

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

---

James Van Etten Publications

Plant Pathology Department

---

6-2010

# Identification of an L-Rhamnose Synthetic Pathway in Two Nucleocytoplasmic Large DNA Viruses

Madhu Parakkottil Chothi  
*University of Genova*

Garry A. Duncan  
*Nebraska Wesleyan University, [gduncan@nebrwesleyan.edu](mailto:gduncan@nebrwesleyan.edu)*


Andrea Armirotti  
*University of Genova*

Chantal Abergel  
*Structural and Genomic Information*

James R. Gurnon  
*University of Nebraska - Lincoln*

*See next page for additional authors*

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

 Part of the [Genetics and Genomics Commons](#), [Plant Pathology Commons](#), and the [Viruses Commons](#)

---

Chothi, Madhu Parakkottil; Duncan, Garry A.; Armirotti, Andrea; Abergel, Chantal; Gurnon, James R.; Van Etten, James L.; Bernardi, Cinzia; Damonte, Gianluca; and Tonetti, Michela, "Identification of an L-Rhamnose Synthetic Pathway in Two Nucleocytoplasmic Large DNA Viruses" (2010). *James Van Etten Publications*. 13.  
<http://digitalcommons.unl.edu/vanetten/13>

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

---

**Authors**

Madhu Parakkottil Chothi, Garry A. Duncan, Andrea Armirotti, Chantal Abergel, James R. Gurnon, James L. Van Etten, Cinzia Bernardi, Gianluca Damonte, and Michela Tonetti

## Identification of an L-Rhamnose Synthetic Pathway in Two Nucleocytoplasmic Large DNA Viruses<sup>∇</sup>

Madhu Parakkottil Chothi,<sup>1,3</sup> Garry A. Duncan,<sup>2</sup> Andrea Armirotti,<sup>1,3</sup> Chantal Abergel,<sup>4</sup>  
James R. Gurnon,<sup>5</sup> James L. Van Etten,<sup>5</sup> Cinzia Bernardi,<sup>1</sup>  
Gianluca Damonte,<sup>1,3</sup> and Michela Tonetti<sup>1,3\*</sup>

Department of Experimental Medicine, University of Genova, Genova, Italy<sup>1</sup>; Biology Department, Nebraska Wesleyan University, Lincoln, Nebraska 68504-2794<sup>2</sup>; Center of Excellence for Biomedical Research, University of Genova, Genova, Italy<sup>3</sup>; Structural and Genomic Information, UPT2589-CNRS, Marseille, France<sup>4</sup>; and Department of Plant Pathology and Nebraska Center for Virology, University of Nebraska, Lincoln, Nebraska 68583-0900<sup>5</sup>

Received 12 April 2010/Accepted 2 June 2010

**Nucleocytoplasmic large DNA viruses (NCLDVs) are characterized by large genomes that often encode proteins not commonly found in viruses. Two species in this group are *Acanthocystis turfacea* chlorella virus 1 (ATCV-1) (family *Phycodnaviridae*, genus *Chlorovirus*) and *Acanthamoeba polyphaga* mimivirus (family *Mimiviridae*), commonly known as mimivirus. ATCV-1 and other chlorovirus members encode enzymes involved in the synthesis and glycosylation of their structural proteins. In this study, we identified and characterized three enzymes responsible for the synthesis of the sugar L-rhamnose: two UDP-D-glucose 4,6-dehydratases (UGDs) encoded by ATCV-1 and mimivirus and a bifunctional UDP-4-keto-6-deoxy-D-glucose epimerase/reductase (UGER) from mimivirus. Phylogenetic analysis indicated that ATCV-1 probably acquired its UGD gene via a recent horizontal gene transfer (HGT) from a green algal host, while an earlier HGT event involving the complete pathway (UGD and UGER) probably occurred between a protozoan ancestor and mimivirus. While ATCV-1 lacks an epimerase/reductase gene, its *Chlorella* host may encode this enzyme. Both UGDs and UGER are expressed as late genes, which is consistent with their role in posttranslational modification of capsid proteins. The data in this study provide additional support for the hypothesis that chloroviruses, and maybe mimivirus, encode most, if not all, of the glycosylation machinery involved in the synthesis of specific glycan structures essential for virus replication and infection.**

The nucleocytoplasmic large DNA viruses (NCLDVs) are a heterogeneous group of viruses that infect several eukaryotic organisms (25, 49). They have large genomes that often encode genes not commonly found in viruses. For example, several lines of evidence indicate that *Paramecium bursaria* chlorella virus 1 (PBCV-1) and other chlorovirus members, such as *Acanthocystis turfacea* chlorella virus 1 (ATCV-1), encode at least part, if not all, of the machinery required to glycosylate their structural proteins, including glycosyltransferases (13, 21, 30, 33, 41–43). Furthermore, glycosylation occurs independently of the host endoplasmic reticulum (ER)-Golgi system (33, 42–44). The PBCV-1 major capsid protein located on the viral surface is glycosylated, and the glycan moieties contribute to virus protease resistance and antigenicity. We have previously reported that PBCV-1 and ATCV-1 encode both GDP-D-mannose 4,6-dehydratase (GMD) and GDP-4-keto-6-deoxy-D-mannose epimerase/reductase (GMER), which are involved in “*de novo*” GDP-L-fucose synthesis (14, 40). Because fucose is a component of the glycan portion of the PBCV-1 major capsid protein (43), the viral GMD and GMER enzymes may be necessary to provide the virus with the nucleotide sugar.

The aim of the present study was to identify and analyze additional virus-encoded enzymes involved in glycan produc-

tion. We included chlorella virus ATCV-1, and we extended our study to *Acanthamoeba polyphaga* mimivirus. Mimivirus is a giant DNA virus that infects members of the genus *Acanthamoeba* (6–9, 35). Its 1.2-Mb genome is the largest viral genome described so far, containing more than 900 protein-coding sequences (CDS) (36).

Annotation of ATCV-1 and mimivirus genomes identified genes encoding putative enzymes involved in L-rhamnose production. This 6-deoxyhexose sugar is a common component of surface glycoconjugates such as bacterial lipopolysaccharides (LPS), where it plays an important role in pathogenicity (28, 29). L-Rhamnose also occurs in plant cell wall rhamnogalactoglucans and rhamnosides, such as flavonoids, terpenoids, and saponins (24, 38). Moreover, rhamnose is also present in the virus PBCV-1 glycan(s) attached to its major capsid protein (43). Bacteria and plants synthesize L-rhamnose from glucose by two slightly different pathways (Fig. 1). In bacteria, dTDP-D-glucose is the initial substrate for a dehydratase activity (RfbB/RmlB), which eliminates a water molecule and leads to production of an unstable intermediate compound, dTDP-4-keto-6-deoxy-D-glucose. This compound is then subjected to epimerization at C-3 and C-5 by RfbC/RmlC and finally to an NADPH-dependent reduction of C-4 by RfbD/RmlD. Three separate enzymes catalyze these steps (2, 17, 18, 20). In contrast, in the plant *Arabidopsis thaliana* the initial substrate is UDP-D-glucose, and the three enzymatic activities are fused into a single polypeptide, named RHM enzymes, as depicted in Fig. 1 (34). An enzyme with epimerase/reductase activity

\* Corresponding author. Mailing address: Department of Experimental Medicine, University of Genova, Viale Benedetto XV, 1, 16132 Genova, Italy. Phone: 39-010-3538131. Fax: 39-010-354415. E-mail: tonetti@unige.it.

<sup>∇</sup> Published ahead of print on 10 June 2010.

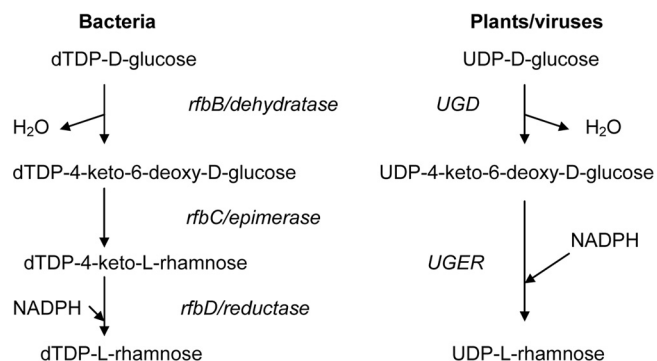


FIG. 1. L-Rhamnose biosynthetic pathways in bacterial and plant cells. In bacteria, the three enzymatic activities are on separate polypeptides. In plants, UGD and UGER are fused into a single polypeptide (RHM isoforms) (34).

(NSR/ER) leading to UDP-L-rhamnose was also identified in *A. thaliana* (44).

