)

TABLE 1 (Continued)

Protein (CDS)	Mass (Da)	pI	Expression stage	Function or putative function	Proteomic method(s)	TM prediction[a]		
						T	H	P
A625R	49,945	10.7	Late	COG0675 (transposase and inactivated derivatives) [1e−06]; PfamA, PF12323.1 (helix-turn-helix domain) [1.4e−06]/PF07282.4 (putative transposase DNA binding domain) [6.7e−18]	1	0	0	0
A627R	49,629	11.1	Late	Unknown protein	1, 2	1	3	0
A629R	86,292	7.5	Early-late	PfamA, PF03477.9 (ATP cone domain) [8.5e−15]/PF00317.14 (ribonucleotide reductase all-alpha domain) [7.9e−19]/PF02867.8 (ribonucleotide reductase barrel domain) [2e−194]	1	0	0	0
A631L	10,392	9.9	NA	Unknown protein	1	0	0	0
A643R	53,097	11.3	Late	Unknown protein	1, 2	0	0	0
A644R	19,207	6.0	Late	Unknown protein	1, 2	0	0	0
A655L	12,002	11.4	NA	Unknown protein	1	0	1	0
A676R	42,432	10.6	Late	Unknown protein; PfamA, PF08789.3 (PBCV-specific basic adaptor domain) [1.9e−17]/PF08793.3 (2-cysteine adaptor domain) [1.8e−15]	1, 2	0	0	0
A678R	41,287	10.3	Late	Unknown protein	1, 2	0	3	0
A686L	18,316	6.9	Early	Unknown protein	1	0	1	0

[a] Transmembrane (TM) regions of the protein were predicted by TMHMM (T) (29), HMMTOP (H) (49), and Phobius (P) (19) methods. For all the methods, default parameters were used for prediction. The numbers are the numbers of helices predicted by the method.
[b] NA, not applicable.

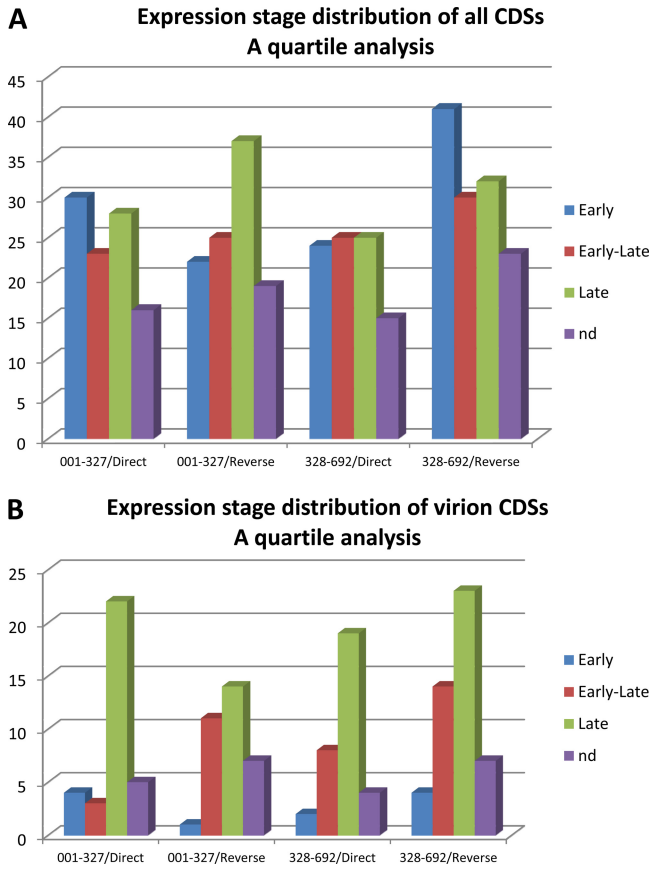

FIG 3 Expression stage distribution of PBCV-1 CDSs as a quartile analysis. (A) Number of all coding CDSs expressed either during the early, early-late, or late stage or not determined shown as a function of the genome map position. The genome map is divided into four regions, both direct (R genes) and reverse (L genes) on each half of the genome (left-half gene numbers, 001 to 327; right-half gene numbers, 328 to 692). (B) Distribution of virion-associated CDSs with respect to expression stage and genome position.

shown). However, we never obtained good resolution of the proteins using 2-D gels, even though many protocols were tried, because the MCP dominated the gel.

