Virology 513 (2018) 114–128

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Sequence analysis of malacoherpesvirus proteins: Pan-herpesvirus capsid module and replication enzymes with an ancient connection to "*Megavirales*"

Arcady Mushegian[a,*], Eli Levy Karin[b], Tal Pupko[b]

[a] Division of Molecular and Cellular Biosciences, National Science Foundation, 2415 Eisenhower Avenue, Alexandria, VA 22314, USA
[b] Department of Cell Research and Immunology, George S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel

ARTICLE INFO

Keywords:
Herpesvirales
Virus evolution
Virus phylogeny

ABSTRACT

The order *Herpesvirales* includes animal viruses with large double-strand DNA genomes replicating in the nucleus. The main capsid protein in the best-studied family *Herpesviridae* contains a domain with HK97-like fold related to bacteriophage head proteins, and several virion maturation factors are also homologous between phages and herpesviruses. The origin of herpesvirus DNA replication proteins is less well understood. While analyzing the genomes of herpesviruses in the family *Malacohepresviridae*, we identified nearly 30 families of proteins conserved in other herpesviruses, including several phage-related domains in morphogenetic proteins. Herpesvirus DNA replication factors have complex evolutionary history: some are related to cellular proteins, but others are closer to homologs from large nucleocytoplasmic DNA viruses. Phylogenetic analyses suggest that the core replication machinery of herpesviruses may have been recruited from the same pool as in the case of other large DNA viruses of eukaryotes.

1. Introduction

The order *Herpesvirales* consists of animal viruses with large linear double-strand (ds) DNA genomes of 125–145 kbp, encoding between 70 and 200 proteins. Herpesviruses are characterized by the following phenotypic attributes (Davison et al., 2009; Pellett et al., 2011): the mature particle contains the genome within a T = 16 icosahedron capsid, composed of 162 capsomers, of which one (the portal) has a distinct structure, specialized for mediating virus DNA entry and exit from the capsid; the capsid is surrounded by a proteinaceous tegument and further wrapped in a lipid envelope that contains several virus-encoded transmembrane proteins; the viral genome contains directed or inverted DNA repeats; the virus DNA replication mechanism generates head-to-tail concatemers that are later cleaved and individually packed into pre-made capsids.

These characteristic molecular and morphological features have been used to assign members to the order *Herpesvirales* even in the absence of the genome sequence information (reviewed in Davison, 2010). In the genomic era, however, the state-of-the-art taxonomy of herpesviruses is based on phylogenetic analyses of virus nucleotide and amino acid sequences (Davison et al., 2009; Davison, 2010; Pellett et al., 2011). For certain subsets of herpesviruses, additional phylogenetic signal can be extracted from the information about gene synteny and by estimating the minimal amount of rearrangements required to convert the gene order between two species (Hannenhalli et al., 1995; Bourque and Pevzner, 2002; Larget et al., 2005).

Analysis of these and other molecular features resulted in establishing three families within *Herpesvirales* that are currently recognized by the International Committee for Taxonomy of Viruses (ICTV; Davison et al., 2009; Pellett et al., 2011). The *Herpesviridae* family (GenBank TaxID 10292, with 73 completely sequenced virus genomes as of May 2017) includes viruses that infect amniotes; typically, a pair of virus species from this family shares dozens of protein-coding genes recognized by high amino acid sequence similarity (Davison, 2010). The family *Alloherpesviridae* (TaxID 548682, 8 completely sequenced genomes) includes viruses of fishes and amphibians. These viruses have been reported to share only 13 core orthologous genes with each other, suggesting that this group is more divergent than the family *Herpesviridae* (van Beurden and Engelsma, 2012). The third family, *Malacoherpesviridae* (TaxID 548685), includes two ICTV-approved members, *Ostreid herpesvirus 1* (OsHV-1) and *Haliotid herpesvirus 1* (HaHV-1), which infect bivalve and gastropod molluscs, respectively. In addition to the herpesvirus-like morphology of their virions and capsids, as well as the genomic repeat features, these two viruses were reported to have 39 orthologous genes in common (Savin et al., 2010), some of which are related to enzymes that play essential roles in other herpesviruses, for example, the large subunit of virus terminase ATPase and family B replicative DNA polymerase. However, no gene products with similarity

* Corresponding author.
E-mail address: mushegian2@gmail.com (A. Mushegian).

http://dx.doi.org/10.1016/j.virol.2017.10.009
Received 20 August 2017; Received in revised form 8 October 2017; Accepted 9 October 2017
Available online 21 October 2017
0042-6822/ Published by Elsevier Inc.

to herpesvirus structural proteins have been reported in *Malacoherpesviridae*. Interestingly, a putative virus related to mollusc herpesviruses has been detected by sequence database searches, among the contigs co-assembled with the portions of the draft genome of a cephalochordate *Branchiostoma floridae* (Savin et al., 2010). Taken together, these genomic data suggest that the order *Herpesvirales* is characterized by higher genomic diversity, broader host range, and, likely, deeper phylogeny than was assumed just a decade ago.

In contrast to our understanding of the phylogenetic relationships within *Herpesivirales*, the evolutionary origin of this order as a whole remains uncertain. The picture is clearer in the case of structural proteins. One domain, "the floor domain", in the large multi-domain capsid proteins of vertebrate herpesviruses, whose structure has been determined by a cryoEM approach, is structurally similar to bacteriophage capsid proteins with the HK97 fold (Baker et al., 2005; Yu et al., 2017). Moreover, mechanistic similarities in the maturation pathways of these diverse virus particles have been noted (Casjens and King, 1975; Booy et al., 1991; Davison, 2002; Mettenleiter et al., 2009; Veesler and Johnson, 2012), and sequence comparisons have established the ancestral relationship of the capsid maturation protease between phages and *Herpesviridae* (Cheng et al., 2004; Liu and Mushegian, 2004). That relationship is further supported by the demonstration that the high-resolution structure of phage protease is folded very similarly to the herpesvirus assemblin proteases, whose structure has been determined earlier (Fokine and Rossmann, 2016). Along with the long-known monophyly of the large (ATPase) subunit of phage and virus terminases (Mitchell et al., 2002), these observations point at the common origin of the entire head/capsid formation module in *Herpesviridae* and in a subset of tailed viruses of bacteria and archaea.

The state of affairs is different for genes involved in virus DNA replication and expression. For several established families of eukaryotic dsDNA viruses, which, unlike herpesviruses, replicate partly or fully in the cytoplasm (Nucleocytoplasmic Large DNA Viruses, or NCLDV), considerable evidence of the monophyletic origin has been derived from comparative genome analyses (Iyer et al., 2006; Yutin et al., 2009; Koonin and Yutin, 2010). The assembly of large DNA viruses consisting of NCLDV and several related lineages has been proposed as a candidate order "*Megavirales*" (Colson et al., 2013), not yet accepted by ICTV. Large dsDNA viruses replicating in the nucleus, such as herpesviruses and baculoviruses, have several replication enzymes homologous to those of NCLDV (see below), but the evolutionary scenario linking them together has not been established.

In this work, we analyzed the gene repertoire of malacoherpesviruses, using predicted open reading frames (ORFs) of HaHV-1 (synonym *Abalone herpesvirus* Victoria/AUS/2009) as the starting point. Our study characterized the enlarged repertoire of conserved malacoherpesvirus genes, and for many of them we identified similar sequences in other viruses or in the genomes of cellular organisms. Analysis of sequence similarities and phylogenetic inference on those gene families expand the common gene core of *Herpesvirales,* support the hypothesis of the common ancestry of their morphogenetic module, and suggest that the herpesvirus replication module may have been recruited from the same ancient gene pool as the replication genes of other large DNA viruses.

2. Methods

Sequence database searches were mostly performed in July-December of 2016, except for the results presented in Fig. 1 and the accompanying text, for which the searches were repeated in June-September of 2017. Protein queries representing every predicted ORF in HaHV-1 (GenBank Taxonomy ID 860344) were used to query the nucleotide sequence databases (NT and dbEST) and the non-redundant protein sequence database (NR) at NCBI. Searches were done with the BLAST family of programs (Altschul et al., 1997) with the "Composition-based statistics" option. The HHPred server (Söding et al., 2005; Alva et al., 2016) was used to compare profile-Hidden Markov Models (HMM) of virus proteins to profile-HMMs built from the entries in the NCBI CDD database (Marchler-Bauer et al., 2017), with multiple sequence alignment generation method set to PSI-BLAST, re-aligning the results with MAC set to 0.3, and the remaining options set to default values. Multiple sequence alignments were computed using the MUSCLE program (Edgar, 2004), joining closely related (family level and below) sequences, and combining the alignments obtained at that first step using the -profile option of MUSCLE. When distantly related proteins were to be aligned with structural constraints, the PROMALS3D server was used (Pei and Grishin, 2014). Spatial structures were visualized with the PyMOL legacy build 0_99rc6 (https://sourceforge.net/projects/pymol/files/Legacy/).

