Comprehensive identification of protein orthologs facilitated an understanding of the phylogenomics, protein conservation and phosphorylation of *Ascoviridae* species

Yanhua Shi
Weifang Medical University - Fuyanshan Campus: Weifang Medical University
Weiping Lin
Weifang Medical University
Guohui Wang
Weifang Medical University
Punan Zhao
Weifang Medical University - Fuyanshan Campus: Weifang Medical University
Guo-hua Huang
Hunan Agricultural University
Dianhai Hou (houdh@wfmc.edu.cn)
Weifang Medical University https://orcid.org/0000-0003-0282-0240

Research Article

Keywords: Ascoviridae, phosphorylation, phylogenomic inference

Posted Date: November 2nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1023975/v1

Abstract

Analysis of orthology is important for understanding protein conservation, function and phylogenomics. This study performed a comprehensive identification of *Ascoviridae* orthology based on identification of 366 ascoviridae protein homologue groups and phylogenetic analysis of 34 non-single copy proteins. Our fondings revealed 90 newly annotated proteins, five new identified *Ascoviridae* core proteins and 14 *Ascovirus* core proteins. Moreover, a phylogenomic tree of 11 ascoviridae species was inferred based on the concatenation of 35 of 45 *Ascoviridae* ortholog groups. In combination with phosphoproteomic results and conservation estimations, 30 conserved phosphorylation sites on 17 phosphoproteins were identified from a total of 176 phosphosites on 57 phosphoproteins from *Heliothis virescens ascovirus 3h* (HvAV-3h), supplying potential research targets for exploration of the detailed role of these protein in the regulation of viral infection mechanisms. This study would facilitates further *Ascoviridae* genome annotation and comparison and other functional genomic investigations.

Introduction

The family *Ascoviridae* is a group of insect-specific large DNA viruses that have been isolated from members of the orders Lepidoptera, Diptera and Neoptera, and is transmitted by endoparasitic wasps. *Ascoviridae* is divided into two genera, *Ascovirus* and *Toursvirus* [2]. The *Ascoviridae* genome is a double stranded circular DNA genome with a length of about 110–200 kb encoding 119–194 genes [2] (Table S1). *Ascoviridae* virions initiate replication in the nuclei of infected cells, which results in the enlargement of the nucleus rupture and then cleavage of the cell into a cluster of virion-containing vesicles. Infection by this virus results in decreased food intake and retarded growth of infected larvae [6, 9]. Thus *Ascoviridae* species represent as a new candidate for the biological control of pest insect species [9].

To date, only a few of investigations have focused on the *Ascoviridae* infection mechanism. The viral caspase has been demonstrated to function in inducing cell apoptosis [5]. The viral caspase [5, 55] and viral lipase [44] play important roles in cell cleavage and thus facilitate the formation of viral vesicles and function in viral reproduction. Transcriptomic analyses have demonstrated that *Ascoviridae* infection affects host molting and inhibits host innate immunity *in vivo* [54–56], and it also stimulates host larval lipid and protein syntheses, including mitochondrial ATP6, ATP synthase, and NADH dehydrogenase subunits, and inhibits carbohydrate synthesis [31, 56], which promotes the formation of viral vesicles. It has been also demonstrated that viral RNase III functions in suppressing host cell RNAi (RNA interference) defense responses [25] and the virion-associated P64 [10, 46] is a DNA binding protein that is incorporated into the viral DNA core during virion assembly [47]. Nevertheless, functional analysis of *Ascoviridae* genomics has been hindered due to a lack of molecular operation platforms, such as cloning viral genomes into infectious bacterial artificial chromosomes [58].

Comprehensive identification of the conservation of genes and proteins promotes the investigation and understanding of their biological functions, particularly those involved in DNA replication, RNA transcription, virion assembly and regulation of host metabolism. This also hold true for the protein

interaction network for virus reproduction [18, 52], and this fact facilitates the construction of artificial chromosomes and further medication targeting viral vectors [4, 21, 37, 43, 49]. To date, 11 *Ascoviridae* genomes have been fully sequenced, including nine ascoviruses and two toursviruses (Table S1). A total of 37 core genes in all of these sequenced *Ascoviridae* genomes have been identified [50]. Nevertheless, the identification of homologs and orthologs of core genes is often obscured by the presence of paralogs [3, 16, 20, 30] (for non-single copy genes) and missing genes during the annotation of sequenced genomes [27] (see Glossary relevant to gene homology relationships such orthology and homology in [16, 17]). Previously, Garavaglia *et al*. optimized several BLAST approaches for the identification of baculovirus orthology, and supplied a comprehensive view of protein conservation, leading to increased identification of novel core genes [18]. Furthermore, for comparative genomic and phylogenomic analyses, identification of orthologs is a key step that can be achieved through differentiating gene homology relationships and selection of orthologs based on constructed gene trees with different branch lengths [3, 17, 34, 48].

To facilitate understanding protein conservation and species phylogenomics, in this study, we performed comprehensive identification of *Ascoviridae* orthology and phylogenomic inference, and updated the global view of *Ascoviridae* protein conservation. Further identification of phosphorylation sites on HvAV-3h proteins with orthologs were found to be virion-associated, and can be used to promote functional investigation of orthologs in the future.

Materials And Methods

Insects, viral infection and production of viral vesicles

Laboratory stocks of *H. armigera* were reared on an artificial diet and the third-instar larvae were infected with a laboratory stock HvAV-3h strain [9]. Virus-containing hemolymph was collected from infected larvae 10 days after infection [10], diluted with four volume of 0.1 × TE (TE, 10 mM Tris, 1.0 mM EDTA, pH 7.4) containing protease and phosphatase inhibitor cocktail (Sigma, Germany), and centrifuged at 3000 g for 10 min at 4°C to collect virion-containing vesicles [15].

Genome and protein databases

Individual genomes from nine ascoviruses and two toursvirus were collected from the NCBI GenBank database (Table S1). The non-redundant annotated protein dataset of each ascoviridae genome was also acquired from NCBI. Proteins were repredicted for all genome sequences using getorf (http://emboss.bioinformatics.nl/cgi-bin/emboss/getorf) and proteins ≥ 50 aa were kept. Thus, each *Ascoviridea* species had three different datasets: a) the Individual Genome DataSet (IGD), containing the corresponding genome sequence, b) the Individual Proteome DataSet (IPD), containing the corresponding annotated proteins, and c) the Individual Repredicted Proteome DataSet (IRPD), containing the corresponding repredicted proteins.

Identifification of protein homologue groups

An overview of the identification of *Ascoviridae* homologue groups is demonstrated in a schematic diagram (Fig. S1). In the first round, OrthoFinder (v2.5.1) [13] was used to all-versus-all BLAST search among all IPDs for the identification of homologue groups, using the settings -S blast -M msa -os. To decrease the disability of orthology prediction tools for identification of all orthologs [18], the OrthoFinder identified homologue groups were used as a Primary Homologous Group Database (PHGD), and were then subjected to an all-versus-all search using OrthoFinder against the IRPD in the second round to screen candidate homologues that were not annotated in the IPD. The candidate homologues were then manually checked using BLASP [32] in NCBI to search homologues with an inclusion threshold of 1e-5. This resulted in a Revised Homologous Group Database (RHGD). Further, local TBLASTN was performed by searching against the corresponding viral genome sequences using BioEdit 7.5.2 [1] with an expect threshold of 1e-5 to identify candidate homologues, and all the selected candidate homologues were manually checked using BLASP [32] as described above. Thus, a Final Homologous Group Database (FHGD) was obtained, which contained all the sequences of individual homologue groups and statistical matrix of the presence of homologous proteins in each virus was used for further analysis below.

Identification of orthologs in non-single-copy protein homologue groups

For non-single-copy protein homologue groups, orthologs were determined by screening single representative proteins or paralogs for each genome [34]. Protein homologue groups were then subjected to MAFFT (v7.310-1) alignment using the G-INS-i progressive method [35], and maximum likelihood (ML) trees of proteins were inferred using IQ-TREE (v1.6.2) [28, 33] with the best model being screened using the standard algorithm while 10,000 ultrafast bootstrap replicates were used for BROs (baculovirus repeat ORFs).