Membrane proteins. The virion proteins were evaluated for potential transmembrane domains by three independent methods (19, 29, 49); the results suggest that at least 26% of the proteome may be associated with a membrane structure (Table 1), presumably the internal membrane of the virion. Two-thirds of the CDSs with predicted transmembrane domains (3 out of 3 programs used) were detected by both proteomic methods. The remaining 1/3 of the CDSs were detected equally by method 1, biased toward somewhat larger (mean, 23.8 kDa) and more basic (mean pI = 9.2) proteins, and method 2, biased toward smaller (mean, 10.3 kDa) and less basic (mean pI = 7.8) proteins.

The origin of the PBCV-1 internal membrane is unknown. If all, or at least most, of the PBCV-1 internal membrane contains virus-encoded proteins and no host-encoded proteins, it suggests extensive modification of the host membrane to form the virus membrane.

PBCV basic adaptor domain-containing proteins. Eight PBCV-1 CDSs have at least one copy of a small, highly positively charged C-terminal domain referred to as the PBCV basic adaptor domain (18): A092/093L, A176L, A205R, A278L, A282L, A436L, A571R, and A676R. All of these CDSs were detected in the virion (Table 1). These proteins range in size from 6.9 to 69 kDa, but their pI values are very basic, 10.6 to 13.0. Five of the proteins contain a single copy of the basic adaptor domain; however, A092/093L and A278L have 2 copies and A282L has 3 copies. A278L and A282L are S/T protein kinases (50). The A676R protein contains both the PBCV basic adaptor domain and a 2-cysteine domain (Pfam 08793), which is a virus-specific domain fused to OUT/A20-like peptidases and S/T protein kinases that is suggested to function as a targeting device for specific substrates (18). The PBCV-1 basic adaptor domain is found only in the chloroviruses, and A176L is found only in PBCV-1. The function of the PBCV-1 basic adaptor domain is unknown.

MCP paralogs. The initial understanding of the architectural

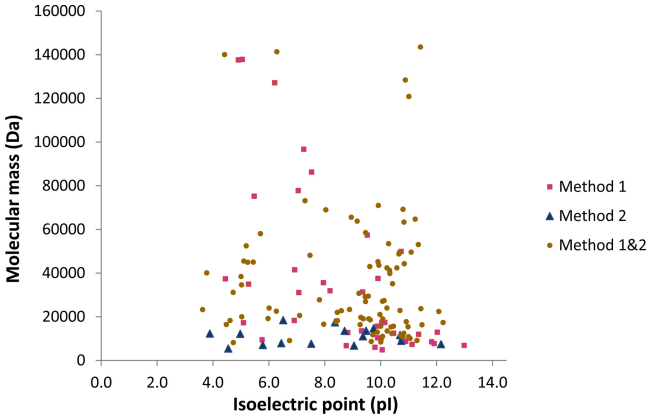

FIG 4 Mass-versus-pI distribution of PBCV-1 virion CDSs identified by two independent proteomic methods. The virion proteins are displayed as a function of their intrinsic molecular masses and isoelectric points. The results of each method are shown. Note that method 2 was especially useful for discovering a set of low-molecular-mass proteins that were not detected by method 1.

makeup of the PBCV-1 virion was a simple quasi-icosahedral particle consisting of a single MCP (Vp54) (63). This picture has evolved to the present 8.5-Å-resolution complex particle with several surface features, including a unique vertex with a spike structure and fiber-like structures associated with some capsomers in the trisymmetrons (5, 65). Genome sequencing revealed genes encoding 6 additional capsid-like proteins (25). Previously, these paralogs were not considered relevant because at least two of them (genes *a010r* and *a011l*) could be deleted from the genome without loss of virion formation (23). However, the proteome presented here indicates that all of the capsid-like proteins are present in the virion (Table 1) and that they fall into 5 paralog classes (Fig. 5A). Each of the proteins contains 2 conserved domains (D1 and D2) (Fig. 5B), consistent with the Vp54 structure (Fig. 5C). The relative abundances of the proteins, as estimated by their emPAI values, ranged from 1 (A384dL and A383R) to 13 (A430L and A011L). These abundance ratios support the hypothesis that the architecture of the PBCV-1 virion is composed of a complex mix-

