AMERICAN SOCIETY FOR MICROBIOLOGY Microbiology Spectrum

RESEARCH ARTICLE

RNA Sequencing of Medusavirus Suggests Remodeling of the Host Nuclear Environment at an Early Infection Stage

Ruixuan Zhang,[a] Hisashi Endo,[a] Masaharu Takemura,[b] Hiroyuki Ogata[a]

[a]Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Japan

[b]Laboratory of Biology, Institute of Arts and Sciences, Tokyo University of Science, Shinjuku, Tokyo, Japan

ABSTRACT Viruses of the phylum *Nucleocytoviricota*, or nucleo-cytoplasmic large DNA viruses (NCLDVs), undergo a cytoplasmic or nucleo-cytoplasmic cycle, the latter of which involves both nuclear and cytoplasmic compartments to proceed viral replication. Medusavirus, a recently isolated NCLDV, has a nucleo-cytoplasmic replication cycle in amoebas during which the host nuclear membrane apparently remains intact, a unique feature among amoeba-infecting NCLDVs. The medusavirus genome lacks most transcription genes but encodes a full set of histone genes. To investigate its infection strategy, we performed a time course RNA sequencing (RNA-seq) experiment. All viral genes were transcribed and classified into five temporal expression clusters. The immediate early genes (cluster 1, 42 genes) were mostly (83%) of unknown functions, frequently (95%) associated with a palindromic promoter-like motif, and often (45%) encoded putative nucleus-localized proteins. These results suggest massive reshaping of the host nuclear environment by viral proteins at an early stage of infection. Genes in other expression clusters (clusters 2 to 5) were assigned to various functional categories. The virally encoded core histone genes were in cluster 3, whereas the viral linker histone H1 gene was in cluster 1, suggesting they have distinct roles during the course of the virus infection. The transcriptional profile of the host *Acanthamoeba castellanii* genes was greatly altered postinfection. Several encystment-related host genes showed increased representation levels at 48 h postinfection, which is consistent with the previously reported amoeba encystment upon medusavirus infection.

IMPORTANCE Medusavirus is an amoeba-infecting giant virus that was isolated from a hot spring in Japan. It belongs to the proposed family "*Medusaviridae*" in the phylum *Nucleocytoviricota*. Unlike other amoeba-infecting giant viruses, medusavirus initiates its DNA replication in the host nucleus without disrupting the nuclear membrane. Our RNA sequencing (RNA-seq) analysis of its infection course uncovered ordered viral gene expression profiles. We identified temporal expression clusters of viral genes and associated putative promoter motifs. The subcellular localization prediction showed a clear spatiotemporal correlation between gene expression timing and localization of the encoded proteins. Notably, the immediate early expression cluster was enriched in genes targeting the nucleus, suggesting the priority of remodeling the host intranuclear environment during infection. The transcriptional profile of amoeba genes was greatly altered postinfection.

KEYWORDS NCLDV, RNA-seq, giant virus, medusavirus

Giant viruses are characterized by their large viral particles and complex genomes and are found worldwide (1–6). They have been classified within the phylum *Nucleocytoviricota* (also referred to as nucleo-cytoplasmic large DNA viruses [NCLDVs]) (7). Phylogenetic analyses suggested that the diversification of this group of viruses predated the emergence of modern eukaryotic lineages (8, 9), which revived the

Citation Zhang R, Endo H, Takemura M, Ogata H. 2021. RNA sequencing of medusavirus suggests remodeling of the host nuclear environment at an early infection stage. Microbiol Spectr 9:e00064-21. https://doi.org/10.1128/Spectrum.00064-21.

Editor Samuel K. Campos, University of Arizona

Address correspondence to Masaharu Takemura, giantvirus@rs.tus.ac.jp, or Hiroyuki Ogata, ogata@kuicr.kyoto-u.ac.jp.

Received 23 April 2021
Accepted 16 August 2021
Published 29 September 2021

debate about their evolutionary origin (10, 11) and their relationship to the genesis of the eukaryotic nucleus (12, 13). Genomic analysis revealed a large number of genes (referred to as orphan genes) without detectable homology to any known genes. The abundance of orphan genes or lineage-specific genes has been considered evidence that supports the ongoing *de novo* creation of genes in these viruses (14, 15). In addition to the efforts to isolate and characterize new giant viruses, environmental genomics has revealed their ubiquitous nature, extensive gene transfers with eukaryotes, and complex metabolic capabilities (16–18).

Medusavirus, a giant virus that infects the amoeba *Acanthamoeba castellanii*, was isolated from a hot spring in Japan (2). Recently, a related virus, medusavirus stheno, was isolated from fresh water in Japan (19) and another distantly relate virus, clandestinovirus, was isolated from wastewater in France (20). These three viruses represent the proposed family "*Medusaviridae*" (2), which is distantly related to other giant virus families and forms an independent branch in the tree of the phylum *Nucleocytoviricota*. During the infection cycle of medusavirus, its genome enters the host nucleus to initiate DNA replication, and particle assembly and DNA packaging are carried out in the cytoplasm. Of note, the host nuclear membrane remains intact until near the end of the viral replication cycle, which represents a unique feature of medusavirus among currently characterized amoeba-infecting giant viruses. The viral replication cycles of other amoeba-infecting giant viruses are characterized as either a cytoplasmic replication by establishing cytoplasmic viral factories (e.g., mimiviruses [21], marseilleviruses [22], pithoviruses [5], cedratvirus [23], and orpheovirus [24]) or a nucleo-cytoplasmic replication, like in medusavirus, but with a degradation of the host nucleus (e.g., pandoraviruses [15] and molliviruses [1]). For medusavirus, no visible cytoplasmic virus factory has been observed by transmission electron microscopy (2). Thus, it appears that the host nucleus is transformed into a virus factory, from which mature and immature medusavirus virions emerge. It has also been reported that some of the host amoeba cells display encystment upon medusavirus infection as early as 48 h postinfection (hpi) (2). Medusavirus has a 381-kb genome that encodes 461 putative proteins; 86 (19%) have their closest homologs in *A. castellanii*, whereas 279 (61%) are orphan genes. Compared with other amoeba-infecting giant viruses, medusaviruses have fewer transcriptional and translational genes and have no genes that encode RNA polymerases and aminoacyl-tRNA synthetases, suggesting that medusaviruses are heavily reliant on the host machinery for transcription and translation. In contrast to their paucity in expression-related genes, medusaviruses are unique among known viruses in encoding a complete set of histone domains, namely, the core histones (H2A, H2B, H3, and H4) and the linker histone H1. A virion proteomic study detected proteins encoded by the four core histone genes in medusavirus particles (2). Given these unique features, medusavirus is expected to have a characteristic infection strategy among known amoeba-infecting giant viruses. However, the dynamics of gene expression during the medusavirus infection cycle has not been investigated so far.

Previous RNA sequencing (RNA-seq) studies of giant viruses detected viral genes that were expressed in a coordinated manner during the viral infection. Viral genes that belong to different functional categories tend to show different expression patterns and can be grouped as, for instance, early, intermediate, or late. Different viruses also have different gene expression programs; for example, the transcription order of informational genes (those involved in replication, transcription, translation, and related processes) can differ among viruses. The expression of DNA replication genes (starting from 3 hpi) precedes the expression of transcription-related genes (6 hpi) in mimivirus, whereas this order is reversed in marseillevirus (i.e., transcription-related genes from <1 hpi and DNA replication genes at 1 to 2 hpi) (25, 26). Putative promoter motifs associated with temporal expression groups have been identified in mimiviruses

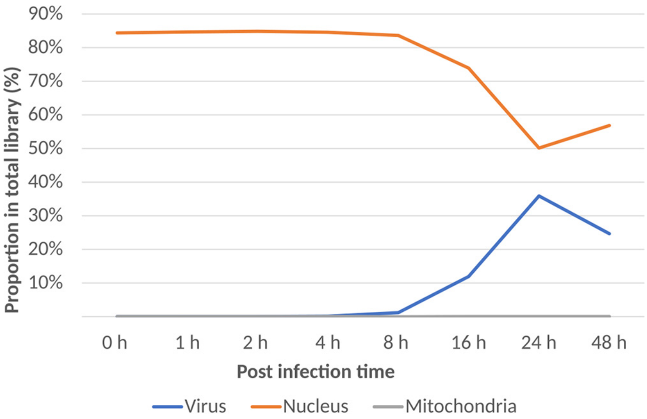

FIG 1 Proportions of viral and host mRNA reads at different time points during the course of medusavirus infection.

and marseilleviruses (25–28). The expression patterns of host genes during infection of giant viruses have been investigated by RNA-seq and proteomics (1, 26, 28).

We performed a time-series RNA-seq analysis of infected amoeba cells to investigate the transcriptional program and infection strategy of medusavirus. We report expression clusters of medusavirus genes, putative viral promoter motifs, and changes in host gene expression.

RESULTS

Transcription profile of medusavirus genes. The overall composition of the mRNA library during the course of the virus infection is shown in Fig. 1 (see Data set S1 in the supplemental material). Until 8 hpi, viral reads were less than 1% of the total reads, and then they increased and reached a peak at 24 hpi. The proportion of host reads stayed at a high level during the first 8 h and then decreased rapidly and reached a minimum at 24 h, which still accounted for approximately half of the total reads in the library.