Phylogenomic analysis of 11 ascoviridae species

Ortholog groups present in all 11 *Ascoviridae* genomes were subjected to MAFFT alignment using the G-INS-i Progressive method [35]. The alignment proteins from each genome were filtered using the Gblocks (v0.91b) [45] with the settings -t=p -b1=6 -b2 =7 -b3 =10 -b4 =2 -b5 = h -e =-gb. These were then concatenated in a stationary order (HvAV-3h ORF order) using TBtools (v1.046) [8] and PhyloSuite (v1.2.2) [57]. Furthermore, a full tree search for protein models and phylogenetic inference was performed using IQ-TREE [28, 33]. FigTree (http://tree.bio.ed.ac.uk/software/figtree/) was used for viewing the acquired ML tree files and further construction of a phylogenetic tree.

In-solution digestion, IMAC-enrichment of phosphopeptides and mass spectrometry analysis

The pellets of vesicles above were resuspended in 1% Triton X-100 in 0.1×TE and sonicated [15]. After centrifugation at 10,000 g for 15 min at 4°C, the supernatant was collected and precipitated with three volumes of 50% acetone/50% methanol/0.1% acetic acid. The precipitated proteins were resuspended in lysis buffer (8 M urea, 100 mM TEAB [pH 8.0]), reduced, alkylated, and subjected to in-solution trypsin

digestion as described previously [22, 53]. Peptides digested from 2 mg of proteins were then dissolved in 1% acetic acid and loaded onto immobilized metal affinity chromatography (IMAC)-Fe^{3+} resin to enrich phosphopeptides [22, 53]. Theses peptides were then subjected to LC-MS/MS analysis using an Orbitrap Exploris™ 480 mass spectrometer (Thermo Scientific, USA) coupled with an Easy-nLC 1200 system. Peptides were loaded using an auto-sampler and separated in a C18 analytical column (75μm × 25cm, C18, 1.9μm, 100Å). Mobile phase A (0.1% formic acid) and mobile phase B (80% ACN, 0.1% formic acid) were used to establish the separation gradient. A constant flow rate was set at 300 nL/min. For the DDA mode analysis, each scan cycle is consisted of one full-scan mass spectrum (R = 60 K, AGC = 300%, max IT = 20 ms, scan range = 350–1500 m/z) followed by 20 MS/MS events (R = 15 K, AGC = 100%, max IT = auto, cycle time = 2 s). The HCD collision energy was set to 30. The isolation window for precursor selection was set to 1.6 Da. Former target ion exclusion was set for 35 s.

Database search for protein identification and phosphorylation site identification

The acquired raw MS/MS data were formatted in Mascot generic format (.mgf) using MSconvert from ProteoWizard (v3.0) [7, 29] and then the processed MS/MS spectra were searched using the Mascot search engine (Version 2.3, Matrix Science, London, UK) [39] against a database including 21038 host proteins from the annotated *Helicoverpa armigera* genome [38], 185 annotated HvAV-3h proteins [23], and seven newly identified HvAV-3h proteins. Trypsin was specified as the digestion enzyme, allowing up to two missed cleavage sites. The carbamidomethylation of cysteine was set as a fixed modification, while protein N-terminal acetylation, oxidation of methionine, deamidation of asparagine and glutamine, pyro-glutamate formation from glutamine, and phosphorylation of tyrosine, serine and threonine were considered as variable modifications. The mass tolerance was set to 20 ppm and 0.1 Da for peptide precursors and MS/MS fragment ions, respectively. Data were filtered using a 1% false discovery rate (FDR) on peptide spectrum matches (PSMs) using automatic decoy searching. Peptides were considered if the matching score was higher than the score threshold. Phosphorylation Site Localization was manually checked and labeled using pLabel (version 2.4) [51].

Bioinformatic analysis

The conserved motifs in proteins were searched using the MEME suite (Bailey and Elkan, 1994). The minimum motif was set to 10 and the number of motifs that MEME was allowed to find was set to 5, while the other parameters were set to defalt. Putative signal peptide (SP) and transmembrane (TM) domains were predicted using SignaIP 5.0 (Almagro Armenteros et al., 2019) and TMHMM 2.0 (Krogh et al., 2001) respectively. And the PHOSIDA [19] was used for searching phosphopeptide sequence matches against annotated motifs targeted by eukaryotic kinases.

Results

Comprehensive identification of Ascoviridae homologue groups

A total of 366 homologue groups were identified, covering 1934 proteins from 11 genomes (Table S2, and Table S3 with accession numbers of previous annotated proteins and locations of newly annotated proteins). A total of 90 proteins with distribution in 63 homologue groups were newly identified in 10 genomes (Table S2). This facilitated identification of five new protein homologues that were present in all *Ascoviridae* genomes, including HvAV-3h ORF46, HvAV-3h ORF46a (DjTV-2a ORF125), HvAV-3h ORF56, Myristylated membrane protein-like protein (HvAV-3h ORF62) and Transcription elongation factor S-II (HvAV-3h ORF112), and the identification of 14 new protein homologues that were present in all *Ascovirus*,genomes, including HvAV-3h ORF12, HvAV-3h ORF29, HvAV-3h ORF44, GIY-YIG-like endonuclease (HvAV-3h ORF49), HvAV-3h ORF68, HvAV-3h ORF69, Phenylalanyl-tRNA synthetase subunit (HvAV-3h ORF72), Nicotinate-nucleotide pyrophosphorylase (HvAV-3h ORF100), HvAV-3h ORF115, Hemolysin-like protein (HvAV-3h ORF137), HvAV-3h ORF140, HvAV-3h ORF141, HvAV-3h ORF142 and DEAD-like helicase (HvAV-3h ORF147). Furthermore, 46 homologue groups were found to be present in all ascoviridae genomes, among which, 33 were single-copy homologue groups. Among all the 336 homologue groups, including 106 single-copy species-specific homologue groups, 260 were identified with a taxon occupancy of ≥ 2 species, and most of them (226/260) were single-copy homologue groups.

Comprehensive identification of Ascoviridae orthologs

For all 34 non-single copy homologue groups, the groups were manually divided into five categories (A–E). For the 15 homologue groups of category A in which only one genome contained two copies of the homologue (Table S4), manual checking of the MAFFT alignment files of homologue groups was performed and it was found that the two homologues of Major capsid protein (HvAV-3h ORF53) in the TnAV-6a (*Trichoplusia ni ascovirus 6a*) genome were completely the same, facilitating the clarification of a representative protein ortholog in the TnAV-6a genome and the selection of the orthologous group of the Major capsid protein (HvAV-3h ORF53) for the construction of a species tree. A representative protein ortholog for each of the remaining homologue groups in category A was selected (indicated in red in Table S2) based on a phylogenetic analysis [3]. Taking the HvAV-3h ORF46a as an example, for the paralogs of DpTV (*Diadromus pulchellus toursvirus*) ORF56 and DpTV ORF57, DpTV ORF57 was not orthologous with DjTV (*Dasineura jujubifolia toursvirus*) ORF125 (i.e., not on a terminal branch), while the ortholog in DjTV-2a on the same terminal branch with DpTV ORF57 was missing. DpTV ORF56 was observed on the same terminal branch with DjTV ORF125, and both were considered to be orthologous with each other and single-copy homologues in other *Ascoviridae* species. Thus, this facilitated the selection of orthologous groups from HvAV-3h ORF46a/DjTV-2a ORF125 (Fig. 1A) for the construction of a species tree. Nevertheless, the representative protein ortholog and its paralogs were the truncated homologues of RNaseIII (HvAV-3h ORF24), HvAV-3h ORF48, HvAV-3h ORF122, Serine/threonine protein kinase (HvAV-3h ORF124), and ATPase (HvAV-3h ORF125) in the SfAV-1a genome, and Myristylated

membrane protein-like protein (HvAV-3h ORF62), Serine/threonine protein kinase (HvAV-3h ORF75) and RNA polymerase subunit (HvAV-3h ORF143) in the TnAV-6a genome, were not selected for construction of a species tree.