Phylogenetic inference was done using a local installation of the PhyML program as well as the PhyML 3.0 web server (Guindon et al., 2010) with the LG substitution model, most other parameters estimated from the data, and bootstrap replicates performed to assess the support of the internal partitions in the tree. The iTOL v.3 server (Letunic and Bork, 2016) was employed for tree examination and visualization.

Statistical tests for relaxed selective constraints in the lineages leading to *B.floridae* and *Capitella teleta* were conducted using RELAX (Wertheim et al., 2015). For each set of orthologous proteins, corresponding nucleotide sequences were extracted from GenBank, codons were aligned using the PRANK program (Löytynoja, 2014), and for each multiple sequence alignment, two p-values were computed (one for a relaxation signal in the lineage leading to *B. floridae* and one for a relaxation signal in the lineage leading to *C. teleta*). These p-values were computed using a chi-square distribution in a likelihood ratio test, as proposed in RELAX. In order to account for multiple hypotheses testing and control the false discovery rate, p-values were adjusted using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), as implemented in the R function p.adjust.

3. Results and discussion

3.1. Nearly half of malacoherpesvirus gene products are conserved

We compared 118 putative proteins encoded by the genome of HaHV-1 to the protein and nucleotide sequence databases, as well as the databases of known conserved protein domains (the CDD database at NCBI). The identities of the best statistically supported matches in viruses and cellular organisms were recorded; some of the matches had low statistical significance, but could be validated by observing the conservation of characteristic sequence motifs. The combined results of these analyses are presented in Table 1.

Nearly 60 proteins in HaHV-1 have homologs in the better-studied malacoherpesvirus, OsHV-1 (and often also in the unclassified, but highly similar, malacoherpesvirus infecting scallops, *Chlamys acute necrobiotic virus;* data not shown). The BLASTP pairwise sequence identity within malacoherpesvirus sequences was generally between 20% and 40%, suggesting considerable divergence of these viruses. Some amount of synteny between the two virus genomes was observed, usually in blocks of 2–4 genes.

As reported before (Savin et al., 2010), substantial portions of the malacoherpesvirus genomes also show homology to a part of the draft genome assembly of the cephalochordate *B. floridae.* Most of the herpesvirus-like matches in the *B. floridae* genome are mapped to a single contig; some of these genes have been predicted by the original genome annotation, while others have not been reported by Savin et al. (2010), but could be identified by matching malacoherpesvirus protein sequences to the translations of the putative intergenic regions. We also detected a collection of homologous genes in a draft genome assembly of another marine invertebrate, annelid *C. teleta*. In this case too, at least some of the genes appear to be located on the same contig (Table 1).

Taken together, these matches provide evidence for evolutionary

Fig. 1. Multiple sequence alignments of malacoherpesvirus morphogenesis proteins. GenBank accession numbers as well as ICTV-approved, sometimes truncated, but not italicized, virus species names are shown before each sequence. PDB ID numbers, when available, are also shown. The color codes of conserved amino acids are as follows: bold type and yellow shade, residues with bulky hydrophobic side chains (I, L, V, M, F, Y, W); red type, residues with negatively charged or amido side chains (D, E, N, Q); blue type, residues with positively charged side chains (K, R); white type and red shade, residues with turn- or kink-prone side chains (A, G, P, S). Other conserved residue colors are ad hoc. In the secondary structure lines, h and s indicate helix and strand, respectively. **A. Floor and buttress domains of the major capsid proteins.** The number of the leftmost amino acid residue in each block, counting from the N-terminus of the precursor protein, is shown before each sequence block. In the secondary structure line, l stands for a loop, and different colors match the loop elements highlighted in Additional Files 1 and 2. **B. Procapsid proteases.** The number of the leftmost amino acid residue in each block, counting from the N-terminus of the precursor protein, is shown before each sequence. Numbers in parentheses indicate distances between conserved sequence regions, in amino acids. Asterisks indicate the amino acids that form the catalytic triad of the protease. The wild-type sequence of protease from phage T4 (*Escherichia virus T4*) is shown, though two of the three catalytic residues and a non-catalytic methionine residue have been mutated in the sequence of 5jbl.

conservation at the family level for at least half of the genes in each malacoherpesvirus genome. Each of these herpesvirad genomes, in addition, has genes that were not found in other herpesviruses but are distantly related to gene products of a different evolutionary origin. In more than 35 of the conserved malacoherpesvirus gene products, we identified specific sequence motifs suggestive of their molecular

Table 1
Putative protein functions in malacoherpesviruses and orthologous relationships in four species. Gray shade, two types of repeats within the HaHV-1 genome. Blue shade, gene products with predicted structural and morphogenetic roles. Yellow shade, gene products with predicted roles in genome replication. Bold type, novel predictions.

HaHV-1 Protein ID	Functions or sequence similarities	OsHV-1	*Branchiostoma*	*Capitella*
YP_006908651.1	ORF001_1 inhibitor of apoptosis	AKM21020.1 ORF87	XP_002594758.1	ELT95174.1
YP_006908652.1	ORF002_1	YP_024556.1 ORF11		
YP_006908653.1	ORF003 eukaryotic translation initiation factor		BRAFLscaffold_28_Cont5495	CAPTEscaffold_570_Cont12099
YP_006908654.1	**ORF004 DUF1335 as in *White spot syndrome virus***			
YP_006908655.1	ORF005	YP_024584.1 ORF40	XP_002591176.1	ELT95154.1
YP_006908656.1	ORF006			
YP_006908657.1	ORF007	YP_024630.1 ORF91	BRAFLscaffold_217_Cont31171	
YP_006908658.1	ORF008			
YP_006908659.1	ORF009	YP_024632.1 ORF93		
YP_006908660.1	ORF010	YP_024633.1 ORF94	XP_002591165.1	ELT95190.1
YP_006908661.1	ORF011			
YP_006908662.1	ORF012			
YP_006908663.1	ORF013			
YP_006908664.1	**ORF014 capsid protein VP23 family**	YP_024621.1 ORF82	XP_002591173.1	ELT95179.1
YP_006908665.1	ORF015	YP_024620.1 ORF81		ELT95179.1
YP_006908666.1	ORF016	YP_024619.1 ORF80		
YP_006908667.1	ORF017	ASK05601.1 ORF78 (partial match)		
YP_006908668.1	ORF018	KY242785.1 (TBLASTN, partial match)		
YP_006908669.1	ORF019			

(continue on next page)

Table 1 *(continued)*

YP_006908670.1	ORF020	YP_024627.1 ORF88		
YP_006908671.1	**ORF021 DUF1335 as in *White spot syndrome virus***			
YP_006908672.1	ORF022 ribonucleotide reductase large subunit RNR1	YP_024594.1 ORF51	XP_002589901.1 ---- host?	ELT88174.1 ----host?
YP_006908673.1	ORF23	YP_024595.1 ORF52		
YP_006908674.1	ORF24	YP_024596.1 ORF53		
YP_006908675.1	**ORF025 envelope fusion protein?**	YP_024597.1 ORF54	XP_002591197.1	ELT95181.1
YP_006908676.1	**ORF026 helicase C domain**	YP_024657.1 ORF115		
YP_006908677.1	**ORF027 HUH endonuclease fused to an ATPase**			
YP_006908678.1	ORF028			
YP_006908679.1	**ORF029 winged helix domain**			
YP_006908680.1	ORF030 DNA packaging terminase large subunit	YP_024647.1 ORF109	XP_002591195.1	ELT95188.1
YP_006908681.1	ORF031putative ribonucleotide reductase small subunit RNR2	YP_024565.1 ORF20	BRAFLscaffold_135 _Cont18938	
YP_006908682.1	ORF032	YP_024602.1 ORF59 YP_024616.1 ORF77 YP_024585.1 ORF41		
YP_006908683.1	ORF033	YP_024573.1 ORF28	XP_002591172.1	ELT95172.1
YP_006908684.1	ORF034 dUTPase **herpesvirus-typical internal duplication**	YP_024614.1 ORF75		
YP_006908685.1	ORF035	YP_024627.1 ORF88		
YP_006908686.1	**ORF036 herpesvirus helicase**	YP_024591.1 ORF47	XP_002591194.1	ELT95177.1
YP_006908687.1	ORF037	YP_024590.1 ORF46		
YP_006908688.1	ORF038 DNA replication origin-binding helicase	YP_024552.1 ORF7 YP_024593.1 ORF49	XP_002591168.1	
YP_006908689.1	ORF039			
YP_006908690.1	ORF040			

(continue on next page)