All viral genes were gradually expressed and continuously increased up to 16 hpi (Fig. 2A). We identified five clusters of viral gene expression profiles using the *k*-means method (Fig. 2B) and named these clusters as follows: cluster 1 (immediate early) genes showed a gradual increase in expression from 0 hpi; cluster 2 (early) genes showed a gradual increase in expression from 1 hpi; clusters 3 and 4 (intermediate) genes showed a gradual increase in expression from 2 hpi; and cluster 5 (late) genes showed a gradual increase in expression from 4 hpi. The expression patterns of genes in clusters 3 and 4 were only slightly different; genes in cluster 3 showed higher Z-score scaled reads per kilobases of transcript per Million mapped reads (RPKM) values at 8 hpi than those in cluster 4. In the following text, both of these clusters were referred to as "intermediate" genes.

The distribution of genes with annotated functions showed characteristic patterns among the five expression clusters (Fig. 3A to C; Data set S8). Among the 42 genes in cluster 1 (i.e., immediate early), 35 (83%) were unknown genes. Of those annotated in this cluster, there were a linker histone H1 gene and a poly-A polymerase regulatory subunit gene. The proteins encoded by these two genes were not detected in a previous virion proteomic study of medusavirus (2). Cluster 2 included genes that were classified in the "nucleotide metabolism" and "DNA replication, recombination, and repair" categories, including a DNA helicase, a DNA primase, and ribonucleotide reductase large/small subunits. Cluster 3 contained genes in various functional categories, including histone genes (the four core histone genes H2A, H2B, H3, and H4), "DNA replication, recombination, and repair" category (e.g., two of five nuclease genes, a Yqaj viral recombinase gene, and a Holliday junction resolvase gene), "Transcription and RNA

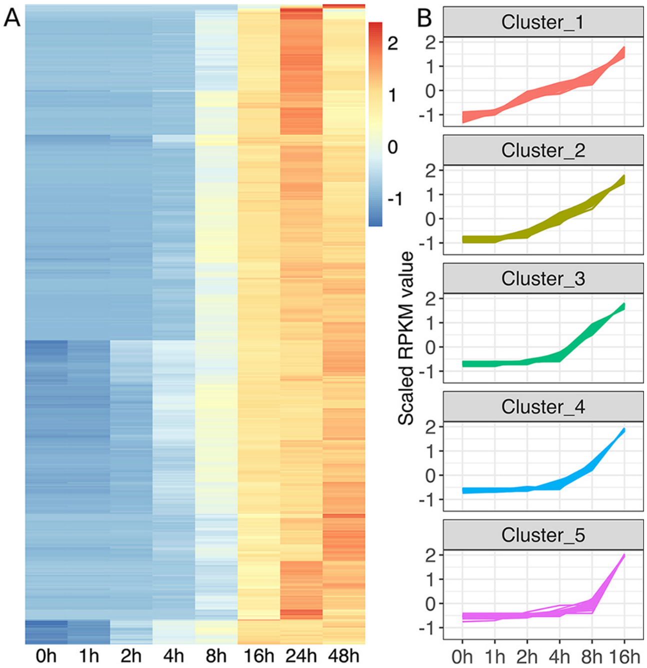

FIG 2 Expression of medusavirus genes at different time points during the course of medusavirus infection. (A) Heatmap of medusavirus gene expression profiles. Each column represents one time point; each row represents a viral gene; the color scale indicates Z-score scaled RPKM values. (B) Medusavirus temporal gene expression profiles in the five clusters. *x* axis, time points postinfection; *y* axis, Z-score scaled RPKM value for each gene. Each line represents a viral gene.

processing" category (e.g., putative VLTF-2 transcription factor, putative late transcription factor 3, and transcription elongation factor S-II), "virion structure" (e.g., major capsid protein and putative membrane protein), and "translation" (e.g., translation initiation factor eIF1 and a tRNA^{His} guanylyltransferase). Clusters 4 and 5 also contained genes under various functional categories, including those related to transcription, translation, and virion structure, but many genes in these clusters were functionally unannotated (Fig. 3A). Our data indicate relatively late transcription of the 80 genes that encode proteins that are known to be packaged in viral particles (Fig. 3B), and most of them (73 genes, 92%) were in the intermediate or late expression clusters (i.e., clusters 3 to 5) (2).

Subcellular location of viral gene products. A large majority of viral gene products were predicted to be transported to the nucleus (131 genes, 28.4%), cytoplasm (170 genes, 36.9%), mitochondrion (51 genes, 11.1%), or extracellular components (37 genes, 8.0%) (Fig. 4). We combined this subcellular localization information with previously identified clusters. The proportion of nucleus-localized proteins showed a clear descending trend in the order of expression clusters, with the highest proportion in cluster 1 (45.2%) and lower proportions (19.7% to 32.6%) in other clusters. The proportion of nucleus-localized proteins in the virion-packaged group (i.e., proteins that are known to be packaged inside the virion) was 21.3%. The proportion of cytoplasm-localized proteins increased from cluster 1 (28.6%) to cluster 3 (43.7%) and then decreased to cluster 5 (20.7%). The proportion of cytoplasm-localized proteins in the virion-packaged group was high (38.8%) and was the biggest category among the virion-packaged proteins. The proportions of mitochondrion- and extracellular-localized proteins increased in cluster 4 (extracellular, 9.5%; mitochondrion, 16.6%) and cluster 5 (extracellular, 13.8%; mitochondrion, 17.2%); however, their proportions in the virion-packaged group was low (5 genes, 6.3% of the total virion proteins). Few proteins were predicted to localize to the cell membrane (22 genes), endoplasmic reticulum (16 genes), peroxisome (12 genes), Golgi apparatus (6 genes), and lysosome/vacuole (4 genes).

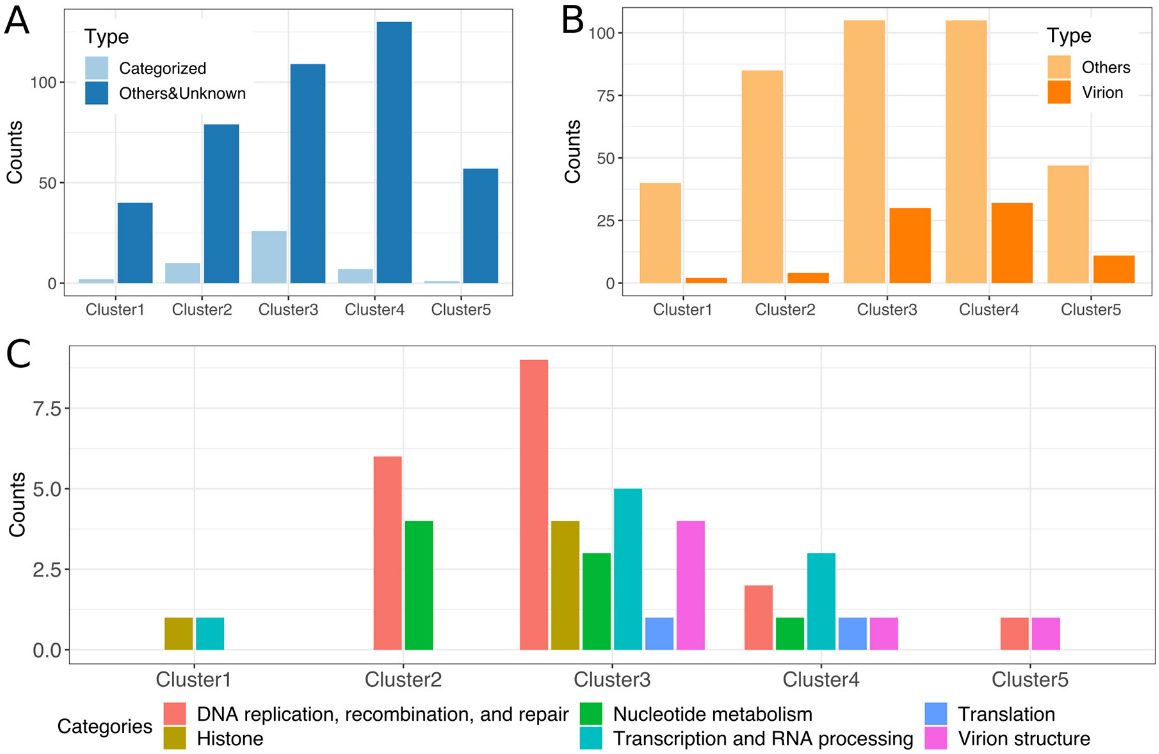

FIG 3 Distribution of genes with annotated functions among the five expression clusters. (A) Numbers of unknown and annotated genes in the expression clusters. Light blue indicates genes with categorized functions; dark blue indicates genes with uncategorized or unknown function. (B) Numbers of genes in the expression clusters. Light orange indicates genes that encode proteins not packaged in virions; dark orange indicates genes that encode proteins packaged in virions (2). (C) Numbers of functionally annotated genes in each of the expression clusters.

Their proportions and absolute numbers increased in the intermediate or late expression clusters (i.e., clusters 3 to 5) (Data set S6). In addition, half of the peroxisome-localized (6 out of 12 genes) and cell membrane-localized (11 out of 22 genes) proteins were in the virion-packaged groups.