In this study we identified and characterized three functional enzymes involved in UDP-L-rhamnose synthesis in both ATCV-1 and mimivirus. ATCV-1 encodes only a UDP-D-glucose 4,6-dehydratase (UGD), whereas mimivirus encodes the complete pathway, i.e., UGD and a bifunctional UDP-4-keto-6-deoxy-D-glucose 3,5-epimerase/4-reductase (UGER). The virus-encoded UDP-L-rhamnose pathway and the enzymatic properties are similar to those described for plants. Sequence and phylogenetic analyses indicate that ATCV-1 likely acquired UGD from its chlorella host through a recent horizontal gene transfer (HGT). In contrast, both UGD and UGER were transmitted much earlier, probably between mimivirus and a protozoan ancestral host. Thus, these results support the hypothesis that both ATCV-1 and mimivirus encode at least part of a host-independent glycosylation system, which may be essential for virus replication and infection.

#### MATERIALS AND METHODS

**Sequence and phylogenetic analyses.** BLASTP searches of the NCBI (www.ncbi.nlm.nih.gov) and DOE-JGI (http://genome.jgi-psf.org) databases were used to identify UGD homologs of ATCV-1 (CDS Z544R, YP\_001427025.1) and mimivirus (CDS R141, YP\_142495.1). These same databases were used to identify UGER homologs in mimivirus (CDS L780, Q5UPS5.1). Bacterial sequences were used to root the phylogenetic trees for both proteins. Phylogenetic analyses were performed on the phylogeny.fr web tool (10), which implements MUSCLE for multiple-sequence alignments (12), GBLOCK for automatic curation of multiple alignments (5), and PhyML for maximum-likelihood (ML) phylogenetic reconstruction (22). Accession numbers for all sequences are given in the figures of both phylogenetic trees.

**Expression and purification of the recombinant proteins.** Proteins were expressed in *Escherichia coli* BL21-Gold cells (Stratagene) as glutathione S-transferase (GST) fusion proteins as previously described, using the pGEX-6P-1 vector (GE Healthcare) (40). Viral DNA was purified by standard procedures. The ATCV-1 putative UGD sequence corresponds to CDS Z544R in the viral genome, and the coding region was amplified by PCR using the following primers: forward, 5'-ACTGGATCCATGAATTCTCAGGAA; reverse, 5'-TGATATTC~~CCCGGGTTAAAATTTCCGGGAC~~ (BamHI [forward] and SmaI [reverse] restriction sites are underlined). The mimivirus putative UGD sequence, corresponding to open reading frame (ORF) R141, was amplified with the following primers: forward, 5'-AATTGGATCCATGAAGAATATTCTCGTTAC; reverse, 5'-AATTC~~TCGAGTTATTGTGATCCGGGGA~~. The mimivirus putative UGER sequence, corresponding to ORF L780, was amplified with the following primers: forward, 5'-AATTGGATCCATGAAGTGGCTCATTTTTGG; reverse, 5'-AATTC~~TCGAGTTATTGTGAAATTTCC~~

ATTTTTGG. BamHI restriction sites in the forward primers and XhoI sites in the reverse ones are underlined.

Amplification was conducted with *Pfu* DNA polymerase (Promega). PCR products were cloned into the pGEX-6P-1 vector using standard procedures. DNA sequencing was performed by TibMolBiol (Genova, Italy). Protein expression, purification, and GST cleavage were done as previously described (40).

The proteins were concentrated to 4 to 6 mg/ml using a Centricon YM-10 system (Amicon-Millipore) and stored at 4°C in 50 mM Tris-HCl-150 mM NaCl-1 mM EDTA, pH 7.5 (TBSE), containing 1 mM dithiothreitol (DTT). Concentrations were determined by UV absorbance at 280 nm of the purified proteins in water, using calculated extinction coefficients of 52,830 M<sup>-1</sup> cm<sup>-1</sup> for ATCV-1 UGD, 40,340 M<sup>-1</sup> cm<sup>-1</sup> for mimivirus UGD, and 28,420 M<sup>-1</sup> cm<sup>-1</sup> for mimivirus UGER (15). Protein purity, monitored by SDS-PAGE, exceeded 95% in all preparations.

**Size exclusion chromatography.** The molecular masses of UGD and UGER proteins were determined by size exclusion chromatography using a TSKgel G3000SWXL column (7.8 by 300 mm; 5-μm particle size) (Tosoh Biosciences). The mobile phase was 0.1 M sodium phosphate buffer, pH 6.7, containing 0.1 M Na<sub>2</sub>SO<sub>4</sub>. The eluate was monitored at 220 nm. Protein standards, obtained from Sigma, were cytochrome *c* (12 kDa), carbonic anhydrase (29 kDa), ovalbumin (44 kDa), bovine serum albumin (66 kDa), gamma globulin (157 kDa), and thyroglobulin (670 kDa). Blue dextran and DTT were used to determine the void and total column volumes, respectively, to enable calculation of the distribution coefficient  $K_{av}$  according to the equation (19, 45)  $K_{av} = (V_e - V_o)/(V_t - V_o)$ , where  $V_e$  is the elution volume of the protein,  $V_o$  is the column void volume, and  $V_t$  is the total bed volume. The  $K_{av}$  for each protein standard was plotted against the logarithm of the corresponding molecular mass.  $K_{av}$ s of samples were used to calculate the molecular mass by linear regression analysis.

**Analysis of enzyme activities.** UGD activity was assayed in TBSE-DTT, using various concentrations of UDP-D-glucose or dTDP-D-glucose and protein. Assays were conducted at 25°C, and 100-μl aliquots were withdrawn at various times (maximum of 15 min). Reactions were stopped by heating at 80°C for 3 min, followed by centrifugation at 12,000 × *g* for 5 min. The conversion of UDP (dTDP)-D-glucose to UDP (dTDP)-4-keto-6-deoxy-D-glucose measured dehydratase activity; the sugar nucleotides were detected by anion-exchange high-pressure liquid chromatography (HPLC) as reported previously (14). Effects of pH were analyzed using the buffers 50 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES)-HCl (pH 5.5, 6.0, 6.5, 7.0, and 7.5) and 50 mM Tris-HCl (pH 7.0, 7.5, 8.0, 8.5, and 9.0).  $K_m$  and  $V_{max}$  parameters were determined with the Michaelis-Menten equation using nonlinear regression (GraphPad Prism). To determine thermal stability, the concentrated proteins were preincubated at the indicated temperatures for 30 min, and enzyme activity was assayed as indicated above. UDP-L-rhamnose, used to test feedback inhibition of UGD, was obtained in millimolar amounts using UGER (see below) after purification by anion-exchange chromatography.

Epimerase/reductase activity was analyzed using UDP-4-keto-6-deoxy-D-glucose as the substrate. This compound was produced by incubation of 1 mM UDP-D-glucose in the presence of ATCV-1 UGD. Complete conversion of UDP-D-glucose to the intermediate compound UDP-4-keto-6-deoxy-D-glucose was monitored by HPLC. The sugar nucleotide was separated from the protein by ultrafiltration using the Microcon YM-10 system (Amicon-Millipore). Heat inactivation was avoided to prevent possible degradation of the intermediate compound. Enzymatic activity of UGER in the presence of either NADH or NADPH was determined by anion-exchange HPLC. Enzymatic parameters were obtained as indicated above for UGD.

**Characterization of UDP-sugars.** The products of ATCV-1 and mimivirus UGD and mimivirus UGER were analyzed by electrospray ionization-mass spectrometry (ESI-MS) and gas chromatography (GC)-MS. Electrospray analysis was performed on an Agilent1100 MSD ion trap instrument (Agilent Technologies, Palo Alto, CA) by flow injection of the samples with an infusion pump (KD Scientific, New Hope, PA). The reaction mixtures were diluted 1:50 with a water-acetonitrile-ammonia solution (49.5/49.5/1); spectra were acquired in the negative-ion mode in the mass range of the expected *m/z* ratios. The ion source parameters were set to obtain optimal signal-to-noise ratio for molecules of interest.

The UGER product was purified by anion-exchange HPLC. The eluate was subjected to several cycles of lyophilization to remove NH<sub>4</sub>HCO<sub>3</sub>. Purified UDP-L-rhamnose was then derivatized as described by Henry et al. (23) and Merkle and Poppe (31) with slight modifications. Briefly, the UGER product or standards were treated with 2 N HCl for 90 min at 100°C and dried in a Speed-Vac. The resulting compounds were reduced to their corresponding alditols with NaBH<sub>4</sub> (10 mg/ml in 1 M ammonia solution). The products were acetylated in 1:1 pyridine-acetic anhydride at 90°C for 20 min, dried, and resuspended in 100 μl of

ethyl acetate, and 5- $\mu$ l aliquots of this solution were injected into an HP 5890 series II gas chromatograph coupled to an HP 5889A engine mass spectrometer equipped with an electron impact ionization source (Hewlett Packard). Separation was performed on an S.E.-54 (Alltech) capillary column; the helium gas flow was 2 ml/min. Sample injection was performed in splitless mode. The oven temperature gradient was as follows: initial temperature of 80°C, isothermal for 5 min, 80 to 190°C (rate, 15°C/min), 190 to 250°C (rate, 5°C/min), and isothermal for 5 min. The MS analysis was performed in full-scan mode.