For each of the seven homologue groups of category B (Table S4) with species occupancy ≥ 4 and with homologue number ≥ 2 in ≥2 genomes, different orthologous groups could be classified from the homologous groups of category B due to the presence of orthologous pairs. For the homologous group of P64 (HvAV-3h ORF58), both the orthologous pair DjTV ORF98 and DpTV ORF8 and the pair DjTV ORF95 and DpTV ORF80 could be combined with the orthologs of all other *Ascoviridae* species, and thus resulted in orthologous groups 1 and 2 respectively. It should be noted that DpTV ORF80 was selected as a representative protein ortholog for the paralogous pair of DpTV ORF80 and DpTV ORF115 (Fig. 1B). The orthologous group with species occupancy equal to that of these homologue groups could be chosen for the homologous groups of Virion S/T kinase cofactor (HvAV-3h ORF58) (Fig. 1B), Putative S1/P1 nuclease (HvAV-3h ORF88/ORF139) (Fig.S2A), HvAV-3h ORF31/ORF34 (Fig.S2B), HvAV-3g (*Heliothis virescens ascovirus 3g*) ORF10 (Fig.S2C), and TnAV-6a ORF71 (Fig.S2D), and even for the homologous repeat (HR) (Fig.S2E) and inhibitor of apoptosis (IAP)–like protein (Fig. 2) homologues with multiple copies. Two orthologous groups (Fig. 2A) could be selected from the IAP homologue group among which all the members shared the Ringer domains (Fig. 2B). Nevertheless, orthologous groups for P64 (HvAV-3h ORF58) and IAP were not selected for further construction of a species tree.

For each of the five homologue groups of category C (Table S4) with the number of ≥ 4 homologues present only in two genomes, the homologue group was further divided into two subgroups and orthologous pairs were determined from these homologues (Fig. S3) including the homologue group of TnAV-6a ORF36 (Fig. S3A), TnAV-6a ORF27 (Fig. S3B), TnAV-6a ORF66 (Fig. S3C), TnAV-6a ORF70 (Fig. S3D) and TnAV-6a ORF73 (Fig. S3E). An ortholog was not further identified through the construction of a species tree due to there being fewer than four homologues for the three homologue groups of DjTV-2a ORF134, TnAV-6a ORF48 and TnAV-6a ORF63 in category D (Table S4). Moreover, based on sequence alignment, the two homologues of TnAV-6b ORF55 in TnAV-6a and the two homologues of TnAV-6a ORF63 in TnAV-6b seems to be truncated forms. In addition, we failed to screen out an orthologous group with species occupancy equal to that of the homologue groups for the four homologue groups of category E (Table S4), including that of HvAV-3h ORF95 (Fig. 3A), RING-finger-containing E3 ubiquitin ligase (HvAV-3h ORF167/ORF175) (Fig. 3B), ARO (ascovirus repeat ORF) (Fig. S4), and BRO (Fig. S5).

Phylogenomic inference of Ascoviridae species

A total of 35 selected orthologs were used for the construction of our species tree, covering 33 single-copy homologue groups and two orthologous groups selected from the non-single-copy homologue groups of Major capsid protein (HvAV-3h ORF53) and the HvAV-3h ORF46a/DjTV-2a ORF that were present among all the 11 genomes. Phylogenomic inference with concatenation of these 35 orthologs was performed (Fig. 4A), using the bootstrap values for the branches of 100 with the exception of only one, which was 98. The phylogenomic inference in this study was consistent with previous tree inference

using DNA polymerase [24, 50] and major capsid protein[24], despite the minor adjustment of phylogenetic positions between all of the species from HvAV-3i (*Heliothis virescens ascovirus 3e*) to HvAV-3j (*Heliothis virescens ascovirus 3j*).

Identification and classification of conserved proteins in HvAV-3h

Based on identification and phylogenetic inference of homologous groups, orthologs of 45 HvAV-3h ORFs were found among all other eight *Ascovirus* and two *Toursvirus* genomes (Table 1), and among the 45 HvAV-3h ORFs conserved in all *Ascoviridae* species, two copies of IAP-like proteins (HvAV-3h ORF25 and ORF47) were orthologous to single-copy ones (DjTV-2a ORF108 and DpTV-1a ORF7) in *Toursvirus* species (Fig. 2). These proteins were involved in DNA/RNA replication/transcription/metabolism, viral packaging and assembly, signaling, and sugar and lipid metabolism (Table 1). The HvAV-3i orthologs of all the 45 HvAV-3h ORFs were found to be associated with HvAV-3i virions, with the exception of ORF118 and RNA polymerase subunit (ORF143) (Table 1). In addition, orthologs of 37 HvAV-3h ORFs were also found among all other eight *Ascovirus* species. It should be noted that GIY-YIG-like endonuclease (HvAV-3h ORF49) and Helicase (HvAV-3h ORF129) were also found in *Toursvirus* species. Nevertheless, both of these proteins were not conserved in *Toursvirus* species. Further investigations on proteins with unknown functions are be also needed to unveil their functions in *Ascoviridae* infection (Table 1).

Table 1
Classification of the 45 ortholog groups found in all *Ascoviridae* species and 36 ortholog groups found in all *Ascovirus* species.

Classification	Protein in all *Ascoviridae* species	Protein in all *Ascovirus* species
Replication	DNA polymerase (ORF1)[#], DNA repair exonuclease (ORF67)[#], ATPases (ORF116[#], ORF125[#] and ORF126[#])	Helicase (ORF129)
Transcription	RNA polymerase subunit (ORF9)[#], DEAD-like helicase (ORF13)[#], RNaseIII (ORF24)[#], Poxvirus late transcription factor VLTF3 like protein (ORF30)[#], RNA polymerase subunit (ORF61)[#], RNA polymerase subunit (ORF80)[#], Transcription elongationfactor S-II (ORF112)[#], CDT phosphatase transcription factor (ORF117)[#], Yabby-like transcription factor (ORF135)[#], RNA polymerase subunit (ORF143)	
DNA/RNA metabolism	Thymidine kinase (ORF52)[#]	GIY-YIG-like endonuclease (ORF49), Nuclease (ORF78), Phenylalanyl-tRNA synthetase subunit (ORF72), Nicotinate-nucleotide pyrophosphorylase (ORF100), Putative S1/P1 nuclease (ORF139)[#, SP]
Packaging and assembly	Major capsid protein (ORF53)[#], P64 (ORF58)[#], Sulfhydry1 oxidase Erv1 like protein (ORF71)[#]	
Sugar and lipid metabolism	Lysophospholipid acetyltransferase (ORF113)[#,TM], Patatin-like phospholipase (ORF133)[#, TM], Fatty acid elongase (ORF146)[#, TM]	
Signaling	Zinc-dependent metalloprotease (ORF17)[#], IAP-like proteins (ORF25 and ORF47) *****, Serine/threonine protein kinase (ORF75[#] and ORF124[#])*****	BRCA1 like protein (ORF127)[TM], Caspase-like protein (ORF171)[#],
Cell lysis	Cathepsin B (ORF109)[#, TM]	Hemolysin-like protein (ORF137)[#]

Note:

#: homologue associated with HvAV-3i virions.

TM: protein with transmembrane domain(s).

SP: protein with signal peptide.

Classification	Protein in all *Ascoviridae* species	Protein in all *Ascovirus* species
Others	Myristylated membrane protein-like protein (ORF62)[#, TM], Thioredoxin-like protein (ORF107)[#], Putative zinc-finger DNA binding protein (ORF111)[#], Lipopolysaccharide modifying enzyme (ORF136)[#], Dynein-like chain (ORF152)[#]	
Unknown	ORF46[#], ORF46a[#], ORF48[#, TM], ORF50[#], ORF54[#], ORF56[#], ORF63[#], ORF64[#, TM], ORF65[#], ORF76[#], ORF118[TM], ORF134[#, TM]	ORF5, ORF12, ORF14[SP], ORF19[#], ORF21[#], ORF22, ORF23, ORF26[#], ORF27[#], ORF29[#], ORF31[SP], ORF34[TM], ORF43, ORF44[TM], ORF55[#, TM], ORF66[#], ORF68, ORF69[#, TM], ORF74, ORF78, ORF81, ORF115[#], ORF122[#], ORF128[#, TM], ORF140[#], ORF141, ORF142, ORF147[TM]
Note:		
#: homologue associated with HvAV-3i virions.		
TM: protein with transmembrane domain(s).		
SP: protein with signal peptide.		