Putative regulatory elements. To investigate the regulatory mechanisms of medusavirus gene expression, we analyzed the genomic localizations of the temporal gene expression clusters and associated gene functions. However, this analysis did not detect any definitive features related to the organization of genes in the genome and their temporal or functional groups (Fig. 5). *De novo* motif searches in the 5′ region upstream of the viral genes previously identified two motifs, a palindromic motif (GCCATRTGAVKTCATRTGGYSRSG, 53 occurrences) and a poly-A motif (VMAAMAAMARMAAMA, 251 occurrences) (19). We used the same method and found 3 additional putative promoter motifs, which were statistically significantly overrepresented in the analyzed sequences (E value, $<1 \times 10^{-5}$)—GCCRYCGYCGH (GC-rich motif, 134 occurrences), NRAAWAAA (AATAAA-like motif, 123 occurrences), and GTGTKKGTGGTGGTG (GT-rich motif, 37 occurrences) (Fig. 6; Fig. S2; Tables S1 to S3). In the following paragraphs, we investigate these five motifs with respect to their genomic locations and associations with the expression clusters.

The palindromic motif was preferentially found in the region approximately 40 to 70 bp upstream of the start codon. The poly-A motif was preferentially found in the region approximately 0 to 40 bp upstream of the start codon. The GC-rich motif had no obvious preferred position upstream of the start codon, but it often overlapped upstream genes. The AATAAA motif was preferentially found in the region approximately 0 to 60 bp upstream of the start codon, which is similar to the preferred positions of the poly-A motif. The GT-rich motif was preferentially found close to the start codon (Fig. 6).

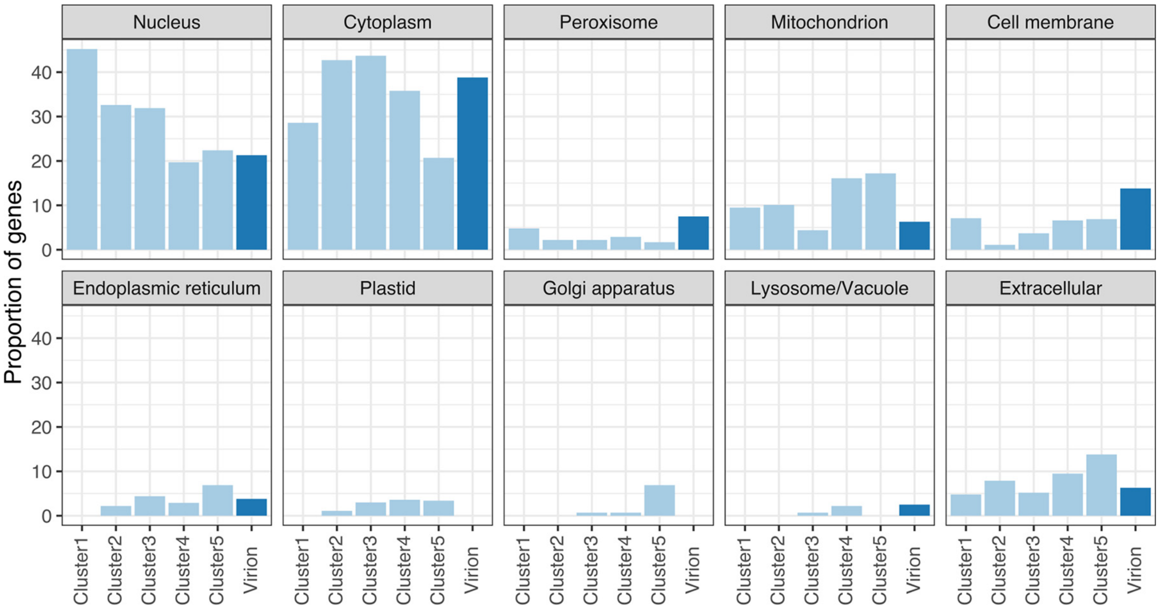

FIG 4 Predicted subcellular locations of the products of viral genes. The height of each bar indicates the proportion of genes in each cluster. Light blue indicates the proportion of genes in the expression clusters; dark blue indicates the proportion of genes among the viral genes whose products are known to be packaged inside the virion. Amoebas do not possess plastids, but the plastid predictions were retained (see Materials and Methods).

The palindromic motif was highly associated with the genes in cluster 1 (immediate early) (Fig. 7). Among the 53 viral genes with the upstream palindromic motif, 40 (75%) were in cluster 1, and they made up 95% of the genes in this cluster. The 13 other genes with the palindromic motif were distributed among other clusters, 8 in cluster 2 and 5 in clusters 3, 4, and 5 (Fig. 7). Furthermore, among the 56 genes detected at 1 hpi, 49 (88%) had the palindromic motif. We also found that 81.0% of the genes in cluster 1 and 69.7% of the genes in cluster 2 had the upstream poly-A motif, whereas they

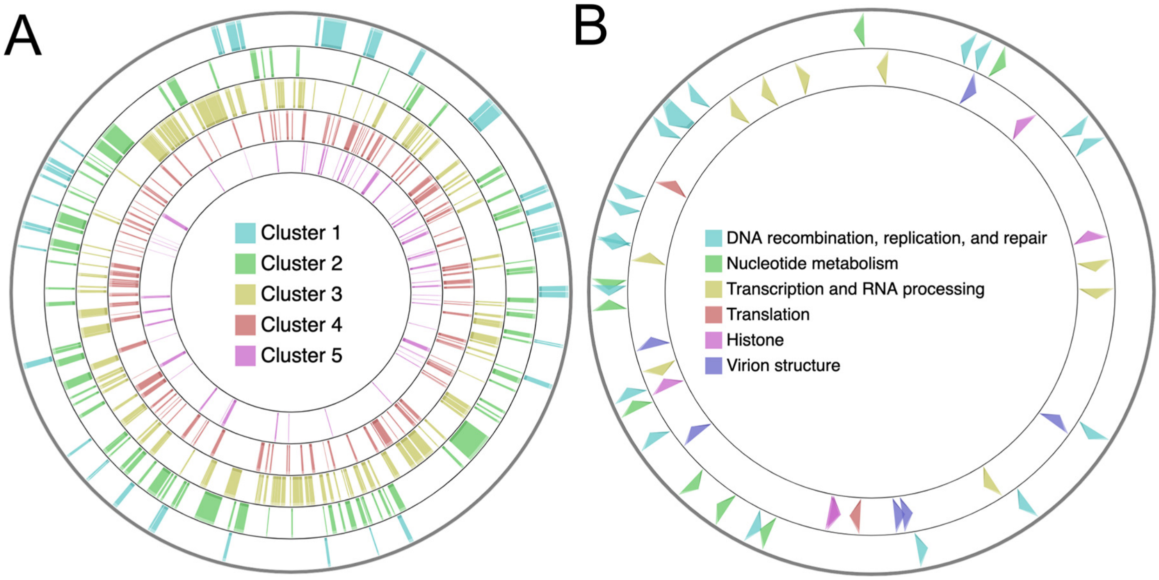

FIG 5 Organization of genes in the medusavirus genome. (A) Organization of the five expression clusters on the viral genome. (B) Organization of functional groups of genes on the viral genome. (Outside layer) Genes classified in the "DNA replication, recombination, and repair" and "nucleotide metabolism" categories. (Inside layer) Genes classified in the "transcription and RNA processing," "translation," "histones," and "virion structure" categories.

