Long-read transcriptomics of Ostreid herpesvirus 1 uncovers a conserved expression strategy for the capsid maturation module and pinpoints a mechanism for evasion of the ADAR-based antiviral defence

Umberto Rosani, Enrico Bortoletto, Xiang Zhang, Bo-Wen Huang, Lu-Sheng Xin, Mart Krupovic, Chang-Ming Bai

Virus Evolution, 2024, **10(1)**, veae088
DOI: https://doi.org/10.1093/ve/veae088
Advance Access Publication 17 October 2024
Research Article

Long-read transcriptomics of Ostreid herpesvirus 1 uncovers a conserved expression strategy for the capsid maturation module and pinpoints a mechanism for evasion of the ADAR-based antiviral defence

Umberto Rosani[2,*], Enrico Bortoletto[2], Xiang Zhang[1], Bo-Wen Huang[1], Lu-Sheng Xin[1], Mart Krupovic[3], Chang-Ming Bai[1,4,*]

[1]State Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Key Laboratory of Maricultural Organism Disease Control, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, 106 Nanjing Rd, Qingdao 266071, China
[2]Department of Biology, University of Padova, Via U. Bassi, 58/B, Padova 35121, Italy
[3]Institut Pasteur, Université Paris Cité, CNRS UMR6047, Archaeal Virology Unit, 25 rue du Dr. Roux, Paris 75015, France
[4]Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, 168 Wenhai Rd, Qingdao 266237, China

*Corresponding authors. Department of Biology, University of Padova, Via U. Bassi, 58/B, Padova 35121, Italy. E-mail: umberto.rosani@unipd.it; State Key Laboratory of Mariculture Biobreeding and Sustainable Goods, Yellow Sea Fisheries Research Institute, CAFS, 106 Nanjing Rd, Qingdao 266071, China. Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, 168 Wenhai Rd, Qingdao 266237, China. E-mail: baicm@ysfri.ac.cn.

Abstract

Ostreid herpesvirus 1 (OsHV-1), a member of the family *Malacoherpesviridae* (order *Herpesvirales*), is a major pathogen of bivalves. However, the molecular details of the malacoherpesvirus infection cycle and its overall similarity to the replication of mammalian herpesviruses (family *Orthoherpesviridae*) remain obscure. Here, to gain insights into the OsHV-1 biology, we performed long-read sequencing of infected blood clams, *Anadara broughtonii*, which yielded over one million OsHV-1 long reads. These data enabled the annotation of the viral genome with 78 gene units and 274 transcripts, of which 67 were polycistronic mRNAs, 35 ncRNAs, and 20 natural antisense transcripts (NATs). Transcriptomics and proteomics data indicate preferential transcription and independent translation of the capsid scaffold protein as an OsHV-1 capsid maturation protease isoform. The conservation of this transcriptional architecture across *Herpesvirales* likely indicates its functional importance and ancient origin. Moreover, we traced RNA editing events using short-read sequencing and supported the presence of inosine nucleotides in native OsHV-1 RNA, consistent with the activity of adenosine deaminase acting on dsRNA 1 (ADAR1). Our data suggest that, whereas RNA hyper-editing is concentrated in specific regions of the OsHV-1 genome, single-nucleotide editing is more dispersed along the OsHV-1 transcripts. In conclusion, we reveal the existence of conserved pan-*Herpesvirales* transcriptomic architecture of the capsid maturation module and uncover a transcription-based viral counter defence mechanism, which presumably facilitates the evasion of the host ADAR antiviral system.

Keywords: OsHV-1; Nanopore DRS; Malacoherpesviruses; capsid maturation protease; inosine; RNA editing; ADAR

Introduction

Ostreid herpesvirus 1 (OsHV-1, family *Malacoherpesviridae*) is one of the two described herpesviruses infecting invertebrates, including scallop, clam, and oyster species (Davison et al. 2005). OsHV-1 is a major pathogen of bivalves, threatening aquaculture production due to its environmental persistence and multi-host infection (Whittington et al. 2018). Massive mortalities of blood clam *Anadara broughtonii* (previously known as *Scapharca broughtonii*) associated with OsHV-1 have been reported in China since 2012 (Xia et al. 2015), whereas the OsHV-1 micro-variant has been linked to mortalities in spat and juvenile *Crassostrea gigas* reported in France since 2008, with devasting economic and social impacts (Jenkins et al. 2013, Whittington et al. 2018). The recent report of malacoherpes-like viruses circulating in coastal waters poses concerns for possible future viral outbreaks and calls for better understanding of virus–host interactions in the megadiverse group of molluscs (Rosani et al. 2023). Thus, a better understanding of the molecular details of the virus replication cycles, including the interplay between antiviral defence and counter defence mechanisms (Jacobson et al. 2021, Prasad and Prasad 2021), is of utmost importance. One way to achieve this goal is by integrating different state-of-the-art "omics" approaches (Depledge et al. 2019, Neogi et al. 2022), as demonstrated for SARS-CoV-2 (Stephenson et al.

2021), human immunodeficiency virus type 1 (Ivanov et al. 2020), and Crimean–Congo haemorrhagic fever virus (Neogi et al. 2022).

The OsHV-1 infection process has been investigated in *C. gigas* and *A. broughtonii* using classical and high-throughput sequencing approaches (Corporeau et al. 2014, Green et al. 2015, Rosani et al. 2015, Bai et al. 2018, Martenot et al. 2019, Delisle et al. 2020). Soon after the first mortality episode reported in France, the complete OsHV-1 genome was sequenced and used to substantiate its taxonomic classification among herpesviruses; however, the high divergence from all known members of the order *Herpesvirales* suggested that it belongs to a new family, subsequently named *Malacoherpesviridae*. The family was expanded to include a second member, Haliotid herpesvirus 1 (HaHV-1), which infects gastropod species (Davison et al. 2005, 2009). The evolutionary history of *Malacoherpesviridae* and their relationship to vertebrate herpesviruses remain obscure due to high-sequence divergence and paucity of mechanistic studies on their infection cycle (Iranzo et al. 2016, Mushegian et al. 2018). A low-coverage PacBio transcriptome map of OsHV-1 and HaHV-1 revealed an unexpected complexity of viral gene arrays (Bai et al. 2021), suggesting that part of the viral transcription program has been previously overlooked. The RNA of both OsHV-1 and HaHV-1 resulted heavily impacted by the A-to-G variations, compatible with the activity of enzymes of the adenosine deaminase acting on dsRNA (ADAR) family (Bai et al. 2021). These enzymes can deaminate the adenosines to inosines and, as a result, restrict the replication of diverse RNA and DNA viruses, including mammalian herpesviruses (Pfaller et al. 2021). The phylogenetic conservation of ADAR homologs among bivalves and the impact of this system on the RNA editing during OsHV-1 infection in the oyster *C. gigas* has been described recently (Rosani et al. 2019). The conservation of RNA editing across metazoans suggested that this activity is important for animal's physiology (Bass 2002, Keegan et al. 2005, Grice and Degnan 2015). Notably, in humans, the alterations in the ADAR system can result in serious diseases (Bajad et al. 2017). However, given that the ADAR activity can be expected to exert either pro- or antiviral effects depending on the host–virus combination (Pfaller et al. 2021), its functional role during malacoherpesviruses infection remains to be elucidated.

Advances in long-read RNA sequencing are now providing the technological platform to resolve the expression of complex gene arrays, typical of the densely coding genomes of some viruses (Boldogkői et al. 2019). Additionally, the possibility to sequence native RNA directly (direct RNA sequencing, DRS) opens new avenues for epitranscriptomic analysis in virology (Tan et al. 2018, Imam et al. 2020).

Aiming to gain insights into the OsHV-1 biology, we constructed six native RNA libraries and sequenced them at high coverage to characterize the OsHV-1 transcriptional program. The long-read data were coupled with short-read and proteomic data to (i) reveal the complexity of the OsHV-1 transcriptional program, including alternative transcription start and stop sites, read-through transcription events and natural antisense transcripts (NATs), (ii) identify transcriptional architecture with possible functional roles for viral replication and/or interaction with the host, and (iii) validate the extent and distribution of RNA editing mediated by ADAR.

Materials and methods

Experimental infection of *A. broughtonii* with OsHV-1

The blood clam (*A. broughtonii*) specimens were collected from a wild population (size range: 56.48–68.74 mm) in the Rizao sea area, Shandong Province, China. The clams were acclimated for a period of 24 days in 60 l aquaria tanks (about 30 clams per tank) supplied with aerated, sand-filtered seawater maintained at a temperature of 15.1–15.8°C. The blood clams were subjected to a complete replacement of seawater and were provided with a diet consisting of a mixture of microalgae, including species of *Isochrysis* and *Chlorella*. Subsequently, 30 animals were randomly selected and tested negative for OsHV-1 by qPCR, as previously described by Bai et al. (2019a). The relevant experiments were approved by the local animal care and use committee and conducted in accordance with the government regulations.

A total of 120 blood clams were randomly divided into two groups: an infected group (90 animals) and a control group (30 animals). For each group, the clams were randomly assigned to three tanks (each containing 30 and 10 animals for the challenged and control groups) with a volume of 2 l of water per individual, which corresponded to three replicates. A viral inoculum was prepared from an infected blood clam collected in a local hatchery, as previously described by Bai et al. (2018) and 100 μl (adjusted at 10^4 copies of viral DNA/μl) of the homogenate was injected into the foot of the clams, with paired controls injected with seawater. The infection trial was followed for 72 h, with physiological conditions monitored every 6–12 h. One clam was sampled from each challenged tank at 6, 12, 24, 36, 48, 60, and 72 h post-injection (hpi), which was used for hemolymph collection and DNA, RNA and protein extraction as described further. At approximately 36 hpi, the water of the challenged group began to form froth, which gradually increased in intensity over the following days. This was a result of the excessive secretion of hemolymph and mucus. At 60 hpi, animals displaying typical disease signs, such as opening valves and gill erosion, were identified. During the monitoring period, no mortality was recorded, although nearly half of the remaining clams exhibited clinical signs at 72 hpi. No clinical signs or bubbly water were observed in the control group.

