Heliyon 10 (2024) e32811

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Genomic and molecular characterization of a cyprinid herpesvirus 2 YC-01 strain isolated from gibel carp

Jia Yang [a,1], Simin Xiao [a,1], Liqun Lu [a,b], Hao Wang [a,b,*], Yousheng Jiang [a,b,**]

[a] National Pathogen Collection Center for Aquatic Animals, Shanghai Ocean University, Shanghai, 201306, China
[b] Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture, Shanghai Ocean University, Shanghai, 201306, China

ARTICLE INFO

Keywords:
CyHV-2
Carassius gibelio
Whole-genome sequencing
Virulence factor
B cell epitope

ABSTRACT

Cyprinid herpesvirus 2 (CyHV-2) is the pathogen of herpesviral hematopoietic necrosis (HVHN), causing the severe economic losses in farmed gibel carp (*Carassius gibelio*). Further exploration of the genome structure and potential molecular pathogenesis of CyHV-2 through complete genome sequencing, comparative genomics, and molecular characterization is required. Herein, the genome of a CyHV-2 YC-01 strain isolated from diseased gibel carp collected in Yancheng, Jiangsu Province, China was sequenced, then we analyzed the genomic structure, genetic properties, and molecular characterization. First, the complete YC-01 genome comprises 275,367 bp without terminal repeat (TR) regions, with 151 potential open reading frames (ORFs). Second, compared with other representative published strains of the genus *Cyvirus*, several evident variations are found in YC-01, particularly the orientation and position of *ORF25* and *ORF25B*. *ORF107* and *ORF156* are considered as potential molecular genetic markers for YC-01. *ORF55* (encoding thymidine kinase) might be used to distinguish YC-01 and ST-J1 from other CyHV-2 isolates. Third, phylogenetically, YC-01 clusters with the members of the genus *Cyvirus* (together with the other six CyHV-2 isolates). Fourth, 43 putative proteins are predicted to be functional and are mainly divided into five categories. Several conserved motifs are found in nucleotide, amino acid, and promoter sequences including *cis*-acting elements identification of YC-01. Finally, the potential virulence factors and linear B cell epitopes of CyHV-2 are predicted to supply possibilities for designing novel vaccines rationally. Our results provide insights for further understanding genomic structure, genetic evolution, and potential molecular mechanisms of CyHV-2.

1. Introduction

Carassius species, as important freshwater fish species, are widely farmed, and the annual worldwide production reached 2748.6 thousand tons in 2020 (FAO, 2022). The gibel carp (*Carassius gibelio*) as a novel aquaculture target has been cultured on a large scale in China [1]. Diseases cause serious economic losses to the gibel carp farming industry. Cases of cyprinid herpesvirus 2 (CyHV-2) infection were reported in gibel carp farms in central China from 2011 to 2012 [2]. Afterwards, similar cases of this severe disease were

* Corresponding author. National Pathogen Collection Center for Aquatic Animals, Shanghai Ocean University, Shanghai 201306, China.
** Corresponding author. National Pathogen Collection Center for Aquatic Animals, Shanghai Ocean University, Shanghai, 201306, China.
E-mail addresses: h-wang@shou.edu.cn (H. Wang), ysjiang@shou.edu.cn (Y. Jiang).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e32811
Received 14 March 2024; Received in revised form 27 May 2024; Accepted 10 June 2024
Available online 16 June 2024

reported by other investigations [3,4].

CyHV-2, a double-stranded DNA virus with a genome size of approximately 290 kb, is a member of the genus *Cyvirus* of the family *Alloherpesviridae* [5]. CyHV-2 infection was originally identified as herpesviral hematopoietic necrosis (HVHN) in goldfish, with high mortality, in Japan in 1995 [6]. Subsequently, cases of CyHV-2 infection have been reported from many countries and regions worldwide [7]. CyHV-2 infections in crucian carp [8], goldfish [9], gibel carp [3], and other inbred hybrids of *Carassius* [10] causing HVHN with different susceptibilities and mortalities, have been reported in China. CyHV-2 appears to be widespread in goldfish throughout the world, causing high mortality among juvenile goldfish in water at 15–25 °C [11,12]. Notably, the symptoms of CyHV-2 infection in crucian carp are significantly more severe than those in goldfish, including pale gills and swollen spleens and kidneys [11]. To date, four species, namely cyprinid herpesvirus 1 (CyHV-1), CyHV-2, cyprinid herpesvirus 3 (CyHV-3), and anguillid herpesvirus 1 (AngHV-1), have been classified into the genus *Cyvirus* (ICTV, 2023). Phylogenetically, the three cyprinid herpesviruses (CyHVs) are closely associated [13], in which CyHV-2 is more closely related to CyHV-3 than CyHV-1 while AngHV-1 is separate from CyHVs; other alloherpesviruses are much more distantly related to CyHVs [14–16]. In addition, a new *Carassius auratus* herpesvirus (CaHV) strain, which is closely related to CyHV-2 evolutionarily, and whose classification status has not been clarified according to the ICTV (2023), was reported to cause high mortality in crucian carp, and is closely related to known CyHVs despite several differences in its genome structures [17].