**Expression of UGD and UGER after viral infection.** *Chlorella* strain SAG 3.83 was grown at 25°C as described previously (13). Cells ( $1 \times 10^8$ /ml) were infected with ATCV-1 at a multiplicity of infection (MOI) of 3, and 30 ml of cells was withdrawn between 45 min and 8 h postinfection (p.i). Cells were pelleted by centrifugation, and total RNA was isolated by adding 1 ml of Trizol (Invitrogen). Cells were disrupted with a TissueLyser from Qiagen (twice for 3 min, 30 Hz). The following steps are described in the Trizol protocol. Removal of residual genomic DNA was achieved using the gDNA Wipeout buffer of the QuantiTect reverse transcription protocol (Qiagen); this step was followed by reverse transcription using the same kit. Real-time PCR was performed on a Bio-Rad IQ5 cyclor using the iQSYBR green supermix (Bio-Rad). Primers for ATCV-1 UGD were as follows: forward, 5'-AACTCGCTGGCATTACC; reverse, 5'-GTAGA CACATTGATGAACC.

Sample standardization was achieved by densitometry of 28S and 18S RNAs on agarose gels. Reverse transcription was omitted in control RNA samples, which were then used for normalization of the expression levels.

Mimivirus production and purification were performed as described previously (4). Briefly, cells (*Acanthamoeba castellanii* Neff purchased from ATCC [ATCC 30871]) were infected with mimivirus at a MOI of 1,000. After 30 min of incubation in Page's amoeba saline (PAS) medium (2.5 mM NaCl, 1 mM  $\text{KH}_2\text{PO}_4$ , 0.5 mM  $\text{Na}_2\text{HPO}_4$ , 40 mM  $\text{CaCl}_2$ , and 20 mM  $\text{MgSO}_4$ ) with gentle stirring, infected cells were centrifuged ( $500 \times g$ , 5 min) and the supernatant containing excess virus discarded. The cell pellet was washed once with PAS medium and once with PPYG (2% [wt/vol] proteose peptone, 0.5% [wt/vol] yeast extract, 0.5% [wt/vol] glucose, pH 7.2). A total of  $2.5 \times 10^7$  cells were kept as the time zero sample, while the rest of the cells were distributed among 175-cm<sup>2</sup> flasks containing 25 ml PPYG medium. For each time point,  $2.5 \times 10^7$  infected cells (two 175-cm<sup>2</sup> flasks) were harvested and centrifuged ( $500 \times g$ , 5 min).

RNA extraction, quantification, and double-stranded cDNA production were performed as previously described (4). cDNAs were synthesized using a Clontech SMART protocol optimized for 454 GS FLX sequencing from the polyadenylated RNA fraction of *A. castellanii* cells at -15 min, 0 min, 90 min, 3 h, 6 h, 9 h, and 12 h after mimivirus infection. Pyrosequencing was performed on the French National sequencing platform ("Genoscope") according to the manufacturer's protocol using at least 4  $\mu$ g of double-stranded cDNA (260/280-nm absorbance ratio of >1.6) for each time point.

**Nucleotide sequence accession number.** The 454 RNA-Seq data used in this study have been deposited in the NCBI Short Read Archive under accession no. SRA010763.

## RESULTS

**Sequence analysis.** The viral CDSs encoding UGD and UGER were identified by a BlastP survey of ATCV-1 and mimivirus genomes using *A. thaliana* RHM2 protein (NP\_564633.2) (34). ATCV-1 Z544R had about 60% amino acid identity (207/345) with the N-terminal region of RHM2 proteins; this region is responsible for UGD activity (34). The maximum-likelihood tree (Fig. 2) and the high amino acid identity with *Chlorella* sp. strain NC64A (297/357, 83% identity) (G. Blanc et al., submitted for publication) suggest that ATCV-1 Z544R was acquired from its host (*Chlorella* SAG 3.83) via HGT. Indeed, all other members of the ATCV-1 clade are green algae. This alga-virus HGT event is not surprising or unique, since there is evidence that algae and their viruses have shared some other genes via HGT (32). It is worth noticing that two other green algal viruses, *Ostreococcus tauri* virus 1 and *Ostreococcus* virus OsV5, encode putative UGDs; however, they are closer to the one found in metazoa. Amino acid identity with the *E. coli* RfbB

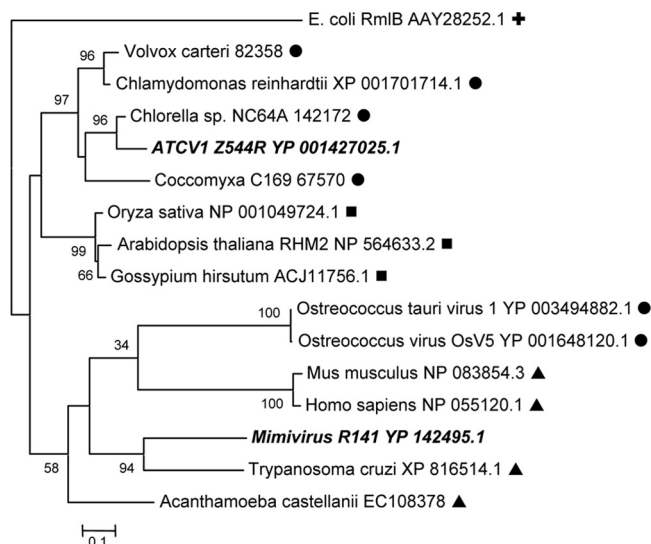


FIG. 2. Maximum-likelihood tree of UDP-D-glucose 4,6-dehydratase (UGD) for 14 taxa, including the viruses ATCV-1 and mimivirus. Only posterior probabilities of  $\geq 50\%$  are indicated to the left of each node. The distance bar represents 0.1 amino acid substitution per site. Accession numbers follow each of the taxon names. Accession numbers with five or six digits can be obtained at the JGI Genome portal at <http://genome.jgi-psf.org> for the respective species. ■, plant species; ●, green algal species; ▲, nonplant/algal species; +, bacterial species (used as the outgroup, TDP-glucose 4,6-dehydratase). The two viruses are in italics.

(RmlB) protein, which catalyzes the dehydration of dTDP-D-glucose in bacteria (2), is 140/339 (41%).

The putative UGDs from mimivirus (CDS R141) and ATCV-1 (CDS Z544R) have 46% amino acid identity. The best hit for mimivirus R141 is with a putative *Trypanosoma cruzi* protein (166/327, 50% amino acid identity), while identity with the N-terminal region of the *A. thaliana* RHM2 protein is 45%. The complete *A. castellanii* genome is not present in GenBank. However, a putative *A. castellanii* UGD sequence was retrieved from 454 RNA-Seq data deposited in NCBI Short Read Archive (accession number SRA010763). The *A. castellanii* and mimivirus UGDs have 50% amino acid identity (159/315). The mimivirus UGD gene is 31% GC, which is similar to the case for the mimivirus genome (38). These observations, together with the phylogenetic analysis (Fig. 2), suggest that mimivirus did not acquire its UGD gene through a recent HGT from its host. Interestingly, the mimivirus and ATCV-1 UGDs also have about 40% amino acid identity with some mammalian proteins (NP083854.3 from *Mus musculus* and NP055120.1 from *Homo sapiens*, as indicated in the phylogenetic tree in Fig. 2) and with several others found in metazoa such as *Nematostella vectensis*. The function of these proteins in metazoa is still unknown, with no data on the possible presence of L-rhamnose in lower animals and clear evidence that this sugar is absent from higher vertebrates. We can speculate that, at least in mammals, UGD homologs have evolved a new specificity. Such a case has already been reported for a protein sharing significant homology with the bacterial RlmD and corresponding to the regulatory subunit of the methionine adenosyl transferase 2 (MAT2B) in mammals (48).