RNA extraction and high-throughput sequencing

Clam hemolymph was collected from the adductor muscle sinus of three specimens per time point using a 23 G needle attached to a 5 m syringe. The hemolymph samples were centrifuged at 800 rpm for 5 min at 4°C and, after removing the supernatants, 1 m of Trizol (Invitrogen, Carlsbad, CA, USA) was added and samples were stored at −80°C for RNA and DNA extraction. DNA extraction was performed using a TIANamp™ Marine Animals DNA Kit (Tiangen Biotech, Beijing, China), according to the manufacturer's protocol and used to quantify the OsHV-1 DNA load in the collected samples by quantitative PCR (qPCR), as described previously (Bai et al. 2019a). Total RNA was extracted using Trizol reagent, according to the manufacturer's protocol, and the RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). For short-read libraries, 1 μg of total RNA was used, rRNA was removed using Ribo-Zero (Illumina, San Diego, CA, USA) and the remaining RNA was fragmented and reverse transcribed into cDNA with random primers. Second-strand cDNA were synthesized, purified with QiaQuick PCR extraction kit (Qiagen, Venlo, The Netherlands), end repaired, poly(A) added, ligated to Illumina sequencing adapters, and sequenced using an Illumina HiSeq 4000 with PE150 read layout. According to OsHV-1 load measurements, the three biological replicates collected at 60 and 72 hpi were also selected for Nanopore direct RNA sequencing. In detail, 800 ng of total RNA was used to produce six libraries (SQK-RNA001, Oxford Nanopore Technologies), according to the manufacturer's instructions. The poly(T) adapter was

ligated to the mRNA using the T4 DNA ligase (New England Biolabs) in the Quick Ligase reaction buffer (New England Biolabs) for 15 min at room temperature. First-strand cDNA was synthesized by SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) using the oligo(dT) adapter and RNA–cDNA hybrids were purified using Agencourt RNAClean XP magnetic beads (Beckman Coulter). The sequencing adapter was ligated to the mRNA using the T4 DNA ligase (New England Biolabs) in the Quick Ligase reaction buffer (New England Biolabs) for 15 min at room temperature, followed by a second purification step using Agencourt beads. Nanopore sequencing of the six libraries was carried out on 6 GridIon flow-cells, to generate approximately 5 million (M) reads per sample.

Preliminary analysis of high throughput sequencing reads

Nanopore raw electric signals were converted into nucleotides using *guppy v6.2.1* in high confidence mode, merging the resulting fastq files per sample. Critical parameters, such as read length and average quality of the readouts, were evaluated using *Nanoplot v1.40.2*. Nanopore reads were aligned against the OsHV-1 (ID: MF509813) and *A. broughtonii* (Bai et al. 2019b) reference genomes separately, using *minimap2 v2.22* (Li and Birol 2018) with the following parameters, -ax splice -k14 -uf -secondary = no. The number of mapped reads was counted with *samtools* (Li et al. 2009) to evaluate the percentages of on- and off-target reads for each reference. Raw short reads were quality-trimmed using *fastp* (Chen et al. 2018) and mapped on the OsHV-1 and ark clam reference genomes using the CLC Genomic Workbench (Qiagen) mapping tool applying 0.8 of both similarity and mapped read fraction parameters. The read mapping files were also exploited to detect gene expression levels, counted as transcript per millions (TPMs), and to perform a preliminary analysis of single-nucleotide variations (SNVs) after removing duplicated mapped reads. SNVs were called applying a cutoff of 1% of minimal frequency, a minimal count of 10 reads, and a minimal coverage of 100 reads. Quality thresholds of PHRED30 for the variable position and of PHRED25 for 5-nt window around the SNV were used. A-to-G and T-to-C SNPs were considered ADAR-compatible counted per reference genome.

Annotation of the OsHV-1 genome with transcripts

The uncorrected DRS reads were initially used to annotate the OsHV-1 genome with transcripts. In detail, *samtools* was used to extract the reads mapped on each OsHV-1 genome strand and separate coverage graphs were generated with *BEDtools* (Quinlan and Hall 2010). Then, only the 3′ and 5′ mapping boundaries of each read were used to identify transcription start sites (TSS) and transcription termination sites (TTS) using the HOMER *findPeaks* module (Heinz et al. 2010), adjusting -localSize and -size parameters to 50–300 and 20–100 for TSS and TTS, respectively. TSS and TTS occurrences were counted and used to generate a 'theoretical transcriptome' annotation file, by combining TSS and TTS found on each strand and populating these intervals with all the possible pseudo-transcripts in a length range of 250–15,000 bp. These pseudo-transcripts were used as reference for mapping DRS read. In detail, the DRS reads not mapped on the clam genome were error-corrected using *RATTLE v1.0* (de la Rubia et al. 2022) and clustered into isoforms The corrected reads per sample were mapped on the theoretical OsHV-1 pseudo-transcripts with *minimap2* using the following parameters: -ax map-ont -p 0 -N 10. *Samtools* was used to sort and index the alignments in bam format, which were then used as input for *NanoCount* (Gleeson et al. 2022). *NanoCount* was run with the -d 50 and -u 50 flags, in order to generate transcript abundance counts considering only reads mapped within 50 bp from TSS and TTS sites. The subset of pseudo-transcripts supported by at least two full-length reads was annotated on the OsHV-1 genome as "mRNA" and used for expression analysis, which was carried out with *NanoCount*, as described earlier. *Prodigal v2.6.3* (Hyatt et al. 2010) was used to predict the ORFs encoded on the validated pseudo-transcripts and generate an OsHV-1 proteome, which was then compared with the available OsHV-1 proteome by reciprocal *blastp* searches. The predicted transcripts were compacted into genes using the *flattenGTF* script (Liao et al. 2019) and OsHV-1 genome regions with overlapping antisense transcripts were identified with *bedtools intersect* and annotated as "dsRNA." Short-read Illumina datasets were mapped on the OsHV-1 genome annotated with transcripts using the CLC mapped with 0.8 and 0.8 of similarity and length fraction.

Analysis of ADAR-mediated hyper-editing

The six Illumina datasets paired to the Nanopore DRS libraries were used to investigate the extension of ADAR-mediated hyper editing impacting OsHV-1. The *hyperediting* tool (Porath et al. 2014) was applied after minimal modifications implemented to overcome software incompatibilities of the original version. The parameters were adapted applying: 5 for Minimum of edited sites at Ultra-Edit read (%); 60 for Minimum fraction of edit sites/mismatched sites (%); 25 for Minimum sequence quality for counting editing event (PHRED); 60 for Maximum fraction of same letter in cluster (%); 20 Minimum of cluster length (%); and imposing that the hyper-editing clusters should not be completely included in the first or last 20% of the read. The obtained reads, representing putatively hyper-edited reads, were realigned to the reference genome using the CLC mapper with 0.8 and 0.8 of similarity and length fraction (Qiagen, USA). The editing sites were also detected through SNP calling using the SPRINT software (Zhang et al. 2017) with custom parameters. The first six bases of each read were trimmed to reduce potential mapping errors. The reads were then aligned with *bwa* v.0.7.12 (Li and Durbin 2009). Paired-end reads were mapped separately with the command options of "bwa aln fastqfile" and "bwa samse -n4," as suggested (Ramaswami et al. 2012). *Samtools v.1.2* was used for all the format conversions (Li et al. 2009), and *picard-tools* v.1.119 was then used to remove PCR duplicates in the sorted BAM files with the command option of "MarkDuplicates.jar REMOVE_DUPLICATES = true" (http://broadinstitute.github.io/picard). The reads with high mapping quality (≥20) were considered as mapped reads and the resulting editing sites were filtered out if they possessed less than 5 supporting reads and a frequency lower than 1%. The reads with high mapping quality (≥20) were considered as mapped reads and the resulting editing sites were kept only if they possessed both 5 supporting reads and a frequency higher than 1%.

The detection of "inosine-like" modifications in the DRS data was performed with the *DeepEdit* tool (Chen et al. 2023), with a few modifications described further. The ids of the DRS reads mapped on to the OsHV-1 genome were extracted with samtools from the sam file and used to extract the corresponding data from the fast5 files only using the *ont_fast5_api* (https://github.com/nanoporetech/ont_fast5_api). The fast5 subset was then converted to single fast5 reads files with *multi_to_single_fast5* using standard parameters and used for genome mapping in order to

assign the raw electric signals to the final DRS reads using the *tombo re-squiggle* algorithm (Stoiber et al. 2017). Finally, the alignment file in sam format was processed using the *DeepEdit* tool (Chen et al. 2023) in order to obtain a list of validate RNA editing sites with their editing frequencies as well as the positions of inosine-like bases along the DRS reads. A SNV was considered validated if covered by at least four DRS reads gathering the inosine base and showing at least 2-fold more modified reads on the predicted editing strand.

Proteomic data production and analysis

Proteins were extracted from the samples collected at 6, 24, 48, and 72 hpi. Proteins were extracted using the chloroform–isopropanol method and the final precipitate was redissolved with 8 M urea (containing 1% protease inhibitor), and protein concentration was determined by using the BCA kit. Trypsin digestion was performed on an equal amount of each sample protein with a ratio of 1:50 (protease: protein, m/m) overnight. Dithiothreitol was added to a final concentration of 5 mM and reduced at 56°C for 30 min, after which iodoacetamide was added to a final concentration of 11 mM, and incubated for 15 min at room temperature. High-performance liquid chromatography (HPLC) classification was performed by grading the peptides by high pH reversed-phase HPLC on an Agilent 300Extend C18 column (5 μm particle size, 4.6 mm inner diameter, 250 mm length) in 8%–32% gradient of acetonitrile, pH 9. The peptides were then dissolved in liquid chromatography mobile phase A and then separated using an EASY-nLC 1000 ultra-HPLC (UHPLC) system. The peptide fragments were separated on the UHPLC system and then injected into the NSI Ion Source for ionization and then analyzed by QE mass spectrometry. The ion source voltage was set at 2.2 kV, and the peptide parent ion and its secondary fragments were detected and analyzed using a high-resolution Orbitrap. The primary mass spectrometry scanning range was set from 400 to 1500 m/z with a scanning resolution of 70,000.00, while the secondary mass spectrometry scanning range was set at a fixed starting point of 100 m/z and the secondary scanning resolution was set at 17,500.00. The data acquisition mode used a data-dependent scanning procedure.

Secondary mass spectrometry data were searched using Maxquant 1.5.2.8. The search parameters were set as follows: the database was composed by the blood clam and OsHV-1 proteomes (24,172+ 132 sequences), an inverse library was added to calculate the false discovery rate (FDR) caused by random matches, and common contaminant libraries were added to the database for eliminating the influence of contaminant proteins in the identification results; the digestion method was set as Trypsin/P; the number of missed cut sites was set as 2; the minimum length of peptide was set as 7 amino acid residues; the maximum number of peptide modifications was set to 5; the mass error tolerance of primary parent ions for First search and Main search was set to 10.0 ppm and 5 ppm, respectively, and that of secondary fragment ions was set to 0.02 Da. Carbamidomethyl(C), a cysteine alkylation, was set as a fixed modification. The quantification method was set to TMT-6plex, and the FDR for protein identification and PSM identification were set to 1%. A total of 798,016 secondary spectra were obtained by mass spectrometry analysis. After searching the protein theoretical data, the number of available effective spectra was 140,218, with utilization rate of 17.6%. A total of 72,876 peptides were identified by spectral analysis, including 71,478 specific peptides originating from 5083 blood clam and 41 OsHV-1 proteins.

Table 1. Summary of the performed analyses.