Research in the past decade has mainly focused on identification of CyHV-2 through detection and genome sequencing, including the development of vaccines and drugs targeting CyHV-2 [18–21]. However, so far, the whole genomes of only seven CyHV-2 isolates have been sequenced that are available for study in the GenBank (https://www.ncbi.nlm.nih.gov/genbank/accessed on May 15, 2024), and some were classified into two genotypes, Japanese (J) and Chinese (C) genotypes, respectively, based on differences in their genomes and isolation sites, which ST-J1 was classified as J genotype while SY-C1, YZ-01, and CNDF-TB2015 are C genotypes [22,23]. The availability of complete genome sequences greatly benefits studies on morphological features and functional proteins as well as the geographical distribution and genetic evolution of aquatic viruses [24]. Although CyHV-2 is distributed worldwide, different isolates have not been comprehensively compared and analyzed, including the deletions or mutations in their genomes or the genes in different isolates that result in virulence change. The lack of data as mentioned above is most likely an important reason that the pathogenesis and genetic evolution of CyHV-2 have not been fully elucidated. It would be beneficial to further explore and understand the genome structure and potential molecular pathogenic mechanisms of CyHV-2 through complete genome sequencing, comparative genomics, and molecular characterization.

In a previous work, we isolated and purified a CyHV-2 strain from samples of diseased gibel carp cultured in Yancheng, Jiangsu Province, China in 2012, which was identified and named as CyHV-2 YC-01 [25]. The moribund gibel carps displayed typical external symptoms within 24 h after infection by YC-01, including hyperemia at the base of the fins, abdomen hemorrhage, anus hyperemia, sunken eyes, and severe gill hemorrhage [25]. The identification of new isolates will assist studies to elucidate the diversity of CyHV-2 and the molecular evolution relationships among the species or different isolates. Here, we present the complete genome structure and potential molecular characterization analysis of CyHV-2 isolate YC-01 and a comparison with other representative published strains of the genus *Cyvirus*. Furthermore, we predicted potential virulence factors and linear B cell epitopes of CyHV-2 to provide possibilities for the rational design of novel vaccines. The insights provided by our study form the basis for further understanding the genomic structure, genetic evolution, and potential molecular pathogenesis of CyHV-2.

2. Materials and methods

2.1. Virus isolation, propagation, purification, and examination by transmission electron microscope

Previously, members of our laboratory isolated a CyHV-2 strain from diseased gibel carp samples collected in Yancheng, Jiangsu Province, China in 2012, and the strain was named YC-01 [25]. The virus (MOI = 1) was propagated in the *C. auratus gibelio* caudal fin (GiCF) cells (isolated and cultured by this laboratory) [26] cultured in sterile plastic flasks (75 cm^2; seed 10^7 cells per flask and incubate until cells reach 85–95 % confluency) in 1 × Medium 199 (M199; Gibco, Invitrogen, Grand Island, NY, USA), supplemented with 2 % (v/v) fetal bovine serum (FBS; Gibco) and 1 % (v/v) 100 × antibiotics containing 10,000 U/mL penicillin and 10,000 μg/mL streptomycin (Gibco) at 25 °C. The cytopathic effect (CPE) appeared at 72 h post-infection (hpi) and after the appearance of moderate CPE at 4 days post-infection (dpi), infected cells were harvested together with the supernatant when complete cytolysis occurred to prepare CyHV-2 stocks and were stored at −150 °C.

The virions were purified from the collected infected GiCF cells and supernatant (approximately 200 mL). Briefly, the mixture was frozen (−150 °C) and thawed (25 °C) three times quickly to disrupt cell membranes and release intracellular virus. The cell debris were removed by centrifugation at 3500 ×*g* for 20 min at 4 °C, and the cell-free supernatant containing virus was centrifuged at 3,2000 ×*g* for 90 min at 4 °C, followed by resuspension of the precipitate in sterile water. Purified virions were then collected using sucrose gradient centrifugation (ultracentrifuge: himac, CP70MX; rotor: Hitachi-Himac, P40ST-1516, Hitachi, Tokyo, Japan), resuspended in 200 μL of sterile water, and stored at −150 °C as previously described [27]. Then we use copper grids covered by a Formvar (400 mesh; Agar Scientific) to adsorb the purified virions (10 μL) for 10 min followed by negatively staining with 3 % phosphotungstic acid (pH 7.2–7.4) for 10 s. Finally, dried samples were examined by the transmission electron microscopy (TEM; FEI Tecnai Spirit) at 75 kV.