Mimivirus also encodes CDS L780, which has about 40%

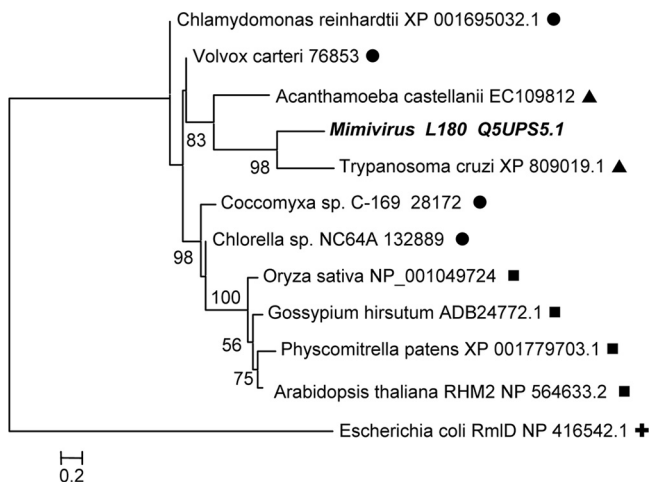


FIG. 3. Maximum-likelihood tree of UDP-4-keto-6-deoxy-D-glucose epimerase/reductase (UGER) for 12 taxa, including mimivirus. Only posterior probabilities of  $\geq 50\%$  are indicated to the left of each node. The distance bar represents 0.2 amino acid substitution per site. Accession numbers follow each of the taxon names. Accession numbers with five or six digits can be obtained at the JGI Genome portal at <http://genome.jgi-psf.org> for the respective species. ■, plant species; ●, green algal species; ▲, nonplant/algal species; +, bacterial species (used as the outgroup, dTDP-4-deoxyrhamnose reductase). Mimivirus is in italics.

amino acid identity with the C-terminal region of *A. thaliana* RHM2, a well-documented epimerase/reductase (UGER) that forms UDP-L-rhamnose (Fig. 1) (34). Phylogenetic analysis and sequence alignment indicate that L780 also was likely obtained by HGT from a *Trypanosoma* sp. (150/284 [52%] amino acid identity for *T. cruzi*) (Fig. 3). The putative *A. castellanii* UGER (SRA010763) shares 37% (106/284) amino acid identity with the mimivirus UGER, while bacterial RfbD/RmlD proteins (which have only dTDP-4-keto-L-rhamnose reductase activity) have less than 23% amino acid identity (18, 20). Thus, mimivirus probably acquired the complete pathway for the conversion of UDP-D-glucose to UDP-L-rhamnose from a protozoan ancestor. In contrast, ATCV-1 lacks an epimerase/reductase gene; however, its *Chlorella* host may encode this enzyme, which could complete UDP-L-rhamnose synthesis. Indeed, *Chlorella* sp. NC64A, whose genome was recently sequenced and annotated (Blanc et al., submitted), encodes a putative epimerase/reductase (*Chlorella* sp. NC64A 132889). The RHM2 proteins from *A. thaliana* and *Chlorella* sp. NC64A (132889) share 68% amino acid identity (191/281) and 81% amino acid similarity (228/281).

#### Expression and structural analysis of UGD and UGER.

ATCV-1 and mimivirus proteins were expressed in *E. coli* as GST fusion proteins. After removal of the GST tag by protease cleavage, the proteins had the expected molecular masses and were  $\sim 95\%$  pure, as indicated by SDS-PAGE analysis (results not shown). Size exclusion chromatography of ATCV-1 UGD (monomer, 39 kDa) indicated a molecular mass of approximately 75 kDa, which suggests a dimeric structure (Fig. 4A). Incubation of this protein with NAD or NADP, in either the oxidized or reduced forms, had no effect on the apparent molecular mass of ATCV-1 UGD. In contrast, chromatographic analysis of mimivirus UGD (monomer, 37 kDa) re-

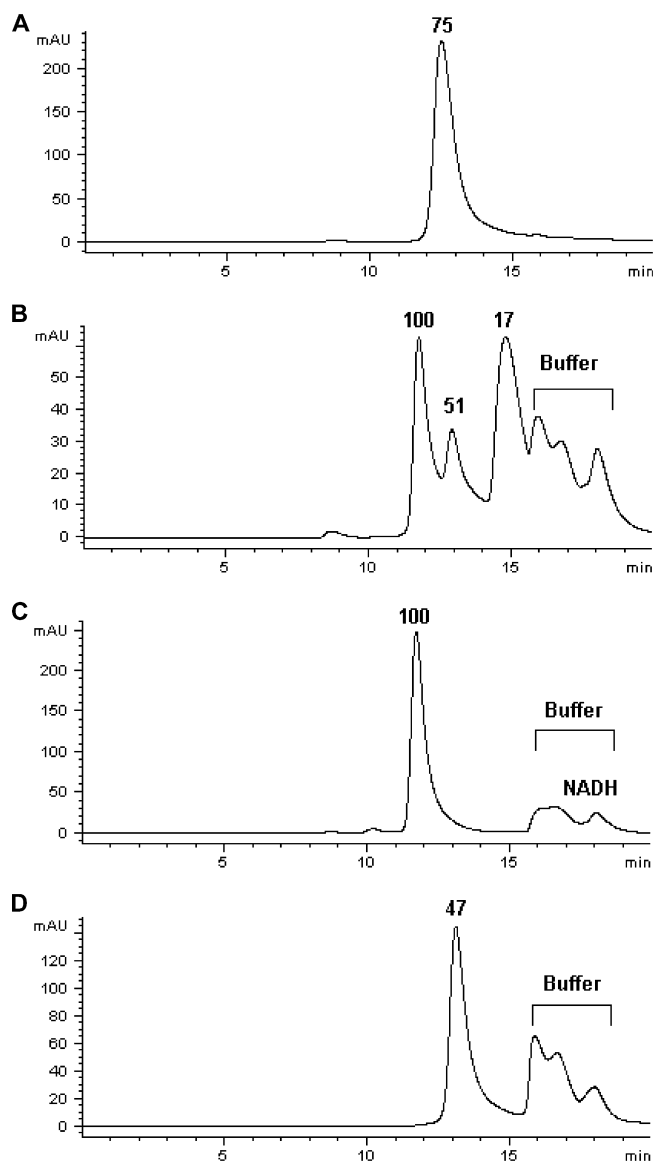


FIG. 4. Size exclusion chromatography of UGDs and UGER. (A) ATCV-1 UGD; (B) mimivirus UGD; (C) mimivirus UGD preincubated with 10  $\mu\text{M}$  NADH; (D) mimivirus UGER. Proteins were kept in 50 mM Tris-HCl-150 mM NaCl, pH 7.5. Mimivirus UGD and UGER proteins also contained 1 mM Na EDTA and 1 mM DTT (indicated as buffer). Detection was performed at 220 nm. Molecular masses (kDa) were calculated after column calibration using known standards, as indicated in the Materials and Methods section.

vealed three forms with apparent molecular masses of 100 kDa, 51 kDa, and 17 kDa (Fig. 4B). Incubation with either  $\text{NAD}^+$  or NADH converted the lower-molecular-mass peaks to the 100-kDa peak of mimivirus UGD. However, NADH was 10-fold more efficient in this conversion than  $\text{NAD}^+$  in inducing mimivirus UGD dimer formation. Figure 4C shows a chromatogram after incubation of mimivirus UGD with 10  $\mu\text{M}$  NADH; 100  $\mu\text{M}$   $\text{NAD}^+$  was required to achieve the same effect (results not shown). No change occurred with either  $\text{NADP}^+$  or NADPH, indicating that the protein has a high specificity for its coenzyme. A 100-kDa molecular mass is

higher than that expected for the dimer and the mass observed for ATCV-1 UGD. However, retention times in size exclusion chromatography can be influenced by the shape of the protein (19). We cannot presently draw any definitive conclusion about the 17-kDa species. No additional low-molecular-mass bands were observed on SDS-PAGE, thus excluding contamination from bacterial proteins. Moreover, this peak completely disappears when mimivirus UGD is incubated with either  $\text{NAD}^+$  or  $\text{NADH}$ . We postulate that this species may be a partially folded monomer which is converted to the dimeric form by  $\text{NAD(H)}$  binding, but more-specific analyses (i.e., circular dichroism, light scattering, or equilibrium sedimentation) will be required for conclusive identification. These results clearly indicate that, like for the PBCV-1 and ATCV-1 GMDs (14), a coenzyme is essential for both catalysis (see below) and structural integrity of UGD. Moreover, the higher ability of the reduced coenzyme to induce dimer formation resembles results obtained for viral GMDs (14).

Size exclusion chromatography of mimivirus UGER indicated an apparent molecular mass of 47 kDa (Fig. 4D). These results do not unequivocally support either a monomeric (33-kDa) or a dimer structure. Addition of a molar excess of either  $\text{NADP}^+$  or  $\text{NADPH}$  and/or bivalent cations did not alter the retention time of the protein.