Sequencing platform	Sequencing strategy	No. of samples	No. of total reads	No. of OsHV-1 reads
Nanopore	DRS (native RNA)	6 (3 at 60 hpi; 3 at 72 hpi)	32.1 M	1 M
Illumina	total RNA (cDNA)	21 (6, 12, 24, 36, 48, 60, 72 hpi)	2,258 M	25.1 M
HPLC-MS	Trypsin digestions	12 (6, 24, 48, 72 hpi)	798 k	147

Database searches

For the annotation of OsHV-1 proteins, similarity searches in databases have been performed as follows. Standard and iterative blastp searches (PSI-BLAST) against the NCBI nr clustered database were performed using 10^{-10} as cutoff e-value to consider the hits for the subsequent iterations. To detect distant homology, the *HHblits* v3.3.0 tool (Remmert et al. 2012) was used in combination with the pdb70 database. Functional annotations have been retrieved using *InterProScan* v.5.57-90.0 (Jones et al. 2014).

Data rendering and statistical analysis

OsHV-1 genomic features, annotations and coverage levels were inspected and rendered using the GVIZ package (Hahne and Ivanek 2016) implemented in *R v4.2.1* (Giorgi et al. 2022). All the plots and statistical analyses have been performed using the *tidyverse, ggplot2, ggpubr*, and *smplot2* packages of *R* and with CLC Genomics Workbench (Qiagen).

Results

An OsHV-1 homogenate was injected in a batch of wild *A. broughtonii* specimens, and hemocyte samples were collected from three animals at 6, 12, 24, 36, 48, 60, and 72 hpi to extract RNA and proteins. Short-read data (described in Bai et al. 2021) showed that OsHV-1 RNA reads are barely detectable until 24–36 hpi, slightly increase in number at 48 hpi, and become prominent, occupying a considerable part of the sequenced transcriptome at 60 and 72 hpi before the appearance of mortality signs (Supplementary Fig. S1). Accordingly, to ensure the abundance of viral reads in the DRS data, we selected three biological replicates at 60 and 72 hpi and produced more than 30 million long reads in total (Table 1). Majority of them (92.6–95.6%) are "on-target," i.e. could be mapped to either the Ark clam or OsHV-1 genomes, whereas "off-target" reads were possibly produced from other biotic components, sequencing artifacts or from the transcription of host genes not present in the reference genome (Fig. 1a). DRS data included fewer off-target reads compared to the paired Illumina datasets, possibly because of the enrichment for polyadenylated transcripts, a procedure not applied for Illumina total RNA sequencing. An average of 2.2 and 4.2% of the DRS reads at 60 and 72 hpi, respectively, mapped to the OsHV-1 genome (Fig. 1b), showing a good correlation with Illumina data ($R^2 = 0.83$, Fig. 1c). Mass spectrometry (MS) analysis performed on protein extracts collected throughout the course of infection generated a total of 798,086 spectra, of which 0.09% (147 unique peptides) mapped to the predicted OsHV-1 proteome.

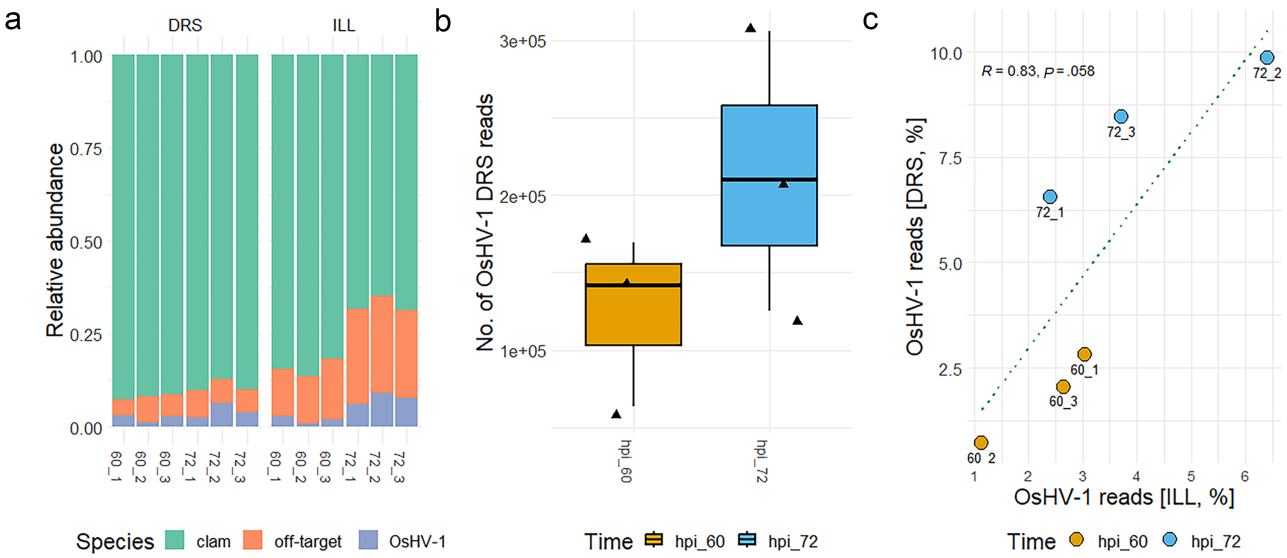

Figure 1. Sequencing of the OsHV-1 transcriptome using different platforms. (a) Relative abundance of the on-target and off-target reads in the six DRS and Illumina (ILL) datasets. (b) Number of DRS reads mapped on the OsHV-1 genome at 60 and 72 hpi. (c) Distribution of the fraction of Illumina and DRS reads mapped on the OsHV-1 genome in the six samples. The Spearman correlation value and associated *P*-value are reported on the graph.

Long-read transcriptomics of the OsHV-1 infection

The sequenced transcriptome of OsHV-1 included a total of 1,015,000 DRS reads. The detection of sudden coverage changes (increase or decrease) on each strand of the viral genome was used to identify 292 TSS and 201 TTS (Fig. 2a). These TSSs and TTSs were *in silico* combined per strand to obtain 2274 pseudo-transcripts, with a length range of 250–15,000 nt. Pseudo-transcripts were used as mapping references for the DRS data and 274 of them were supported by at least three full-length DRS reads and were considered as bonafide transcripts (Fig. 2a and Supplementary File S1–S2).

Thirty-five of them did not contain ORFs and were classified as noncoding RNAs (ncRNAs), with a size ranging from 0.26 to 3.6 kb (Fig. 2b and c). Additionally, 67 transcripts were identified as polycistronic, encoding between two and five ORFs each. By collapsing overlapping transcripts for each genome strand, 78 gene units were defined, characterized by longer sizes compared to transcripts or coding regions (Fig. 2b). A total of 328 proteins were predicted to be encoded by the 274 OsHV-1 transcripts, reduced to 132 after removing identical hits originating from gene isoforms. Our analysis confirmed the expression of 113 out of the 131 ORFs previously annotated in the reference OsHV-1 genome (Supplementary Table S1). The genes for which transcripts were not obtained include ORF1 and ORF2, located within the repeated terminal region of the genome and therefore present in two copies, and ORF5, ORF16, ORF18, ORF27, ORF36, ORF69, ORF74, ORF81, ORF105, ORF116, ORF118, ORF119, ORF120, and ORF121, all encoding functionally uncharacterized proteins. Notably, 13 ORFs located on 9 transcripts are not annotated in the reference genome, although some of these were annotated in nonreference OsHV-1 strains (e.g. ORF60_x reported in OsHV-1-PT), whereas ORFs encoded antisense with regard to the reference annotations were missed previously altogether.

The size distribution of the OsHV-1 DRS reads is reproducible among five of the six samples (Fig. 2d), with all the samples showing a bimodal size distribution. The first peak matched the average length of the most expressed transcripts (Fig. 2d), whereas the second peak, corresponding to shorter transcripts, likely represented RNA degradation products or incomplete transcription events. Notably, only sample 60-hpi_3 possessed a higher "full-length peak" compared to the short one. In order to verify this observation, we mapped the DRS reads to the 78 gene units separately and evaluated the distribution of the read lengths compared to the nominal length of each gene. As expected, for most of the genes, the DRS size distribution discretely peaked near the annotation boundaries or, occasionally, before them if shorter gene isoforms are expressed (Supplementary Fig. S2). However, for 25 genes, we obtained evidence of read-through-transcription (RTT), with fractions of the reads exceeding gene boundaries. RTT generally produced molecules of variable lengths, except for a few cases in which we detected peaks of discrete size, possibly corresponding to overlooked isoforms (e.g. gene 174989_+ in Fig. 2e).

Next, we investigated the dynamic expression of the OsHV-1 transcriptome based on DRS (60–72 hpi) and short-read data (6–72 hpi). Analysis of the DRS data showed that late in infection, most of the highly expressed transcripts were the same in five samples, whereas sample 60_3 had a higher number of transcripts with a minimal expression level of 10,000 TPMs, yielding a different global expression profile (Fig. 3a). Two of the top5 expressed transcripts per sample are present in all six datasets (127600_128424, a gene encoding the putative capsid protein VP23 and 64787_65654 with unknown function), two in five datasets (23934_24291 and 166004_166879, the first being a ncRNA and the second encoding a short isoform of the capsid maturation protease), one in four datasets (135399_136225, unknown function), and four transcripts are in the top5 list of a single sample only. Comparison of the six short-read datasets paired to DRS data showed that the correlation of expression levels between the two technologies is generally good, in particular for the transcripts with mid-to-high expression levels (Supplementary Fig. S3). For 48 transcripts, short-read data did not support the expression observed with DRS data, the latter being at least 10-fold higher, likely due to the intrinsic limitation of short reads when dealing with overlapping genes.