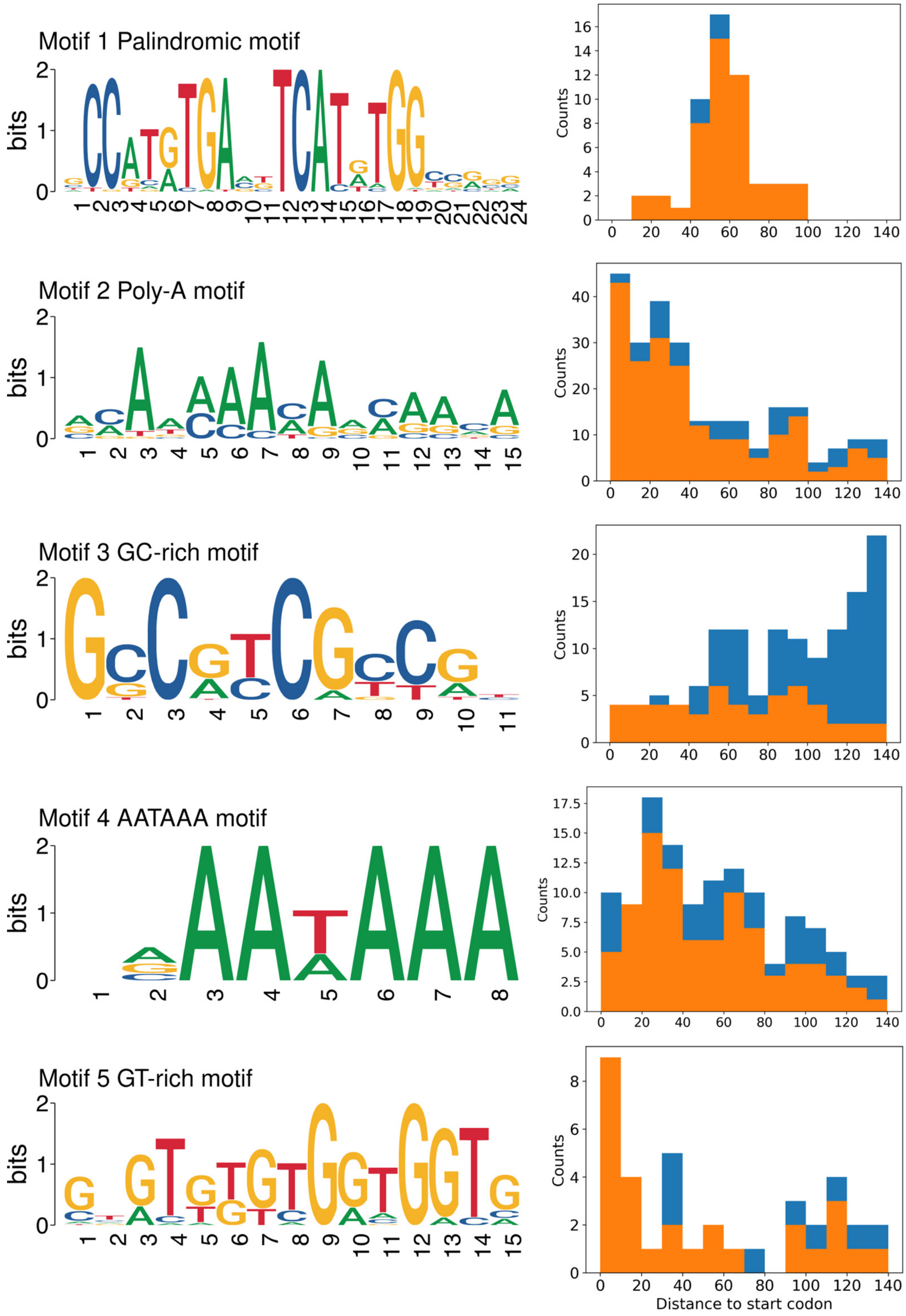

FIG 6 Sequence motifs enriched in the 5′ region upstream of the genes in the medusavirus genome and their distribution relative to the corresponding start codon. (Left panel) Motif name and its logo; (right panel) distance to the corresponding start codon; orange indicates motifs that did not overlap neighboring genes; blue indicates motifs that overlapped with neighboring genes.

made up only 40.2% to 55.2% of the genes in clusters 3 to 5 (Fig. 7). For the other three motifs, we found no specific association with gene clusters.

To investigate if these upstream motifs were promoter motifs, we scanned the medusavirus genome with these motifs. The palindromic and poly-A motifs were statistically

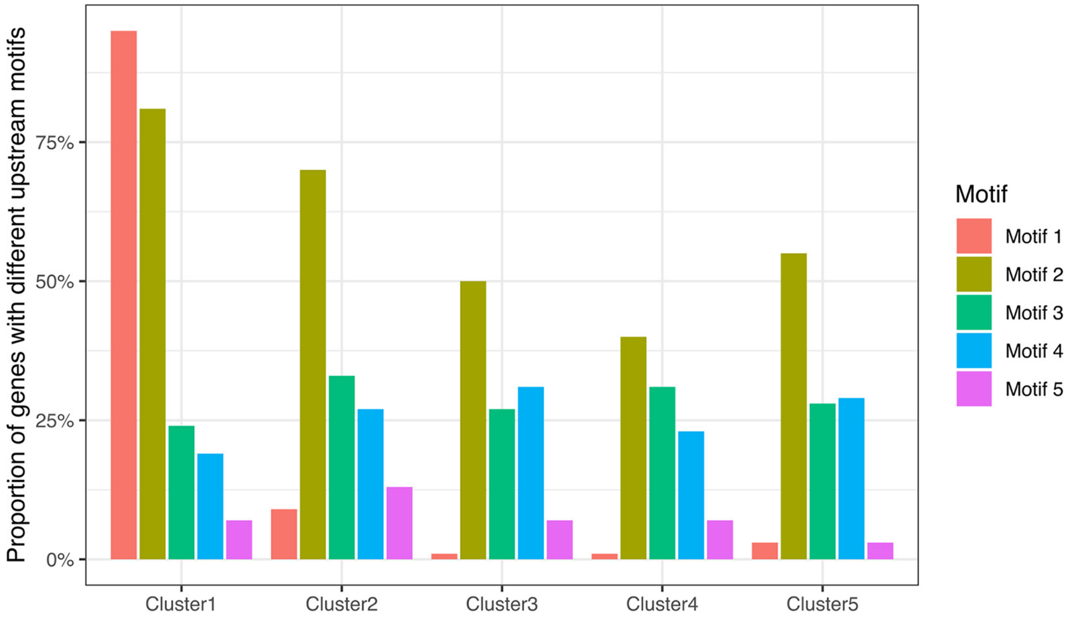

FIG 7 Proportion of genes with the different upstream motifs in each expression cluster.

significantly more abundant in intergenic regions than in coding sequences ($P < 2.2 \times 10^{-16}$; Table 1). Furthermore, these two motifs were more frequent in the upstream intergenic regions of genes than in the downstream intergenic regions ($P < 10^{-5}$; Table 2). The other three motifs showed no preference for either intergenic regions or coding sequences, leaving the putative promoter status of these motifs unclear; they may have other functional or structural roles. We also searched the 3′ downstream regions of the medusavirus genes for hairpin structures but failed to identify any. This is different from the presence of hairpin structures in *Acanthamoeba polyphaga mimivirus* (stem length, ≥13 bp; loop, ≤5 bp), *Megavirus chilensis* (stem length, ≥15 bp), *Pithovirus sibericum* (stem length, ≥10 bp; loop, ≤10 bp), and virophage sputnik, noumeavirus, and melbournevirus in the family *Marseilleviridae* (3, 5, 22, 29, 30).

The host nuclear transcriptional profile was greatly altered. The proportion of host mRNA reads and their expression levels assessed by RPKM did not show large changes until 8 hpi (Fig. 1 and 8A). After 8 hpi, the proportion of host reads decreased rapidly, and the proportion of viral reads increased. Our cluster analysis of the data set of 0 to 16 hpi showed that the transcription profile of the host *A. castellanii* genes changed greatly between 8 hpi and 16 hpi, with two expression clusters for the host genes (Fig. 9A). Of the 10,627 *A. castellanii* genes examined, 7,970 (75%) were in cluster 1. Their relative expression levels decreased across time, especially between 8 hpi and 16 hpi. The remaining 2,657 (25%) genes were in cluster 2, and their relative expression levels increased (at 16 hpi, mean $\log_2$ fold changes were −0.445 and 0.364 for cluster 1 and cluster 2, respectively).

TABLE 1 Distribution of upstream motifs in intergenic regions[a] of the medusavirus genome

	Data for:				
	Motif 1	Motif 2	Motif 3	Motif 4	Motif 5
Total count	48	152	807	19	253
Count in IR[b]	30	84	43	19	27
Background frequency[c]	0.105				
P value	4.60e-18	5.62e-42	1.00	0.132	0.495

[a] A binomial test was used to assess whether each motif was preferentially located in intergenic regions (IRs).
[b] Only motifs that did not overlap predicted genes were considered to be located in IRs.
[c] Background frequency was calculated by dividing the sum of the length of all IRs by the length of the whole genome.

TABLE 2 Preference of the palindromic and poly-A motifs for up- or downstream regions of medusavirus genes[a]

	Motif 1 palindromic		Motif 2 Poly-A	
	With motif	Without motif	With motif	Without motif
Divergent[b]	22	93	40	75
Convergent[c]	0	88	4	84
P value	2.76e-06		8.48e-08	

[a] Only motifs predicted to be located in intergenic regions were used to determine their preferred location. The *P* values were calculated by the Fisher exact test.

[b] Divergent cases were defined as motifs being located in the upstream regions of both neighbor genes.

[c] The convergent cases were defined as the motifs being located in the downstream region of both neighbor genes.

Based on the host gene clusters, we performed Gene Ontology GO and KEGG pathway functional enrichment analyses (Fig. 9B to E). Cluster 1 genes were enriched in cellular transportation-related GO terms, such as “localization,” “establishment of localization,” and “transport” (Fig. 9B). Cluster 2 genes were enriched in 60 GO terms that fell into two main categories (Fig. 9D). One category comprised terms related to “cellular protein metabolic process” and “proteolysis involved in cellular protein catabolic process,” and the other category comprised stress-related terms such as “DNA repair” (Fig. S3).

Cluster 1 host genes were also enriched in KEGG pathways, including “biosynthesis of secondary metabolites,” “glycerophospholipid metabolism,” “aminoacyl-tRNA biosynthesis,” “glycerolipid metabolism” and “ABC transporters” (Fig. 9C). Cluster 2 host genes were enriched with “Ribosome,” “Ubiquitin mediated proteolysis,” “Proteasome,” “Base excision repair,” “Homologous recombination,” and “Nucleotide excision repair” (Fig. 9D; Fig. S4). Clearly, the first three of these pathways correspond to the enriched GO terms in cluster 2 related to protein catabolic and metabolic processes, and the latter three pathways correspond to the enriched GO terms related to DNA repair and stress response.