**Enzyme characterization.** Enzyme assays indicated that both UGD proteins preferred UDP-D-glucose as a substrate; activity with dTDP-D-glucose was less than 10% of that with UDP-D-glucose. Figure 5A shows a chromatogram for ATCV-1 UGD, but similar results were obtained with the mimivirus enzyme (results not shown). ATCV-1 UGD was unaffected by adding  $\text{NAD}^+$ , indicating that the protein tightly bound the coenzyme during purification. ATCV-1 UGD had a specific activity of  $23 \text{ nmol/min/mg}$  of protein and a  $K_m$  for UDP-D-glucose of  $18.4 \pm 3.3 \mu\text{M}$ . In contrast, mimivirus UGD required exogenous  $\text{NAD}^+$  for activity (specific activity with  $100 \mu\text{M}$   $\text{NAD}^+$  was  $4.2 \text{ nmol/min/mg}$  of protein). This observation is consistent with the size exclusion chromatography experiments, which indicated that the same concentration of  $\text{NAD}^+$  induced a shift of the lower molecular masses to a 100-kDa species. The optimum pH for ATCV-1 UGD was 7.5, with a sharp decrease at both acidic and basic pHs, while mimivirus UGD had maximal activity over a wider range (between 7.5 and 8.5). Bivalent cations had no effect on the activity of either enzyme.

Feedback inhibition of UGD by the final product of the pathway was examined by monitoring enzyme activity in the presence of UDP-L-rhamnose. Addition of  $100 \mu\text{M}$  UDP-L-rhamnose to equimolar UDP-D-glucose resulted in 50% and 60% reductions in ATCV-1 and mimivirus UGD activities, respectively. These results confirm the inhibitory effect of UDP-L-rhamnose previously reported for *A. thaliana* RHM2 (34). Preincubation of concentrated ATCV-1 and mimivirus UGDs at temperatures higher than  $20^\circ\text{C}$  resulted in a progressive decrease in enzyme activity. Complete inactivation was achieved in 30 min at  $30^\circ\text{C}$ , indicating that both proteins are thermolabile.

To test mimivirus UGER enzyme activity, UDP-4-keto-6-deoxy-D-glucose was produced as a substrate using ATCV-1 UGD (Fig. 5B). When incubated in the presence of mimivirus UGER and  $\text{NADPH}$ , the intermediate compound was con-

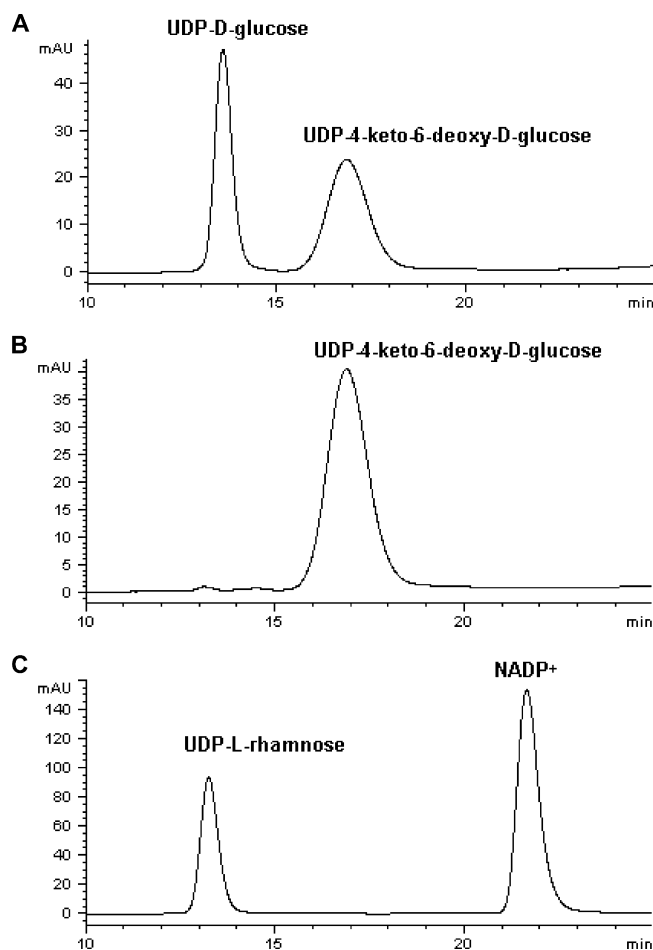


FIG. 5. Products of the enzymatic activity of UGD and UGER determined by anion-exchange chromatography. (A) UGD activity by ATCV-1; (B) production of the intermediate UDP-4-keto-6-deoxy-D-glucose by ATCV-1 UGD; (C) same as panel B but incubated with mimivirus UGER and  $\text{NADPH}$ .

verted to a new peak with a lower retention time (Fig. 5C). The specific activity and  $K_m$  for UDP-4-keto-6-deoxy-D-glucose were  $11 \pm 2.5 \text{ nmol/min/mg}$  and  $183 \pm 51 \mu\text{M}$ , respectively. Mimivirus UGER was specific for  $\text{NADPH}$ , since no activity occurred with  $\text{NADH}$ . Similar to the case for UGD, maximal activity was at pH 7.5. Addition of bivalent cations did not alter enzyme activity. UGER was stable after preincubation for 30 min at up to  $42^\circ\text{C}$ .

**Identification of UGD and UGER products.** The identities of the nucleotide sugars formed by UGD and UGER were confirmed by ESI-MS and ESI-MS/MS analyses. UDP-4-keto-6-deoxy-D-glucose ( $m/z$  547), consistent with UDP-D-glucose (565 Da) lacking a water molecule, was obtained in the UGD reaction (Fig. 6A). UDP-4-keto-6-deoxy-D-glucose, incubated in the presence of  $\text{NADPH}$  and UGER, produced a signal at  $m/z$  549, which agrees with the reduction of the 4-keto group and the formation of a UDP-6-deoxyhexose (Fig. 6B). The identification of these molecules was also confirmed by MS/MS analysis (results not shown).

GC-MS, performed after acid hydrolysis and conversion of the UGER product to the corresponding alditol acetates, es-

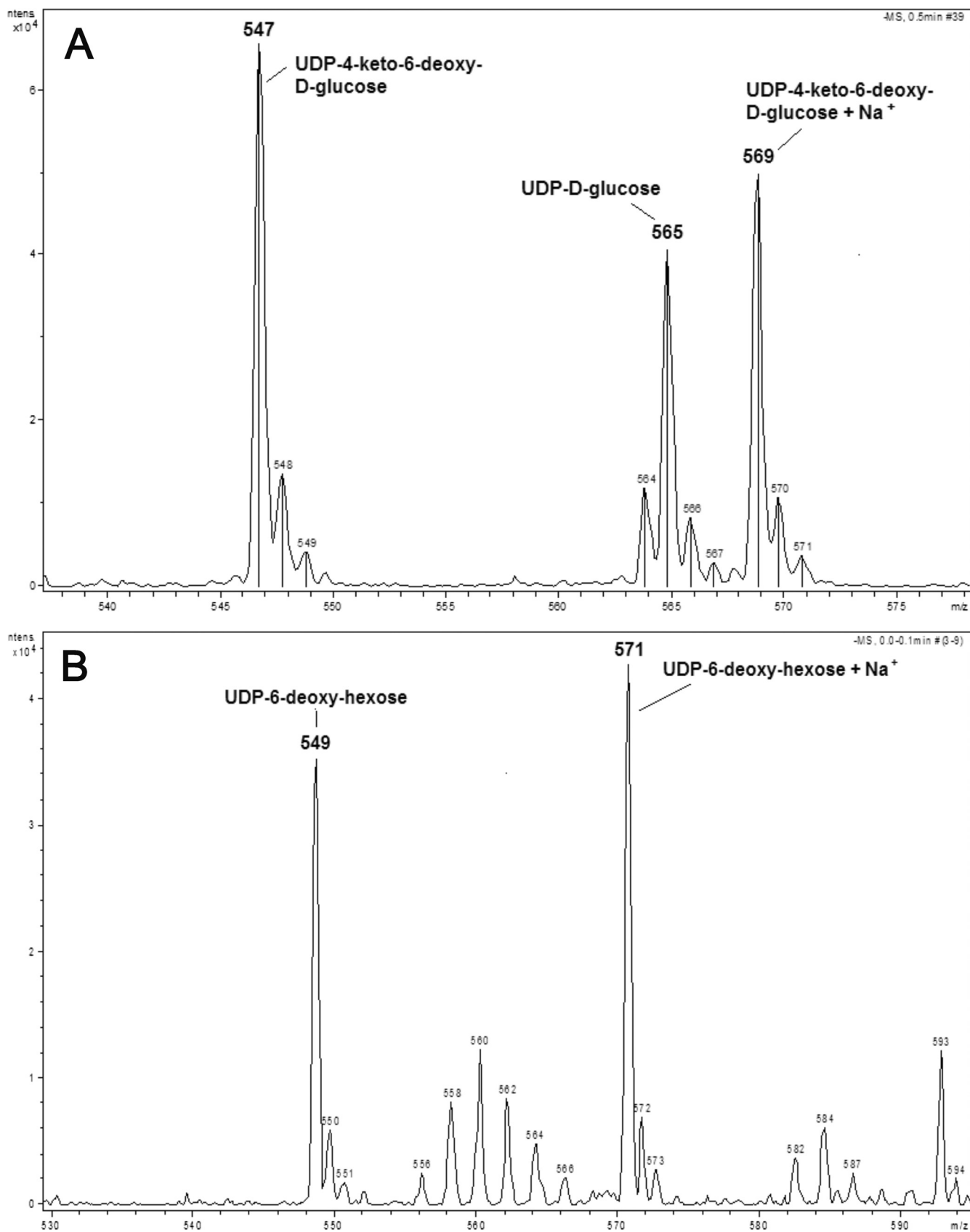


FIG. 6. Electrospray spectrum of UGD and UGER products. (A) Analysis of the UGD product, corresponding to chromatogram shown in Fig. 5A. (B) Analysis of the UGER product, corresponding to Fig. 5C.