Extending the expression analysis to earlier time points (6, 12, 24, 36, and 48 hpi) allowed the detection of OsHV-1 transcripts starting from 36 hpi (Fig. 3b). The expression heatmap clustered

the samples into three groups: the first included the 36-hpi samples with a limited number of expressed transcripts compared to later time points; the second included 48 and 60 hpi samples, and the third cluster accommodated the 72 hpi samples. The second cluster is further organized into two subclusters: one for two 48 hpi and 60-hpi_3 samples and the other subcluster included the rest of the samples (Supplementary Fig. S4). The expression of gene encoding major capsid protein (MCP), capsid protein VP23, scaffold protease, portal protein (identified herein as product of ORF28, transcript 46490_43531; Supplementary Table S1), DNA

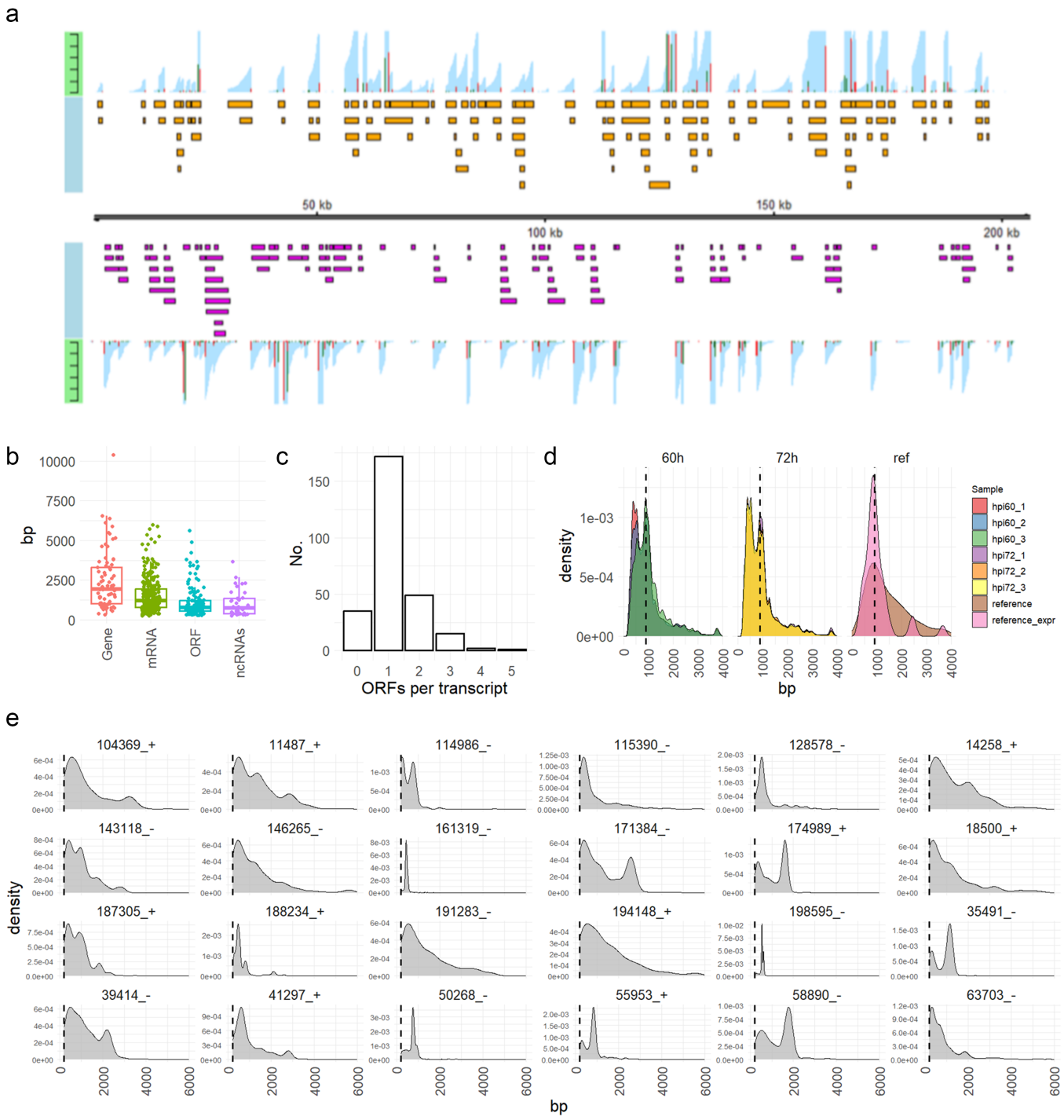

Figure 2. The OsHV-1 transcription map. (a) The identified TSSs (green bars) and TTSs (red bars) are depicted on each strand of the OsHV-1 genome, together with the 274 transcripts reported by orange (forward strand) and magenta (reverse strand) boxes. The height of the TSS and TTS bars represented the frequency of the events and the light-blue overlay represented the coverage of each transcript based on the DRS data. (b) Length distributions of the annotated mRNAs, ORFs, ncRNAs, and gene units of OsHV-1. (c) Distribution of the number of predicted ORFs per transcript. Hits with no ORF referred to ncRNAs. (d) Distribution of the sizes of the OsHV-1 DRS reads at 60 and 72 hpi reported together with the sizes of all the 274 predicted transcripts ("reference" in legend) and of the most expressed ones ("reference_expr"). A dotted line denotes the size of the main peak (976 bp). (e) Density plot referring to the lengths of the DRS reads mapped to 8 selected genes among the 25 showing RTT evidence. The length range reported in the graph started from 200 nt after the TTS. The panel is a selection of the results reported in Supplementary Fig. S2.

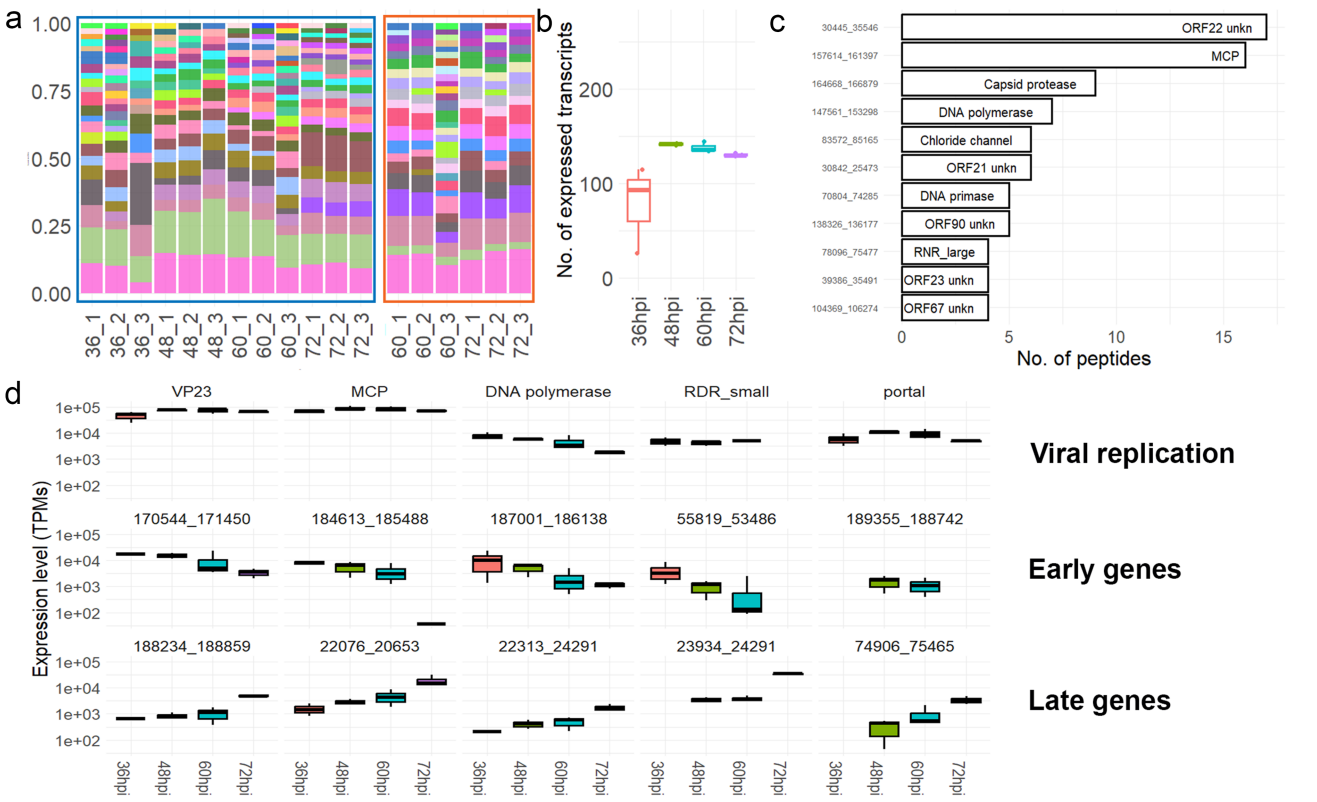

Figure 3. Expression dynamic of OsHV-1. (a) The relative abundances of the OsHV-1 transcripts with expression levels higher than 10,000 transcripts per millions (TPMs) are depicted. The part squared in orange is referred to DRS data (60 and 72 hpi), whereas the part squared in blue is referred to short-read data (36–72 hpi). The identity of the samples is indicated in the names of the bars. The expression dataset is included in Supplementary Table S2. (b) Number of expressed transcripts (>100 TPMs) in the samples, measured using short reads. (c) The OsHV-1 proteins with at least four matching peptides are reported, with the protein identities indicated into the bars. (d) Expression trends of selected OsHV-1 transcripts referring to viral genome replication and virion morphogenesis genes [VP23, MCP, DNA polymerase, ribonucleotide-reductase small subunit and the portal protein (first row)], early genes (second row) and late genes (third row) are reported averaged per time point.

polymerase, ribonucleotide reductase subunits, and several proteins of unknown function were detected early in the infection (Fig. 3d and Supplementary Table S2). During the later stages of infection, the identification of the stage-specific viral genes became more challenging because of the incremental number of expressed transcripts, possibly linked to the presence of RNAs destined for degradation rather than active translation. Overall, we identified seven putative early transcripts and five late ones (Fig. 3d). Notably, none of these stage-specific transcripts encoded for proteins with a predictable function. The 147 peptides matching the OsHV-1 genome found by MS proteomics originated from 41 proteins. The top 3 most abundant detected proteins were ORF22 (unknown function), MCP (ORF104), and capsid maturation protease (ORF107) (Fig. 3c).

A transcriptional architecture of the capsid maturation protease is conserved among distant herpes viruses

Virion formation is a quintessential feature, which distinguishes viruses from cellular organisms and all other types of mobile genetic elements (Koonin et al. 2021). Given the importance of virion morphogenesis genes and our observation that they are among the most strongly expressed, we investigated in detail the OsHV-1 regions coding for MCP (ORF104), capsid proteins VP19 (ORF112) and VP23 (ORF82), portal protein (ORF28), and capsid maturation protease (ORF107). The functions of these proteins have not been experimentally validated, thus similarity searches were used to support their identities.