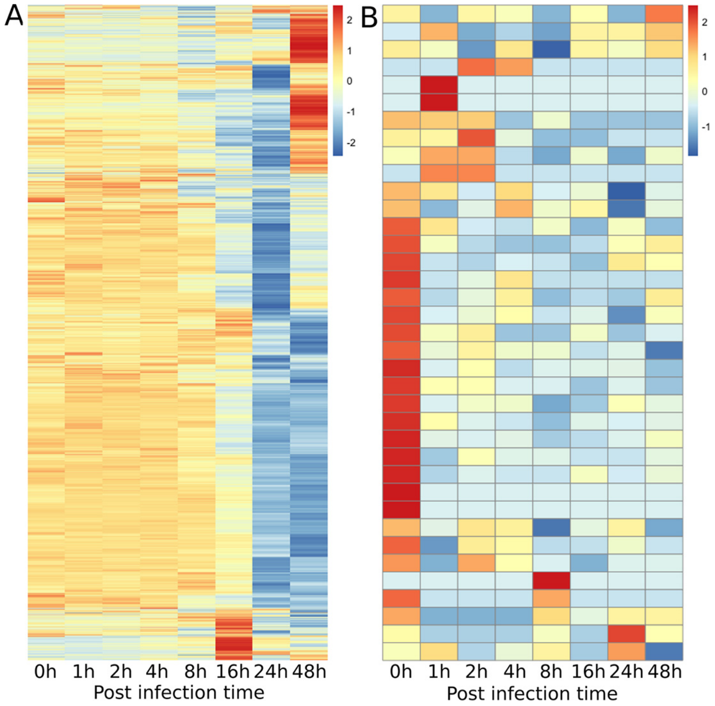

FIG 8 Transcription profiles of the *Acanthamoeba castellanii* (host) genes. (A) Host nuclear genes. (B) Host mitochondrial genes. *x* axis, time points of the infection cycle; *y* axis, different genes in the host genome and its mitochondrial genome. The color scale indicates Z-score scaled RPKM values.

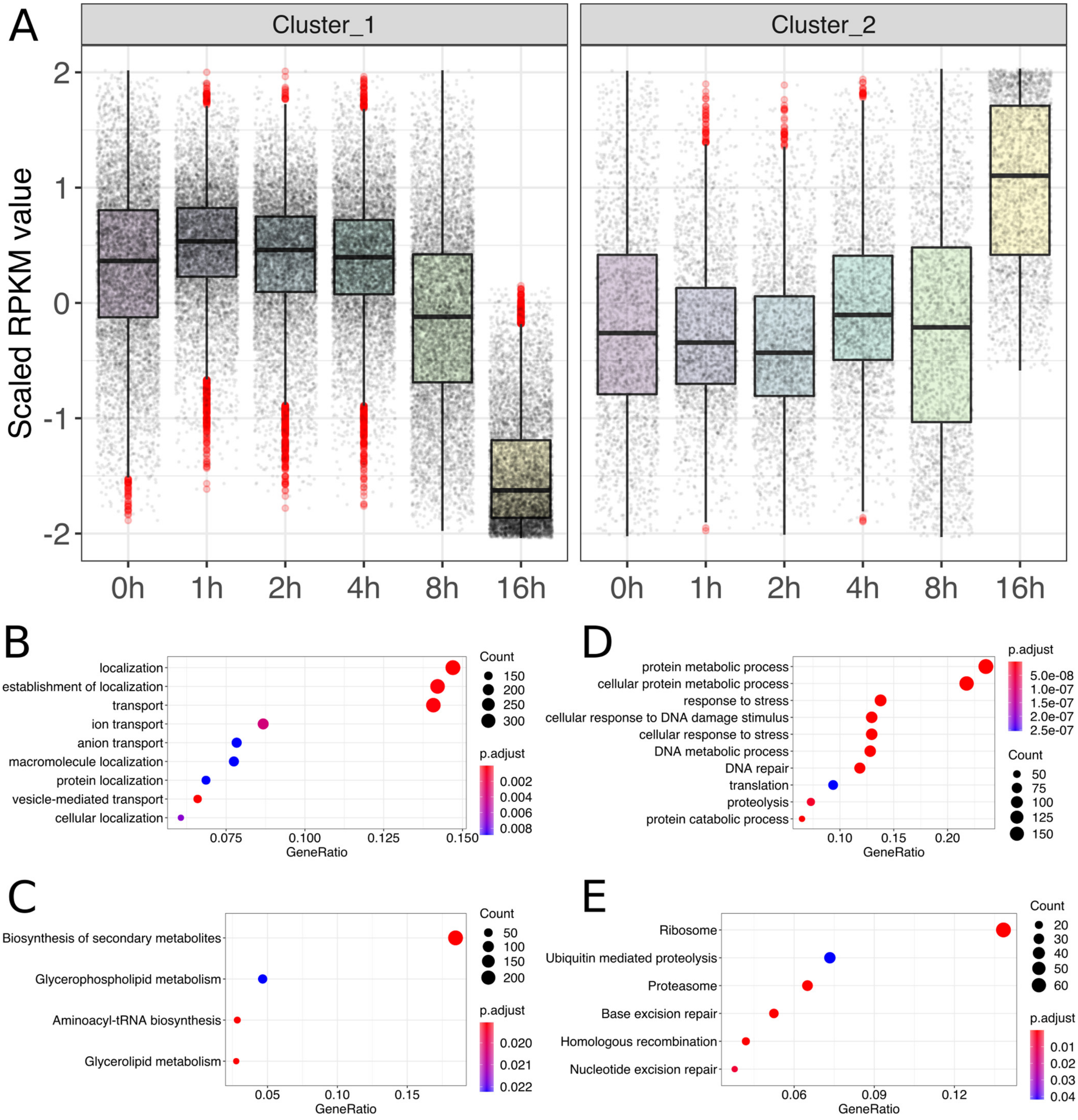

FIG 9 *Acanthamoeba castellanii* (host) nuclear gene expression clusters and their predicted functions. (A) Two expression clusters were identified for the host nuclear genes. (B) Enriched GO terms for the genes in cluster 1. (C) Enriched KEGG pathways for the genes in cluster 1. (D) Enriched GO terms for the genes in cluster 2. (E) Enriched KEGG pathways for the genes in cluster 2.

The host nuclear gene expression pattern changed after 16 hpi. Clusters identified in the first 16 hpi did not maintain their expression patterns after 16 hpi (Fig. S5). The expression levels of some genes annotated with the GO term "transport" were increased greatly at 48 hpi. In contrast, cluster 2 genes, which were activated at 16 hpi, were suppressed at 24 hpi and then recovered to some extent at 48 hpi. We found that some of the genes that were activated at 48 hpi were encystment-mediating genes, which included an encystation-mediating serine proteinase (EMSP), eight cysteine protease proteins, cyst-specific protein 21 (CSP21), and two cellulose synthases (31–35) (Fig. S6).

Mitochondrial expression was maintained during medusavirus infection. In our RNA-seq data, 37 of the 53 genes encoded in the *A. castellanii* mitochondrial genome had at least one read count during the course of the infection (Fig. 8B). The numbers of mitochondrial mRNA reads were low (4,365 reads at 0 hpi and maintained at <2,500 at the later time points). However, the numbers of mitochondrial reads were more stable than the numbers of reads that were mapped onto the nuclear genome, the latter of which decreased by about 30% during the course of the infection.

Mitochondrial genes with the highest read counts were related to rRNA genes (AccaoMp41, 23S-like rRNA and AccaoMp42, 16S-like rRNA), followed by energy metabolism [AccaoMp13, H (+)-transporting ATPase subunit 9]. These genes showed a sudden decrease at 1 hpi (~2-fold decrease in RPKM) but then maintained at later time points. The proportions of the other mapped mitochondrial mRNA reads were small. Among the 16 mitochondrial genes that were not detected in the RNA-seq data, 11 were tRNA genes, 3 were ribosomal protein genes, and 2 genes were of unknown function.

DISCUSSION

We performed RNA-seq to dissect the transcriptional program of medusavirus. Medusavirus has been reported to initiate its genome replication in the host nucleus and maintain the nuclear membrane intact during the infection cycle, with occasional induction of the encystment of the host amoeba *A. castellanii* at approximately 48 hpi (2). We found that transcription began in less than 1 h after the start of infection even though medusavirus has no RNA polymerase genes. Compared with other amoeba-infecting viruses, the speed of the medusavirus infection cycle was slow and apparently weak, because it took approximately 24 h for the virus to reach its expression peak, which was still only about 35% of the total mRNAs (at a multiplicity of infection [MOI] of 2.88). In contrast, mimivirus and marseillevirus genes occupy 80% of the total mRNA library at less than 6 hpi (at an MOI of 100 and 1,000 for marseillevirus and mimivirus, respectively) (25, 26). The slow and mild medusavirus infection may be explained by the different MOI used in the infection experiments, as a higher MOI has been reported to accelerate the infection course (36, 37). An additional explanation may be a slow start of medusavirus replication. Mimivirus carries its RNA polymerase in the viral particles to initiate its transcriptional process as soon as the viral particles open up in the cytoplasm (37). In contrast, medusavirus encodes no RNA polymerase and thus depends largely on host transcriptional machinery, which may account for its slow infection.