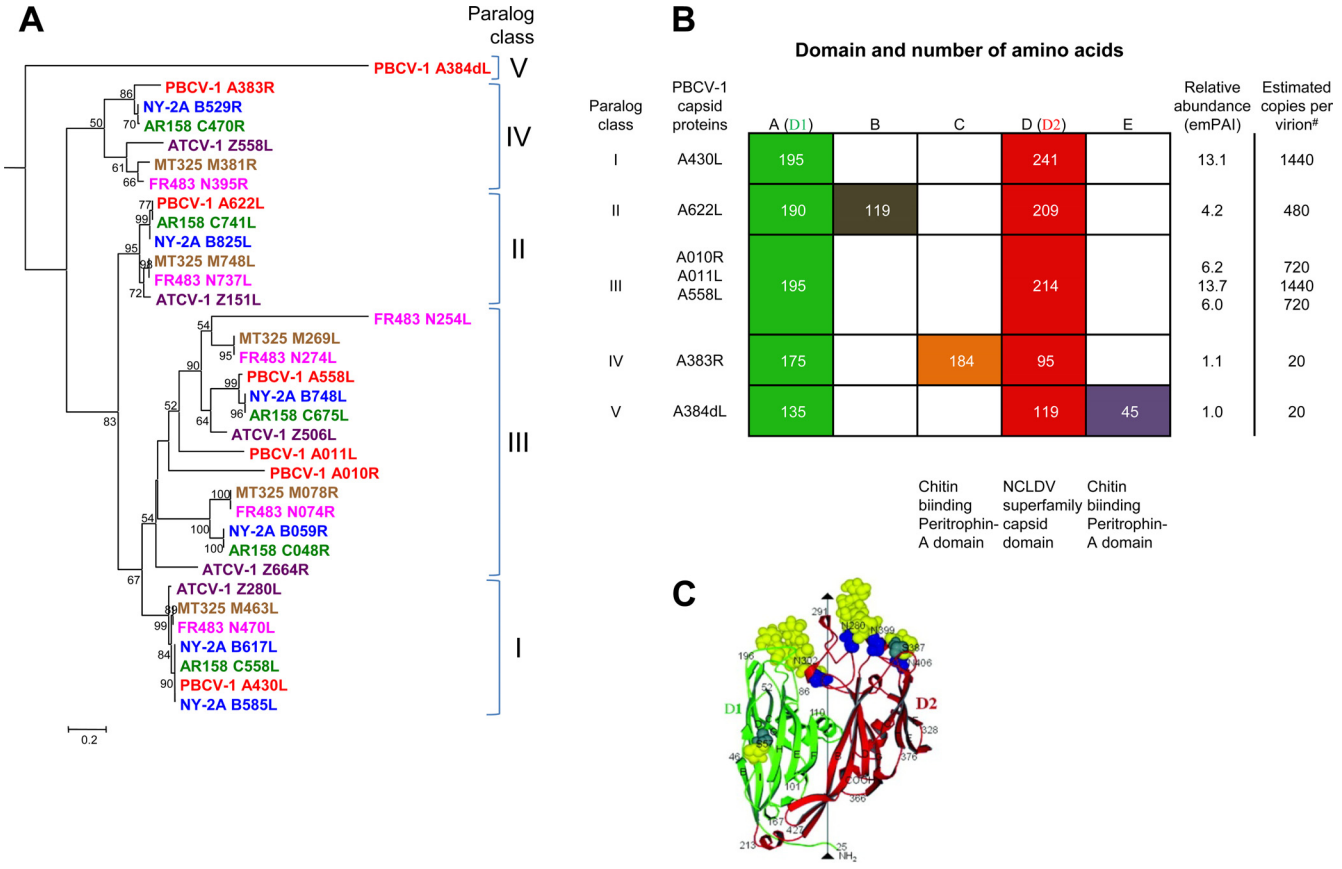

FIG 5 Capsid protein paralog classes and relative abundances in PBCV-1. (A) The seven capsid-like proteins detected in the PBCV-1 virion were evaluated against a data set of chloroviruses, including PBCV-1 (RefSeq NC_000852.5), NY-2A (RefSeq NC_009898.1), AR158 (RefSeq NC_009899.1), MT325 (GenBank DQ491001.1), FR483 (RefSeq NC_008603.1), and ATCV-1 (RefSeq NC_008724.1). These 7 proteins had homologs in each of the viruses that separated into 5 distinct paralog classes (I to V), as shown in the neighbor-joining tree (see Table S3 in the supplemental material for CDS accession numbers). The sequence for PBCV-1 A384dL, a member of paralog class V, which is distantly related, was used as the outgroup to root the phylogenetic analysis using the website http://www.phylogeny.fr (10). Muscle was used to align the sequences. Bootstrap analysis was used to construct the tree. Similar tree topologies were produced by maximum-likelihood and maximum-parsimony analyses. The values on the branches are the percentages of bootstrap support (200 replicates). Only bootstrap values of >50% are shown. The distance bar represents 0.2 amino acid substitution per site. (B) PBCV-1 capsid proteins grouped into 5 paralog classes within their two conserved domains. The D1 domain (column A) and the D2 NCLDV superfamily capsid domain (column D) were previously determined by structure analysis of the Vp54 MCP (30) (shown in panel C). The relative abundances, as determined by the emPAI method for the method 1 data, are listed on the right, along with the hypothetical estimated number of copies of each capsid protein per virion. Note that the two proteins at relatively low abundance contain chitin binding peritrophin A conserved domains (columns C and E). Column B is a domain with function.

ture of capsids and that the capsomers are composed of heteromeric proteins with a conserved structure. Additionally, the 2 minor capsid-like proteins, A383R and A384dL, contain an additional domain that is similar to the chitin binding peritrophin A domain (Pfam 01607.17) (see Table S1 in the supplemental material) and may contribute to the attachment of the virion to the algal cell surface. The relative abundance of these proteins is consistent with the frequency of fiber structures found in each trisymmetron, but the composition of the structures is unknown.