2.2. Viral DNA extraction, sequencing, genome assembly, and annotation

The complete genome sequence of CyHV-2 YC-01 was commercially determined by LC-Bio Technology Co., Ltd. (Hangzhou, China)

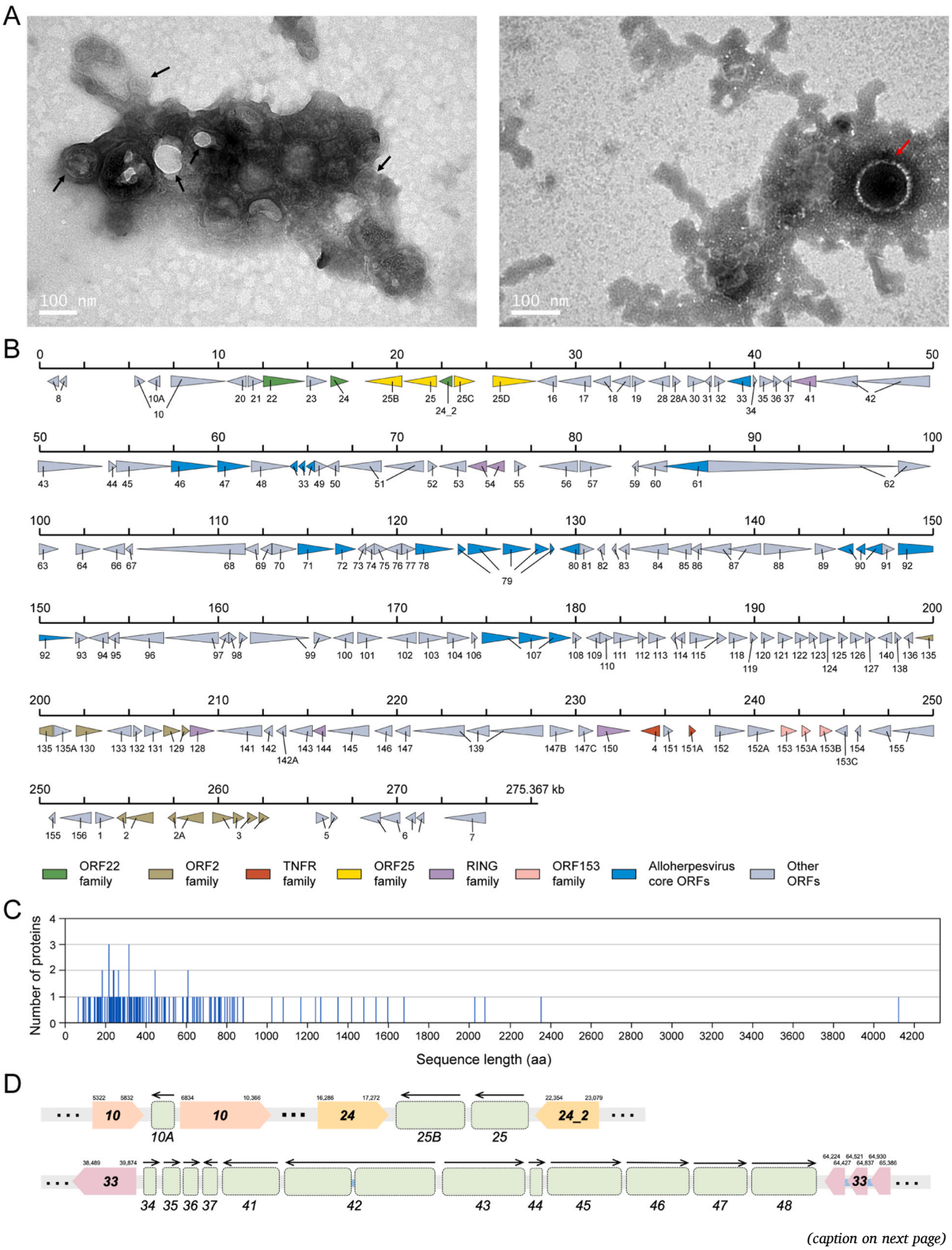
using a whole-genome shotgun (WGS) strategy through the next-generation sequencing (NGS). To degrade host-sourced nucleic acids, the viral genomic DNA from purified virus was extracted for sequencing using a FastPure Host Removal and Microbiome DNA Isolation Kit (Vazyme, Nanjing, China) according to the manufacturer's instructions (viral DNA concentration is approximately 20 ng/μL). The viral DNA samples were subjected to quality control and 1 μg of quality-checked DNA was randomly sheared into 300–500 bp-sized fragments based on the characteristics of the genome using an ultrasonicator (M220, Covaris, Woburn, MA, USA) followed by DNA library construction using a TruSeq Nano DNA Sample Prep Kit (Illumina, San Diego, CA, USA) and sequencing from both ends using a TruSeq SBS Kit v3-HS (Illumina). The WGS data were analyzed on an Illumina Hiseq Xten platform (Illumina). The raw data were base-called using Phred [28,29] (https://www.phrap.com/phred/accessed on April 15, 2019) and some low-quality reads with adapters were filtered out using fastp v0.20.0 [30] (https://github.com/OpenGene/fastp accessed on April 15, 2019) to obtain clean data, followed by fastqc [31] (https://github.com/s-andrews/FastQC accessed on April 15, 2019) analysis for quality control. Then, the optimized sequences were spliced with multiple Kmer parameters using ABySS v2.0.2 [32] (https://www.bcgsc.ca/resources/software/abyss accessed on April 15, 2019) to obtain the optimal assembly results. Meanwhile, local inner hole filling and base correction were carried out using SOAP GapCloser v1.12 [33,34] (https://sourceforge.net/projects/soapdenovo2/files/GapCloser/ accessed on April 15, 2019). Then, ATG-initiated open reading frames (ORFs) as coding gene sequences of the newly assembled genome were predicted using GeneMarkS v4.6b [35] (http://topaz.gatech.edu/GeneMark/accessed on April 16, 2019), yielding complete coding sequences (CDSs). The protein sequences encoded by the predicted genes were compared with entries in the Non-redundant protein sequence (nr) database [36] (https://www.ncbi.nlm.nih.gov accessed on April 16, 2019) using BLAST+ 2.7.1 [37] (https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/accessed on April 15, 2019; comparison criteria: E-value $\leq$ 1e-5) for annotation.