Table 1 *(continued)*

YP_006908691.1	ORF041			
YP_006908692.1	ORF042	YP_024567.1 ORF22		
YP_006908693.1	ORF043			
YP_006908694.1	ORF044	KY242785.1 (TBLASTN, partial match)		
YP_006908695.1	**ORF045 archaeo-eukaryotic primase small subunit**	YP_024606.1 ORF66	XP_002591170.1	ELT95184.1
YP_006908696.1	ORF046 ATP-dependent RNA helicase	YP_024607.1 ORF67		
YP_006908697.1	ORF047 Herpesviridae UL92 family	YP_024575.1 ORF30		
YP_006908698.1	ORF048	YP_024616.1 ORF47	XP_002591196.1	
YP_006908699.1	ORF049	YP_024587.1 ORF43		
YP_006908700.1	**ORF050 PD-(DE)XK exonuclease**	YP_024634.1 ORF95		
YP_006908701.1	ORF051			
YP_006908702.1	ORF052	YP_024611.1 ORF71		
YP_006908703.1	ORF053			
YP_006908704.1	ORF054	YP_024604.1 ORF61	BRAFLscaffold_217 _Cont31174	ELT95186.1
YP_006908705.1	ORF055			
YP_006908706.1	ORF056			
YP_006908707.1	ORF002_2	YP_024556.1 ORF11		
YP_006908708.1	ORF001_2 inhibitor of apoptosis			
YP_006908709.1	ORF057			
YP_006908710.1	ORF058			
YP_006908711.1	ORF059			
YP_006908712.1	ORF060			
YP_006908713.1	ORF061			

(continue on next page)

Table 1 *(continued)*

YP_006908714.1	ORF062	YP_024556.1 ORF11		
YP_006908715.1	ORF063			
YP_006908716.1	ORF064			
YP_006908717.1	**ORF065 DUF1335 as in *White spot syndrome virus***			
YP_006908718.1	ORF066	YP_024627.1 ORF88		
YP_006908719.1	ORF067	YP_024622.1 ORF83		
YP_006908720.1	**ORF068 major capsid protein**	YP_024643.1 ORF104	XP_002591178.1	ELT95197.1
YP_006908721.1	**ORF069 inhibitor of apoptosis**	YP_024635.1 ORF96		
YP_006908722.1	ORF070	YP_024551.1 ORF6 YP_024652.1 ORF114		ELT95166.1
YP_006908723.1	ORF071	YP_024628.2 ORF89		
YP_006908724.1	ORF072			
YP_006908725.1	**ORF073 prohead protease assemblin, possibly scaffold domain at the C-terminus**	YP_024645.1 ORF107	XP_002591175.1	CAPTEscaffold_559_Cont11963
YP_006908726.1	ORF074			
YP_006908727.1	ORF075			
YP_006908728.1	ORF076			
YP_006908729.1	ORF077			
YP_006908730.1	ORF078			
YP_006908731.1	ORF079 inhibitor of apoptosis		BRAFLscaffold_1_Cont354	CAPTEscaffold_559_Cont11961
YP_006908732.1	ORF080	YP_024649.1 ORF111	BRAFLscaffold_217_Cont31173	
YP_006908733.1	**ORF081 capsid protein VP19 family**	YP_024650.1 ORF112	XP_002591198.1	ELT95164.1
YP_006908734.1	**ORF082 HUH endonuclease fused to an ATPase**			
YP_006908735.1	ORF083	YP_024651.1 ORF113		

(continue on next page)

Table 1 *(continued)*

YP_006908736.1	ORF084			
YP_006908737.1	ORF085			
YP_006908738.1	**ORF086 transposase: HTH-RuvC-Zn ribbon**			
YP_006908739.1	ORF087	YP_024641.1 ORF102	XP_002591174.1	ELT95155.1 ELU16009.1
YP_006908740.1	ORF088	YP_024642.1 ORF103	BRAFLscaffold_217 _Cont31171	
YP_006908741.1	ORF089			
YP_006908742.1	ORF090 putative chloride channel	YP_024600.1 ORF57	XP_002610653.1	
YP_006908743.1	ORF091			
YP_006908744.1	ORF092 Family B DNA polymerase	YP_024639.1 ORF100	XP_002591163.1	ELT95180.1
YP_006908745.1	ORF93			
YP_006908746.1	**ORF094 helicase-primase**	YP_024552.1 ORF7 and YP_024593.1 ORF49	XP_002591168.1	
YP_006908747.1	ORF095			
YP_006908748.1	ORF096	YP_024601.1 ORF58		
YP_006908749.1	ORF097 tRNA ligase	YP_024605.1 ORF64		
YP_006908750.1	**ORF098 YqaJ-like endonuclease**	YP_024615.1 ORF76	XP_002591179.1	ELT95196.1
YP_006908751.1	ORF099			
YP_006908752.1	ORF100	YP_024570.1 ORF25		
YP_006908753.1	ORF101	YP_024569.1 ORF24		
YP_006908754.1	**ORF102 envelope fusion protein**	YP_024608.1 ORF68 YP_024597.1 ORF54	XP_002591190.1 XP_002591197.1	ELT95157.1 ELT95181.1
YP_006908755.1	ORF103	YP_024624.1 ORF85	XP_002591189.1	
YP_006908756.1	ORF104			
YP_006908757.1	**ORF105 Small molecule kinase**	YP_024588.1 ORF44		
YP_006908758.1	ORF106	YP_024609.1 ORF69	XP_002591200.1	ELT95169.1

(continue on next page)

Table 1 *(continued)*

YP_006908759.1	ORF107 ATP-dependent DNA ligase		XP_002595387.1	ELT99220.1
YP_006908760.1	ORF108 Zn finger protein; EVH1-domain E3 ubiquitin-protein ligase			
YP_006908761.1	ORF109	YP_024627.1 ORF88		
YP_006908762.1	ORF110			
YP_006908763.1	ORF111 guanylate kinase		BRAFLscaffold_158_Cont23243	ELU10045.1
YP_006908764.1	ORF112			
YP_006908765.1	ORF062_2	YP_024556.1 ORF11		
YP_006908766.1	ORF061_2			
YP_006908767.1	ORF060_2			
YP_006908768.1	ORF059_2			

function (Table 1 and see below).

3.2. Structural proteins: herpesvirus orthologs and phage connections

The major capsid protein in *Herpesviridae* (MCP; also known as VP5/UL19 gene product in *Human herpesvirus 1*, HHV-1, synonym Herpes simplex virus) is a large multidomain protein. Structural data for the entire protein within the assembled capsid, obtained by cryoEM, indicate that VP5 consists of the "floor" domain, which forms the icosahedral shell and is the only part of the protein that interacts with the capsid interior, as well as the "middle" and "top" domains protruding above the surface of the shell (Zhou et al., 2000). More recently, an assembly of additional domains has been recognized in the MCPs of HHV-1 (Hui et al., 2013) and of human cytomegalovirus (*Human herpesvirus 5*; Yu et al., 2017). A high-resolution X-ray structure is still available only for the top, or tower, domain of HHV-1, which is contiguous in the sequence of VP5, starting approximately at amino acid 500 (Bowman et al., 2003).

Our analysis detected a match between a profile-HMM seeded with the HaHV-1 ORF68 and the probabilistic model specified by the top domain of HHV1 VP5 (PDB ID 1no7; Bowman et al., 2003). The HHPred search initiated with HaHV-1 ORF68 matched the sequence family model of PDB identifier 1no7 with p-value of 10^{-4}. The match is about 100 amino acids long, and is located close to the N-terminus of the proteolytically isolated top domain of VP5, whereas in ORF68 it is found in the middle, approximately between amino acids 570 and 670. In addition, a PSI-BLAST search against the complete NR database at NCBI (mined on June 14th, 2017) was initiated with the full-length ORF68. This database search retrieved the *Malacoherpesviridae* homologs at the first round, followed at the next round by the putative capsid protein sequence from *Macacine betaherpesvirus 3* (synonym Rhesus cytomegalovirus; protein ID APT40174.1; p-value $< 10^{-4}$) as well as their homologs from several other betaherpesviruses and gammaherpesviruses. The matches covered two-thirds of the protein lengths, starting close to the N-termini and extending to the same region of similarity as above.

To refine the alignment, we used the PROMALS3D approach (Pei and Grishin, 2014), which collects homologs of all input sequences by PSI-BLAST and realigns them using the compatibility of highly-scoring alignment segments, with the additional constraint imposed by the known tertiary structure of the homologous sequences. There were three regions of high sequence similarity between herpesvirus and malacoherpesvirus major capsid proteins. Two of these regions map to the terminal sequence segments of the tower domain, and the third region is located adjacent to the C-terminal end of the tower domain sequence (Fig. 1A).

The structure of the tower domain of herpesvirus MCP is described as a pyramid with a square base (Bowman et al., 2003). Much attention has been given to the structural elements constituting the sides and the apex of the pyramid, as these sites appear to be involved in multiple protein-protein interactions (Hui et al., 2013). Less is known about the

Table 2
Testing for relaxation of selection in the herpesvirus genes deposited with the assemblies of *C. teleta* and *B. floridae*. For the identity of HaHV-1 orthologs in the annelid and cephalochordate, refer to Table 1. Fourteen statistical tests for relaxation were conducted separately, and the p-values are shown before correcting for multiple testing. After correcting for multiple testing, none of the tests was statistically significant.