Identification of phosphorylated proteins and phosphorylation sites of HvAV-3h proteins

A total of 68 HvAV-3h proteins were identified using LC-MS/MS, among which, 63 proteins have orthologs in HvAV-3i and 55 of the 63 orthologs had been found to be associated with HvAV-3i virions (Table S5). Of these 68 proteins, 18 were conserved among all *Ascovirus* genomes and 25 were conserved among all *Ascoviridae* genomes, respectively (Table S5). A total of 175 phosphorylation sites in 154 phosphopeptides from 58 viral proteins were identified (Table S5), yielding a Ser/Thr/Tyr phosphorylation ratio of 126:46:3 (72.0% / 26.3% / 1.7%). Eighty-nine phosphorylation sites and their surrounding motifs were matched to consensus substrate sequences (motifs) for specific eukaryotic protein kinases using PHOSIDA (Phosphorylation Site Database) analysis [19] (Table S6). Taken conservation into consideration, 28 phosphorylation sites were conserved among all *Ascovirus* species, while two phosphorylation sites were conserved among all *Ascoviridae* species (Table 2). For example, five conserved phosphorylation sites were identified on HvAV-3h ORF50 with unknown function. Among the five phosphorylation sites, three sites (S125, S150 and T153) were conserved among all *Ascovirus* proteins, one site (Y126) was conserved among all *Ascoviridae* proteins, and the phosphorylation site S133 was also considered to be conserved among all *Ascoviridae* proteins, despite that the corresponding site on *Tousvirus* proteins changed from serine to threonine (Fig. 4B).

Table 2
Summary of conserved phosphorylation sites.

Protein	Peptide sequence	Phosphorylation site	Motif (targeted by eukaryotic kinase)
ORF19[a]	lRpSHVDDNEDLAPSK	S433[a]	PKA (K-R-X-X-S/T), CAMK2 (R-X-X-S/T), PKD (L/V/I-X-R/K-X-X-S/T), CHK1/2 (L-X-R-X-X-S/T), CHK1 (M/I/L/V-X-R/K-X-X-S/T)
	LVDCpSTTR	S556[a]	
ORF21[a]	YNNDpSIR	S33[a]	
ORF48[b]	MGSpSISTAVSK	S4[a]	
	ATQNpTVQTLR	T16[a]	
ORF50[b]	CCpSpYRDYFLSpSK	S125[a]	
		Y126[b]	
		S133[b]	
	ApSCGpTIPDAR	S150[a]	
		T153[a]	CK1 (S-X-X-S/T)
ORF55[a]	IGQpTPKAYELANER	T10[a]	CDK1 (S/T-P-K/R)
	LPQpSFTR	S214[a]	NEK6 (L-X-X-S/T)
	DPLFRNpSV	S224[a]	
P64 (ORF58)[b]	AVpSPLTNKPIDVHGR	S26[a]	
	SPSPRPYTATpSVTR	S272[a]	
	LGGLpSPYR	S515[a]	

Note:

a: conserved among all the *Ascovirus* genomes.

b: conserved among all the *Ascoviridae* genomes.

*: newly annotated protein in this study.

Protein	Peptide sequence	Phosphorylation site	Motif (targeted by eukaryotic kinase)
Myristylated membrane protein-like protein (ORF62)[b]	DECWYKPCTSDGAMTLpSTQK	S212[a]	
ORF63[b]	GpSVSDLEFK	S104[a]	Aurora (R/K-X-S/T-I/L/V)
ORF64[b]	ESGNNTNIpSPLNR	S355[a]	
ORF76[b]	SLWSpTDETLPR	T136[a]	CK1 (S-X-X-S/T), CK2 (S/T-X-X-E)
Lysophospholipid acetyltransferase (ORF113)[b]	TTLGGSTDpSVMR	S189[a]	CK1 (S-X-X-S/T)
ORF115[a]	pSVGDDpSAPVQHK	S245[a]	
	pTVTVANR	T268[a]	PKA (R-X-S/T, R-R/K-X-S/T, K-R-X-X-S/T), CAMK2 (R-X-X-S/T, R-X-X-S/T-V), AKT (R-R/S/T-X-S/T-X-S/T), Aurora (R/K-X-S/T-I/L/V), Aurora-A (R/K/N-R-X-S/T-M/L/V/I)
ATPase (ORF116)[b]	IECpSPLPEPGR	S82[a]	
ORF128[a]	QTTTSRpSPSK	S90[a]	CK1 (S/T-X-X-X-S), CDK2 (S/T-P-X-K/R), CDK1 (S/T-P-X-K/R)
ORF140[a]	STGTTVGYTGpTTSNAAMK	T17[a]	
	pSGDEMTLSR	S155[a]	CK1 (S/T-X-X-X-S), CK2 (S/T-X-X-E)
ORF142[a]	TPAIpSVTR	S128[a]	CK1 (S/T-X-X-X-S)
Caspase-like protein (ORF171)[a]	HApSCQCLMYPSSER	S146[a]	
	VFGHSSpSPR	S179[a]	CDK1 (S/T-P-K/R)

Note:

a: conserved among all the *Ascovirus* genomes.

b: conserved among all the *Ascoviridae* genomes.

*: newly annotated protein in this study.

Discussion

Functional genomics studies of large DNA viruses have been facilitated by the construction of infectious bacterial artificial chromosomes [58], which supply a useful molecular manipulation platform for serial knock out and knock in of viral genes and examination of the subsequent effects on viral infection [11, 12, 36, 41], such as been done for human hepersviruses and baculoviruses. With increase in the number of sequenced ascoviridae genomes, identification of *Acoviridae* orthology and protein phosphorylation is expected to facilitate an understanding *Acoviridae* phylogenomics and functional genomics which has been hindered by a lack of constructed bacterial artificial chromosomes for the molecular manipulation of *Ascoviridae* species.

The new annotation of 90 proteins missing in previous annotated genomes facilitates understanding of the protein conservation among *Ascoviridae* genomes, and the comprehensive identification of 366 homologous groups supplies a checklist for future annotation of *Ascoviridae* genomes. In a comprehensive view, most of the core genes commonly found in *Ascoviridae* species or *Ascovirus* species are single-copy and orthologous genes as well. Nevertheless, the orthology of non-single copy core genes needs further differentiation based on phylogenetic inference of gene/protein trees. This prompted the first construction of a phylogenetic tree of *Ascoviridae* species using concatenated orthologous genes in this study, rather than using individual genes [14, 40, 50].

Conserved genes seem to be important for viral infection process, such as viral DNA replication, RNA transcription and viral regulation by host (signaling and protein-protein interaction network) and these facilitate virus reproduction. Among the 45 core genes in HvAV-3h, 27 proteins were involved in DNA/RNA replication/transcription/metabolism, packaging and assembly, signaling, sugar and lipid metabolism, and cell lysis (Table 1), while the involvement of viral proteins in viral entry into a host cell is unknown. Apart from the membrane proteins of Cathepsin B (ORF109), Lysophospholipid acetyltransferase (ORF113), Patatin-like phospholipase (ORF133), and Fatty acid elongase (ORF146), which can facilitate formation and release of viral vesicles [10], the five HvAV-3h proteins including ORF48 (486aa), Myristylated membrane protein-like protein (ORF62) (292aa), ORF64 (427aa), ORF118 (168aa), ORF134 (219aa) conserved in *Ascoviridae* genome, and three proteins including ORF55 (225aa), ORF69 (78aa) and ORF128 (268aa), were conserved among *Ascovirus* genome, and were screened out as conserved membrane proteins with unknown functions. Taking protein conservation and association with virions into consideration [10], the involvement of these membrane proteins in viral entry needs to be further elucidated.