The estimated relative abundances of virion proteins were determined using the emPAI method (17) for the method 1 data set. The distribution of the capsid proteins suggests a more complex assembly of PBCV-1 capsids than was previously assumed for a single MCP (Vp54) responsible for the particle architecture. We assumed the MCP (A430L) is present in 1,440 copies per virion for these calculations, and other protein abundances were estimated from this value (Fig. 5B). The data indicate there are two capsid proteins with relatively high abundance (A430L and A011L), while two capsid proteins were present at approximately one-half that abundance (A010R and A558L), one capsid protein was present at one-third that abundance (A622L), and two capsid proteins were present at relatively low abundance (A383R and A384dL). Assuming these ratios, icosahedral symmetry, and the fact that the virion is composed of 1,680 capsids (63), each of the triangular facets of the icosahedron would contain seven proteins at a 72:72:36:36:24:1:1 ratio. Recent structural analysis of PBCV-1 at 8.5-Å resolution indicates the capsomer volumes are more varied than previously thought (65), but how these capsids are arranged is not known. The trimeric capsomers may be homomeric (as previously thought) or possibly heteromeric, utilizing the conserved beta-barrel domains as binding surfaces. This higher complexity of virion structure is consistent with several other large DNA viruses in which multiple capsid proteins have been detected; herpesviruses have 4 to 7 capsid proteins (21, 32), and mimivirus has at least 5 capsid proteins (36). The emPAI method was used to estimate the abundances of intracellular mature virion proteins of vaccinia virus (7) and indicated a dynamic range of 1 to 1,000, with certain core proteins being most abundant (i.e., A4L, A10L, F17R, and A3L) while one had low abundance (i.e., E11L).

PBCV-1 proteome functionalities. The 148 virion proteins were grouped into 11 functional/structural categories (see Fig. S3A in the supplemental material) and compared to the distribution of CDSs of the overall genome (see Fig. S3B in the supplemental material). The majority (72%) of virion proteins are in the unknown-function category. However, several functions were inferred by sequence similarity analyses, and 13 of the 148 proteins have demonstrated functions that include DNA binding, cell signaling via phosphorylation, DNA degradation, virus structure, cell attachment, and polyamine biosynthesis, such as homospermidine synthase. Among the identified CDSs are the restriction endonucleases R.CviAII (A252R) and R.CviAI (A579L) thought to be responsible for host DNA degradation early in the infection cycle (3).