2.3. Analysis of the genome structure and molecular characterization

The genome map of CyHV-2 YC-01 was drawn using Adobe Illustrator 2021 (Adobe, San Jose, CA, USA). A graph of the lengths of the proteins encoded by ORFs on the X-axis and the number of proteins per length on the Y-axis was constructed using Geneious Prime v2022.0.2 (Biomatters, Auckland, New Zealand). The frequency of codon usage in the YC-01 genome was analyzed using the CUSP program (https://www.bioinformatics.nl/cgi-bin/emboss/cusp accessed on January 10, 2023). The conserved motifs of all ORFs, potential proteins, and promotor regions of YC-01 were scanned and constructed using the Multiple Em for Motif Elicitation (MEME) Suite v5.5.2 [38] (https://meme-suite.org/meme/accessed on February 5, 2023) with an E-value $<$ 0.05 as a cut-off. In the figure, the logo represents each column of the alignment by a stack of base or amino acid letters, with the height of each base or amino acid letter proportional to the observed frequency of the corresponding nucleotide or amino acid, and the overall height of each stack proportional to the sequence conservation at that position, measured in bits. The base or amino acid letters of each stack were ordered from most to least frequent, such that the consensus sequence from the tops of the stacks was obviously obtained [39]. The incongruent conserved bases (base mix) of some columns were named using degenerate bases (A/C to M; A/T to W; A/G to R; C/T to Y; G/T to K; C/G to S; A/C/T to H; A/G/T to D; A/C/G to V; C/G/T to B; and A/C/G/T to N) [40] instead of bases in the consensus sequence in the description of results for convenience, i.e., the base composition of all sequences are described using consistent sequences containing degenerate bases. The conserved motifs of proteins were named using the most frequent amino acid letters in each stack. The signal peptide (SP) sequences and transmembrane domains (TMDs) of proteins encoded by YC-01 were predicted using SignalP-5.0 [41] (https://services.healthtech.dtu.dk/service.php?SignalP-5.0 accessed on February 10, 2023; reference organism group: Eukarya) and TMHMM-2.0 (https://services.healthtech.dtu.dk/service.php?TMHMM-2.0 accessed on February 10, 2023), respectively. The functional features of all potential proteins encoded by YC-01 were predicted and analyzed using Swiss-Prot [42] (https://www.expasy.org/resources/uniprotkb-swiss-prot accessed on February 10, 2023), EggNOG v4.5.1 [43] (http://eggnog45.embl.de/#/app/home accessed on February 10, 2023), and the Kyoto Encyclopedia of Genes and Genomes [44] (KEGG; https://www.genome.jp/kegg/ accessed on February 10, 2023) databases, as well as using Batch CD-Search (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi accessed on February 10, 2023) in the Conserved Domains Database (CDD; https://www.ncbi.nlm.nih.gov/cdd/ accessed on February 10, 2023) with an E-value $<$ 1e-5 as a cut-off [45,46]. The conserved domains of the gene families of YC-01 were predicted using Simple Modular Architecture Research Tool [47] (SMART; https://smart.embl.de accessed on February 20, 2023). The nucleotide or amino acid sequence alignments of certain ORFs were analyzed through Multiple Alignment using Fast Fourier Transform (MAFFT), Clustal Omega or Needle (Needleman-Wunsch algorithm) in the EMBOSS program [48] (https://www.ebi.ac.uk/Tools/emboss/accessed on February 20, 2023) or using SnapGene software v4.3.7 (GSL Biotech LLC, Boston, MA, USA) or Geneious Prime v2022.0.2 (Biomatters) followed by visualization through ESPript 3.0 [49] (https://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi accessed on February 20, 2023) or calculation using Geneious Prime v2022.0.2 (Biomatters). The *cis*-acting elements in the promoter regions (2000 bp upstream of a CDS) were scanned and analyzed using Nsite program [50] (http://www.softberry.com/berry.phtml?topic=nsite&group=programs&subgroup=promoter accessed on February 20, 2023; database: RegsiteAN DB of human or animals) and PlantCARE [51] (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/accessed on February 20, 2023). The number of promoters containing *cis*-acting elements was counted using GraphPad Prism v9.4.1 (GraphPad Software, San Diego, CA, USA). Then, the consensus sequences of some *cis*-acting elements were visualized using WebLogo 3 [39], as well as the putative translation initiation site (TIS) regions.