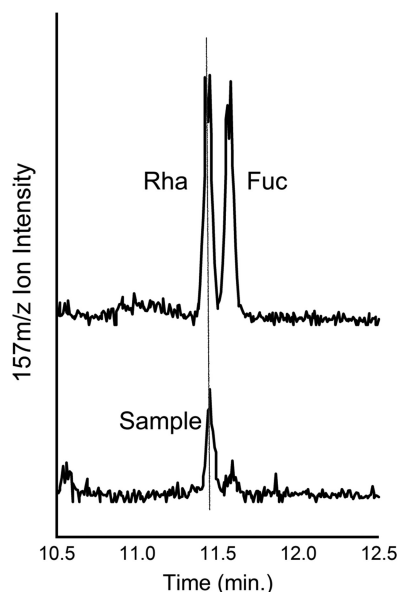


FIG. 7. GC-MS analysis of acetylated alditols. Extracted ion currents (XIC) of the ion at 157  $m/z$  are shown for the standards rhamnose (Rha) and fucose (Fuc) and for the UGER product (sample).

established that L-rhamnose was the product of the reaction (Fig. 7). Unambiguous identification was achieved by mass spectrum interpretation and by comparison of retention times with rhamnose, fucose, and 6-deoxy-glucose standards. The presence of 6-deoxy-D-glucose, which is expected from a 4-reductase activity only, was not detected. These results confirm that the enzyme has both a 3,5-epimerase activity and a stereospecific NADPH-dependent 4-reductase activity leading to rhamnose formation.

**Transcription profiling of ATCV-1 and mimivirus UGD and UGER genes.** ATCV-1 Z544R (UGD) expression was determined by real-time PCR after RNA extraction at different times p.i (Fig. 8A). RNA levels peaked at 6 h p.i, showing a decrease at 8 h. At 8 h, significant cell lysis occurred, indicating that the viral life cycle was completed. Thus, Z544R is a late gene; this differs from the expression profiles of the enzymes involved in GDP-L-fucose synthesis, which are expressed as early genes (40).

For mimivirus, as is customary in sequence tag-based transcriptome studies, a transcription profile for each gene was determined, following a normalization procedure, from the counts of its cognate reads at each time. These numbers were then treated as gene “coordinates” in seven-dimensional space and used in a variety of classical statistical methods allowing pairwise comparison of gene profiles (e.g., distance or correlation indices), their “clustering” into groups sharing similar profiles, and the identification and visualization of the dominant transcriptional patterns (27).

The mimivirus R141 (UGD) and L780 (UGER) genes both belong to the “late” cluster, which includes genes encoding structural components of the virus particles (such as the main capsid protein L425 or the core protein L410), genes encoding enzymes carried by the particle, and genes encoding enzymes most likely involved in the biosynthesis of the LPS-like outer layer of the virus particle (6, 7, 47). The expression profiles of

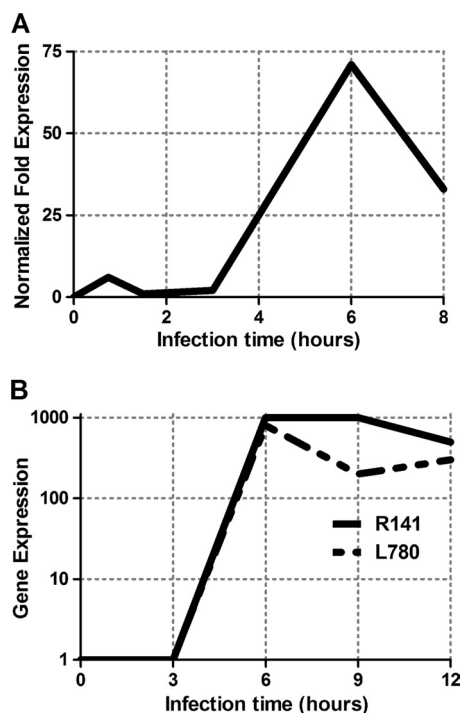


FIG. 8. (A) Expression profiles of the ATCV-1 Z544R (UGD) gene determined by real-time PCR at different times p.i. ATCV-1 DNA replication begins at 60 to 90 min p.i. (B) Expression profiles of mimivirus R141 (UGD) and L780 (UGER) genes.

these two genes are similar (Fig. 8B), suggesting that they are expressed together and could be key enzymes in the biosynthesis of the fiber layer surrounding the mimivirus particle; thus, rhamnose could be one of the sugars in the fibers responsible for the phagocytosis of the host.

## DISCUSSION

The results presented here establish that two NCLDV viruses, ATCV-1 and mimivirus, members of the *Phycodnaviridae* and *Mimiviridae*, respectively, encode and express functional proteins involved in L-rhamnose biosynthesis. Thus, virus-encoded enzymes involved in sugar nucleotide synthesis and glycan production are not restricted to the chloroviruses as previously reported (14, 26, 40); this property is shared with some other large DNA viruses. These enzymes presumably contribute to the formation of specific glycan structures, which differ from those in host cells and which contribute to virus infectivity. This property differs from those of most viruses, which use the host ER/Golgi system for glycosylation of their surface proteins and whose glycan structure and composition are completely dependent on their host cells (30). The glycans produced by the virus PBCV-1, which infects *Chlorella* sp. NC64A, have features not found in other organisms. For example, N-linked glycans in the PBCV-1 major capsid protein are bound to Asn located in sequences different from those typically found in eukaryotic cells (33). Moreover, the final glycan structure probably differs from those of eukaryotic N-linked glycans (43). Similarly, several proteins in mimivirus particles are glycosylated, and genes encoding putative en-

zymes involved in glycan synthesis (including glycosyltransferases) are present in its genome (36, 37). Thus, mimivirus may also encode some, if not all, of its glycosylation machinery. Studies on the enzymes involved in glycoconjugate production encoded by NCLDV s could provide important insights, not only about their role in viral life cycles but also about the predicted long evolutionary history of these viruses, including relationships with bacterial glycans and the evolution of the eukaryotic glycosylation machinery (42).

L-Rhamnose is common in plants, where it is an important component of surface polysaccharides as well as small molecules. However, L-rhamnose is not present in animals, at least in higher vertebrates. Very little is known about the presence of L-rhamnose in Protista and lower animals. Enzymes involved in the UDP-L-rhamnose pathway in *Arabidopsis thaliana* have been identified and characterized (34, 44). *A. thaliana* encodes three isoforms, named RHM proteins, which result from a fusion between a dehydratase and an epimerase/reductase; indeed, fusion of enzymes in the same pathway is a common strategy used for metabolic channeling. Enzymes involved in UDP-L-rhamnose biosynthesis were described over 40 years ago in a *Chlorella* sp., and L-rhamnose is a major component of its cellular polysaccharides (3, 46). Recently, the *Chlorella* sp. NC64A genome was sequenced and annotated (Blanc et al., submitted); UGD- and UGER-like proteins with high amino acid identities to the *A. thaliana* enzymes were identified. Even though the genome sequence of the ATCV-1 host, *Chlorella* sp. SAG 3.83, is not available, the ATCV-1-encoded UGD has 80% amino acid identity with the *Chlorella* sp. NC64A UGD. Thus, it is likely that ATCV-1 acquired the gene from an algal host by a recent HGT. It should be noted that genes encoding UGD-like proteins are not present in five other sequenced chloroviruses, suggesting that this enzymatic activity is not essential for all *Chlorella* viruses. On the other hand, both putative enzymes for L-rhamnose synthesis (UGD and UGER) are encoded by virus OsV5, another member of the *Phycodnaviridae*, which infects *Ostreococcus tauri*, the smallest known marine photosynthetic eukaryote (11).