Virion morphogenesis is one of the last events in the PBCV-1 replication cycle, and it is reasonable that virion proteins are synthesized during the late phase. Most of the proteome (87%) is from genes expressed either late or early-late (64); however, the time of expression has not been determined for 23 new CDSs discovered during the resequencing and annotation (see Fig. S2 in the supplemental material). Eleven proteins are from genes transcribed in the early phase of replication; 7 of these proteins were detected by a single proteomic method with a relatively low number of unique peptides detected. Therefore, these 7 proteins require further verification. Three of these early proteins, A171R, A440L, and A443R, have unknown functions. The A456L protein has two conserved domains, a D5 N superfamily domain found in certain viral DNA primases (PfamA, PF08706.4) and a phage/plasmid primase P4 family C-terminal domain with predicted ATPase activity. The A548L protein has two conserved P-loop NTPase domains that are associated with DEXDc, DEAD, and DEAH box proteins, including the hepatitis C virus NS3 helicases (PfamA, PF00176.16). Thus, these proteins might contribute to early transcriptional events that occur within minutes of infection.

PBCV-1 packaged host protein. The PBCV-1 proteome contains one protein (101 amino acids) derived from the host (GenBank EFN53917.1) (4); the protein was detected by both proteomic methods. This protein is most similar to a fungal 93-amino-acid *Naumovozyma dairenensis* CBS 421 nucleosome binding protein (NCBI reference sequence: XP_003667927.1) and similar to the HMGB-UBF_HMG box class II and III members of the HMG box superfamily of DNA binding proteins. It has no similarity to any PBCV-1-encoded protein. HMG box-containing proteins bind non-B-type DNA conformations with high affinity (44), and they are involved in the regulation of DNA-dependent processes, such as transcription, replication, and DNA repair, all of which require changing the conformation of chromatin (48). Thus, this host protein may be important in initiating PBCV-1 gene expression, which occurs within minutes of infection (64). At least two other large DNA viruses contain chromosomal proteins in the virion. An HMG box protein (HMG1) and a histone H2B.q protein occur in the Western Reserve strain of vaccinia virus (37), and murine cytomegalovirus virions have a histone H2A protein (21), suggesting that large DNA viruses utilize host-derived proteins for DNA binding functions.

Presumed virion proteins that were not detected. A few proteins that were expected to be packaged in PBCV-1 were absent in the proteome analysis. As noted previously, PBCV-1 packages one or more enzymes involved in digesting the host cell wall during infection (28). Annotation of the PBCV-1 genome identified 5 enzymes that might be involved in this process—two chitinases, a chitosanase, a β-1-3-glucanase, and a β- or α-1,4-glucuoronic lyase (see Table S1 in the supplemental material). Recombinant proteins indicated that all of these enzymes are functional (45, 46), and Western blots suggested that one of the chitinases and the chitosanase were in the virion (46). However, none of the five proteins were detected in the proteome analysis. Consequently, the enzyme(s) involved in digesting the host cell wall is unknown.

Circumstantial evidence suggests that PBCV-1 and other chloroviruses package a small virus-encoded K^+ channel protein named Kcv (12). It has been hypothesized that Kcv is involved in depolarizing the host membrane, which occurs immediately after virus infection. However, Kcv was not detected in this proteome study. On the other hand, at least one putative protein (A201L) with predicted physical/chemical transmembrane properties similar to those of Kcv was detected in the PBCV-1 virion by proteomic method 1. Thus, this methodology can detect small proteins with transmembrane domains, such as Kcv.

Conclusions. Resequencing and annotation of the 331-kbp

chlorovirus PBCV-1 genome revealed that the virus carries 416 predicted CDSs, using a minimum ORF size of 40 codons, and 11 tRNAs. Proteome analysis of highly purified PBCV-1 virions identified 148 virus-encoded proteins (about 35% of the coding capacity of the virus) and one host protein. Some of these proteins appear to be structural/architectural, whereas others have enzymatic, chromatin modification, and signal transduction functions. However, 106 of the proteins have no known function or homologs in the existing gene databases except as orthologs with proteins of other chloroviruses, phycodnaviruses, and NCLDV. The genes encoding these proteins are dispersed throughout the virus genome, and 84% are transcribed late or early-late in the infection cycle, which is consistent with virion morphogenesis.

Probably the biggest surprise is that so many virus-encoded proteins were detected in the virion and only one host-encoded protein. However, except for the MCP Vp54, we cannot definitively assign a protein(s) to any of the structural features of the virus, including the additional 6 major capsid-like proteins, the long spike structure, the surface fibers on the trisymmetrons, and the long glue protein homologs of PRD1 and the membrane protein dimer located at the edges of the trisymmetrons and internal membrane (65). These await further structural analyses. Obviously, one question is, are all of these virion-associated proteins essential for creating an infectious virus, or are some of them the result of "sloppy packaging," i.e., fortuitously associated with the particle? This is a difficult question to answer, but it is clear that PBCV-1 morphogenesis is selective in terms of what it incorporates, e.g., the virus packages 2 virus-encoded restriction endonucleases but not their corresponding DNA methyltransferases. In addition, only one host protein was detected in the virion; no host membrane proteins were detected.