Gene in HaHV	Gene function	*Capitella*	*Branchiostoma*
ORF098	YkaJ-type endonuclease	0.078	0.67
ORF092	Family B DNA polymerase	0.49	0.12
ORF068	Major capsid protein	0.005	0.49
ORF045	Primase small subunit	0.59	0.3
ORF036	Helicase	0.023	0.55
ORF014	Capsid protein VP23 family	0.41	1
ORF030	Terminase large subunit	0.49	0.21

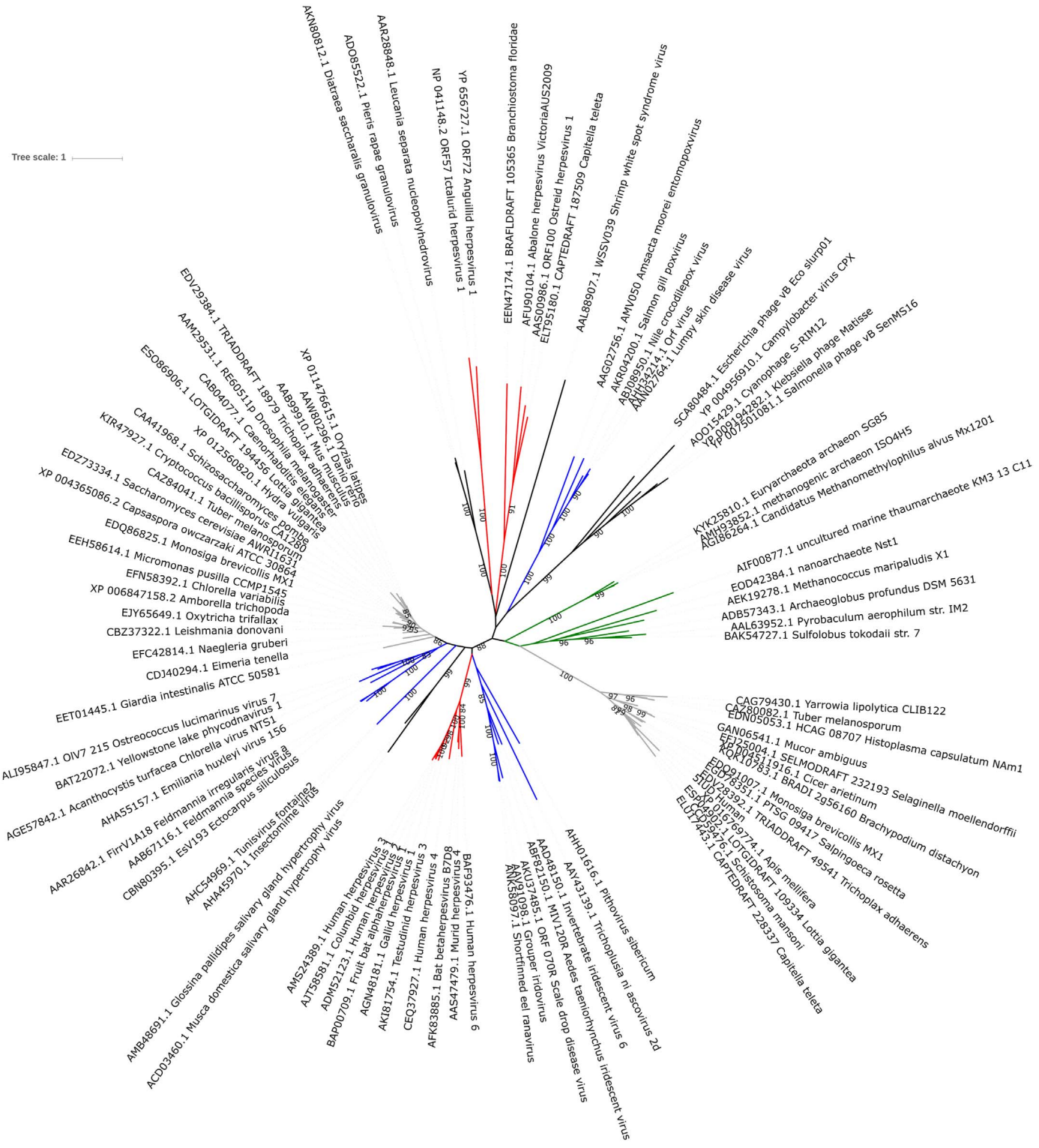

Fig. 2. Unrooted maximum-likelihood tree of viral and cellular Family B DNA polymerases. The percentage of bootstrap support is shown for all internal partitions where it exceeds 80%. Monophyletic clades are indicated by different colors. Blue, lineages within proposed order *Megavirales;* red, families of *Herpesvirales*; black, other viruses and phages; green, *Archaea*; gray, *Eukarya*.

base itself, other than that its structure comprises mostly loops and a few helices (Bowman et al., 2003). Our results, however, indicate that the base is made of the best-conserved portion of the MCP sequence; indeed, conserved regions 1 and 2, distal in the sequence as they are, come together in the three-dimensional structure to account for almost the entirety of the base (visualized in Additional files 1 and 2). Conservation of these regions in the two families of *Herpesvirales* suggests that the foundation of the tower domain plays an important role in maintaining the integrity of the multidomain MCP and, possibly, its interactions with other virion components.

Our efforts to identify a homologous region in *Alloherpesviridae*, where mass-spectrometry approaches have assigned the major capsid protein role to ORF39 in *Ictalurid herpesvirus 1* (Davison and Davison, 1995) and to its ortholog ORF104 in *Anguillid herpesvirus 1* (Van

Beurden et al., 2011) have produced ambiguous results. When ORF39/ORF104 and their orthologs from other alloherpesviruses were used as HHPred queries to search the available HMMs, no matches were found. In a targeted PSI-BLAST search initiated by these sequences against all virus proteins in the NR database, capsid proteins from *Herpesviridae* were detected, but the similarities were not statistically significant.

The N-terminal region of MCPs in *Herpesviridae* was thought to have the HK97-like fold based on low-resolution cryo-EM (Baker et al., 2005; Hui et al., 2013). Very recently, the atomic structure of the entire MCP of HHV-5 has been determined by cryo-EM to 3.9 Å resolution, revealing that a modified HK97 fold is present in the structure, but it is assembled in space from several distal segments of the peptide chain. This can be explained as a series of insertions of several very large domains into the loops of the HK97 "floor" domain. The relative arrangement of the predicted secondary structure elements along much of the lengths of ORF68 and VP5 homologs is similar (Additional file 3), suggesting structural conservation of multiple domains in these proteins; the structural discontinuity of the HK97 domain helps to understand the difficulties of establishing sequence homology in these regions of herpesvirus MCPs. The conserved region 3 (Fig. 1A) corresponds to the "buttress domain", which appears to play a role in interactions between MCP subunits within the capsomer, as well as between MCP and "triplexes" that stabilize the capsomers (Yu et al., 2017). A better understanding of the role of conserved sequence elements within the buttress domain shown in Fig. 1A awaits the atomic model of this region.

Another connection to the morphogenesis module of bacteriophages is provided by HaHV-1 ORF73, which appears to belong to the herpesvirus capsid maturation protease family (assemblins), previously found to be homologous to capsid proteases from various viruses of bacteria and archaea (Liu and Mushegian, 2004 and Fig. 1B). Interestingly, the configuration of putative catalytic residues in malacoherpesvirus proteases is reminiscent of phage capsid proteases, in that the most C-terminal member of the predicted catalytic triad is an acidic residue as in phage enzymes, rather than by histidine as in assemblins of *Herpesviridae*. The C-terminal part of the ORF73 product has a reduced compositional complexity; this is similar to the scaffold protein domains, which in some other herpesviruses are likewise compositionally biased and form a fusion with the protease that self-cleaves following the capsid assembly (Newcomb et al., 2000; Veesler et al., 2014). Here again, sequences from *Alloherpesviridae* are outliers: we did not recognize any sequence homolog of assemblin among alloherpesvirid genes, including ORF78 product of *Cyprinid herpesvirus 3* and its orthologs, for which such role has been assigned based on several lines of circumstantial evidence (Michel et al., 2010; Davison et al., 2013).

Other putative structural proteins were also identified in HaHV-1. ORF14 is homologous to the herpesvirus VP23 family, and ORF81 appears to belong to the VP19C family (Table 1). A heterotrimer ("triplex") of two molecules of VP19C and VP23 stabilizes the adjacent capsomers of VP5 (Newcomb et al., 2003; Okoye et al., 2006). We detected neither the homologs of the accessory assembly factor VP3 or of the capsid outside-surface protein VP26, nor the components of the capsid portal complex that is required for letting the scaffold fragments out of the capsid and for importing virus DNA. Nonetheless, identification of sequence similarities in three structural components of the capsid and in the capsid maturation protease, together with the catalytic subunit of terminase ATPase detected earlier, suggests that the structural module of malacoherpesviruses is mostly or entirely homologous to other herpesviruses, and the ancient connections to ancestors from viruses of prokaryotes are retained in morphogenetic modules of diverse *Herpesvirales*.