2.4. Comparison of genomic structures and evolutionary relationships

The sequence identities of the genomes, CDS, and proteins among YC-01 and six closely related strains in the genus *Cyvirus* and



(caption on next page)

Fig. 1. Purified virions image, genome map, and sequence lengths of encoded proteins of CyHV-2 YC-01. (A) Purified virions of YC-01 under transmission electron microscope (TEM). Black and red arrows indicate YC-01 virions with intact and broken envelopes, respectively, scale bar = 100 nm. (B) Arrows indicate the size, location, and orientation of the 151 open reading frames (ORFs), with nomenclature lacking the "ORF" prefix given below. The number of arrows in an ORF indicates the number of exon-formed coding sequences (CDSs) it contains because of the presence of introns. Three ORFs belonging to the ORF22 family are indicated by green arrows (ORF22 family). Six ORFs belonging to the ORF2 family are indicated by brown arrows (ORF2 family). Two ORFs belonging to the tumor necrosis factor receptor (TNFR) family are indicated by red arrows (TNFR family). Four ORFs belonging to the ORF25 family are indicated by yellow arrows (ORF25 family). Five ORFs with a really interesting new gene (RING)-finger domain are indicated by purple arrows (RING family). Three ORFs belonging to the ORF153 family are indicated by pink arrows (ORF153 family). Twelve alloherpesvirus core ORFs are indicated by blue arrows (Alloherpesvirus core ORFs). The other 117 unclassified ORFs are indicated by grey arrows (Other ORFs). (C) Graph of the sequence lengths of the proteins encoded by the 151 ORFs of YC-01 (x-axis) *vs*. the number of proteins per length (y-axis). (D) Thick arrows with starting and ending positions above indicate the relative size, location, and orientation of *ORF10*, *ORF24* (*ORF24_2*), and *ORF33*. The green squares indicate the relative sizes of the other nearby ORFs with black arrows showing their orientations. Other ORFs are replaced by ellipses.

CaHV (the classification status has not been clarified according to the ICTV 2023 and it is probably considered as a novel isolate of CyHV-2) were aligned using MAFFT by Geneious Prime v2022.0.2 (Biomatters). The graph of the genomic structure comparison among the eight strains was drawn using Adobe Illustrator 2021 (Adobe). Furthermore, a phylogenetic tree among YC-01 and 18 other strains in the family *Alloherpesviridae* was constructed based on the amino acid sequences of the DNA polymerase catalytic subunit using the neighbor-joining method in MEGA v11 (https://www.megasoftware.net accessed on May 18, 2024) with bootstrap values of 1000 replications.

2.5. Prediction of B cell epitopes of CyHV-2

Linear B cell epitope prediction was carried out on eight CyHV-2 proteins, ORF25 (Protein ID: QIV66841.1), ORF57 (QIV66875.1), ORF66 (QIV66882.1), ORF72 (QIV66888.1), ORF92 (QIV66908.1), ORF115 (QIV66930.1), ORF131 (QIV66949.1), and ORF132 (QIV66948.1) that were identified as the immunogenic proteins of CyHV-2 in previous work [27]. We used the BepiPred-2.0 algorithm [52] embedded in the B cell epitope prediction analysis tool available in the Immune Epitope Database (IEDB) [53] (http://tools.iedb.org/bcell/accessed on March 16, 2023). For each protein, the epitope probability score for each amino acid residue was retrieved. The residues with scores above the threshold (the default value was 0.500) were predicted to be part of an epitope and marked with "E". Then, the potential B cell epitopes were predicted using a cutoff of 0.550 (corresponding to a specificity of 0.817 and a sensitivity of 0.292, approximately) and considering sequences having not less than seven amino acid residues. In addition, the B cell epitopes of the above eight proteins were also predicted using ABCpred server [54] (https://webs.iiitd.edu.in/raghava/abcpred/index.html accessed on March 16, 2023), in which a threshold of 0.50 showed 65.93 % accuracy with equal sensitivity and specificity using a default window length of 16. The secondary structures of the above eight proteins were predicted in PSIPRED server v4.0 [55] (http://bioinf.cs.ucl.ac.uk/psipred accessed on March 16, 2023), from which the peptide sequences predicted above containing helix and strand structures were removed.