The PBCV-1 capsid protein composition may be somewhat flexible, because the genes encoding 2 of the capsid proteins (A010R and A011L) can be deleted (6, 23), yet the deletion mutants are viable. This finding suggests some type of compensation in capsid protein utility. Among large DNA viruses, the number of capsid proteins ranges from 4 to 7, and these homologs are virion associated (21, 32, 36); thus, the discovery of 7 putative capsid proteins in the PBCV-1 virion is consistent with this theme, yet little is known about how these proteins contribute to virion structure or function.

These data are essential for understanding the immediate-early events of infection, such as binding, the entrisome, macromolecular synthetic shutoff, DNA degradation, and viral transcription (47). PBCV-1 is representative of the giant viruses in that they are large, complex, and highly diverse (11). When characterizing these viruses, the traditional meaning of the term "structural protein" (i.e., virion associated) has been lost. Giant-virus virions incorporate both structural/architectural proteins and many other proteins with a wide range of functionalities likely directed at the immediate-early events of infection before viral transcription *de novo* is initiated. Why giant viruses have so many genes continues to intrigue virologists, but as more giant viruses are discovered, there will be more opportunities to explore this question.

ACKNOWLEDGMENTS

The UNL mass spectrometry facility is supported in part by NIH Grant P20 RR15635 from the COBRE Program of the National Center for Research Resources, NCI Cancer Center Support Grant P30 CA36727, and the Nebraska Research Initiative. Additional support was from NIH COBRE Grant P20RR015635 (D.D.D.), NIH R0l GM32441 (J.L.V.E.), NSF-EPSCoR EPS-1004094 (J.L.V.E.), DOE DE-EE0003142 (J.L.V.E.), a Discovery grant from NSERC (Canada) (C.U.), and National Center for Research Resources 5P20RR016469 and NIGMS 8P20GM103427 (G.D.).

We thank Jaehyoung Kim and Joe Nietfeldt at the UNL Core for Applied Genomics and Ecology for the PBCV-1 sequencing.

REFERENCES

1. **Abrescia NG, et al.** 2004. Insights into assembly from structural analysis of bacteriophage PRD1. Nature **432**:68–74.
2. **Agarkova I, et al.** 2008. Chlorovirus-mediated membrane depolarization of *Chlorella* alters secondary active transport of solutes. J. Virol. **82**:12181–12190.
3. **Agarkova IV, Dunigan DD, Van Etten JL.** 2006. Virion-associated restriction endonucleases of chloroviruses. J. Virol. **80**:8114–8123.
4. **Blanc G, et al.** 2010. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. Plant Cell **22**:2943–2955.
5. **Cherrier MV, et al.** 2009. An icosahedral algal virus has a complex unique vertex decorated by a spike. Proc. Natl. Acad. Sci. U. S. A. **106**:11085–11089.
6. **Chuchird N, et al.** 2002. A variable region on the chlorovirus CVK2 genome contains five copies of the gene for Vp260, a viral-surface glycoprotein. Virology **295**:289–298.
7. **Chung C-S, et al.** 2006. Vaccinia virus proteome: identification of proteins in vaccinia virus intracellular mature virion particles. J. Virol. **80**:2127–2140.
8. **Cockburn J, et al.** 2004. Membrane structure and interactions with protein and DNA in bacteriophage PRD1. Nature **432**:122–125.
9. **Craig R, Beavis RC.** 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics **20**:1466–1467.
10. **Dereeper A, et al.** 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. **36**:W465–W469.
11. **Dunigan DD, Fitzgerald LA, Van Etten JL.** 2006. Phycodnaviruses: a peek at genetic diversity. Virus Res. **117**:119–132.
12. **Frohns F, et al.** 2006. Potassium ion channels of chlorella viruses cause rapid depolarization of host cells during infection. J. Virol. **80**:2437–2444.
13. **Graves MV, Meints RH.** 1992. Characterization of the major capsid protein and cloning of its gene from algal virus PBCV-1. Virology **188**:198–207.
14. **Higdon R, Hogan J, Kolker N, van Belle G, Kolker E.** 2007. Experiment-specific estimation of peptide identification probabilities using a randomized database. OMICS **11**:351–356.
15. **Higdon R, Hogan J, Van Belle G, Kolker E.** 2005. Randomized sequence databases for tandem mass spectrometry peptide and protein identification. OMICS **9**:364–379.
16. **Higdon R, Kolker E.** 2007. A predictive model for identifying proteins by a single peptide match. Bioinformatics **23**:277–280.
17. **Ishihama Y, et al.** 2005. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics **4**:1265–1272.
18. **Iyer L, Balaji S, Koonin E, Aravind L.** 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. **117**:156–184.
19. **Käll L, Krogh A, Sonnhammer EL.** 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. **338**:1027–1036.
20. **Kang M, Dunigan DD, Van Etten JL.** 2005. Chloroviruses: a genus of Phycodnaviridae that infects certain chlorella-like green algae. Mol. Plant Pathol. **6**:213–224.
21. **Kattenhorn L, et al.** 2004. Identification of proteins associated with murine cytomegalovirus virions. J. Virol. **78**:11187–11197.
22. **Kuznetsov YG, Gurnon JR, Van Etten JL, McPherson A.** 2005. Atomic force microscopy investigation of a chlorella virus, PBCV-1. J. Struct. Biol. **149**:256–263.
23. **Landstein D, Burbank DE, Nietfeldt JW, Van Etten JL.** 1995. Large deletions in antigenic variants of the chlorella virus PBCV-1. Virology **214**:413–420.
24. **Lane L.** 1978. A simple method for stabilizing protein-sulfhydryl groups during SDS-gel electrophoresis. Anal. Biochem. **86**:655–664.