The identities of tegument proteins as well as virion outer surface proteins in malacoherpesviruses remain to be determined. Interestingly, a pair of paralogous proteins (products of ORF25 and ORF102 in HaHV-1) is found in all malacoherpesviruses. Database searches demonstrated that homologs of ORF25 and ORF102 are found in many animal genomes, including diverse aquatic invertebrates and terrestrial arthropods. In addition, an initial BLASTP search seeded with ORF25 recovered a match (p-value $< 10^{-6}$) to a putative glycoprotein of *Hubei myriapoda virus 8*, an RNA virus with a negative-strand RNA genome that has been characterized in a large-scale arthropod metagenomic effort (Shi et al., 2016). Further searches suggest that these proteins are related to structural proteins from several other animal viruses with different types of genome organization. Their distant homologs include spike proteins of positive-strand RNA toroviruses and coronaviruses, as well as baculovirus F proteins (see full list of matches in Additional File 4). All these genes encode virus membrane envelope proteins that facilitate virus contact with the host receptor, virus-cell membrane fusion and virus entry into the cell (Belouzard et al., 2012; Baquero et al., 2015; Palgen et al., 2015). More distant relationships (Additional File 4) link a sub-region within these proteins to the envelope proteins of the Ty3-copia-class retrotransposons in insects (virus family *Metaviridae*), as well as their domesticated homologs of those proteins, such as the *Iris* gene product in *Drosophila* (Malik and Henikoff, 2005). Conservation of this region in integrated copies of metaviruses would explain the matches between ORF25/ORF102 family to other animal genomes.

3.3. Non-structural proteins: variations within the herpesvirus enzymatic repertoire

Previous analysis has identified a number of malacoherpesvirus genes that contained sequence signatures strongly suggestive of their role in virus genome replication (Davison et al., 2005). These include two subunits of ribonucleoside reductase (ORF22 and ORF31), putative deoxyuridine 5′-triphosphate nucleotidohydrolase ORF34, two related helicases (most likely, DNA helicases) ORF38 and ORF46, as well as family B DNA polymerase ORF92 and ATP-dependent DNA ligase ORF107.

In addition to these identifications, we predict, on the basis of sequence similarity, molecular functions of two putative nucleases, ORF50 and ORF98, and three additional helicase-related proteins, ORF26, ORF36 and ORF94. Of these, the ORF26 product appears to include only the putative non-catalytic C-terminal domain, whereas ORF36 and ORF94 contain the apparently complete sets of helicase catalytic residues in their N-terminal domains, as well as the C-terminal domain homology. In ORF94, the helicase is fused to the archaeo-eukaryotic-type primase (AEP) domain. One more protein that contains an AEP domain is ORF45; it is adjacent to a helicase, ORF46, on the virus chromosome.

Other genes with possible roles in the replication and expression of the virus genome are two small-molecule kinases ORF105 and ORF111, which could play a role in nucleotide salvage and their delivery to the replication complexes, as has been shown for several phages and viruses (Kim et al., 2005; Jeudy et al., 2006; Guevara-Hernandez et al., 2012). One gene product, ORF105, appears to belong to the herpesvirus UL92 family, which in betaherpesviruses and gammaherpesviruses is essential for the expression of the true-late transcripts encoding the virion proteins (Omoto and Mocarski, 2014). Another gene product, ORF86, is erroneously annotated in the database as a methyltransferase, but appears to be a homolog of the TnpB protein, known to occur in the IS*200/IS605* class of transposons in prokaryotes, as well as in various eukaryotic transposable elements and NCLDV genomes (Bao and Jurka, 2013); it contains an N-terminal helix-turn-helix domain and RuvC-like putative nuclease domain with inserted Zn ribbon (Bao and Jurka, 2013; Makarova and Koonin, 2015). Finally, ORF27 and ORF82 each encodes a fusion of an HUH-motif endonuclease and an ATPase-helicase domains. Fusions of those domains are common in proteins involved in replication of plasmids, transposons and small DNA viruses (Koonin et al., 2015). None of the latter three ORFs have homologs in other

malacoherpesviruses, and it is unclear whether they are involved in virus genome propagation or represent selfish genetic elements parasitizing virus DNA.

3.4. Evolutionary origin of Herpesvirales replication module: sampling from the same gene pool as NCLDV?

Gene sharing between viruses and many groups of virus-related mobile elements have been studied recently using the network analysis techniques, with virus genomes represented as nodes and genes shared between pairs of genomes modeled as weighted edges (Iranzo et al., 2016). A giant module in that network comprises *Megavirales* – a recently proposed order that joins several families of NCLDV, including *Poxviridae*, *Asfarviridae*, *Iridoviridae*, *Ascoviridae*, *Phycodnaviridae*, *Mimiviridae* and *Marseilleviridae* (Colson et al., 2013). This order is further linked to several groups of viruses with moderate-sized genomes and to cellular mobile elements. Most viruses in that super-module share subsets of structural and morphogenetic proteins, including a double jelly-roll major capsid protein, whereas the sets of genes involved in DNA replication and transcription differ between modules within the super-module.

Herpesviruses and tailed bacteriophages comprise a separate large module in the same network, held together by their own conserved assembly of morphogenetic proteins. Thus, the structure/morphogenesis modules in the established order *Herpesvirales* and proposed order *Megavirales* are, for all practical purposes, unrelated. It is not clear, however, whether the genome propagation modules of viruses in the two orders are related or not.

Among the five hallmark proteins shared by all *Megavirales,* three play essential roles in virus genome replication and expression: family B DNA polymerase, helicase-primase, and viral late transcription factor 3 (Yutin et al., 2009). The orthologs of the first two gene products are also found in all *Herpesvirales*. To understand the phylogenetic relationships between these two virus groups, we assessed the phylogenetic position of herpesvirus DNA polymerases. A multiple alignment of viral and cellular Family B polymerases was constructed, and the alignment of 364 amino acid positions was used to infer an unrooted maximum-likelihood tree, shown in Fig. 2 (the source file of the tree in Newick format is available as Additional File 5). The known monophyletic groups of eukaryotic and archaeal DNA polymerases are observed in the tree, and most of the virus taxa approved by ICTV are seen as statistically supported clades, though neither of the two orders, proposed *Megavirales* and approved *Herpesvirales,* resolved as monophyletic. Instead, both orders were split by a well-supported (bootstrap value 0.88) internal partition in the tree. On the one side of the split, DNA polymerases of *Herpesviridae* join an assembly of NCLDV families and, at a longer evolutionary distance, eukaryotic DNA polymerases delta. The other side includes DNA polymerases of *Alloherpesviridae* and *Malacoherpesviridae*, together with their poxvirus, baculovirus and bacteriophage homologs and an assembly of various eukaryotic and archaeal polymerases.

Taking such tree topology at face value, one would have to suggest a very ancient split between *Herpesviridae* and other *Herpesvirales*, as well as a similarly ancient division between poxviruses and the rest of NCLDV, with an *ad hoc* hypothesis required to explain the origin of the eukaryotic polymerases delta. However, the topology of the tree shown in Fig. 2 should be interpreted with caution; viral sequences evolve at a high rate, resulting in long branches in the trees and ensuing topological artifacts, the best-studied of which is long branch attraction (Bergsten, 2005). Indeed, most virus branches in the tree are long, and the branches leading to *Alloherpesviridae* and *Malacoherpesviridae* DNA polymerases are the longest ones; together with generally high divergence in the family, this may obscure the identity of their true nearest evolutionary neighbors. An earlier study (Yutin and Koonin, 2012) also showed that in the DNA polymerase tree *Megavirales* do not resolve as a clade unless the tree topologies are pre-constrained. More recently, another phylogeny of DNA polymerases has been presented, in which *Megavirales* and *Herpesvirales* are seen as two monophyletic lineages (Kazlauskas et al., 2016), albeit the sampling of herpesviruses was much narrower in that study than in the current case.

Recently, a comprehensive analysis of viral and cellular primases of the archaeo-eukaryal type (Prim-Pol superfamily), including virus homologs, has been published (Burroughs and Aravind, 2016). The results of the analysis of that second hallmark gene of dsDNA viruses are strikingly similar to our analysis of DNA polymerases detailed here. In particular, Prim-Pol proteins of both *Herpesvirales* and *Megavirales* resolved as multiple clades, though in that case *Alloherpesviridae* and *Herpesviridae* form a clade to the exclusion of *Malacoherpesviridae* (refer to Fig. 6 in Burroughs and Aravind, 2016, and online material accompanying their paper). All these virus clades are deep in the tree, and, whether coincidentally or not, the closest tree neighbor of *Malacoherpesviridae* is again *Poxviridae*, though their joint clade in the Prim-Pol tree is not well-supported.