3. Results

3.1. Purification, genome structure, and composition of CyHV-2 YC-01

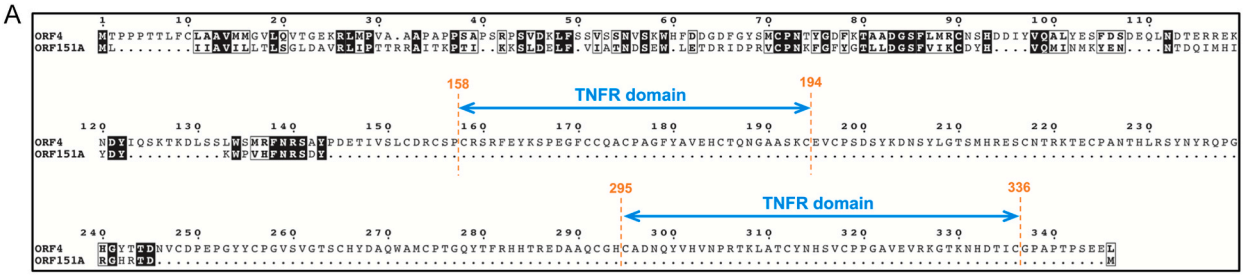
First, we examined the purified CyHV-2 YC-01 virions using the TEM. As shown in Fig. 1A, there is no detectable cell debris in the purified virions and most of them showed the characteristic morphology of mature virions, with an intact envelope surrounding the nucleocapsid, and some virions are with a broken envelope. The mean diameter of the virions is 110–120 nm. Then through whole-genome sequencing, we obtained a total of 3560 MBytes (MB) of raw data using the NGS, and after removing low-quality data, 3396 MB of clean data remained, with an average G + C content of 37.86 %, in which the Q20 (percentage of bases with a Phred value > 20 of the total number of bases) and Q30 (percentage of bases with a Phred value > 30 of the total bases) were 98.79 % and 95.61 %, respectively. Finally, we successfully assembled the complete genome sequence of CyHV-2 YC-01 and submitted it to GenBank (Accession Number: MN593216) together with predicted functional annotation information.

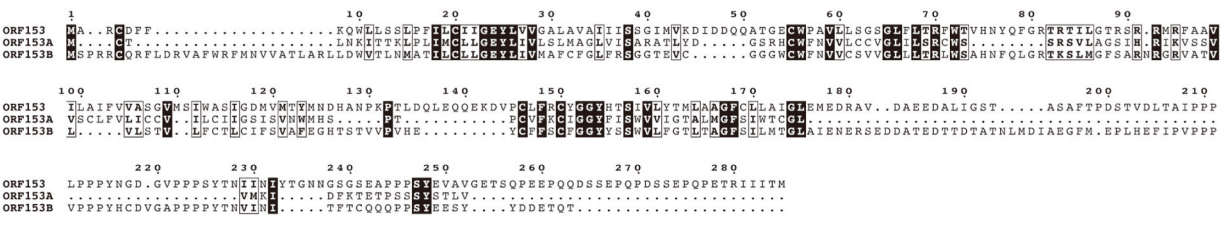
The linear double-stranded DNA genome of YC-01 isolate is 275,367 bp in length, with an overall G + C content of 51.79 %. We analyzed the frequency of codon usage in the YC-01 genome, and the results showed that the coding GC content of the ORFs is 51.79 %, while the GC content of 1st, 2nd, and 3rd letters in the triplet codons are 51.90 %, 51.41 %, and 52.05 %, respectively. After the ergodic analysis of the assembled genome using GeneMarkS and comparing the ORFs and protein sequences encoded by the predicted genes obtained above with entries in the nr database, mainly referring to CyHV-2 ST-J1 (reference strain), we determined that the whole YC-01 genome contains 151 unique protein-coding ORFs, and a tentative gene layout was composed (Fig. 1B). The ORF-encoded proteins range in length from 62 (ORF106) to 4121 (ORF62) aa, with an average length of 519.54 aa (Fig. 1C and Table S1). In terms of genetic regions, 27 ORFs contain introns, among which *ORF79* has four introns, *ORF3* and *ORF6* have three introns, and the other 24 ORFs have one or two introns (Table S1). Eighty-one ORFs are located on the positive strand and 70 ORFs are on the negative strand in the YC-01 genome. Like reference strains of CyHV-2, *ORF33* is repeatedly present in the YC-01 genome, with one part being located at 38,489–39,874 nt and another three parts clustered at 64,224–64,427, 64,521–64,837, and 64,930–65,386 nt (on the negative strand). Similarly, *ORF10* has two exons separated by introns of 1001 bp and *ORF10A* (on the negative strand) is inserted into it. Interestingly,

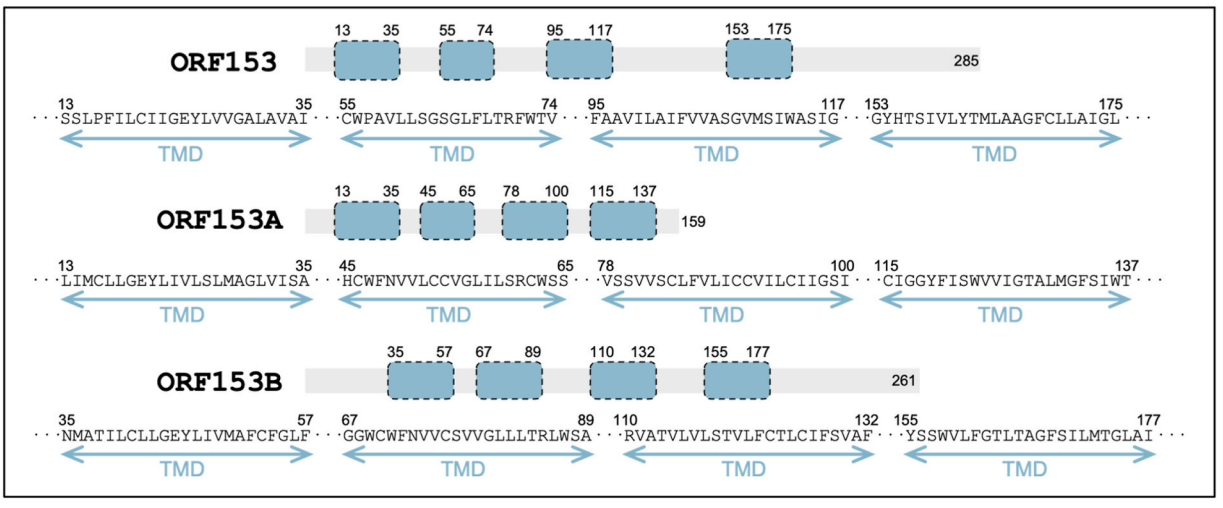
Table 1
Functional features of the predicted proteins encoded by CyHV-2 YC-01.