The genomic region coding for the MCP included two isoforms, with two different TSSs and one shared TTS, resulting in either the full-length MCP (MCP_1) or a 5′-truncated MCP (MCP_2) (Fig. 4a), lacking the first 471 amino acids. Although MCP_1 and MCP_2 showed similar expression levels in the samples (Fig. 4f), the TSS of MCP_1 showed a higher frequency and proteomic data supported the presence of the full-length protein, with both the floor and the tower domains. The putative capsid protein VP19, which we identify here as a homolog of orthoherpesviral triplex 1 protein, is encoded as the first ORF of a bicistronic transcript, characterized by four isoforms coding either for both ORFs (VP19 and ORF113; referred to as VP19_polycist) or for VP19 (1 isoform) and ORF113 (2 isoforms) separately (Fig. 4b). The ORF113 isoforms and VP19_polycist were highly expressed compared to the standalone VP19 isoform (Fig. 4f). The genomic region coding for the capsid protein VP23, a homolog of the orthoherpesviral triplex 2 protein, was expressed as a single isoform with the highest expression levels in all samples, greatly exceeding that of the other considered genes (Fig. 4c). The gene coding for the portal protein included two isoforms, a full-length protein of 853 amino acid residues (named “portal” in Fig. 4d) and an isoform coding for a shorter in-frame ORF, lacking the first 600 residues. Notably, the shorter isoform showed an average 4-fold higher expression than the full-length isoform (Fig. 4f). The capsid maturation protease gene displays an even higher transcriptional complexity, including

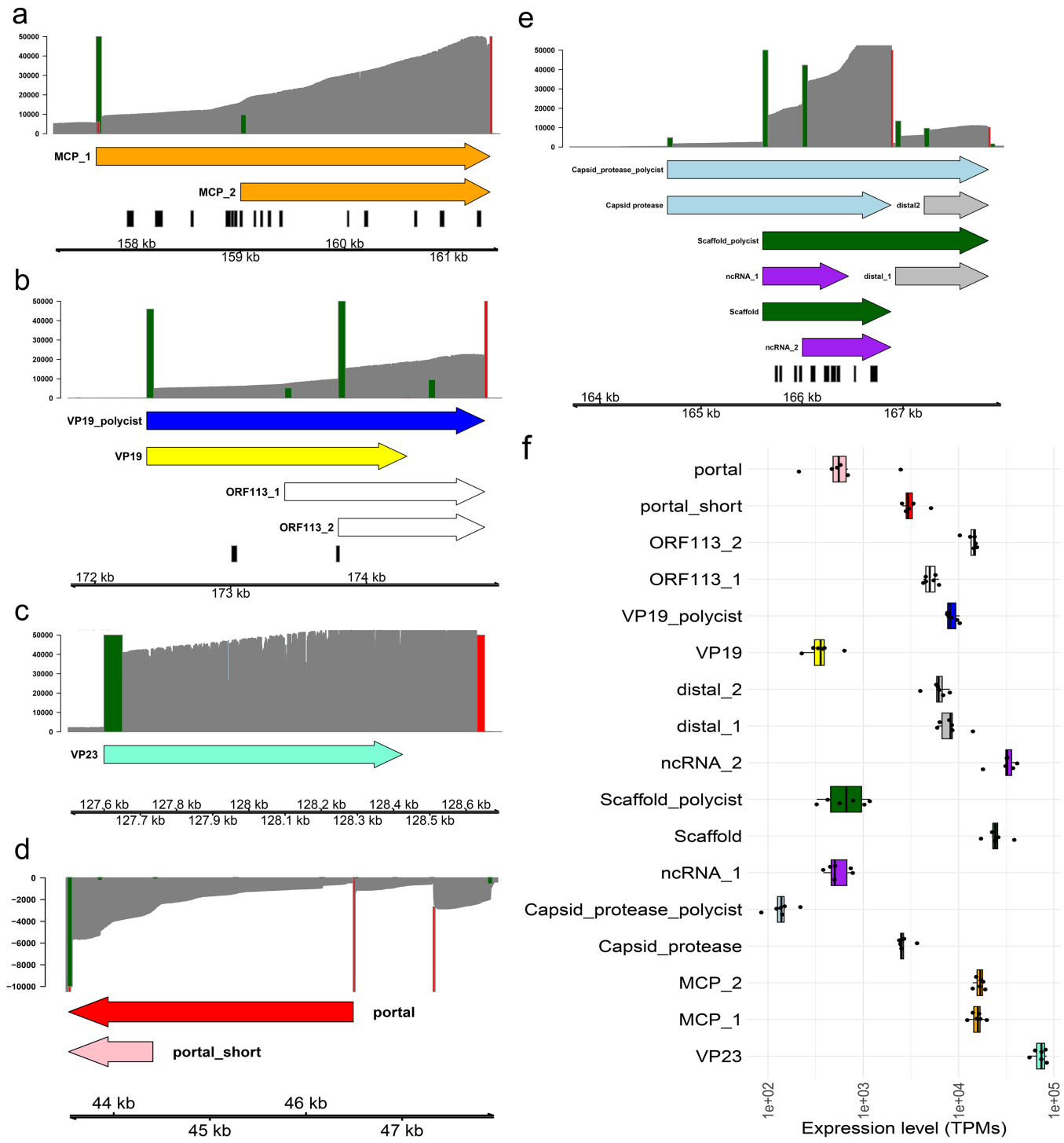

Figure 4. OsHV-1 capsid assembly and maturation. (a) The transcriptional landscape of the OsHV-1 major capsid protein (MCP) is reported, with arrows depicting the isoforms, together with their TSSs and shared TTS (green and red bars, respectively). The DRS coverage is reported as grey background, and the peptide positions identified by proteome analysis are depicted by black bars. (b) Transcriptional landscape of VP19 gene region. (c) Transcriptional landscape of VP23 gene region. (d) Transcriptional landscape of the portal protein. (e) Transcriptional landscape of the capsid protease of OsHV-1 (164,000–168,000, forward strand). (f) Expression levels were reported as TPMs for the considered isoforms in the six samples. The colour code in (f) matches the colours used in the other panels.

eight overlapping transcripts. The transcript encoding the full-length protein (164668_166879, 689 aa, labelled "Capsid protease" in Fig. 4e, light-blue arrow) embedded a shorter in-frame ORF, potentially encoding an N-terminally truncated protein of 367 aa residues (named "Scaffold" in Fig. 4e, green arrow). Both transcripts also showed 3′-RTT, sharing the TTS with a distal gene, also characterized by two isoforms (named "Capsid protease polycist" and "Scaffold polycist," and distal 1 and 2, respectively in Fig. 4e). Finally, two ncRNAs are present (purple arrows in Fig. 4e). Gene expression levels based on DRS data indicated that ncRNA_2 and the scaffold transcript are the most expressed ones (Fig. 4f), as also evident from the high frequencies of their TSSs and shared

TTS (Fig. 4e). The capsid maturation protease showed more modest expression levels, as did the polycistronic transcripts. Notably, all the considered isoforms involved in capsid assembly and maturation showed a limited variation in their relative expression between samples, possibly indicating that they function in stoichiometric quantities. These results suggest that tight regulation of the capsid assembly and maturation genes is ensured by several different mechanisms.

ADAR-mediated editing and hyper-editing along OsHV-1 genome

Analysis of ADAR expression levels along the course of infection revealed ADAR1 as the most expressed gene among the four ADAR homologs, with its expression levels exceeding 70 TPMs at 60 hpi (Supplementary Fig. S5a). However, this homolog is barely upregulated during infection (∼4-fold), compared to another ADAR1 (EVM0005150), which is highly induced (∼100-fold) at 72 hpi (Supplementary Fig. S5b). A preliminary analysis of the SNVs profile on OsHV-1 revealed a prevalence of ADAR-compatible SNVs (A-to-G or T-to-C) at 72 hpi, compared to the earlier time points (Supplementary Fig. S5c). Overall, these results suggest that ADAR is active during the OsHV-1 infection in blood clam.

The genomic positions of the 274 predicted OsHV-1 transcripts, collapsed into the 78 gene units, were mined to identify 20 NATs and 10 regions of overlapping sense–antisense transcription. Regions 4, 5, 6, 7, 8, and 9 are characterized by divergent overlaps, i.e. overlaps of the 5′-ends of the transcripts, with regions 1 and 2 containing antisense embedded genes. In contrast, regions 3 and 10 corresponded to convergent overlaps, i.e. overlaps of transcripts' 3′-ends (Fig. 5a). Detailed characterization of the expression levels of these NATs revealed that 9 out of 10 NAT pairs showed medium-to-high expression levels in the six samples, with three pairs exceeding an average of 10,000 TPMs (Fig. 5b).

To gain insights into the potential roles of the NATs, we considered the distribution of polymorphisms across different OsHV-1 transcripts, which could signify editing by host defence systems, and of the extent of ADAR-mediated adenosine-to-inosine modifications in OsHV-1. This was obtained by analyzing the distribution of inosines and of hyper-edited reads. The mapping of short-read data on the OsHV-1 genome was also used to identify 465 ADAR-compatible single nucleotide variations (SNVs, A-to-G or T-to-C), located predominantly in noncoding regions (65.9%), whereas 20.2% caused nonsynonymous changes and 13.9% synonymous changes (Supplementary Table S3). A total of 226 SNVs were validated for the presence of inosine, using at least five DRS reads with a minimal strandedness of 2-fold (i.e. A-to-G variations supported by DRS reads mapping in forward and T-to-C variations with DRS reads mapping in reverse). Overall, 60,420 RNA molecules sequenced by DRS contained at least one inosine (5.9% of total reads), with a slight reduction at 72 hpi (Supplementary Fig. S6b). Notably, sample 72_1 comes across as an outlier, with a higher hyper-editing level and a lower editing level, in comparison with the paired samples. In parallel, we extracted from the six Illumina datasets 21,180 hyper-edited reads, accounting for 0.1% of the total reads, with an increase observed from 60 to 72 hpi (Supplementary Fig. S6a). The distribution of hyper-edited reads and of DRS reads containing inosines was comparable and mostly matched sense–antisense transcription regions 1, 5, 8, 9, and 10 (Fig. 5a). An additional hotspot of editing is found around position 150 kb, but this is not supported by hyper-edited reads.

Notably, seven pairs of the described NATs, which showed balanced expression between sense and antisense transcripts (Fig. 5c), were the most impacted by editing (regions 1, 4, 5, 7, 8, 9, and 10). We also found that hyper-editing preferentially impacted sense–antisense transcription regions, whereas this preference was less pronounced for the single-nucleotide editing traced by the DRS data (Fig. 5d and e).

Focusing on the edited DRS reads, most of them possessed a single inosine (52%), with a minority having multiple inosines, up to 36, with no difference among samples (Fig. 6a). The average inosine frequency at the edited positions was 10%, with the exception of sample 60-hpi_2 which showed a mean frequency of 25% (Fig. 6b). Based on the distribution of inosines along the reads, only 733 DRS reads (1.2%) showed a hyper-editing profile with at least 5 inosines within 150 nt, and all these reads are located within antisense regions 9 and 10. Eleven transcripts have at least 100 modified reads, but notably only one of them (55953_58943) is a NAT, while most of the others are located near antisense transcription regions (Fig. 6c). None of these 11 transcripts encoded proteins with predictable functions. Finally, no significant size differences were found between edited and nonedited reads (Fig. 6d).

In conclusion, the presence of both RNA editing and hyper-editing supports the active role of ADAR during OsHV-1 infection. However, hyper-editing and editing seem to follow different trajectories, with the former occurring within regions of perfect sense–antisense matches only, and the latter being more diffuse across the OsHV-1 transcripts.