Notwithstanding the concerns about tree artifacts, there appears to be a common theme in the phylogenies of the two hallmark families from the proposed order *Megavirales* and approved order *Herpesvirales.* In both cases, proteins from viruses that belong to the same order tend not to resolve as a single clade, and in both cases subsets of each order are intermingled. These results are compatible with the notion that the DNA polymerase genes in *Megavirales* as well as in *Herpesvirales* are roughly of the same evolutionary age and have been drawn from the same source – probably, from the ancestral pool of phage replication enzymes.

Unexpected tree topologies may be caused not only by the noisy data or methodological artifacts, but also by true past events, such as horizontal gene transfer (HGT) and gene loss, both of which are well-documented in viruses. Whereas no credible scenario involving HGT can be proposed for the two hallmark genes, DNA polymerase and Prim-Pol, a transfer between distant clades may cause unusual evolutionary affinities of several other malacoherpesvirus replication proteins. For example, of the two subunits of HaHV-1 ribonucleoside reductase, the large subunit ORF022 is most similar to eukaryotic homologs, and the small subunit ORF037 is more closely related to phage and bacterial homologs; and HaHV-1 ATP-dependent ligase ORF107 is most similar to the homologs from echinoderms and other invertebrates while lacking homologs in the genome of OsHV-1 (Table 1 and data not shown). Interestingly, exactly these trends, i.e., the pattern of frequent HGTs and repeated loss events, have been observed in the orthologs of those same genes in NCLDV (Yutin and Koonin, 2009, 2012).

On a separate note, a relatively recent HGT event between a malacoherpesvirus and other viruses is suggested by the presence in HaHV-1 of two copies of the uncharacterized conserved domain DUF1335, which thus far was thought to be restricted to *White spot syndrome virus*, the sole member of the dsDNA virus family *Nimaviridae* (Table 1).

3.5. Expanded host range and conserved gene core of malacoherpesvirids

In this study, we report a novel genome related of malacoherpesviruses, inadvertently sequenced by the genome project of the annelid *C. teleta.* This, as well as the earlier discovery of a nearly-complete virus-like DNA scaffold co-deposited with the genome of amphioxus *B. floridae* (Savin et al., 2010), suggest that herpesviruses may be consorting with diverse marine invertebrates. Putative herpesvirus-like isolated from other aquatic *Metazoa*, such as two species of stony corals and wild-caught *Hydra sp.*, have been reported (Vega Thurber et al., 2008; Grasis et al., 2014; Correa et al., 2016). Caution has been advised in interpreting these results (Houldcroft and Breuer, 2015; Wood-Charlson et al., 2015; Sweet and Bythell, 2016), and our reanalysis of the three publicly available short sequences from these studies annotated as fragments of anthozoan herpesvirus DNA suggests that they are in fact of bacterial and fungal provenance (Additional File 6). The existence of herpesviruses of coelenterates is nevertheless likely,

given that particles with herpesvirus-like morphology have been observed in coral *Acropora* (Correa et al., 2016). Another recent report of a virus with herpes-like virion morphology in blue king crab *Paralithodes platypus* (Ryazanova et al., 2015) appears to be corroborated by sequence similarity analysis, which suggests that the DNA fragment specifically amplified from the diseased tissue is closely related to alloherpesvirus DNA polymerase (Additional file 6).

All told, close to 50% of all HaHV-1 virus genes have orthologs in the *Ostreid herpesvirus 1* genome, and the majority of these genes also have orthologs in either annelid or amphioxus herpesviruses. Thus, at least 30 genes are conserved within the expanded *Malacoherpesviridae* family, and their sequence conservation models may serve as this family genetic signature. Two-thirds of these genes represent a variety of essential molecular functions, both structural and informational, and the remaining ones are uncharacterized.

Several gene products with the roles in capsid formation are conserved throughout the family and in *Herpesvirales*, suggesting that the sequences deposited with the two invertebrate genomes may represent non-defective, particle-forming viruses. To study this possibility further, we tested whether the herpes-like contigs found within assembled genomes of *Branchiostoma* and *Capitella* show evidence of relaxed selective constraints. We used the RELAX method (Wertheim et al., 2015), which contrasts two probabilistic evolutionary models, one in which relaxation of selection is allowed in a specific branch and the other in which relaxation is not allowed. The test then compares the log-likelihoods of the two competing models. The results of this analysis, summarized in Table 2, show no evidence of relaxed selection in the *Capitella* and *Branchiostoma* virus-like genes. This may suggest that these sequences represent bona fide viruses that were sequenced together with their host. If, on the other hand, this DNA is integrated into the host genome (the possibility supported in the case of *B. floridae* by the recent deposition in the database of shotgun sequence of the gonad-derived genomic DNA of another cephalochordate, *B. belcheri*, which also includes a complement of herpesvirus genes, including for example the major capsid protein homolog XP_019628941), the integration event must have taken place recently enough so that a change in selective constraints had no time to leave an imprint on the protein-coding sequences.

Improved detection of sequence similarity – as reflected by 33 malacoherpesvirus gene homologs in *B. floridae* detected here, as compared to only 19 reported in an earlier study (Savin et al., 2010) – provides a new lower bound of the size of the common gene complement in herpesviruses. We expect that re-sequencing of virus-like replicons associated with the amphioxus and annelid genomes will reveal more gene products shared with other herpesviruses. Based on these data and the evidence in the literature, the fraction of widely conserved genes (in the first approximation, having homologs beyond the family level) in herpesviruses may be at least 30% and perhaps closer to 50%, similarly to what has been suggested for bacteriophages with similarly-sized DNA genomes (Kristensen et al., 2011; Grazziotin et al., 2017). Given the observations of the putative herpesviruses associated with diverse marine invertebrates, the family name *Malacoherpesviridae* may be worth replacing with a more inclusive one, perhaps "*Ectoherpesviridae*".

4. Concluding remarks

Recent studies have provided important advances in our understanding of early evolutionary history of eukaryotic viruses. A general principle of virus evolution has emerged, i.e., that the hallmark genes in major groups of viruses of eukaryotes are related, sometimes distantly, to phage genes with similar functions, with additional contribution of genes acquired from mobile elements and host genomes. The genome replication gene modules in many groups of viruses are derived from different phage sources than the structural genes. Putative order *Megavirales*, all RNA viruses, and certain classes of DNA viruses with smaller genomes represent distinct variations on this general theme (Koonin et al., 2015; Iranzo et al., 2017).

Here we see the same trend in the evolution of the order *Herpesvirales,* where the morphogenesis and replication gene sets have distinct evolutionary histories. A core set of morphogenetic genes appears to be conserved in at least two families of *Herpesvirales* and is more distantly related to homologs in phages with HK97-type capsids. In contrast, several essential herpesvirus genome replication factors, such as Family B DNA polymerase and DNA Prim-Pol primase, appear to have an ancient origin and possible ancestral relationship to the hallmark genes of *Megavirales,* whose structural module is distinct from that of herpesviruses.

Taken together, this evidence indicates that the ancestors of two orders, *Herpesvirales* and putative *Megavirales*, may have recruited their genome propagation module from the same source – most likely, from the replication genes in the community of phages and prokaryotic mobile elements. These genes may have been confined to each lineage of eukaryotic viruses ever since; unlike several other enzymes with roles in DNA synthesis in the same virus orders, these essential replication genes do not appear to have been horizontally transferred between these viruses and their hosts, except for occasional integration of the whole virus DNA into the host genome.

Also at an early stage, the two emerging virus lineages have acquired distinct morphogenetic modules, i.e., the jelly-roll type in *Megavirales* and the HK97 type in *Herpesvirales*. Curiously, the capsid proteins of the former type are relatively narrowly distributed in the known phages, but are found, either in single jelly-roll or in an internally duplicated form, in eukaryotic viruses that infect nearly all major clades of eukaryotes (Krupovic and Koonin, 2017). The latter type of capsid displays the opposite trend, i.e., it is found in many groups of bacterial and archaeal viruses, but among eukaryotic viruses it has been identified in just one order infecting *Metazoa*. The reasons for these asymmetries and for the association of a specific capsid type with nuclear vs. nucleocytoplasmic replication cycle are unknown. It is quite likely, however, that better sampling of viruses from diverse hosts and from the environment will improve virus phylogeny and provide better understanding of the emergence of the two major lineages of eukaryotic large dsDNA viruses with their distinct replication strategies.

Acknowledgments

We thank Haim Ashkenazy and all members of Pupko lab for help and support, and to the two anonymous reviewers for helpful criticisms of the initial version of this manuscript. ARM is a Program Director at the US National Science Foundation (NSF); his work on this project was supported by the NSF Independent Research and Development Program, but the statements and opinions expressed herein are made in the personal capacity and do not constitute the endorsement by NSF or the government of the United States. TP is supported by Israel Science Foundation grant 802/16. ELK is a fellow of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2017.10.009.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.