Protein phosphorylation targeted by various kinases are usually involved in regulation of protein function, cellular location and interaction with other proteins [26]. Of the 175 phosphorylation sites identified here, 89 sites and their surrounding motifs were matched with consensus motifs targeted by eukaryotic kinases, especially by the CK1 CAMK2, and PKA kinases, and the eukaryotic kinase CK1 targeted six conserved phosphorylation sites, indicating the involvement of eukaryotic kinases in the phosphorylation of viral proteins. Furthermore, the presence of more phosphorylation sites with no matched motif targeted

by eukaryotic kinases will enrich the set of known phosphorylation motifs of kinases [42]. For the phosphoproteins conserved among all *Ascoviridae* species, HvAV-3h ORF50 of 189 amino acids in length was highly phosphorylated with 10 phosphorylation sites, and contained five conserved phosphorylation sites including one tyrosine site (Y126), supplying a potential research target for investigation of a protein of unknown function. In addition, two serine/threonine protein kinases were found among all *Ascoviridae* species and need further work to unveil their involvement in the phosphorylation of host and viral proteins [26].

In sum, this study updated our understating of *Ascoviridae* protein conservation and species phylogeny based on a comprehensive identification of protein orthologs from homologue groups and phylogenetic analysis of non-single copy proteins, and provides multiple phosphorylation sites, in particular the conserved sites for virion-associated phosphoproteins. This should facilitate future annotation of increasing numbers of *Ascoviridae* genomes and lays a foundation and identifies potential research targets for further understanding of virus infection mechanism especially viral entry into host cells, despite the present circumstances of limited function genomic analysis of *Ascoviridae* species due to lack of molecular manipulation platforms.

Declarations

Acknowledgements

This research was funded by the Shandong Provincial Natural Science Foundation (No.ZR2020QC014), the National Natural Science Foundation of China (No.31872027) and the Doctoral Fund of Weifang Medical University (No.02181801). We also thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Ethical approval

This article does not contain any studies of involving human participants or animals.

Author's Contributions

Conceptualization, W.L., D.H. and G.H.; formal analysis, W.L., Y.S. and D.H.; writing—original draft preparation, W.L., Y.S., P.Z. and D.H.; writing—review and editing, G.W., D.H. and G.H.; supervision, D.H. and G.H; funding acquisition, D.H. and G.H. All authors have read and agreed to the published version of the manuscript.

References

1. Alzohairy AM (2011) BioEdit: An important software for molecular biology. GERF Bull Biosci 2:60–61
2. Asgari S, Bideshi DK, Bigot Y, Federici BA, Cheng XW, Consortium IR (2017) ICTV Virus Taxonomy Profile: Ascoviridae. J Gen Virol 98:4–5
3. Baldauf SL (2003) Phylogeny for the faint of heart: a tutorial. Trends Genet 19:345–351
4. Berger I, Fitzgerald DJ, Richmond TJ (2004) Baculovirus expression system for heterologous multiprotein complexes. Nat Biotechnol 22:1583–1587
5. Bideshi DK, Tan Y, Bigot Y, Federici BA (2005) A viral caspase contributes to modified apoptosis for virus transmission. GENE DEV 19:1416–1421
6. Carner GR, Hudson JS (1983) Histopathology of Virus-Like Particles in Heliothis Spp. J Invertebr Pathol 41:238–249
7. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30:918–920
8. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant 13:1194–1202
9. Chen G, Liu H, Mo BC, Hu J, Liu SQ, Bustos-Segura C, Xue J, Wang X (2020) Growth and Development of Helicoverpa armigera (Lepidoptera: Noctuidae) Larvae Infected by Heliothis virescens ascovirus 3i (HvAV-3i). Front Physiol 11:93
10. Chen ZS, Cheng XW, Wang X, Hou DH, Huang GH (2019) Proteomic analysis of the Heliothis virescens ascovirus 3i (HvAV-3i) virion. J Gen Virol 100:301–307
11. Donofrio G, Martignani E, Sartori C, Vanderplasschen A, Cavirani S, Flammini CF, Gillet L (2007) Generation of a transposon insertion mutant library for bovine herpesvirus 4 cloned as a bacterial artificial chromosome by in vitro MuA based DNA transposition system. J Virol Methods 141:63–70
12. Dunn W, Chou C, Li H, Hai R, Patterson D, Stolc V, Zhu H, Liu F (2003) Functional profiling of a human cytomegalovirus genome. Proc Natl Acad Sci U S A 100:14223–14228
13. Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238
14. Federici B (2003) Origin and evolution of polydnaviruses by symbiogenesis of insect DNA viruses in endoparasitic wasps. J Insect Physiol 49:419–432
15. Federici Ba, Vlak JM, Hamm JJ (1990) Comparative study of virion structure, protein composition and genomic DNA of three ascovirus isolates. J Gen Virol 71:1661–1668
16. Fitch WM (2000) Homology: a personal view on some of the problems. Trends Genet 16:227–231
17. Gabaldon T, Koonin EV (2013) Functional and evolutionary implications of gene orthology. Nat Rev Genet 14:360–366

18. Garavaglia MJ, Miele SAB, Iserte JA, Belaich MN, Ghiringhelli PD (2012) The ac53, ac78, ac101, and ac103 Genes Are Newly Discovered Core Genes in the Family Baculoviridae. J Virol 86:12069–12079
19. Gnad F, Gunawardena J, Mann M (2011) PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res 39:D253–D260
20. Hellmuth M, Wieseke N, Lechner M, Lenhof HP, Middendorf M, Stadler PF (2015) Phylogenomics with paralogs. Proc Natl Acad Sci U S A 112:2058–2063
21. Hitchman RB, Locanto E, Possee RD, King LA (2011) Optimizing the baculovirus expression vector system. Methods 55:52–57
22. Hou DH, Zhang LK, Deng F, Fang W, Wang RR, Liu XJ, Guo L, Rayner S, Chen XW, Wang HL, Hu ZH (2013) Comparative Proteomics Reveal Fundamental Structural and Functional Differences between the Two Progeny Phenotypes of a Baculovirus. J Virol 87:829–839
23. Huang GH, Hou DH, Wang M, Cheng XW, Hu Z (2017) Genome analysis of Heliothis virescens ascovirus 3h isolated from China. Virol Sin 32:147–154
24. Huang GH, Hou DH, Wang ML, Cheng XW, Hu ZH (2017) Genome analysis of Heliothis virescens ascovirus 3h isolated from China. Virol Sin 32:147–154
25. Hussain M, Abraham AM, Asgari S (2010) An Ascovirus-encoded RNase III autoregulates its expression and suppresses RNA interference-mediated gene silencing. J Virol 84:3624–3630
26. Jacob T, Van den Broeke C, Favoreel HW (2011) Viral serine/threonine protein kinases. J Virol 85:1158–1173
27. Javed MA, Biswas S, Willis LG, Harris S, Pritchard C, van Oers MM, Donly BC, Erlandson MA, Hegedus DD, Theilmann DA (2017) Autographa californica Multiple Nucleopolyhedrovirus AC83 is a Per Os Infectivity Factor (PIF) Protein Required for Occlusion-Derived Virus (ODV) and Budded Virus Nucleocapsid Assembly as well as Assembly of the PIF Complex in ODV Envelopes. J Virol 91:e02115–16
28. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589
29. Kessner D, Chambers M, Burke R, Agus D, Mallick P (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24:2534–2536
30. Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet 39:309–338
31. Li ZQ, Yu H, Huang GH (2018) Changes in lipid, protein and carbohydrate metabolism in Spodoptera exigua larvae associated with infection by Heliothis virescens ascovirus 3h. J Invertebr Pathol 155:55–63
32. McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20–W25
33. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol 37:1530–1534