Clustering of medusavirus gene expression profiles showed clear temporal expression patterns akin to those observed for other giant viruses (25, 26, 38, 39). Of isolated viruses, medusaviruses are the only viruses that encode the linker histone in addition to the core histone domains (2, 19). Therefore, the functional relationship between the linker histone H1 and the core histones was a focus of this study. We found that linker histone H1, which is not packaged in viral particles (2), was transcribed immediately after the beginning of transcription. In contrast, the four core histones, which are carried in virions (2), started to be transcribed later. The different transcriptional profiles between the linker histone H1 and core histones suggest different functional roles between them. Histone H1 may cooperate with high-mobility group proteins in viral particles to regulate the accessibility of the viral genome for the subsequent transcription process (40, 41), or it may function to regulate the host chromatin. Regarding viral core histone proteins, the core histone proteins of marseilleviruses have been shown to bind DNA and form a structure resembling eukaryotic nucleosomes (42, 43). Marseillevirus histones have been also shown to localize the cytoplasmic viral factories and mature virions in the end of infection (43). Medusavirus core histones may function in a similar way for viral genome packaging as in marseilleviruses.

The predicted subcellular localization of viral gene products showed that cluster 1 had a higher proportion of nucleus-localized proteins than the other clusters. The predicted proportions of nucleus-localized proteins in the medusavirus and medusavirus stheno genomes were ranked 7th and 4th, respectively, among all known amoeba-infecting NCLDVs, indicating the importance of remodeling the nuclear environment immediately after medusavirus infection (Fig. S1). The remodeling probably contributes to subsequent viral gene transcription and DNA replication within the host nucleus. Putative cytoplasm-localized proteins were enriched in the virion-packaged group (31 genes, 38.8% of genes in virions), and almost half of the cell membrane and peroxisome-localized proteins were also packaged inside virions, suggesting that there may be interactions between virion-packaged proteins and the host cytoplasm and other subcellular membrane-bound compartments at an early phase of infection. The increasing expression of genes targeting the mitochondrion, endoplasmic reticulum, and Golgi apparatus suggests that medusavirus synthesizes these genes, probably to maintain or reprogram the functions of these organelles, after starting infection, rather than bringing them within the virion.

The enrichment of the palindromic motif in the upstream region of genes that were transcribed immediately after infection suggests that this motif may be the immediate early promoter of medusavirus genes. The poly-A motif that we detected in the upstream region of early expressed genes is reminiscent of the A/T-rich early promoter motifs found in other giant viruses in the phylum *Nucleocytoviricota* that have been proposed to have a common ancestral promoter motif, TATATAAAATTGA (44–47). The poly-A motif in the upstream regions of the medusavirus genes may have evolved from this common ancestral motif. Although the AATAAA motif was not preferentially located in the intergenic regions, it is similar to the 3′-end motif in the polyadenylation signal sequence in eukaryotes (48). The AATAAA motif also was detected in mimivirus, but it did not function as a polyadenylation signal (29). Regarding the 3′-end processing mechanism of giant viruses, A/T-rich hairpin structures have been identified after stop codons (3, 5, 22, 29, 30), and proteins that can recognize these structure have been studied (49). However, we did not find any A/T-rich hairpin structures in the 3′ downstream regions of medusavirus genes.

We identified two temporal clusters for host genes during viral replication. The fact that a majority (75%) of host genes showed decreases in their relative expression level at 16 hpi suggests that the host genes experienced global suppression. GO terms related to localization and transport were enriched in host cluster 1, suggesting that decreased transport activity occurred within the host cell during the course of the virus infection. In addition, the increased representation of the KEGG pathways "ribosome" and "proteosome" and the GO terms "cellular protein metabolic process" and "proteolysis involved in cellular protein catabolic process" suggests an increased activity of viral protein synthesis and degradation of host proteins, which needs experimental validation. We found enriched homologous recombination and DNA repair related GO and KEGG terms at 16 hpi (Fig. 9C and D; Fig. S4). Their increased representation, which has been reported to aid polyomavirus reproduction (simian virus 40 and JC polyomavirus) (50–52), may actively help medusavirus reproduction, although it may be due to a host response against virus infection. We also found an overrepresentation of encystment-related genes at 48 hpi (Fig. S6). As the culture may be a mixture of infected and uninfected amoeba cells at this time point with the initial MOI of 2.88, determining the cause of this overrepresentation (i.e., due to either healthy or infected cells) requires further investigation. Of note, encystment of both infected and healthy *Veramoeba vermiformis* cells has been observed upon infection by Faustovirus meriensis and has been suggested as an antiviral mechanism of the host trapping the viruses inside the cyst walls (53). A similar host strategy may be working for the *A. castellanii*-medusavirus infection system.

The expression pattern of the *A. castellanii* mitochondrial genes during the course of medusavirus infection was similar to the pattern found in marseillevirus (26). All tRNA-encoding genes had low expression levels, possibly because transcripts with a poly-A tail were used to build the RNA library and tRNA genes do not have poly-A tails. The genes with the highest expression levels included genes involved in energy metabolism and rRNA genes. Unlike the host nuclear genes, the transcriptional activity of these mitochondrial genes was maintained after 1 hpi, suggesting that host mitochondria may stably supply energy for viral replication.

In summary, our transcriptome data clearly delineated five temporal expression clusters for viral genes. Most of the immediate early genes (cluster 1) were of unknown function and had a palindromic promoter-like motif upstream of their start codons. Many of the immediate early gene products were predicted to localize in the host nucleus, suggesting that medusavirus modifies the host nuclear environment right after the start of infection by involving the action of dozens of viral genes. The genes that were expressed later (clusters 2 to 5) have various functions. The viral histone H1 gene is in the cluster 1, whereas the four core histone genes are in cluster 3, suggesting that they have distinct roles in viral replication. The transcriptional landscape of host nuclear genes was altered during infection, especially after 8 hpi. At 16 hpi, the host nuclear transcription showed a great alteration. Our transcriptome data will serve as a fundamental resource for further investigation of the infection strategies of medusaviruses, which are a group of amoeba-infecting giant viruses that have no close relatives among the diverse NCLDVs.

MATERIALS AND METHODS

Amoeba culture, virus infection, and sequencing. *Acanthamoeba castellanii* strain Neff (ATCC 30010) cells were purchased from the American Type Culture Collection (ATCC; Manassas, VA, USA). The *A. castellanii* cells were cultured in eight 75-cm^2 flasks with 25 ml of peptone-yeast-glucose (PYG) medium at 26°C for 1 h and then infected with purified medusavirus as previously described (2), at a multiplicity of infection (MOI) of 2.88. The titer of medusavirus was measured by 50% tissue culture infective dose (TCID$_{50}$) by inoculating fresh amoeba solution on a 96-well plate with a serially diluted virus solution (54). In a previous study, infection of medusavirus was associated with the appearance of the host amoebas forming cysts at an MOI of about 1 to 2 (2). With the aim of investigating this phenomenon, we performed our infection experiment with a similar MOI level. After addition of medusavirus to 7 of the 8 flasks (1 was the negative control), cells were harvested from each flask at 1, 2, 4, 8, 16, 24, and 48 hpi. Each cell pellet was washed with 1 ml of phosphate-buffered saline (PBS) by centrifugation (500 × *g*, 5 min at room temperature). Total RNA extraction was performed with an RNeasy minikit (Qiagen, Inc., Japan) and quality checked by agarose gel electrophoresis. The extracted RNA was sent to Macrogen Corp., Japan, for cDNA synthesis and library construction.

The cDNA synthesis and library construction were done using a TruSeq stranded mRNA low-throughput (LT) sample prep kit (Illumina, Inc., San Diego, CA, USA) following the manufacturer's protocol. Briefly, the poly-A-containing mRNAs were purified using poly-T oligonucleotide attached magnetic beads. Then, the mRNA was fragmented using divalent cations under elevated temperature. First-strand cDNA was obtained using reverse transcriptase and random primers. After second-strand synthesis, the cDNAs were adenylated at their 3′ ends, and adaptors were added. The DNA fragments were amplified by PCR and purified to create the final cDNA library. The RNA-seq was performed on a NovaSeq 6000 platform (Illumina, Inc.).

Read mapping and count normalization. The quality of the obtained reads was checked using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which showed that the overall quality was above the threshold (quality threshold, ≥20; no known adapters). Thus, we did no further trimming of the reads. The mRNA reads were mapped to a merged data set composed of the nuclear genome of *A. castellanii* (GCF_000313135.1_Acastellanii.strNEFF_v1), the medusavirus genome (GenBank accession number AP018495.1), and the mitochondria genome of *A. castellanii* (GenBank accession number NC_001637.1) using HISAT 2 (55) with a maximum intron size of 1,000 bp. The number of reads mapped on each gene was calculated using HTSeq in union mode (56). The transcriptional activity of genes was estimated by reads per kilobases of transcript per million mapped reads (RPKM) (Data sets S1 to 4).

Clustering. To discover the transcriptional patterns during medusavirus infection, we clustered the transcription profiles of viral and amoeba nuclear genes using the *k*-means method. We chose the library from 0 to 16 hpi to cluster viral genes, because a previous study indicated that replicated viral DNA was first observed in the cytoplasm at approximately 14 hpi and new virions were also observed to be released at the same time point (2), which indicated the termination of a cycle of infection at this time point. Genes with at least one mapped read across the 0 to 16 hpi libraries were included in the downstream analysis. To define the optimal number of clusters without prior biological information, we used the R packages NbClust and clusterCrit, which use different clustering indices to estimate the quality of clusters (57, 58). For virus genes, most indices gave 5 as the optimal number of clusters, and for amoeba nuclear genes, most indices gave 2 as the optimal number of clusters. Therefore, we performed the *k*-means clustering with *k* = 5 and 2 for viral and amoeba nuclear genes, respectively (Data set S5). We did not perform clustering of the expression of mitochondrial genes but analyzed expression of individual genes based on RPKM values.