Alva, V., Nam, S.Z., Söding, J., Lupas, A.N., 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res. 44 (W1), W410–W415. http://dx.doi.org/10.1093/nar/gkw348.

Baker, M.L., Jiang, W., Rixon, F.J., Chiu, W., 2005. Common ancestry of herpesviruses

and tailed DNA bacteriophages. J. Virol. 79 (23), 14967–14970. http://dx.doi.org/10.1128/JVI.79.23.14967-14970.2005.
Bao, W., Jurka, J., 2013. Homologues of bacterial TnpB_*IS605* are widespread in diverse eukaryotic transposable elements. Mob. DNA 4, 12. http://dx.doi.org/10.1186/1759-8753-4-12.
Baquero, E., Albertini, A.A., Gaudin, Y., 2015. Recent mechanistic and structural insights on class III viral fusion glycoproteins. Curr. Opin. Struct. Biol. 33, 52–60. http://dx.doi.org/10.1016/j.sbi.2015.07.011.
Belouzard, S., Millet, J.K., Licitra, B.N., Whittaker, G.R., 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 4 (6), 1011–1033. http://dx.doi.org/10.3390/v4061011.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300. http://dx.doi.org/10.2307/2346101.
Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193. http://dx.doi.org/10.1111/j.1096-0031.2005.00059.x.
Booy, F.P., Newcomb, W.W., Trus, B.L., Brown, J.C., Baker, T.S., 1991. Steven AC. Liquid-crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. Cell 64, 1007–1015. http://dx.doi.org/10.1016/0092-8674(91)90324-R.
Bourque, G., Pevzner, P.A., 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. Genome Res. 12 (1), 26–36.
Bowman, B.R., Baker, M.L., Rixon, F.J., Chiu, W., Quiocho, F.A., 2003. Structure of the herpesvirus major capsid protein. EMBO J. 22 (4), 757–765. http://dx.doi.org/10.1093/emboj/cdg086.
Burroughs, A.M., Aravind, L., 2016. RNA damage in biological conflicts and the diversity of responding RNA repair systems. Nucleic Acids Res. 44 (18), 8525–8555. http://dx.doi.org/10.1093/nar/gkw722.
Casjens, S., King, J., 1975. Virus assembly. Annu. Rev. Biochem. 44, 555–611.
Cheng, H., Shen, N., Pei, J., Grishin, N.V., Double-stranded, D.N.A., 2004. bacteriophage prohead protease is homologous to herpesvirus protease. Protein Sci. 13 (8), 2260–2269. http://dx.doi.org/10.1110/ps.04726004.
Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D.K., Cheng, X.W., Federici, B.A., Van Etten, J.L., Koonin, E.V., La Scola, B., Raout, D., 2013. "Megavirales", a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. Arch. Virol. 158, 2517–2521. http://dx.doi.org/10.1007/s00705-013-1768-6.
Correa, A.M., Ainsworth, T.D., Rosales, S.M., Thurber, A.R., Butler, C.R., Vega Thurber, R.L., 2016. Viral outbreak in corals associated with an in situ bleaching event: atypical herpes-like viruses and a new megavirus infecting symbiodinium. Front. Microbiol. 7, 127. http://dx.doi.org/10.3389/fmicb.2016.00127.
Davison, A.J., Davison, M.D., 1995. Identification of structural proteins of channel catfish virus by mass spectrometry. Virology 206, 1035–1043. http://dx.doi.org/10.1006/viro.1995.1026.
Davison, A.J., 2002. Evolution of the herpesviruses. Vet. Microbiol. 85, 69–88.
Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C., Le Deuff, R.M., Renault, T., 2005. A novel class of herpesvirus with bivalve hosts. J. Gen. Virol. 86, 41–53. http://dx.doi.org/10.1099/vir.0.80382-0.
Davison, A.J., Eberle, R., Ehlers, B., Hayward, G.S., McGeoch, D.J., Minson, A.C., Pellett, P.E., Roizman, B., Studdert, M.J., Thiry, E., 2009. The order herpesvirales. Arch. Virol. 154 (1), 171–177. http://dx.doi.org/10.1007/s00705-008-0278-4.
Davison, A.J., 2010. Herpesvirus systematics. Vet. Microbiol. 143 (1), 52–69. http://dx.doi.org/10.1016/j.vetmic.2010.02.014.
Davison, A.J., Kurobe, T., Gatherer, D., Cunningham, C., Korf, I., Fukuda, H., Hedrick, R.P., Waltzek, T.B., 2013. Comparative genomics of carp herpesviruses. J. Virol. 87 (5), 2908–2922. http://dx.doi.org/10.1128/JVI.03206-12.
Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform. 5, 113. http://dx.doi.org/10.1186/1471-2105-5-113.
Fokine, A., Rossmann, M.G., 2016. Common evolutionary origin of procapsid proteases, phage tail tubes, and tubes of bacterial Type VI secretion systems. Structure 24 (11), 1928–1935. http://dx.doi.org/10.1016/j.str.2016.08.013.
Grasis, J.A., Lachnit, T., Anton-Erxleben, F., Lim, Y.W., Schmieder, R., Fraune, S., et al., 2014. Species-specific viromes in the ancestral holobiont hydra. PLoS One 9 (10), e109952. http://dx.doi.org/10.1371/journal.pone.0109952.
Grazziotin, A.L., Koonin, E.V., Kristensen, D.M., 2017. Prokaryotic virus orthologous groups (pVOGs): a resource for comparative genomics and protein family annotation. Nucleic Acids Res. 45 (D1), D491–D498. http://dx.doi.org/10.1093/nar/gkw975.
Guevara-Hernandez, E., Garcia-Orozco, K.D., Sotelo-Mundo, R.R., 2012. Biochemical characterization of thymidine monophosphate kinase from white spot syndrome virus: a functional domain from the viral ORF454. Protein Pept. Lett. 19 (11), 1220–1224. http://dx.doi.org/10.2174/092986612803217033.
Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59 (3), 307–321. http://dx.doi.org/10.1093/sysbio/syq010.
Hannenhalli, S., Chappey, C., Koonin, E.V., Pevzner, P.A., 1995. Genome sequence comparison and scenarios for gene rearrangements: a test case. Genomics 30 (2), 299–311. http://dx.doi.org/10.1006/geno.1995.9873.
Houldcroft, C.J., Breuer, J., 2015. Tales from the crypt and coral reef: the successes and challenges of identifying new herpesviruses using metagenomics. Front. Microbiol. 6, 188. http://dx.doi.org/10.3389/fmicb.2015.00188.
Iranzo, J., Krupovic, M., Koonin, E.V., 2016. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. mBio 7 (4). http://dx.doi.org/10.1128/mBio.00978-16.
Iranzo, J., Krupovic, M., Koonin, E.V., 2017. A network perspective on the virus world. Commun. Integr. Biol. 10 (2), e1296614. http://dx.doi.org/10.1080/19420889.2017.1296614.
Iyer, L.M., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res. 117, 156–184. http://dx.doi.org/10.1016/j.virusres.2006.01.009.
Hui, W.H., Tang, Q., Liu, H., Atanasov, I., Liu, F., Zhu, H., Zhou, Z.H., 2013. Protein interactions in the murine cytomegalovirus capsid revealed by cryoEM. Protein Cell 4 (11), 833–845. http://dx.doi.org/10.1007/s13238-013-3060-7.
Jeudy, S., Claverie, J.M., Abergel, C., 2006. The nucleoside diphosphate kinase from mimivirus: a peculiar affinity for deoxypyrimidine nucleotides. J. Bioenergy Biomembr. 38 (3–4), 247–254. http://dx.doi.org/10.1007/s10863-006-9045-y.
Kazlauskas, D., Krupovic, M., Venclovas, Č., 2016. The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. Nucleic Acids Res. 44 (10), 4551–4564. http://dx.doi.org/10.1093/nar/gkw322.
Kim, J., Wheeler, L.J., Shen, R., Mathews, C.K., 2005. Protein-DNA interactions in the T4 dNTP synthetase complex dependent on gene 32 single-stranded DNA-binding protein. Mol. Microbiol. 55 (5), 1502–1514. http://dx.doi.org/10.1111/j.1365-2958.2004.04486.x.
Koonin, E.V., Yutin, N., 2010. Origin and evolution of eukaryotic large nucleo-cytoplasmic DNA viruses. Intervirology 53 (5), 284–292. http://dx.doi.org/10.1159/000312913.
Koonin, E.V., Dolja, V.V., Krupovic, M., 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. Virology 479–480, 2–25. http://dx.doi.org/10.1016/j.virol.2015.02.039.
Kristensen, D.M., Cai, X., Mushegian, A., 2011. Evolutionarily conserved orthologous families in phages are relatively rare in their prokaryotic hosts. J. Bacteriol. 193 (8), 1806–1814. http://dx.doi.org/10.1128/JB.01311-10.
Krupovic, M., Koonin, E.V., 2017. Multiple origins of viral capsid proteins from cellular ancestors. Proc. Natl. Acad. Sci. USA. http://dx.doi.org/10.1073/pnas.1621061114.
Larget, B., Kadane, J.B., Simon, D.L., 2005. A Bayesian approach to the estimation of ancestral genome arrangements. Mol. Phylogenet. Evol. 36 (2), 214–223. http://dx.doi.org/10.1016/j.ympev.2005.03.026.
Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL)v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44 (W1). http://dx.doi.org/10.1093/nar/gkw290.
Liu, J., Mushegian, A., 2004. Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. J. Bacteriol. 186 (13), 4369–4375. http://dx.doi.org/10.1128/JB.186.13.4369-4375.2004.
Löytynoja, A., 2014. Phylogeny-aware alignment with PRANK. Methods Mol. Biol. 1079, 155–170. http://dx.doi.org/10.1007/978-1-62703-646-7_10.
Makarova, K.S., Koonin, E.V., 2015. Annotation and classification of CRISPR-Cas systems. Methods Mol. Biol. 1311, 47–75. http://dx.doi.org/10.1007/978-1-4939-2687-9_4.
Malik, H.S., Henikoff, S., 2005. Positive selection of Iris, a retroviral envelope-derived host gene in Drosophila melanogaster. PLoS Genet. 1 (4), e44. http://dx.doi.org/10.1371/journal.pgen.0010044.
Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Geer, L.Y., Bryant, S.H., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45 (D1), D200–D203. http://dx.doi.org/10.1093/nar/gkw1129.
Mettenleiter, T.C., Klupp, B.G., Granzow, H., 2009. Herpesvirus assembly: an update. Virus Res. 143 (2), 222–234. http://dx.doi.org/10.1016/j.virusres.2009.03.018.
Michel, B., Leroy, B., Stalin Raj, V., Lieffrig, F., Mast, J., Wattiez, R., Vanderplasschen, A.F., Costes, B., 2010. The genome of cyprinid herpesvirus 3 encodes 40 proteins incorporated in mature virions. J. Gen. Virol. 91 (Pt 2), 452–462. http://dx.doi.org/10.1099/vir.0.015198-0.
Mitchell, M.S., Matsuzaki, S., Imai, S., Rao, V.B., 2002. Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases. Nucleic Acids Res. 30 (18), 4009–4021. http://dx.doi.org/10.1093/nar/gkf524.
Newcomb, W.W., Thomsen, D.R., Homa, F.L., Brown, J.C., 2003. Assembly of the herpes simplex virus capsid: identification of soluble scaffold-portal complexes and their role in formation of portal-containing capsids. J. Virol. 77 (18), 9862–9871. http://dx.doi.org/10.1128/JVI.77.18.9862-9871.2003.
Newcomb, W.W., Trus, B.L., Cheng, N., Steven, A.C., Sheaffer, A.K., Tenney, D.J., Weller, S.K., Brown, J.C., 2000. Isolation of herpes simplex virus procapsids from cells infected with a protease-deficient mutant virus. J. Virol. 74 (4), 1663–1673.
Okoye, M.E., Sexton, G.L., Huang, E., McCaffery, J.M., Desai, P., 2006. Functional analysis of the triplex proteins (VP19C and VP23) of herpes simplex virus type 1. J. Virol. 80 (2), 929–940. http://dx.doi.org/10.1128/JVI.80.2.929-940.2006.
Omoto, S., Mocarski, E.S., 2014. Transcription of true late (γ2) cytomegalovirus genes requires UL92 function that is conserved among beta- and gammaherpesviruses. J. Virol. 88 (1), 120–130. http://dx.doi.org/10.1128/JVI.02983-13.
Palgen, J.L., Jurgens, E.M., Moscona, A., Porotto, M., Palermo, L.M., 2015. Unity in diversity: shared mechanism of entry among paramyxoviruses. Prog. Mol. Biol. Transl. Sci. 129, 1–32. http://dx.doi.org/10.1016/bs.pmbts.2014.10.001.
Pei, J., Grishin, N.V., 2014. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol. Biol. 1079, 263–271. http://dx.doi.org/10.1007/978-1-62703-646-7_17.
Pellett, P.E., Davison, A.J., Eberle, R., Ehlers, B., Hayward, G.S., Lacoste, V., Minson, A.C., Nicholas, J., Roizman, B., Studdert, M.J., Wang, F. Order: *Herpesvirales.* Chapter Version: ICTV Ninth Report; 2009 Taxonomy Release. 2011. ⟨https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/89/herpesvirales⟩ (Accessed online March 5 2017).
Ryazanova, T.V., Eliseikina, M.G., Kalabekov, I.M., Odintsova, N.A., 2015. A herpes-like virus in king crabs: characterization and transmission under laboratory conditions. J.