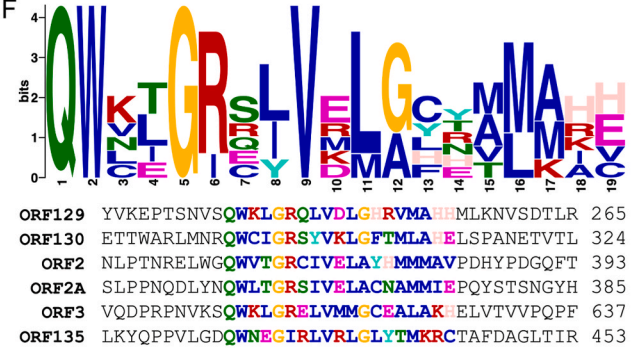
Predicted function category [a]	Putative proteins	Predicted functional domains		Function features
		Putative encoded proteases or protein homologs	Putative domains	
Nucleotide synthesis, metabolism, or transport (7)	ORF19	Deoxyadenosine or deoxycytidine kinase	–	Nucleotide transport and metabolism
	ORF140	Thymidylate kinase	–	
	ORF141	Ribonucleotide reductase (RNR) α subunit	–	Balanced supply of the four dNTPs required for DNA replication and repair
	ORF23	Ribonucleotide reductase (RNR) β subunit	–	
	ORF55	Thymidine kinase	–	Virulence-related protein and nucleoside salvage pathway
	ORF64	Nucleoside transporter as equilibrative nucleoside transporter 1 (ENT1)	–	Nucleotide synthesis
	ORF123	Deoxyuridine triphosphatase (dUTPase)	–	Nucleotide metabolism
DNA replication or damage repair (6)	ORF33	DNA packaging terminase subunit 1	ATPase domain	DNA encapsidation and replication
	ORF47	DNA packaging terminase subunit 2	–	
	ORF46	Helicase-primase primase subunit	–	Viral DNA replication
	ORF71	Helicase-primase helicase subunit	AAA domain	
	ORF79	DNA polymerase catalytic subunit	–	
	ORF98	Uracil-DNA glycosylase (UDG)	–	Repair of uracils in DNA and integrity of genetic information
Viral capsid morphogenesis (4)	ORF66	Capsid triplex subunit 1	–	Viral capsid morphogenesis
	ORF72	Capsid triplex subunit 2	–	
	ORF78	Capsid maturation protease	–	
	ORF92	Major capsid protein (MCP)	–	
Homologs of other viral proteins (5)	ORF32	Homologs of Singapore grouper iridovirus (SGIV) proteins	–	–
	ORF57	Homolog of crocodilepox virus (CRV) protein CRV155	–	–
	ORF64	Homolog of the late lytic protein BDLF3 of Epstein–Barr virus (EBV)	–	Immune evasion
	ORF99	Homolog of *Torovirinae* spike glycoprotein	–	The binding of virions to the host cell receptors and membrane fusion
	ORF139	Homolog of the C-terminal structure of the poxvirus B22R protein	–	–
Analogs of cellular functional proteins (22)	ORF16	G protein-coupled receptors (GPCRs)	–	Intracellular signaling
	ORF28	–	NAD(P)(+)-binding (NADB) Rossmann-fold domain	Numerous dehydrogenases of metabolic pathways such as glycolysis
	ORF25 ORF25B ORF28A ORF146	–	Immunoglobulin (Ig) domain	–
	ORF31	Eukaryotic placenta-associated 8 (PLAC8)	–	–
	ORF48 ORF104	Protein kinases	–	Protein phosphorylation
	ORF41 ORF54 ORF144 ORF150	–	C3HC4 type really interesting new gene (RING) finger domain	E3 ubiquitin ligases with ubiquitination and protein transferase activity
	ORF128	–	SPRY (B30.2) domain	Tripartite motif (TRIM) proteins in anti-retroviral determinant
	ORF68 ORF70	Chromosome segregation ATPase	–	Cell cycle control, cell division, and chromosome partitioning
	ORF62	–	Ovarian tumor (OTU)-like cysteine protease domain	Function as deubiquitinases (DUBs) or ubiquitin thiolesterases
	ORF94	Trypsin-like serine proteases	–	Signal anchor
	ORF112	–	Adenosine deaminase z-alpha domain	Z-DNA binding
	ORF145	–	RNA recognition motif (RRM)	Post-transcriptional gene expression
	ORF4 ORF151A	Tumor necrosis factor receptor (TNFR)	–	Cellular physiological activities

Note.
[a] The numbers in brackets indicate the number of proteins included in the category.