Subcellular localization prediction of viral genes. Subcellular localization prediction of medusavirus genes was performed using DeepLoc 1.0 (59). We also predicted the subcellular localization tendency of other amoeba-infecting NCLDVs using the same method (Data sets S6 and S7; Fig. S1). A minor proportion of genes (0.0 to 5.0% for each virus) were predicted to target the plastid. Although amoebas do not possess plastids, we kept these predictions as they are, because even though these viruses were isolated using amoeba coculture, there remains a possibility that their natural hosts possess plastids.

Sequence motif analysis. MEME 5.1.1 was used for *de novo* motif prediction in the 5′ upstream sequence of medusavirus (60). We extracted 150-bp sequences upstream of the open reading frames. MEME was used in classic mode with motif width ranges of 8 to 10 bp, 6 to 15 bp, and 8 to 25 bp and "zero or more motifs in each intergenic region." We adopted the results with the motif width range of 8 to 25 bp because this was the only range in which the palindromic motif was detected (Fig. S2; Tables S1 to S3). We used the FIMO software tool (61) to scan the medusavirus genome for motifs that were predicted by MEME. The RNAMotif 3.1.1 algorithm (62) was used to find A/T-rich hairpin structures in the region downstream of each stop codon in medusavirus genes.

Functional enrichment analysis. Gene ontology (GO) and KEGG pathway enrichment analysis of the identified clusters of host genes were performed using the ClusterProfiler package in R (63).

Data availability. The sequencing data used in this study have been submitted to the DDBJ under the accession number DRA011802.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, XLSX file, 1.8 MB.

SUPPLEMENTAL FILE 2, PDF file, 0.04 MB.

SUPPLEMENTAL FILE 3, PDF file, 1 MB.

ACKNOWLEDGMENTS

We thank Hiroyuki Hikida for helpful discussion in revising our manuscript. Computational time was provided by the SuperComputer System, Institute for Chemical Research, Kyoto University. This work was supported by JSPS/KAKENHI (no. 18H02279 and 20H03078), The Kyoto University Foundation, and the International Collaborative Research Program of the Institute for Chemical Research, Kyoto University (no. 2018-32, 2019-34, and 2020-31).

We thank Margaret Biswas, Ph.D., from Edanz Group for editing a draft of the manuscript.

REFERENCES

1. Legendre M, Lartigue A, Bertaux L, Jeudy S, Bartoli J, Lescot M, Alempic JM, Ramus C, Bruley C, Labadie K, Shmakova L, Rivkina E, Couté Y, Abergel C, Claverie JM. 2015. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. Proc Natl Acad Sci U S A 112:E5327–E5335. https://doi.org/10.1073/pnas.1510795112.
2. Yoshikawa G, Blanc-Mathieu R, Song C, Kayama Y, Mochizuki T, Murata K, Ogata H, Takemura M. 2019. Medusavirus, a novel large DNA virus discovered from hot spring water. J Virol 93:e02130-18.
3. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie J-M. 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci U S A 108:17486–17491. https://doi.org/10.1073/pnas.1110889108.
4. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM. 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306:1344–1350. https://doi.org/10.1126/science.1101485.
5. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C, Couté Y, Rivkina E, Abergel C, Claverie JM. 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc Natl Acad Sci U S A 111:4274–4279. https://doi.org/10.1073/pnas.1320670111.
6. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C, Garin J, Claverie JM, Abergel C. 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341:281–286. https://doi.org/10.1126/science.1239181.
7. Koonin EV, Dolja VV, Krupovic M, Varsani A, Wolf YI, Yutin N, Zerbini FM, Kuhn JH. 2020. Global organization and proposed megataxonomy of the virus world. Microbiol Mol Biol Rev 84:1–33. https://doi.org/10.1128/MMBR.00061-19.
8. Guglielmini J, Woo AC, Krupovic M, Forterre P, Gaia M. 2019. Diversification of giant and large eukaryotic dsDNA viruses predated the origin of modern eukaryotes. Proc Natl Acad Sci U S A 116:19585–19592. https://doi.org/10.1073/pnas.1912006116.
9. Mihara T, Koyano H, Hingamp P, Grimsley N, Goto S, Ogata H. 2018. Taxon richness of "Megaviridae" exceeds those of bacteria and archaea in the ocean. Microbes Environ 33:162–171. https://doi.org/10.1264/jsme2.ME17203.
10. Yutin N, Wolf YI, Koonin EV. 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology 466–467:38–52. https://doi.org/10.1016/j.virol.2014.06.032.
11. Moreira D, López-García P. 2015. Evolution of viruses and cells: do we need a fourth domain of life to explain the origin of eukaryotes? Philos Trans R Soc Lond B Biol Sci 370:20140327. https://doi.org/10.1098/rstb.2014.0327.
12. Takemura M. 2020. Medusavirus Ancestor in a proto-eukaryotic cell: updating the hypothesis for the viral origin of the nucleus. Front Microbiol 11:1–8. https://doi.org/10.3389/fmicb.2020.571831.
13. Bell PJL. 2020. Evidence supporting a viral origin of the eukaryotic nucleus. Virus Res 289:198168. https://doi.org/10.1016/j.virusres.2020.198168.