Invertebr. Pathol. 127, 21–31. http://dx.doi.org/10.1016/j.jip.2015.02.003.
Savin, K.W., Cocks, B.G., Wong, F., Sawbridge, T., Cogan, N., Savage, D., Warner, S., 2010. A neurotropic herpesvirus infecting the gastropod, abalone, shares ancestry with oyster herpesvirus and a herpesvirus associated with the amphioxus genome. Virol. J. 7, 308. http://dx.doi.org/10.1186/1743-422X-7-308.
Shi, M., Lin, X.D., Tian, J.H., Chen, L.J., Chen, X., Li, C.X., Qin, X.C., Li, J., Cao, J.P., Eden, J.S., Buchmann, J., Wang, W., Xu, J., Holmes, E.C., Zhang, Y.Z., 2016. Redefining the invertebrate RNA virosphere. Nature. http://dx.doi.org/10.1038/nature20167.
Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244–W248. http://dx.doi.org/10.1093/nar/gki408.
Sweet, M., Bythell, J., 2016. The role of viruses in coral health and disease. J. Invertebr. Pathol. http://dx.doi.org/10.1016/j.jip.2016.12.005.
Van Beurden, S.J., Leroy, B., Wattiez, R., et al., 2011. Identification and localization of the structural proteins of anguillid herpesvirus 1. Vet. Res. 42, 105. http://dx.doi.org/10.1186/1297-9716-42-105.
van Beurden, S.J., Engelsma, M.Y., 2012. Herpesviruses of Fish, Amphibians and Invertebrates. InTech, Vancouver, pp. 217–242.
Vega Thurber, R.L., Barott, K.L., Hall, D., Liu, H., Rodriguez-Mueller, B., Desnues, C., Edwards, R.A., Haynes, M., Angly, F.E., Wegley, L., Rohwer, F.L., 2008. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral Porites compressa. Proc. Natl. Acad. Sci. USA 105 (47), 18413–18418. http://dx.doi.org/10.1073/pnas.0808985105.
Veesler, D., Khayat, R., Krishnamurthy, S., Snijder, J., Huang, R.K., Heck, A.J., Anand, G.S., Johnson, J.E., 2014. Architecture of a dsDNA viral capsid in complex with its maturation protease. Structure 22 (2), 230–237. http://dx.doi.org/10.1016/j.str.2013.11.007.
Veesler, D., Johnson, J.E., 2012. Virus maturation. Annu. Rev. Biophys. 41, 473–496. http://dx.doi.org/10.1146/annurev-biophys-042910-155407.
Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., Scheffler, K., 2015. RELAX: detecting relaxed selection in a phylogenetic framework. Mol. Biol. Evol. 32, 820–832.
Wood-Charlson, E.M., Weynberg, K.D., Suttle, C.A., Roux, S., van Oppen, M.J.H., 2015. Metagenomic characterization of viral communities in corals: mining biological signal from methodological noise. Environ. Microbiol. 17, 3440–3449. http://dx.doi.org/10.1111/1462-2920.12803.
Yu, X., Jih, J., Jiang, J., Zhou, Z.H., 2017. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer ofpp150. Science 356 (6345). http://dx.doi.org/10.1126/science.aam6892.
Yutin, N., Wolf, Y.I., Raoult, D., Koonin, E.V., 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. 6, 223. http://dx.doi.org/10.1186/1743-422X-6-223.
Yutin, N., Koonin, E.V., 2009. Evolution of DNA ligases of nucleo-cytoplasmic large DNA viruses of eukaryotes: a case of hidden complexity. Biol. Direct 4, 51. http://dx.doi.org/10.1186/1745-6150-4-51.
Yutin, N., Koonin, E.V., 2012. Hidden evolutionary complexity of nucleo-cytoplasmic Large DNA viruses of eukaryotes. Virol. J. 9, 161. http://dx.doi.org/10.1186/1743-422X-9-161.
Zhou, Z.H., Dougherty, M., Jakana, J., He, J., Rixon, F.J., Chiu, W., 2000. Seeing the herpesvirus capsid at 8.5 Å. Science 288 (5467), 877–880. http://dx.doi.org/10.1126/science.288.5467.877.