Discussion

We performed a deep characterization of the OsHV-1 transcriptome by long-read transcriptomics (DRS), complemented with short-read RNA sequencing and MS proteomics. DRS data support and extend by 2.5-fold the number of transcripts identified in the previous OsHV-1 transcriptome map based on PacBio long reads (Bai et al. 2021). Moreover, our data illuminate interesting new aspects of the OsHV-1 transcription beyond the mere expression analysis. Indeed, we confirm the complexity and high coding density of the OsHV-1 genome, with 274 transcripts grouped into 78 gene units, with evidence of polycistronic transcripts and multiple isoforms involving most of the genes. The high transcriptional complexity has been reported for several representatives of the order *Herpesvirales* (Hale and Moorman 2021), making proteomic and/or ribosome profiling data necessary to further investigate the true coding potential of these viruses (Volkening et al. 2023). Our proteomics analysis supported only a small number of OsHV-1 proteins, most of them linked to important viral functions, such as virion formation and genome replication (e.g. the MCP, capsid maturation protease, and DNA polymerase) or proteins that, despite having unknown functions, were reported as highly expressed in other transcriptomic (Rosani et al. 2015) or proteomic surveys performed in *C. gigas* infected with OsHV-1, such as ORF22 (Leprêtre et al. 2021).

By leveraging the updated transcriptomic annotation and extending the expression analysis to earlier time points, we captured the temporal dynamics of viral genome expression and categorized viral transcripts into those expressed preferentially in early- or late-infection stages. During the early infection stages, we also detected the expression of genes involved in virion assembly and viral genome replication (e.g. capsid proteins and DNA polymerase). However, the transcripts characteristic of the early or late stages only was functionally unknown. The considerable fraction of proteins for which functions cannot be predicted highlights the importance of further functional studies, which may uncover both new, mollusc-specific aspects of viral biology and

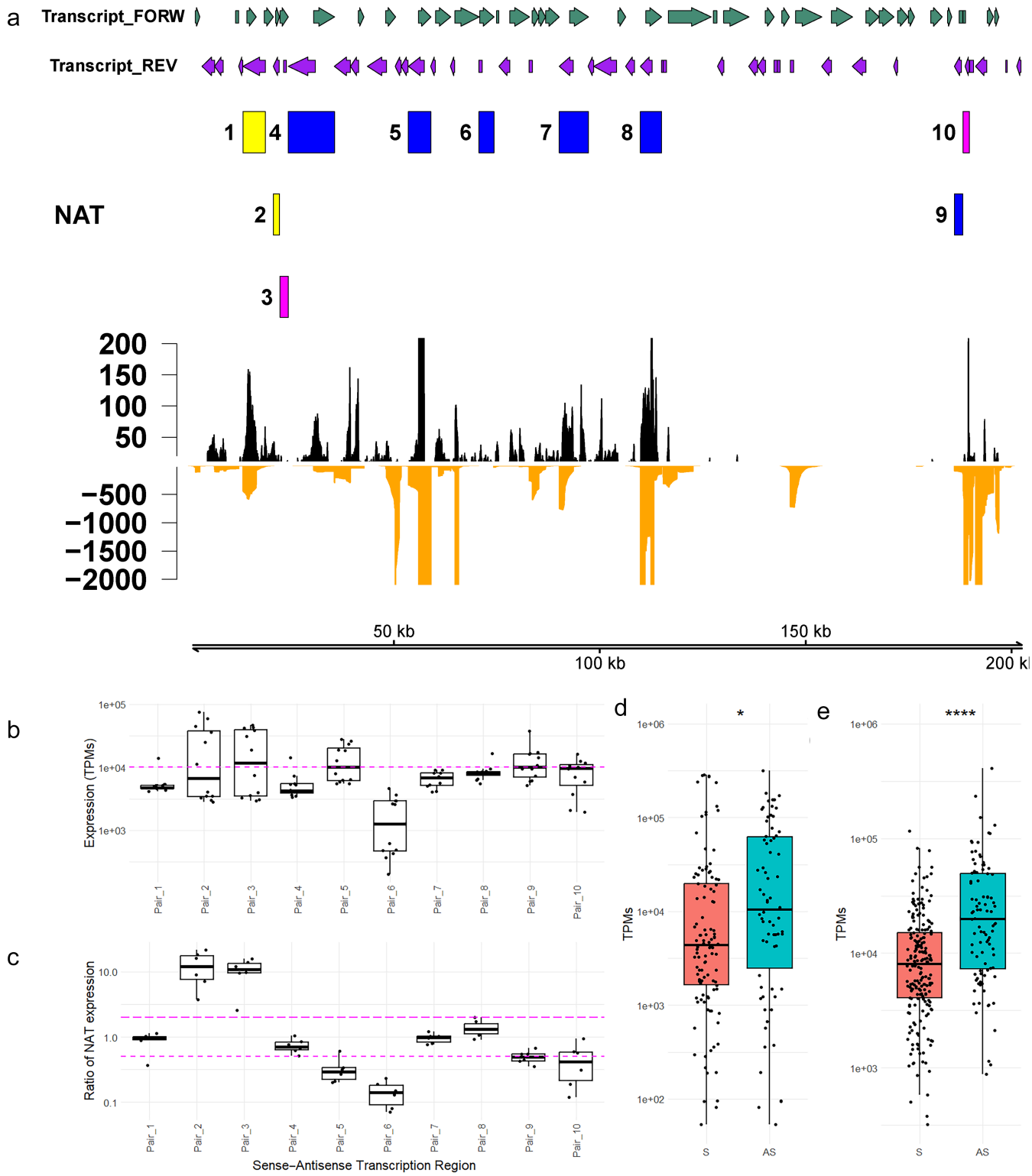

Figure 5. Antisense transcription along the OsHV-1 genome and RNA editing. (a) The figure depicted the transcriptional landscape of OsHV-1 with forward (green arrows) and reverse (purple arrows) transcripts, the 10 sense–antisense transcription regions with a colour code indicating if they are referring to convergent (pink boxes), divergent (blue boxes), or embedded (yellow boxes) NAT overlaps as well as the distribution of hyper-edited short reads (black profile) and of edited DRS reads (orange profile). (b) Expression levels of the NATs measured as TPMs and divided by the 10 sense–antisense transcription pairs. The magenta dotted line indicated the 10,000 TPMs level (c) The expression ratio between the sense and the antisense NATs for each pair is reported. The magenta lines indicated the range of balanced expression between NATs (5-fold). (d) Distribution of the number of edited DRS reads between sense transcripts (S) and NATs (SA). *P-value < .05, t-test. (e) Distribution of the number of hyper-edited reads between sense transcripts (S) and NATs (SA). **** P-value < .001, t-test.

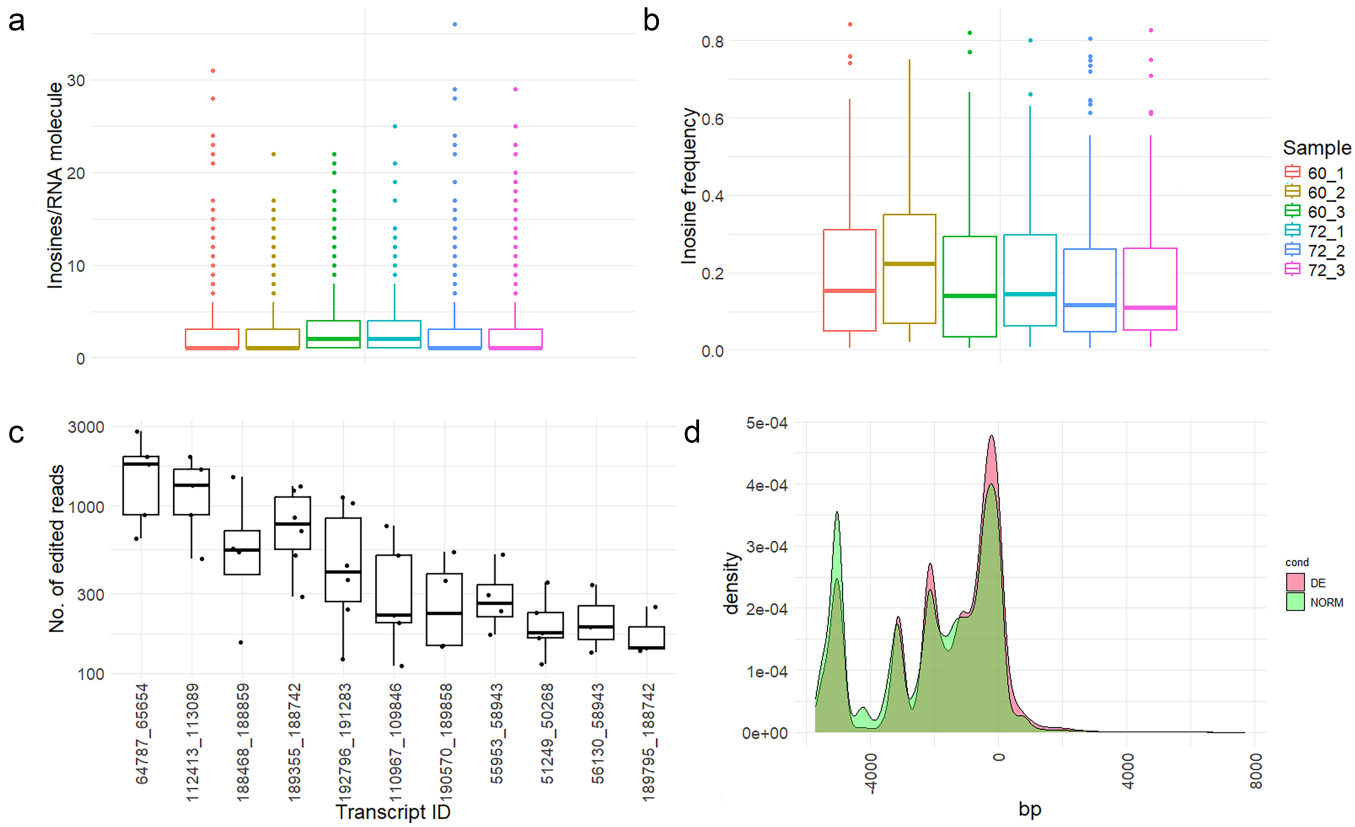

Figure 6. ADAR editing and hyper-editing in OsHV-1. (a) Distribution of the number of inosines nucleotides per RNA molecule in the six samples. (b) Distribution of the frequencies of inosines traced in the 226 validated SNVs measured in the six samples. (c) Number of edited DRS reads per transcripts. Only the 11 transcripts with more than 100 edited reads in at least three samples are reported. (d) Size profiles of the DRS reads (NORM) and of the edited DRS reads (DE), normalized by the nominal lengths of the matching genes.

mechanisms shared with vertebrate herpesviruses and perhaps even more distantly related nonanimal viruses.