25. **Li Y, et al.** 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.
26. **Mehmel M, et al.** 2003. Possible function for virus encoded K^+ channel Kcv in the replication of chlorella virus PBCV-1. FEBS Lett. **552**:7–11.
27. **Meints RH, Burbank DE, Van Etten JL, Lamport DT.** 1988. Properties of the chlorella receptor for the virus PBCV-1. Virology **164**:15–21.
28. **Meints RH, Lee K, Burbank DE, Van Etten JL.** 1984. Infection of a chlorella-like alga with the virus, PBCV-1: ultrastructural studies. Virology **138**:341–346.
29. **Möller S, Croning MDR, Apweiler R.** 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics **17**:646–653.
30. **Nandhagopal N, et al.** 2002. The structure and evolution of the major capsid protein of a large, lipid-containing DNA virus. Proc. Natl. Acad. Sci. U. S. A. **99**:14758–14763.
31. **Neupärtl M, et al.** 2008. Chlorella viruses evoke a rapid release of K^+ from host cells during early phase of infection. Virology **372**:340–348.
32. **O'Connor CM, Kedes DH.** 2006. Mass spectrometric analyses of purified rhesus monkey rhadinovirus reveal 33 virion-associated proteins. J. Virol. **80**:1574–1583.
33. **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.
34. **Onimatsu H, Sugimoto I, Fujie M, Usami S, Yamada T.** 2004. Vp130, a chloroviral surface protein that interacts with the host Chlorella cell wall. Virology **319**:71–80.
35. **Que Q, et al.** 1994. Protein glycosylation and myristylation in chlorella virus PBCV-1 and its antigenic variants. Virology **203**:320–327.
36. **Renesto P, et al.** 2006. Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. J. Virol. **80**:11678–11685.
37. **Resch W, Hixson KK, Moore RJ, Lipton MS, Moss B.** 2007. Protein composition of the vaccinia virus mature virion. Virology **358**:233–247.
38. **Santoni V, Molloy M, Rabilloud T.** 2000. Membrane proteins and proteomics: un amour impossible? Electrophoresis **21**:1054–1070.
39. **Shevchenko A, Wilm M, Vorm O, Mann M.** 1996. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal. Chem. **68**:850–858.
40. **Short CM, Rusanova O, Short SM.** 2011. Quantification of virus genes provides evidence for seed-bank populations of phycodnaviruses in Lake Ontario, Canada. ISME J. **5**:810–821.
41. **Skrdla MP, Burbank DE, Xia Y, Meints RH, Van Etten JL.** 1984. Structural proteins and lipids in a virus, PBCV-1, which replicates in a chlorella-like alga. Virology **135**:308–315.
42. **Song WJ, et al.** 2004. Functional genomics analysis of Singapore grouper iridovirus: complete sequence determination and proteomic analysis. J. Virol. **78**:12576–12590.
43. **Stothard P, Wishart DS.** 2005. Circular genome visualization and exploration using CGView. Bioinformatics **21**:537–539.
44. **Stros M, Launholt D, Grasser KD.** 2007. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. Cell. Mol. Life Sci. **64**:2590–2606.
45. **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.
46. **Sun L, Adams B, Gurnon J, Ye Y, Van Etten JL.** 1999. Characterization of two chitinase genes and one chitosanase gene encoded by chlorella virus PBCV-1. Virology **263**:376–387.