34. Nagy LG, Riley R, Tritt A, Adam C, Daum C, Floudas D, Sun H, Yadav JS, Pangilinan J, Larsson KH, Matsuura K, Barry K, Labutti K, Kuo R, Ohm RA, Bhattacharya SS, Shirouzu T, Yoshinaga Y, Martin FM, Grigoriev IV, Hibbett DS (2016) Comparative Genomics of Early-Diverging Mushroom-Forming Fungi Provides Insights into the Origins of Lignocellulose Decay Capabilities. Mol Biol Evol 33:959–970
35. Nakamura T, Yamada KD, Tomii K, Katoh K (2018) Parallelization of MAFFT for large-scale multiple sequence alignments. Bioinformatics 34:2490–2492
36. Ono C, Kamagata T, Taka H, Sahara K, Asano S, Bando H (2012) Phenotypic grouping of 141 BmNPVs lacking viral gene sequences. Virus Res 165:197–206
37. Palmberger D, Wilson IB, Berger I, Grabherr R, Rendic D (2012) SweetBac: a new approach for the production of mammalianised glycoproteins in insect cells. PLoS ONE 7:e34226
38. Pearce SL, Clarke DF, East PD, Elfekih S, Gordon KHJ, Jermiin LS, McGaughran A, Oakeshott JG, Papanicolaou A, Perera OP, Rane RV, Richards S, Tay WT, Walsh TK, Anderson A, Anderson CJ, Asgari S, Board PG, Bretschneider A, Campbell PM, Chertemps T, Christeller JT, Coppin CW, Downes SJ, Duan G, Farnsworth CA, Good RT, Han LB, Han YC, Hatje K, Horne I, Huang YP, Hughes DST, Jacquin-Joly E, James W, Jhangiani S, Kollmar M, Kuwar SS, Li S, Liu NY, Maibeche MT, Miller JR, Montagne N, Perry T, Qu J, Song SV, Sutton GG, Vogel H, Walenz BP, Xu W, Zhang HJ, Zou Z, Batterham P, Edwards OR, Feyereisen R, Gibbs RA, Heckel DG, McGrath A, Robin C, Scherer SE, Worley KC, Wu YD (2017) Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive Helicoverpa pest species. BMC Biol 15:63
39. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20:3551–3567
40. Piegu B, Asgari S, Bideshi D, Federici BA, Bigot Y (2015) Evolutionary relationships of iridoviruses and divergence of ascoviruses from invertebrate iridoviruses in the superfamily Megavirales. Mol Phylogenet Evol 84:44–52
41. Pomeranz LE, Reynolds AE, Hengartner CJ (2005) Molecular biology of pseudorabies virus: impact on neurovirology and veterinary medicine. Microbiol Mol Biol Rev 69:462–500
42. Schwartz D, Church GM (2010) Collection and motif-based prediction of phosphorylation sites in human viruses. Sci Signal 3:rs2
43. Shang Y, Wang M, Xiao G, Wang X, Hou D, Pan K, Liu S, Li J, Wang J, Arif BM, Vlak JM, Chen X, Wang H, Deng F, Hu Z (2017) Construction and Rescue of a Functional Synthetic Baculovirus. ACS Synth Biol 6:1393–1402
44. Smede M, Hussain M, Asgari S (2009) A lipase-like gene from Heliothis virescens ascovirus (HvAV-3e) is essential for virus replication and cell cleavage. Virus Genes 39:409–417
45. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56:564–577
46. Tan Y, Bideshi DK, Johnson JJ, Bigot Y, Federici B (2009) Proteomic analysis of the Spodoptera frugiperda ascovirus 1a virion reveals 21 proteins. J Gen Virol 90:359–365

47. Tan Y, Spears T, Bideshi DK, Johnson JJ, Hice R, Bigot Y, Federici B (2009) P64, a novel major virion DNA-binding protein potentially involved in condensing the Spodoptera frugiperda Ascovirus 1a genome. J Virol 83:2708–2714
48. Varga T, Krizsan K, Foldi C, Dima B, Sanchez-Garcia M, Sanchez-Ramirez S, Szollosi GJ, Szarkandi JG, Papp V, Albert L, Andreopoulos W, Angelini C, Antonin V, Barry KW, Bougher NL, Buchanan P, Buyck B, Bense V, Catcheside P, Chovatia M, Cooper J, Damon W, Desjardin D, Finy P, Geml J, Haridas S, Hughes K, Justo A, Karasinski D, Kautmanova I, Kiss B, Kocsube S, Kotiranta H, LaButti KM, Lechner BE, Liimatainen K, Lipzen A, Lukacs Z, Mihaltcheva S, Morgado LN, Niskanen T, Noordeloos ME, Ohm RA, Ortiz-Santana B, Ovrebo C, Racz N, Riley R, Savchenko A, Shiryaev A, Soop K, Spirin V, Szebenyi C, Tomsovsky M, Tulloss RE, Uehling J, Grigoriev IV, Vagvolgyi C, Papp T, Martin FM, Miettinen O, Hibbett DS, Nagy LG (2019) Megaphylogeny resolves global patterns of mushroom evolution. Nat Ecol Evol 3:668–678
49. Vijayachandran LS, Thimiri Govinda Raj DB, Edelweiss E, Gupta K, Maier J, Gordeliy V, Fitzgerald DJ, Berger I (2013) Gene gymnastics: Synthetic biology for baculovirus expression vector system engineering. Bioengineered 4:279–287
50. Wang J, Yang ML, Xiao HB, Huang GH, Deng F, Hu ZH (2020) Genome Analysis of Dasineura jujubifolia Toursvirus 2, A Novel Ascovirus. Virol Sin 35:134–142
51. Wang LH, Li DQ, Fu Y, Wang HP, Zhang JF, Yuan ZF, Sun RX, Zeng R, He SM, Gao W (2007) pFind 2.0: a software package for peptide and protein identification via tandem mass spectrometry. Rapid Commun Mass Spectrom 21:2985–2991
52. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J, Quan L, Xia Z, Tan W, Cheng G, Jiang T (2020) Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe 27:325–328
53. Yao Q, Li H, Liu BQ, Huang XY, Guo L (2011) SUMOylation-regulated Protein Phosphorylation, Evidence from Quantitative Phosphoproteomics Analyses. J Biol Chem 286:27342–27349
54. Yu H, Li ZQ, He L, Ou-Yang YY, Li N, Huang GH (2018) Response analysis of host Spodoptera exigua larvae to infection by Heliothis virescens ascovirus 3h (HvAV-3h) via transcriptome. Sci Rep 8:5367
55. Zaghloul H, Hice R, Arensburger P, Federici BA (2017) Transcriptome analysis of the Spodoptera frugiperda ascovirus in vivo provides insights into how its apoptosis inhibitors and caspase promote increased synthesis of viral vesicles and virion progeny. J Virol 91:e00874–e00817
56. Zaghloul HAH, Hice R, Bideshi DK, Arensburger P, Federici BA (2020) Mitochondrial and Innate Immunity Transcriptomes from Spodoptera frugiperda Larvae Infected with the Spodoptera frugiperda Ascovirus. J Virol 94:e01985–19
57. Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. Mol Ecol Resour 20:348–355
58. Zhou FC, Gao SJ (2011) Recent advances in cloning herpesviral genomes as infectious bacterial artificial chromosomes. Cell Cycle 10:434–440