14. Ogata H, Claverie JM. 2007. Unique genes in giant viruses: regular substitution pattern and anomalously short size. Genome Res 17:1353–1361. https://doi.org/10.1101/gr.6358607.
15. Legendre M, Fabre E, Poirot O, Jeudy S, Lartigue A, Alempic JM, Beucher L, Philippe N, Bertaux L, Christo-Foroux E, Labadie K, Couté Y, Abergel C, Claverie JM. 2018. Diversity and evolution of the emerging Pandoraviridae family. Nat Commun 9:2285. https://doi.org/10.1038/s41467-018-04698-4.
16. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh DA, Denef VJ, McMahon KD, Konstantinidis KT, Eloe-Fadrosh EA, Kyrpides NC, Woyke T. 2020. Giant virus diversity and host interactions through global metagenomics. Nature 578:432–436. https://doi.org/10.1038/s41586-020-1957-x.
17. Endo H, Blanc-Mathieu R, Li Y, Salazar G, Henry N, Labadie K, de Vargas C, Sullivan MB, Bowler C, Wincker P, Karp-Boss L, Sunagawa S, Ogata H. 2020. Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. Nat Ecol Evol 4:1639–1649. https://doi.org/10.1038/s41559-020-01288-w.
18. Li Y, Hingamp P, Watai H, Endo H, Yoshida T, Ogata H. 2018. Degenerate PCR primers to reveal the diversity of giant viruses in coastal waters. Viruses 10:1–16. https://doi.org/10.3390/v10090496.
19. Yoshida K, Zhang R, Garcia KG, Endo H, Gotoh Y, Hayashi T, Takemura M, Ogata H. 2021. Draft genome sequence of medusavirus stheno, isolated from the Tatakai River of Uji, Japan. Microbiol Resour Announc 10:e01323-20. https://doi.org/10.1128/MRA.01323-20.
20. Rolland C, Andreani J, Sahmi-Bounsiar D, Krupovic M, La Scola B, Levasseur A. 2021. Clandestinovirus: a giant virus with chromatin proteins and a potential to manipulate the cell cycle of its host Vermamoeba vermiformis. Front Microbiol 12:1–13. https://doi.org/10.3389/fmicb.2021.715608.
21. Yaakov LB, Mutsafi Y, Porat Z, Dadosh T, Minsky A. 2019. Kinetics of mimivirus infection stages quantified using image flow cytometry. Cytometry A 95:534–548. https://doi.org/10.1002/cyto.a.23770.
22. Fabre E, Jeudy S, Santini S, Legendre M, Trauchessec M, Couté Y, Claverie JM, Abergel C. 2017. Noumeavirus replication relies on a transient remote control of the host nucleus. Nat Commun 8:15087. https://doi.org/10.1038/ncomms15087.
23. Silva LKDS, Andrade A, Dornas FP, Rodrigues RAL, Arantes T, Kroon EG, Bonjardim CA, Abrahaõ JS. 2018. Cedratvirus getuliensis replication cycle: an in-depth morphological analysis. Sci Rep 8:4000–4011. https://doi.org/10.1038/s41598-018-22398-3.
24. Souza F, Rodrigues R, Reis E, Lima M, La Scola B, Abrahão J. 2019. In-depth analysis of the replication cycle of Orpheovirus. Virol J 16:158–111. https://doi.org/10.1186/s12985-019-1268-8.
25. Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D, Lartigue A, Lescot M, Bernadac A, Poulain J, Abergel C, Claverie JM. 2010. mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. Genome Res 20:664–674. https://doi.org/10.1101/gr.102582.109.
26. Rodrigues RAL, Louazani AC, Picorelli A, Oliveira GP, Lobo FP, Colson P, La Scola B, Abrahão JS. 2020. Analysis of a marseillevirus transcriptome reveals temporal gene expression profile and host transcriptional shift. Front Microbiol 11:651–617. https://doi.org/10.3389/fmicb.2020.00651.
27. Suhre K, Audic S, Claverie JM. 2005. Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes. Proc Natl Acad Sci U S A 102:14689–14693. https://doi.org/10.1073/pnas.0506465102.
28. Moniruzzaman M, Gann ER, Wilhelm SW. 2018. Infection by a giant virus (AaV) induces widespread physiological reprogramming in Aureococcus anophagefferens CCMP1984: a harmful bloom algae. Front Microbiol 9:1–16. https://doi.org/10.3389/fmicb.2018.00752.
29. Byrne D, Grzela R, Lartigue A, Audic S, Chenivesse S, Encinas S, Claverie JM, Abergel C. 2009. The polyadenylation site of mimivirus transcripts obeys a stringent "hairpin rule". Genome Res 19:1233–1242. https://doi.org/10.1101/gr.091561.109.
30. Claverie JM, Abergel C. 2009. Mimivirus and its virophage. Annu Rev Genet 43:49–66. https://doi.org/10.1146/annurev-genet-102108-134255.
31. Hirukawa Y, Nakato H, Izumi S, Tsuruhara T, Tomino S. 1998. Structure and expression of a cyst specific protein of Acanthamoeba castellanii. Biochim Biophys Acta Gene Struct Expr 1398:47–56. https://doi.org/10.1016/S0167-4781(98)00026-8.
32. Moon EK, Chung DI, Hong YC, Kong HH. 2008. Characterization of a serine proteinase mediating encystation of Acanthamoeba. Eukaryot Cell 7:1513–1517. https://doi.org/10.1128/EC.00068-08.
33. Moon EK, Hong Y, Chung DI, Kong HH. 2012. Cysteine protease involving in autophagosomal degradation of mitochondria during encystation of Acanthamoeba. Mol Biochem Parasitol 185:121–126. https://doi.org/10.1016/j.molbiopara.2012.07.008.
34. Moon EK, Hong Y, Chung DI, Goo YK, Kong HH. 2014. Down-regulation of cellulose synthase inhibits the formation of endocysts in Acanthamoeba. Korean J Parasitol 52:131–135. https://doi.org/10.3347/kjp.2014.52.2.131.
35. Chen L, Orfeo T, Gilmartin G, Bateman E. 2004. Mechanism of cyst specific protein 21 mRNA induction during Acanthamoeba differentiation. Biochim Biophys Acta 1691:23–31. https://doi.org/10.1016/j.bbamcr.2003.11.005.
36. Mallardo M, Leithe E, Schleich S, Roos N, Doglio L, Krijnse Locker J. 2002. Relationship between vaccinia virus intracellular cores, early mRNAs, and DNA replication sites. J Virol 76:5167–5183. https://doi.org/10.1128/jvi.76.10.5167-5183.2002.
37. Mutsafi Y, Zauberman N, Sabanay I, Minsky A. 2010. Vaccinia-like cytoplasmic replication of the giant mimivirus. Proc Natl Acad Sci U S A 107:5978–5982. https://doi.org/10.1073/pnas.0912737107.
38. Blanc G, Mozar M, Agarkova IV, Gurnon JR, Yanai-Balser G, Rowe JM, Xia Y, Riethoven JJ, Dunigan DD, Van Etten JL. 2014. Deep RNA sequencing reveals hidden features and dynamics of early gene transcription in Paramecium bursaria chlorella virus 1. PLoS One 9:e90989. https://doi.org/10.1371/journal.pone.0090989.
39. De Souza FG, Abrah S, Ara R, Rodrigues L. 2021. Comparative analysis of transcriptional regulation patterns : understanding the gene expression profile in Nucleocytoviricota. Pathogens 10:935. https://doi.org/10.3390/pathogens10080935.
40. Štros 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. https://doi.org/10.1007/s00018-007-7162-3.
41. Rajeswari MR, Jain A. 2002. High-mobility-group chromosomal proteins, HMGA1 as potential tumour markers. Curr Sci 82:838–844.
42. Valencia-Sánchez MI, Abini-Agbomson S, Wang M, Lee R, Vasilyev N, Zhang J, De Ioannes P, La Scola B, Talbert P, Henikoff S, Nudler E, Erives A, Armache K-J. 2021. The structure of a virus-encoded nucleosome. Nat Struct Mol Biol 28:413–417. https://doi.org/10.1038/s41594-021-00585-7.
43. Liu Y, Bisio H, Toner CM, Jeudy S, Philippe N, Zhou K, Bowerman S, White A, Edwards G, Abergel C, Luger K. 2021. Virus-encoded histone doublets are essential and form nucleosome-like structures. Cell 184:4237–4250.e19. https://doi.org/10.1016/j.cell.2021.06.032.
44. Oliveira GP, Andrade AC, dos SP, Rodrigues RAL, Arantes TS, Boratto PVM, Silva LKDS, Dornas FP, Trindade G, de S, Drumond BP, La Scola B, Kroon EG, Abrahão JS. 2017. Promoter motifs in NCLDVs: an evolutionary perspective. Viruses 9:1–20. https://doi.org/10.3390/v9010016.
45. Dizman YA, Demirbag Z, Ince IA, Nalcacioglu R. 2012. Transcriptomic analysis of Chilo iridescent virus immediate early promoter. Virus Res 167:353–357. https://doi.org/10.1016/j.virusres.2012.05.025.
46. Fitzgerald LA, Boucher PT, Yanai-Balser GM, Suhre K, Graves MV, Van Etten JL. 2008. Putative gene promoter sequences in the chlorella viruses. Virology 380:388–393. https://doi.org/10.1016/j.virol.2008.07.025.
47. Oliveira GP, Lima MT, Arantes TS, Assis FL, Rodrigues RAL, da Fonseca FG, Bonjardim CA, Kroon EG, Colson P, La Scola B, Abrahão JS. 2017. The investigation of promoter sequences of marseilleviruses highlights a remarkable abundance of the AAATATTT motif in intergenic regions. J Virol 91:1–10. https://doi.org/10.1128/JVI.01088-17.
48. López-Camarillo C, Orozco E, Marchat LA. 2005. Entamoeba histolytica: comparative genomics of the pre-mRNA 3′ end processing machinery. Exp Parasitol 110:184–190. https://doi.org/10.1016/j.exppara.2005.02.024.
49. Priet S, Lartigue A, Debart F, Claverie JM, Abergel C. 2015. MRNA maturation in giant viruses: variation on a theme. Nucleic Acids Res 43:3776–3788. https://doi.org/10.1093/nar/gkv224.
50. Hein J, Boichuk S, Wu J, Cheng Y, Freire R, Jat PS, Roberts TM, Gjoerup OV. 2009. Simian virus 40 large T antigen disrupts genome integrity and activates a DNA damage response via Bub1 binding. J Virol 83:117–127. https://doi.org/10.1128/JVI.01515-08.
51. Orba Y, Suzuki T, Makino Y, Kubota K, Tanaka S, Kimura T, Sawa H. 2010. Large T antigen promotes JC virus replication in G2-arrested cells by inducing ATM- and ATR-mediated G2 checkpoint signaling. J Biol Chem 285:1544–1554. https://doi.org/10.1074/jbc.M109.064311.
52. Weitzman MD, Fradet-Turcotte A. 2018. Virus DNA replication and the host DNA damage response. Annu Rev Virol 5:141–164. https://doi.org/10.1146/annurev-virology-092917-043534.
53. Borges I, Rodrigues RAL, Dornas FP, Almeida G, Aquino I, Bonjardim CA, Kroon EG, La Scola B, Abrahão JS. 2019. Trapping the enemy:

Vermamoeba vermiformis circumvents Faustovirus Mariensis dissemination by enclosing viral progeny inside cysts. J Virol 93:1–19. https://doi.org/10.1128/JVI.00312-19.

54. Hierholzer JC, Killington RA. 1996. Virus isolation and quantitation, p 25–46. *In* Mahy BWJ, Kangro HO (ed), Virology methods manual. Academic Press, London.
55. Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317.
56. Anders S, Pyl PT, Huber W. 2015. HTSeq: a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. https://doi.org/10.1093/bioinformatics/btu638.
57. Charrad M, Ghazzali N, Boiteau V, Niknafs A. 2014. Nbclust: an R package for determining the relevant number of clusters in a data set. J Stat Softw 61:1–36. https://www.jstatsoft.org/article/view/v061i06.
58. Desgraupes B. 2018. clusterCrit: Clustering Indices. R package version 1.2.8. https://CRAN.R-project.org/package=clusterCrit.
59. Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017. DeepLoc: prediction of protein subcellular localization using deep learning. Bioinformatics 33:3387–3395. https://doi.org/10.1093/bioinformatics/btx431.
60. Bailey Timothy LEC. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Second Int Conf Intell Syst Mol Biol 2:28–36.
61. Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. Bioinformatics 27:1017–1018. https://doi.org/10.1093/bioinformatics/btr064.
62. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res 29:4724–4735. https://doi.org/10.1093/nar/29.22.4724.
63. Yu G, Wang LG, Han Y, He QY. 2012. ClusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16:284–287. https://doi.org/10.1089/omi.2011.0118.