Two obvious questions are, what is the role of UGD in the virus ATCV-1 replication cycle, and why does the virus encode only UGD and not the rest of the L-rhamnose biosynthetic pathway. Presumably, the UGD product, UDP-4-keto-6-deoxy-D-glucose, is converted to UDP-L-rhamnose by a host UGER enzyme. (Note that *Chlorella* sp. NC64A, a host for some of the chloroviruses, including PBCV-1, contains a putative UGER [Blanc et al., submitted].) Therefore, *Chlorella* sp. SAG 3.83 also probably encodes a UGER. The resulting L-rhamnose could be one of the component sugars in the ATCV-1 major capsid protein. Indeed, preliminary sugar analysis of the ATCV-1 major capsid protein indicates that rhamnose is present in its glycan(s) (Parakkottil Chothi et al., unpublished results), as previously reported for the PBCV-1 major capsid protein (43). Interestingly, the PBCV-1-encoded GMD enzyme has not only the expected dehydratase activity but also has a reductase activity that produces GDP-D-rhamnose (40). However, the ATCV-1 GMD lacks this reductase activity and does not produce rhamnose (14). Hence, the ATCV-1 enzyme could be required to augment rhamnose production.

Presently, nothing is known about a possible role for L-rhamnose in the mimivirus life cycle. The viral capsid is cov-

ered by dense glycosylated fibers that contribute to the Gram-positive staining of the virus, which resulted in mimivirus initially being identified as a bacterium (35). As indicated above for ATCV-1, preliminary GC-MS analyses indicate that L-rhamnose is present in large amounts in mimivirus particles (Parakkottil Chothi et al., unpublished results). The complete compositions and structures of mimivirus glycans are currently under study, with the oligosaccharide moiety(ies) being important for mimivirus infection of different hosts (7). Mimivirus infects amoebae and other cells by mimicking a bacterial surface, leading to phagocytosis. Nothing is known about the activity of a UDP-L-rhamnose pathway in the *Acanthamoeba* host, and it is not known if and how the host contributes to the glycosylation of mimivirus proteins. Putative UGD and UGER genes are expressed in *A. castellanii* (deposited in the NCBI Short Read Archive, accession no. SRA 010763), but enzymatic activities of their products have not been tested. However, mimivirus infects other phagocytic cell types, including mammalian macrophages, and infectious viruses are produced (16). Since mammalian cells lack an L-rhamnose biosynthetic pathway, a virus-encoded complete UDP-L-rhamnose pathway would be essential to provide this sugar in this host.

The data presented here establish that ATCV-1 UGD and mimivirus UGD and UGER are expressed as late gene products during virus replication. This is consistent with them playing a role in posttranslational modification of capsid proteins. Interestingly, mimivirus genes are coexpressed in the same cluster and at the same time as putative enzymes involved in forming the fibrillar outer layer of the viral particles. These proteins include glycosyltransferases (R139 and L193), acyl transferase (L142), acetyltransferase (L280, L316, R363, and L373), mannose-6-phosphate epimerase (L612), glucosamine-fructose-6-phosphate aminotransferase (L619), and collagen-like proteins (R239, R240, and R241). It is important to note that the enzymes involved in GDP-L-fucose and GDP-D-rhamnose synthesis previously identified in PBCV-1 are expressed as early gene products (40). This differs from the expression pattern of ATCV-1 UGD, which is expressed late during infection. However, this difference could be due to differences in the glycosylation pathway between the two viruses. ATCV-1 is the only *Chlorella* virus identified so far to code for UGD, and it is possible that the role of this enzyme is to increase the L-rhamnose supply already produced by *Chlorella* cells, in particular during the last stages of virus replication. It is also possible that very high UGD activity during the initial phases of virus replication is avoided to prevent a reduction of UDP-D-glucose availability, since this nucleotide-sugar serves both as a donor for its specific glycosyltransferase and as a precursor for UDP-D-galactose. Indeed, preliminary results by GC-MS on the monosaccharide composition of ATCV-1 indicate that glucose and galactose are the main components of the ATCV-1 glycans (Parakkottil Chothi et al., unpublished results). A small peak of UGD expression was also observed at around 45 min p.i., suggesting that UGD production may be biphasic, with a small amount produced in the early phase and higher expression in later stages of virus replication.

The L-rhamnose biosynthetic enzymes identified in both viruses are more closely related to enzymes in the plant pathway (34). In fact, bacteria use dTDP, not UDP, to activate glucose

to initiate the pathway. Another important difference is that the epimerization and reduction steps in bacteria are catalyzed by two proteins, RmlC (RfbC), characterized by a cupin domain, and RmlD (RfbD) (17, 18, 20). This difference could be due to the fact that dTDP-4-keto-6-deoxy-D-glucose can also be used by bacteria to produce other modified sugars (1), while this pathway is probably specific for L-rhamnose production in plants and the viruses reported here. The epimerase/reductase activity found in plants and mimivirus UGER resembles that of the last enzyme in the GDP-L-fucose pathway, the GMER/FX protein, which catalyzes a similar two-step reaction (39).

ATCV-1 and mimivirus UGDs differ in their enzymatic behavior. In particular, mimivirus UGD requires exogenously added NAD<sup>+</sup> for maximal activity, while this coenzyme remains tightly bound to the ATCV-1 enzyme during purification. NAD<sup>+</sup> or NADP<sup>+</sup> is required for all nucleotide-sugar dehydratases to allow the internal oxido-reduction reaction in which the hydride removed from the C-4 is temporarily transferred to the coenzyme and then used to reduce the C-6 (1). We have previously demonstrated that the coenzyme is essential for structure maintenance in the chlorovirus GMDs (14). Indeed, for mimivirus UGD, addition of NAD<sup>+</sup> promotes the catalytic activity and promotes structural changes, confirming that the coenzyme also plays a key role in quaternary assembly of UGD enzymes. We cannot explain the lower affinity of mimivirus UGD for NAD<sup>+</sup>. However, it is worth noting that mimivirus UGD contains an Asn at position 29; the corresponding residue in ATCV-1 UGD and *A. thaliana* RHM2 is a Lys that is well conserved among different species. This residue is located in the coenzyme-binding region and seems particularly important, since the RHM2 K36A mutant has a significantly lower activity than wild-type protein (34). Crystallographic analyses of UGD proteins, which are currently under way, and site-directed mutagenesis will help clarify the role of this residue in coenzyme binding affinity.

#### ACKNOWLEDGMENTS

This research was supported in part by the FIRB 2001 project (M.T.), the FIRB 2008 project (G.L.D.), Public Health Service grant GM32441 (J.L.V.E.), NIH grant P21RR15635 from the COBRE program of the National Center for Research Resources (J.L.V.E.), NIH grant P20 RR106469 from the INBRE program (G.A.D.), and CNRS and ANR grant ANR-08-BLAN-0089 (C.A.).