Figures

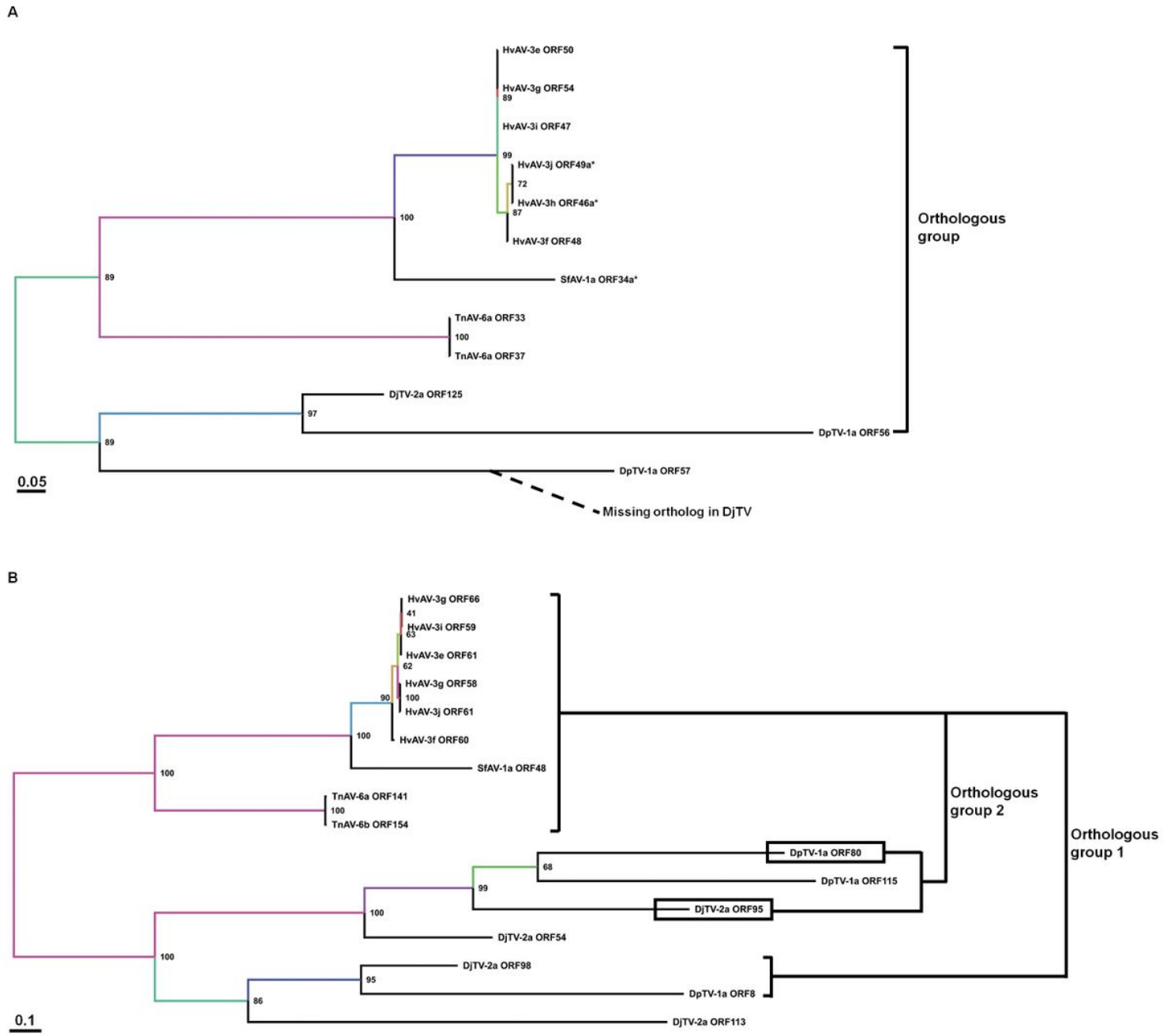

Figure 1

Phylogenetic analysis based on alignment of complete amino acid (aa) sequences of HvAV-3h ORF46a/DjTV-2a ORF125 (A) and P64 (B) homologues. Phylogenetic analysis was performed using the ML method in IQ-TREE with 1000 bootstrap replicates with a protein model of VT and PMB+F+R2 for HvAV-3h ORF46a/DjTV-2a and P64 homologues, respectively. The orthologous group was identified based on the phylogenetic analysis each of homologue group. The scale bar for branch length is shown at the bottom left of each phylogenetic tree. * Newly annotated in this study. See full names of viruses in Table S1.

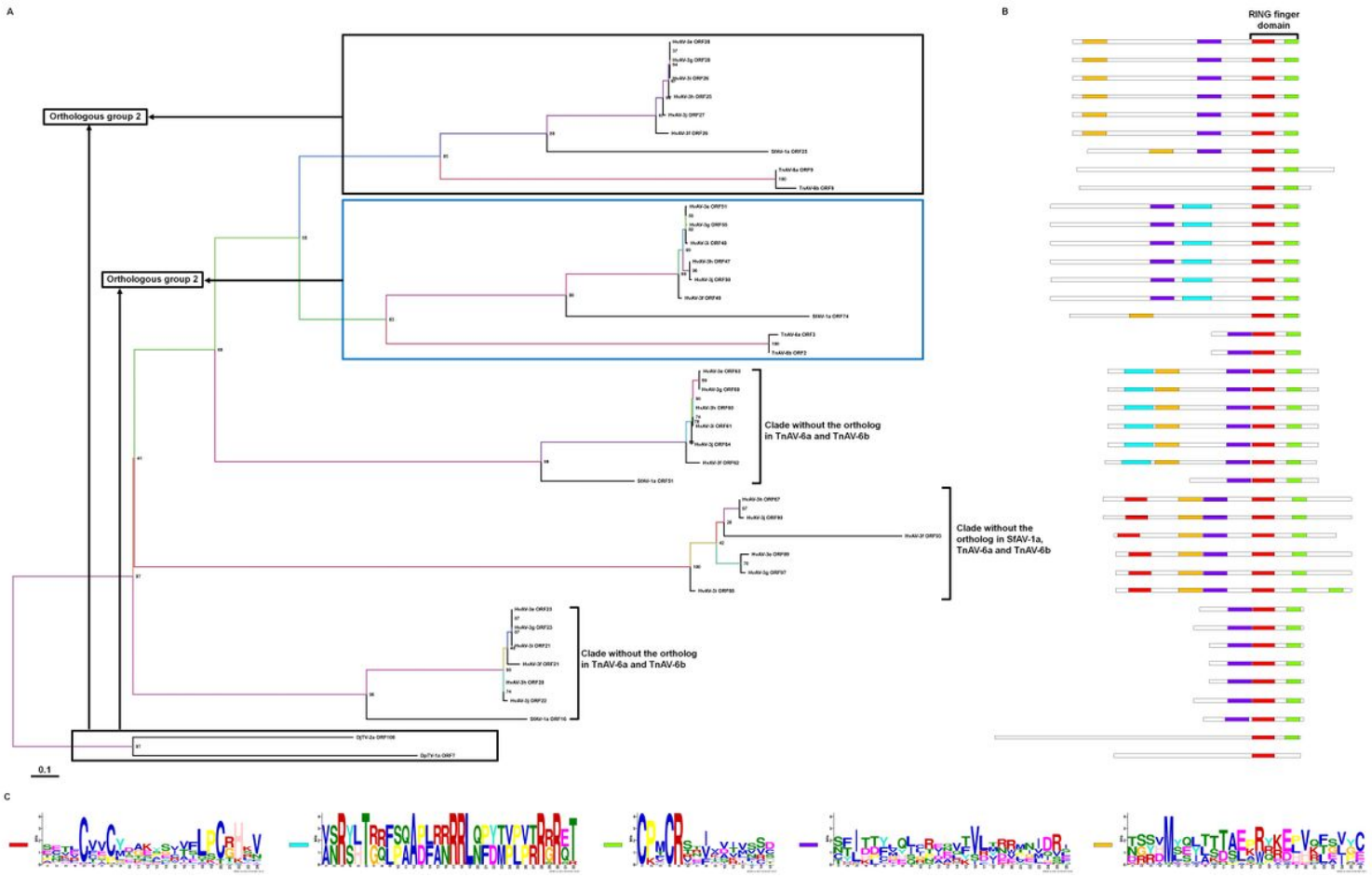

Figure 2

Phylogenetic and structural analysis based on alignment of complete aa sequences of IAP homologues. A Phylogenetic analysis was performed using the maximum-likelihood method in IQ-TREE with 1000 bootstrap replicates with a protein model of WAG+F+I+G4 and the orthologous group was identified based on the phylogenetic analysis. The scale bar for branch length is shown at the bottom left of each phylogenetic tree. See full names of viruses in Table S1. B The corresponding polypeptide sequences of IAPs at the terminal of phylogenetic tree are represented as rectangles, with small rectangles in different colors corresponding to motifs found using MEME. C The motifs corresponding to different colored rectangles.