Malacoherpesviruses together with vertebrate herpesviruses (families *Orthoherpesviridae* and *Alloherpesviridae*), the recently discovered protist-infecting viruses of the proposed phylum 'Mirusviricota' (Gaïa et al. 2023) and prokaryotic viruses of the class *Caudoviricetes* (Duda and Teschke 2019, Baquero et al. 2020) comprise the realm *Duplodnaviria* (Koonin et al. 2020). Viruses in this vast and ancient realm share the morphogenetic module and form icosahedral capsids from structurally conserved MCPs with the HK97-fold (Suhanovsky and Teschke 2015). Virion morphogenesis starts with the assembly of an empty capsid, which undergoes maturation and is subsequently filled with the viral genomic DNA. A scaffolding protein and the capsid maturation protease, two conserved components of the duplodnavirus morphogenetic module, play key roles in virion morphogenesis (Liu and Mushegian 2004, Veesler et al. 2014, Mushegian et al. 2018). The key proteins for capsid assembly and maturation have been identified by similarity searches in HaHV-1 and OsHV-1 (Mushegian et al. 2018). Here, we extended the annotation of the malacoherpesviral morphogenetic module by identifying the OsHV-1 triplex proteins 1 (ORF19) and 2 (ORF23) as well as portal protein (ORF28). The portal is a key component of the virion, which nucleates the capsid assembly and provides a conduit for the entrance of genomic viral DNA (Wang et al. 2020). The two triplex proteins play an important role in stabilizing the icosahedral capsid and are specific to eukaryotic duplodnaviruses, previously described in vertebrate orthoherpesviruses and protist mirusviruses (Gaïa et al. 2023). Notably, similar to other viruses, the two triplex proteins, along with the MCP, are among the most strongly transcribed viral genes (Gaïa et al. 2023). Moreover, our DRS data provide evidence that the transcriptional architecture of the scaffolding gene and capsid maturation protease gene is shared between malacoherpesviruses and vertebrate herpesviruses. We found that the capsid maturation protease is encoded as a fusion with the scaffolding protein, and transcription of this gene yields three types of transcripts corresponding to the capsid protease-scaffolding protein fusion and at least one isoform of the standalone scaffolding protein. Notably, the expression levels of the variants are highly different, with the full-length protease–scaffold fusion transcripts showing 10-fold lower levels of expression compared to the scaffolding protein transcript. A similar feature was also reported in vertebrate herpesviruses, HSV-1, bovine alpha-herpesvirus type 1 (BoHV-1), and Marek's disease virus (MDV) (Depledge et al. 2019, Tombácz et al. 2022, Volkening et al. 2023). In BoHV-1, the short isoform of the UL26 gene, called UL26.5, displayed higher expression levels at 12 h of lytic infection (Tombácz et al. 2022), and independent transcription and translation were reported for MDV (Volkening et al. 2023). Our proteome data supported the presence of the capsid scaffold protein, whereas no peptides were detected for the capsid protease, likely due to the low expression of the full-length isoform and autoproteolytic cleavage of the capsid maturation protease. The conservation of the transcriptional architecture and expression profiles in OsHV-1, HaHV-1 (Bai et al. 2021), and other herpesviruses (Depledge et al. 2019, Tombácz et al. 2022, Volkening et al. 2023) suggest its functional importance. Indeed, the covalent fusion with the scaffolding protein would ensure the timely co-incorporation of the protease into empty procapsids along with

the scaffolding protein. Furthermore, it stands to reason that the amount of the protease necessary for the proteolytic processing of the scaffolding protein would be considerably lower than that of the substrate (i.e. scaffolding protein). Although it is difficult to estimate the divergence time for distantly related viruses, the pan-*Herpesvirales* conservation of transcriptional architecture of the protease-scaffolding gene module has likely evolved at least in the common ancestor of vertebrate and invertebrate herpesviruses, and hence could predate the divergence of arthropods, molluscs, and chordates (Savin et al. 2010, Andrade-Martínez et al. 2019).

Our transcriptomic data has also illuminated a new facet of malacoherpesvirus–host interaction involving the ADAR antiviral defence system. We have previously reported ADAR-mediated RNA editing in HaHV-1 and OsHV-1 infecting different host species (Bai et al. 2021; Rosani et al. 2019), a feature which differentiated malacoherpesviruses from all other known herpesviruses. In fact, in vertebrate herpesviruses, the RNA editing impacts specific genomic features only [e.g. specific miRNAs or transcripts (Ivanišević et al. 2023)], whereas in malacoherpesviruses, RNA editing extensively impacted multiple transcripts. We previously showed that in OsHV-1, the RNA editing is concentrated within genomic hotspots, whereas in HaHV-1 these modifications are more evenly distributed across the genome (Bai et al. 2021). However, up to now, the extension of RNA editing was indirectly deduced tracing highly modified short reads, as a proxy of ADAR hyper-editing, whereas no approaches were undertaken to detect inosines. In this study, through the sequencing of native RNA, it became possible to confirm the presence of inosines. So far, only *Dinopore*, which uses a convolutional neural network approach to predict whether a given site is edited (Nguyen et al. 2022), and DeepEdit, which, according to the authors, can provide a phased read-level detection of RNA editing events (Chen et al. 2023), allowed the identification of inosine from DRS data. Despite certain limitations, such as the use of an exogenous model to train the dataset and the requirement of target positions to verify the presence of inosines, based on *DeepEdit*, we could provide independent support for the presence of ADAR editing in OsHV-1 and an accurate evaluation of the extent of such phenomenon.

As result, we confirmed that ADAR hyper-editing mostly occurred within the antisense transcription regions with balanced expression between NATs, because these regions generated abundant dsRNA able to "capture" the enzymatic capability of ADAR proteins, possibly diverting it from functional transcripts, a phenomenon that we previously described as "ADAR decoys" (Bai et al. 2021). However, here we showed that the distribution of inosines extends beyond NATs, and is mostly associated with single-nucleotide editing rather than hyper-editing. This indicates that single-nucleotide editing might be more "diffuse" than previously considered and can occur due to the formation of local secondary structures of RNAs (hairpins), likely to be relatively short and transient compared to the longer stretches of perfectly matching sense–antisense sequences. Hence, the binding of ADAR would be expected to be weaker on these short hairpins, insufficient for multiple editing. Conceivably, only hyper-edited RNA molecules undergo rapid degradation starting from the 3′-UTR (Mayr 2019), thus hampering their quantitative sequencing using the polyA selection strategies, applied herein for the DRS data. This can explain the near absence of hyper-edited DRS reads, except for a cluster limited to a single OsHV-1 genomic region. Although the activity of ADAR during the OsHV-1 infection is now apparent, its biological role remains unresolved. Besides enzymatic activity, we should consider possible editing-independent roles of ADAR, such as the competitive binding to dsRNAs, preventing the action of other RNA binding molecules. Such editing-independent roles can involve miRNA formation (Nishikura et al. 2013, Deng et al. 2020) and RNA degradation, among others (Shevchenko and Morris 2018).

Conclusions

In this study, we reported a detailed transcriptional map of OsHV-1 based on long Nanopore RNA reads. Undoubtedly, long reads hold a greater potential to provide expression and structural information of overlapping gene architectures, a reason to favor them for (herpes) viral transcriptomics. We also described the first application of DRS for the detection of inosines in viral RNAs. Although the available tools are not fully developed, they already show promise in untangling the inosine modifications. Overall, we showed that the transcriptional architectures of OsHV-1 could play a functional role in infection, here highlighted by the detection of the pan-*Herpesvirales* conservation of the capsid scaffold transcriptomic architecture as well as by the presence of "editing decoys," possibly attracting a considerable part of the enzymatic potential of the host ADAR. These results could also be exploited to investigate, in future studies, the molecular behavior of OsHV-1 in different hosts, which can encompass bivalve and nonbivalve species (Whittington et al. 2018, Martenot et al. 2019, Prado-Alvarez et al. 2021).

Acknowledgements

A part of the computational power required for the analyses was provided by "University of Padova Strategic Research Infrastructure Grant 2017: CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l'Innovazione".

Author's contributions

U.R.: conceptualization, data analysis, writing, discussion, review, and editing. E.B.: data analysis. B.W.H.: methodology. X.Z.: methodology. L.S.X.: validation, resources. M.K.: conceptualization, data analysis, writing, discussion, review and editing. C.M.B.: funding acquisition, conceptualization, writing, and review and editing.

Supplementary data

Supplementary data is available at *VEVOLU Journal* online.

Conflict of interest: None declared.

Funding

C.M.B. was supported by National Natural Science Foundation of China (32073014), Central Public-interest Scientific Institution Basal Research Fund, CAFS (2023TD30), Central Public-interest Scientific Institution Basal Research Fund, YSFRI (20603022023010) and the earmarked fund for CARS (CARS-49); U.R. was supported by the Italian National grant P2022JEEMT (PRIN 2022PNRR).

Data availability

The data described in this work are available in public repositories or are provided within supplementary materials. In detail, long-read and short-read data are available in the NCBI SRA archive

under accession IDs "PRJNA874858." Proteomic data are deposited in the iProX (https://www.iprox.cn/) archive, under project ID IPX0007584002.

References

Andrade-Martínez JS, Moreno-Gallego JL, Reyes A. Defining a core genome for the herpesvirales and exploring their evolutionary relationship with the caudovirales. *Sci Rep* 2019;**9**:11342.

Bai C-M, Rosani U, Li Y-N *et al*. RNA-seq of HaHV-1-infected abalones reveals a common transcriptional signature of Malacoherpesviruses. *Sci Rep* 2019a;**9**:938.

Bai C-M, Rosani U, Xin L-S *et al*. Dual transcriptomic analysis of Ostreid herpesvirus 1 infected Scapharca broughtonii with an emphasis on viral anti-apoptosis activities and host oxidative bursts. *Fish Shellfish Immunol* 2018;**82**:554–64.

Bai C-M, Rosani U, Zhang X *et al*. Viral decoys: the only two herpesviruses infecting invertebrates evolved different transcriptional strategies to deflect post-transcriptional editing. *Viruses* 2021;**13**:1971.

Bai C-M, Xin L-S, Rosani U *et al*. Chromosomal-level assembly of the blood clam, Scapharca (Anadara) broughtonii, using long sequence reads and Hi-C. *GigaScience* 2019b;**8**:giz067.

Bajad P, Jantsch MF, Keegan L *et al*. A to I editing in disease is not fake news. *RNA Biol* 2017;**14**:1223–31.

Baquero DP, Liu Y, Wang F *et al*. Structure and assembly of archaeal viruses. *Adv Virus Res* 2020;**108**:127–64.

Bass BL. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 2002;**71**:817–46.

Boldogkői Z, Moldován N, Balázs Z *et al*. Long-read sequencing – a powerful tool in viral transcriptome research. *Trends Microbiol* 2019;**27**:578–92.