47. **Thiel G, Moroni A, Dunigan D, Van Etten JL.** 2010. Initial events associated with virus PBCV-1 infection of *Chlorella* NC64A. Prog. Bot. **71**:169–183.
48. **Thomas JO.** 2001. HMG1 and 2: architectural DNA-binding proteins. Biochem. Soc. Trans. **29**:395–401.
49. **Tusnády GE, Simon I.** 2001. The HMMTOP transmembrane topology prediction server. Bioinformatics **17**:849–850.
50. **Valbuzzi P.** 2005. Serine/threonine kinases encoded by PBCV-1 virus: characterization and possible role in the phosphorylation of the K+ channel Kcv. Ph.D. dissertation. Universita Degli Studi di Milano, Milan, Italy.
51. **Van Etten JL, Burbank DE, Kuczmarski D, Meints RH.** 1983. Virus infection of culturable chlorella-like algae and development of a plaque assay. Science **219**:994–996.
52. **Van Etten JL, Burbank DE, Xia Y, Meints RH.** 1983. Growth cycle of a virus, PBCV-1, that infects chlorella-like algae. Virology **126**:117–125.
53. **Van Etten JL, Dunigan DD.** 2012. Chloroviruses: not your every day plant virus. Trends Plant Sci. **17**:1–8.
54. **Van Etten JL, Graves MV, Muller DG, Boland W, Delaroque N.** 2002. Phycodnaviridae-large DNA algal viruses. Arch. Virol. **147**:1479–1516.
55. **Van Etten JL, Lane LC, Meints RH.** 1991. Viruses and viruslike particles of eukaryotic algae. Microbiol. Rev. **55**:586–620.
56. **Van Etten JL, Meints RH.** 1999. Giant viruses infecting algae. Annu. Rev. Microbiol. **53**:447–494.
57. **Van Etten JL, Van Etten CH, Johnson JK, Burbank DE.** 1985. A survey for viruses from fresh water that infect a eucaryotic chlorella-like green alga. Appl. Environ. Microbiol. **49**:1326–1328.
58. **Wulfmeyer T, et al.** 2012. Structural organization of DNA in chlorella viruses. PLoS One **7**:e30133. doi:10.1371/journal.pone.0030133.
59. **Yamada T, Furukawa S, Hamazaki T, Songsri P.** 1996. Characterization of DNA-binding proteins and protein kinase activities in chlorella virus CVK2. Virology **219**:395–406.
60. **Yamada T, Higashiyama T, Fukuda T.** 1991. Screening of natural waters for viruses which infect chlorella cells. Appl. Environ. Microbiol. **57**:3433–3437.
61. **Yamada T, Onimatsu H, Van Etten JL.** 2006. Chlorella viruses. Adv. Virus Res. **66**:293–366.
62. **Yan X, et al.** 2005. The structure of a T=169d algal virus, PBCV-1, at 15 Å resolution. Microsc. Microanal. **11**:1056–1057.
63. **Yan X, et al.** 2000. Structure and assembly of large lipid-containing dsDNA viruses. Nat. Struct. Biol. **7**:101–103.
64. **Yanai-Balser GM, et al.** 2010. Microarray analysis of chlorella virus PBCV-1 transcription. J. Virol. **84**:532–542.
65. **Zhang X, et al.** 2011. Three-dimensional structure and function of the *Paramecium bursaria* chlorella virus capsid. Proc. Natl. Acad. Sci. U. S. A. **108**:14837–14842.
66. **Zhang Y, et al.** 2003. Characterization of a chlorella virus PBCV-1 encoded ribonuclease III. Virology **317**:73–83.
67. **Zhang Y, Strasser P, Grabherr R, Van Etten JL.** 1994. Hairpin loop structure at the termini of the chlorella virus PBCV-1 genome. Virology **202**:1079–1082.