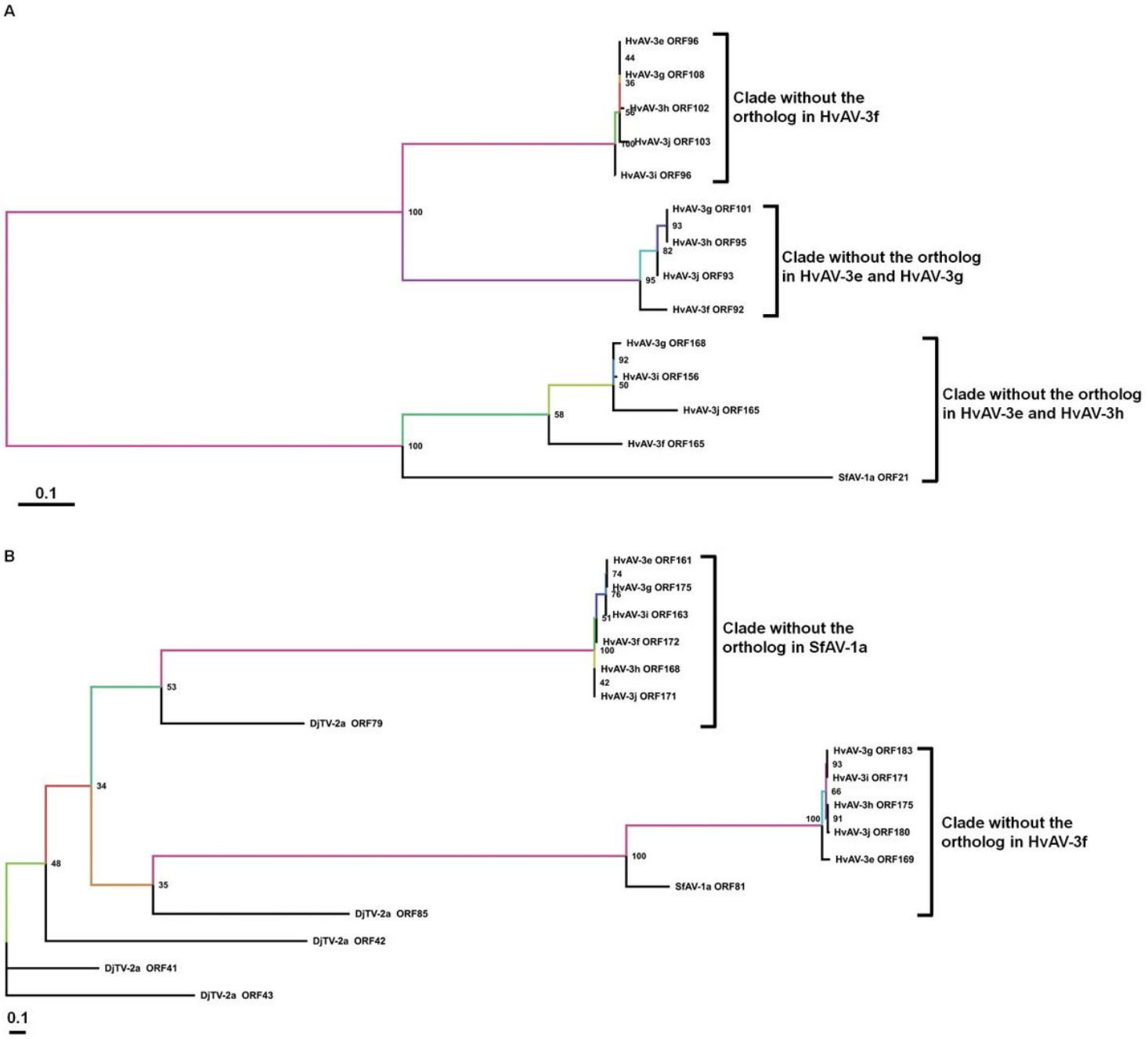

Figure 3

Phylogenetic analysis based on alignment of complete aa sequences of Virokine (HvAV-3h ORF95/ORF102) (A) and RING-finger-containing E3 ubiquitin ligase (HvAV-3h ORF167/ORF175) (B) homologues. Phylogenetic analysis was performed using the maximum-likelihood method in IQ-TREE with 1000 bootstrap replicates with a protein model of WAG+G4 and FLU+I+G4 for Virokine (HvAV-3h ORF95/ORF102) and RING-finger-containing E3 ubiquitin ligase (HvAV-3h ORF167/ORF175) homologues, respectively. The scale bar of branch length was demonstrated at the bottom left of each phylogenetic tree. See full names of viruses in Table S1.

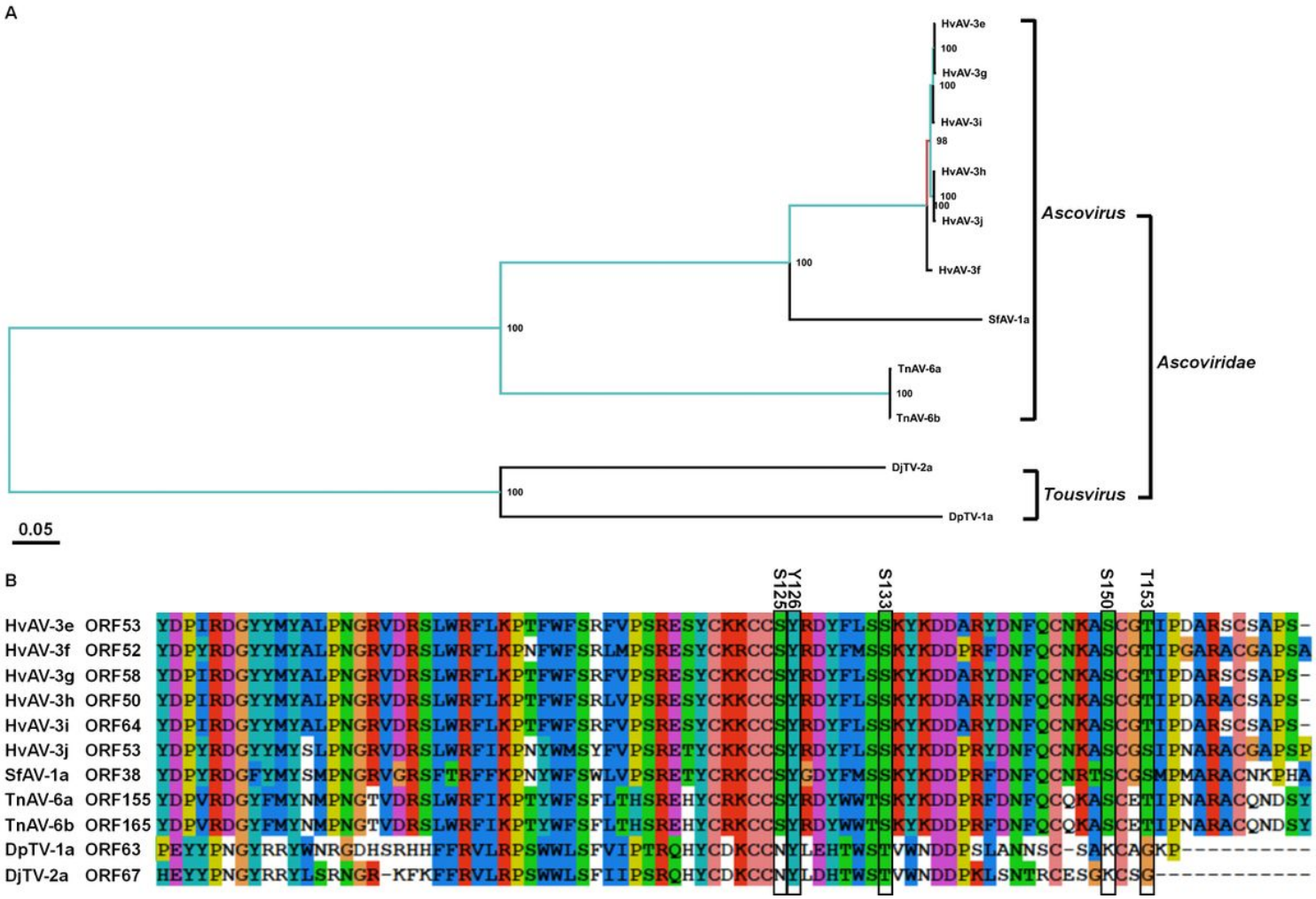

Figure 4

Phylogenomic analysis of Ascoviridae species and multiple alignment of orthologs of HvAV-3h ORF50. A Phylogenetic analysis based on concatenation of 35 orthologous proteins was performed using the maximum-likelihood method in IQ-TREE with 10,000 bootstrap replicates with a protein model for LG+F+R4. The scale bar of branch length is shown at the bottom left of each phylogenetic tree. See full names of viruses in Table S1. B Multiple alignment of the orthologs of HvAV-3h ORF50. Phosphorylation sites are indicated using rectangles. See full names of viruses in Table S1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Fig.S1.jpg
- Fig.S2.jpg
- Fig.S3.jpg
- Fig.S4.jpg
- Fig.S5.jpg

- TableS1.doc
- TableS2.docx
- TableS3.xlsx
- TableS4.docx
- TableS5.docx
- TableS6.docx