Chen L, Ou L, Jing X *et al*. DeepEdit: single-molecule detection and phasing of A-to-I RNA editing events using nanopore direct RNA sequencing. *Genome Biol* 2023;**24**:75.

Chen S, Zhou Y, Chen Y *et al*. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;**34**:i884–90.

Corporeau C, Tamayo D, Pernet F *et al*. Proteomic signatures of the oyster metabolic response to herpesvirus OsHV-1 μVar infection. *J Proteomics* 2014;**109**:176–87.

Davison AJ, Eberle R, Ehlers B *et al*. The order Herpesvirales. *Arch Virol* 2009;**154**:171–77.

Davison AJ, Trus BL, Cheng N *et al*. A novel class of herpesvirus with bivalve hosts. *J Gen Virol* 2005;**86**:41–53.

de la Rubia I, Srivastava A, Xue W *et al*. RATTLE: reference-free reconstruction and quantification of transcriptomes from Nanopore sequencing. *Genome Biol* 2022;**23**:153.

Delisle L, Pauletto M, Vidal-Dupiol J *et al*. High temperature induces transcriptomic changes in Crassostrea gigas that hinder progress of ostreid herpesvirus (OsHV-1) and promote survival. *J Exp Biol* 2020;**223**:20.

Deng P, Khan A, Jacobson D *et al*. Adar RNA editing-dependent and -independent effects are required for brain and innate immune functions in Drosophila. *Nat Commun* 2020;**11**:1580.

Depledge DP, Srinivas KP, Sadaoka T *et al*. Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun* 2019;**10**:1–13.

Duda RL, Teschke CM. The amazing HK97 fold: versatile results of modest differences. *Curr Opin Virol* 2019;**36**:9–16.

Gaïa M, Meng L, Pelletier E *et al*. Mirusviruses link herpesviruses to giant viruses. *Nature* 2023;**616**:783–89.

Giorgi FM, Ceraolo C, Mercatelli D. The R Language: an engine for bioinformatics and data science. *Life* 2022;**12**:648.

Gleeson J, Leger A, Prawer YDJ *et al*. Accurate expression quantification from nanopore direct RNA sequencing with NanoCount. *Nucleic Acids Res* 2022;**50**:e19.

Green TJ, Rolland J-L, Vergnes A *et al*. OsHV-1 countermeasures to the Pacific oyster's anti-viral response. *Fish Shellfish Immunol* 2015;**47**:435–43.

Grice LF, Degnan BM. The origin of the ADAR gene family and animal RNA editing. *BMC Evol Biol* 2015;**15**:4.

Hahne F, Ivanek R. Visualizing genomic data using gviz and bioconductor. *Methods Mol Biol* 2016;**1418**:335–51.

Hale AE, Moorman NJ. The ends dictate the means: promoter switching in herpesvirus gene expression. *Annu Rev Virol* 2021;**8**:201–18.

Heinz S, Benner C, Spann N *et al*. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;**38**:576–89.

Hyatt D, Chen G-L, LoCascio PF *et al*. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf* 2010;**11**:119.

Imam H, Kim G-W, Siddiqui A. Epitranscriptomic(N6-methyladenosine) modification of viral RNA and virus-host interactions. *Front Cell Infect Microbiol* 2020;**10**:584283.

Iranzo J, Krupovic M, and Koonin EV. The double-stranded DNA virosphere as a modular hierarchical network of gene sharing. *mBio* 2016;**7**:e00978–16.

Ivanišević V, Žilić L, Čunko M *et al*. RNA editing-dependent and -independent roles of ADAR proteins in herpesvirus infection – hints for another layer of complexity. *Viruses* 2023;**15**:2007.

Ivanov S, Lagunin A, Filimonov D *et al*. Network-based analysis of OMICs data to understand the HIV–host interaction. *Front Microbiol* 2020;**11**:1314.

Jacobson TB, Callaghan MM, Amador-Noguez D. Hostile takeover: how viruses reprogram prokaryotic metabolism. *Annu Rev Microbiol* 2021;**75**:515–39.

Jenkins C, Hick P, Gabor M *et al*. Identification and characterisation of an ostreid herpesvirus-1 microvariant (OsHV-1 μ-var) in Crassostrea gigas (Pacific oysters) in Australia. *Dis Aquat Organ* 2013;**105**:109–26.

Jones P, Binns D, Chang H-Y *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;**30**:1236–40.

Keegan LP, Brindle J, Gallo A *et al*. Tuning of RNA editing by ADAR is required in Drosophila. *EMBO J* 2005;**24**:2183–93.

Koonin EV, Dolja VV, Krupovic M *et al*. Global organization and proposed megataxonomy of the virus world. *Microbiol Mol Biol Rev* 2020;**84**:e00061–19.

Koonin EV, Dolja VV, Krupovic M *et al*. Viruses defined by the position of the virosphere within the replicator space. *Microbiol Mol Biol Rev* 2021;**85**:e0019320.

Leprêtre M, Faury N, Segarra A *et al*. Comparative proteomics of ostreid herpesvirus 1 and pacific oyster interactions with two families exhibiting contrasted susceptibility to viral infection. *Front Immunol* 2021;**11**:3474.

Li H, Birol I. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100.

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;**25**:1754–60.

Li H, Handsaker B, Wysoker A *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;**25**:2078–79.

Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 2019;**47**:e47.

Liu J, Mushegian A. Displacements of prohead protease genes in the late operons of double-stranded-DNA bacteriophages. *J Bacteriol* 2004;**186**:4369–75.

Martenot C, Faury N, Morga B *et al*. Exploring first interactions between ostreid Herpesvirus 1 (OsHV-1) and its host, crassostrea gigas: effects of specific antiviral antibodies and dextran sulfate. *Front Microbiol* 2019;**10**:1128.

Mayr C. What are 3′ UTRs doing? *Cold Spring Harbor Perspect Biol* 2019;**11**:a034728.

Mushegian A, Karin EL, Pupko T. Sequence analysis of malacoherpesvirus proteins: Pan-herpesvirus capsid module and replication enzymes with an ancient connection to "Megavirales". *Virology* 2018;**513**:114–28.

Neogi U, Elaldi N, Appelberg S *et al*. Multi-omics insights into host-viral response and pathogenesis in Crimean-Congo hemorrhagic fever viruses for novel therapeutic target. *eLife* 2022;**11**:e76071.

Nguyen TA, Heng JWJ, Kaewsapsak P *et al*. Direct identification of A-to-I editing sites with nanopore native RNA sequencing. *Nat Methods* 2022;**19**:833–44.

Nishikura K, Sakurai M, Ariyoshi K *et al*. Antagonistic and stimulative roles of ADAR1 in RNA silencing. *RNA Biol* 2013;**10**: 1240–47.

Pfaller CK, George CX, Samuel CE. Adenosine deaminases acting on RNA (ADARs) and viral infections. *Annu Rev Virol* 2021;**8**: 239–64.

Porath HT, Carmi S, Levanon EY. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat Commun* 2014;**5**: 1–10.

Prado-Alvarez M, García-Fernández P, Faury N *et al*. First detection of OsHV-1 in the cephalopod Octopus vulgaris. Is the octopus a dead-end for OsHV-1? *J Invertebr Pathol* 2021;**183**:107553.

Prasad A, Prasad M. Host-virus interactions mediated by long non-coding RNAs. *Virus Res* 2021;**298**:198402.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–42.

Ramaswami G, Lin W, Piskol R *et al*. Accurate identification of human Alu and non-Alu RNA editing sites. *Nat Methods* 2012;**9**:579–81.

Remmert M, Biegert A, Hauser A *et al*. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods* 2012;**9**:173–75.

Rosani U, Bai C-M, Maso L *et al*. A-to-I editing of Malacoherpesviridae RNAs supports the antiviral role of ADAR1 in mollusks. *BMC Evol Biol* 2019;**19**:149.

Rosani U, Gaia M, Delmont TO *et al*. Tracing the invertebrate herpesviruses in the global sequence datasets. *Front Mar Sci* 2023;**10**:1159754.

Rosani U, Varotto L, Domeneghetti S *et al*. Dual analysis of host and pathogen transcriptomes in ostreid herpesvirus 1-positive Crassostrea gigas. *Environ Microbiol* 2015;**17**:4200–12.

Savin KW, Cocks BG, Wong F *et al*. A neurotropic herpesvirus infecting the gastropod, abalone, shares ancestry with oyster herpesvirus and a herpesvirus associated with the amphioxus genome. *Virol J* 2010;**7**:308.

Shevchenko G, Morris KV. All I's on the RADAR: role of ADAR in gene regulation. *FEBS Lett* 2018;**592**:2860–73.

Stephenson E, Reynolds G, Botting RA *et al*. Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med* 2021;**27**:904–16.

Stoiber M, Quick J, Egan R *et al*. De novo identification of DNA modifications enabled by genome-guided nanopore signal processing. *bioRxiv* 2017;094672.

Suhanovsky MM, Teschke CM. Nature's favorite building block: deciphering folding and capsid assembly of proteins with the HK97-fold. *Virology* 2015;**479-480**:487–97.

Tan B, Gao S-J, Gack MU. The RNA epitranscriptome of DNA viruses. *J Virol* 2018;**92**:e00696–18.

Tombácz D, Kakuk B, Torma G *et al*. In-depth temporal transcriptome profiling of an alphaherpesvirus using nanopore sequencing. *Viruses* 2022;**14**:1289.

Veesler D, Khayat R, Krishnamurthy S *et al*. Architecture of a dsDNA viral capsid in complex with its maturation protease. *Structure* 2014;**22**:230–37.

Volkening JD, Spatz SJ, Ponnuraj N *et al*. Viral proteogenomic and expression profiling during productive replication of a skin-tropic herpesvirus in the natural host. *PLoS Pathog* 2023;**19**:e1011204.

Wang N, Chen W, Zhu L *et al*. Structures of the portal vertex reveal essential protein-protein interactions for Herpesvirus assembly and maturation. *Protein Cell* 2020;**11**:366–73.

Whittington RJ, Paul-Pont I, Evans O *et al*. Counting the dead to determine the source and transmission of the marine herpesvirus OsHV-1 in Crassostrea gigas. *Vet Res* 2018;**49**:34.

Xia J, Bai C, Wang C *et al*. Complete genome sequence of Ostreid herpesvirus-1 associated with mortalities of Scapharca broughtonii broodstocks. *Virol J* 2015;**12**:1–9.

Zhang F, Lu Y, Yan S *et al*. SPRINT: an SNP-free toolkit for identifying RNA editing sites. *Bioinformatics* 2017;**33**:3538–48.

Virus Evolution, 2024, 10(1), veae088 , DOI: https://doi.org/10.1093/ve/veae088, Advance Access Publication 17 October 2024, Research Article