#### REFERENCES

- Allard, S. T., K. Beis, M. F. Giraud, A. D. Hegeman, J. W. Gross, R. C. Wilmouth, C. Whitfield, M. Graninger, P. Messner, A. G. Allen, D. J. Maskell, and J. H. Naismith. 2002. Toward a structural understanding of the dehydratase mechanism. *Structure* **10**:81–92.
- Allard, S. T., M. F. Giraud, C. Whitfield, P. Messner, and J. H. Naismith. 2000. The purification, crystallization and structural elucidation of dTDP-D-glucose 4,6-dehydratase (RmlB), the second enzyme of the dTDP-L-rhamnose synthesis pathway from *Salmonella enterica* serovar typhimurium. *Acta Crystallogr. D Biol. Crystallogr.* **56**:222–225.
- Barber, G. A., and M. T. Chang. 1967. Synthesis of uridine diphosphate L-rhamnose by enzymes of *Chlorella pyrenoidosa*. *Arch. Biochem. Biophys.* **118**:659–663.
- Byrne, D., R. Grzela, A. Lartigue, S. Audic, S. Chenivisse, S. Encinas, J. M. Claverie, and C. Abergel. 2009. The polyadenylation site of Mimivirus transcripts obeys a stringent 'hairpin rule'. *Genome Res.* **19**:1233–1242.
- Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**:540–552.
- Claverie, J. M., and C. Abergel. 2009. Mimivirus and its viroplasm. *Annu. Rev. Genet.* **43**:49–66.
- Claverie, J. M., R. Grzela, A. Lartigue, A. Bernadac, S. Nitsche, J. Vacelet, H. Ogata, and C. Abergel. 2009. Mimivirus and Mimiviridae: Giant viruses with an increasing number of potential hosts, including corals and sponges. *J. Invertebr. Pathol.* **101**:172–180.
- Claverie, J. M., C. Abergel, and H. Ogata. 2009. Mimivirus. *Curr. Top. Microbiol. Immunol.* **328**:89–121.
- Claverie, J. M., H. Ogata, S. Audic, C. Abergel, K. Suhre, and P. E. Fournier. 2006. Mimivirus and the emerging concept of "giant" virus. *Virus Res.* **117**:133–144.
- Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie, and O. Gascuel. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**:W465–W469.
- Derelle, E., C. Ferraz, M. L. Escande, S. Eychenié, R. Cooke, G. Piganeau, Y. Desdèves, L. Bellec, H. Moreau, and N. Grimsley. 2008. Life-cycle and genome of OtV5, a large DNA virus of the pelagic marine unicellular green alga *Ostreococcus tauri*. *PLoS One* **3**:e2250.
- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**:1792–1797.
- Fitzgerald, L. A., M. V. Graves, X. Li, J. Hartigan, A. J. Pfizner, E. Hoffart, and J. L. Van Etten. 2007. Sequence and annotation of the 288-kb ATCV-1 virus that infects an endosymbiotic chlorella strain of the heliozoan *Acanthocystis turfacea*. *Virology* **362**:350–361.
- Fruscione, F., L. Sturla, G. Duncan, J. L. Van Etten, P. Valbuzzi, A. De Flora, E. Di Zanni, and M. Tonetti. 2008. Differential role of NADP<sup>+</sup> and NADPH in the activity and structure of GDP-D-mannose 4,6-dehydratase from two chlorella viruses. *J. Biol. Chem.* **283**:184–193.
- Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. 2005. Protein identification and analysis tools on the Expasy server, p. 571–607. *In* J. M. Walker (ed.), *The proteomics protocols handbook*. Humana Press, Totowa, NJ.
- Ghigo, E., J. Kartenbeck, P. Lien, L. Pelkmans, C. Capo, J. L. Mege, and D. Raoult. 2008. Ameobal pathogen mimivirus infects macrophages through phagocytosis. *PLoS Pathog.* **4**:e1000087.
- Giraud, M. F., G. A. Leonard, R. A. Field, C. Berling, and J. H. Naismith. 2000. RmlC, the third enzyme of dTDP-L-rhamnose pathway, is a new class of epimerase. *Nat. Struct. Biol.* **7**:398–402.
- Giraud, M. F., H. J. McMiken, G. A. Leonard, P. Messner, C. Whitfield, and J. H. Naismith. 1999. Overexpression, purification, crystallization and preliminary structural study of dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD), the fourth enzyme of the dTDP-L-rhamnose synthesis pathway, from *Salmonella enterica* serovar typhimurium. *Acta Crystallogr. D Biol. Crystallogr.* **55**:2043–2046.
- Goetz, H., M. Kuschel, T. Wulff, C. Sauber, C. Miller, S. Fisher, and C. Woodward. 2004. Comparison of selected analytical techniques for protein sizing, quantitation and molecular weight determination. *J. Biochem. Biophys. Methods* **60**:281–293.
- Graninger, M., B. Nidetzky, D. E. Heinrichs, C. Whitfield, and P. Messner. 1999. Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar typhimurium LT2. *J. Biol. Chem.* **274**:25069–25077.
- Graves, M. V., C. T. Bernadt, R. Cerny, and J. L. Van Etten. 2001. Molecular and genetic evidence for a virus-encoded glycosyltransferase involved in protein glycosylation. *Virology* **285**:332–345.
- Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
- Henry, R. J., A. B. Blakney, P. J. Harris, and B. A. Stone. 1983. Detection of neutral and aminosugars from glycoproteins and polysaccharides as their alditol acetates. *J. Chromatogr.* **256**:419–427.
- Ikan, R. (ed.). 1999. *Naturally occurring glycosides*. Wiley, London, United Kingdom.
- Iyer, L. M., S. Balaji, E. V. Koonin, and L. Aravind. 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* **117**:156–184.
- Landstein, D., M. V. Graves, D. E. Burbank, P. DeAngelis, and J. L. Van Etten. 1998. Chlorella virus PBCV-1 encodes functional glutamine: fructose-6-phosphate amidotransferase and UDP-glucose dehydrogenase enzymes. *Virology* **250**:388–396.
- Legendre, M., S. Audic, O. Poirot, D. Byrne, A. Lartigue, L. Lescot, A. Bernadac, J. Poulain, C. Abergel, and J. M. Claverie. 2010. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res.* **20**:664–674.
- Ma, Y., F. Pau, and M. McNeil. 2002. Formation of dTDP-rhamnose is essential for growth of mycobacteria. *J. Bacteriol.* **184**:3392–3395.
- Mäki, M., and R. Renkonen. 2004. Biosynthesis of 6-deoxyhexose glycans in bacteria. *Glycobiology* **14**:1R–15R.
- Markine-Goriaynoff, N. L., L. Gillet, J. L. Van Etten, H. Korres, N. Verma, and A. Vanderplasschen. 2004. Glycosyltransferases encoded by viruses. *J. Gen. Virol.* **85**:2741–2754.
- Merkle, R. K., and I. Poppe. 1994. Carbohydrate composition analysis of glycoconjugates by gas-liquid chromatography/mass spectrometry. *Methods Enzymol.* **230**:1–15.
- Monier, A., A. Pagarete, C. De Vargas, M. J. Allen, B. Read, J. M. Claverie,

- and H. Ogata. 2009. Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Res.* **19**:1441–1449.
33. Nandhagopal, N., A. A. Simpson, J. R. Gurnon, X. Yan, T. S. Baker, M. V. Graves, J. L. Van Etten, and M. G. Rossmann. 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.
34. Oka, T., T. Nemoto, and Y. Jigami. 2007. Functional analysis of *Arabidopsis thaliana* RHM2/MUM4, a multidomain protein involved in UDP-D-glucose to UDP-L-rhamnose conversion. *J. Biol. Chem.* **282**:5389–5403.
35. Raoult, D., B. La Scola, and R. Birtles. 2007. The discovery and characterization of Mimivirus, the largest known virus and putative pneumonia agent. *Clin. Infect. Dis.* **45**:95–102.
36. Raoult, D., S. Audic, C. Robert, C. Abergel, P. Renesto, H. Ogata, B. La Scola, M. Suzan-Monti, and J. M. Claverie. 2004. The 1.2-megabase genome sequence of Mimivirus. *Science* **306**:1344–1350.
37. Renesto, P., C. Abergel, P. Decloquement, D. Moinier, S. Azza, H. Ogata, P. Forquet, J.-P. Gorvel, and J.-M. Claverie. 2006. Mimivirus giant particles incorporate a large fraction of anonymous and unique gene products. *J. Virol.* **80**:11678–11685.
38. Ridley, B. L., M. A. O'Neill, and D. Mohnen. 2001. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **57**:929–967.
39. Tonetti, M., L. Sturla, A. Bisso, U. Benatti, and A. De Flora. 1996. Synthesis of GDP-L-fucose by the human FX protein. *J. Biol. Chem.* **271**:27274–27279.
40. Tonetti, M., D. Zanardi, J. R. Gurnon, F. Fruscione, A. Armirotti, G. Damonte, L. Sturla, A. De Flora, and J. L. Van Etten. 2003. Paramyxovirus 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.
41. Van Etten, J. L. 2003. Unusual life style of giant chlorella viruses. *Annu. Rev. Genet.* **37**:153–195.
42. Van Etten, J. L., J. Gurnon, G. Yanai-Balser, D. Dunigan, and M. V. Graves. 2010. Chlorella viruses encode most, if not all, of the machinery to glycosylate their glycoproteins independent of the endoplasmic reticulum. *Biochim. Biophys. Acta* **1800**:152–159.
43. Wang, I. N., Y. Li, Q. Que, M. Bhattacharya, L. C. Lane, W. G. Chaney, and J. L. Van Etten. 1993. Evidence for virus-encoded glycosylation specificity. *Proc. Natl. Acad. Sci. U. S. A.* **90**:3840–3844.
44. Watt, G., C. Leoff, A. D. Harper, and M. Bar-Peled. 2004. A bifunctional 3,5-epimerase/4-keto reductase for nucleotide-rhamnose synthesis in *Arabidopsis*. *Plant Physiol.* **134**:1337–1343.
45. Welling, G. W., and S. Welling-Wester. 1989. Size-exclusion HPLC of proteins, p. 77–89. *In* R. W. Oliver (ed.), *HPLC of macromolecules, a practical approach*. Oxford University Press, Oxford, United Kingdom.
46. White, R. C., and G. A. Barber. 1972. An acidic polysaccharide from the cell wall of *Chlorella pyrenoidosa*. *Biochim. Biophys. Acta* **264**:117–128.
47. Xiao, C., Y. G. Kuznetsov, S. Sun, S. L. Hafenstein, V. A. Kostyuchenko, P. R. Chipman, M. Suzan-Monti, D. Raoult, A. McPherson, and M. G. Rossmann. 2009. Structural studies of the giant mimivirus. *PLoS Biol.* **7**:e92.
48. Yang, H., A. I. Ara, N. Magilnick, M. Xia, K. Ramani, H. Chen, T. D. Lee, J. M. Mato, and S. C. Lu. 2008. Expression pattern, regulation, and functions of methionine adenosyltransferase 2beta splicing variants in hepatoma cells. *Gastroenterology* **134**:281–291.
49. Yutin, N., Y. I. Wolf, D. Raoult, and E. V. Koonin. 2009. Eukaryotic large nucleocytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol. J.* **6**:223.