(caption on next page)

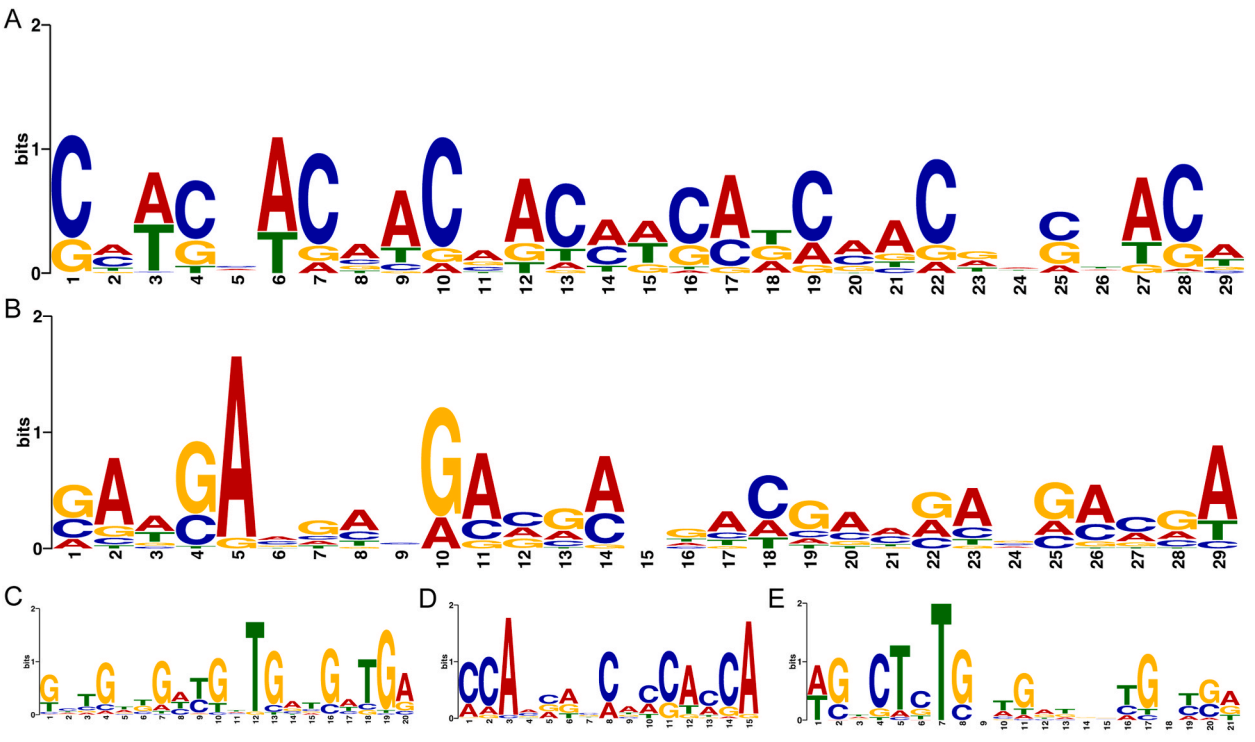
Fig. 2. The predicted conserved regions of gene families encoded by YC-01. (A) The amino acid sequence alignment and conserved domain prediction of ORF4 and ORF151A (TNFR family). The blue bidirectional arrows and orange dashed targeting areas indicate conserved domains, with the starting and ending positions marked above the corresponding dashed lines. (B) The amino acid sequence alignment of ORF153, ORF153A, and ORF153B (ORF153 family). (C) The conserved domain prediction of ORF153 family members. Cyan frames indicate predicted domains, with the start and end positions marked in black, and the corresponding amino acid sequences are listed below, with domain positions restricted by cyan bidirectional arrows (start and end positions are marked above in black). Other sequences of the protein are replaced by ellipses. (D) The amino acid sequence alignment and conserved domain prediction of ORF25 family members. The domain sites are marked in red, with domain positions restricted by red bidirectional arrow (only the sequences of domain positions are shown, and other sequences are not listed). (E) The conserved domain prediction of RING family members. The domain sites are marked in light pink, with the start and end positions marked above in black. The domain positions are restricted by a light pink bidirectional arrow (only sequences of domain positions are shown, and other sequences are replaced by ellipses). The E-value are 1.54e-05 (ORF41), 1.49e-01 (ORF54), 3.62e-01 (ORF128), 6.14e-05 (ORF144), and 4.05e-05 (ORF150), respectively. (F) The conserved motifs of the ORF2 family. The logo represents each column of the alignment by a stack of amino acids, with the height of each amino acid being proportional to its observed frequency, and the overall height of each stack is proportional to the sequence conservation at that position, measured in bits. The width of the motif is indicated in the x-axis. The E-value is 4.4e-019.

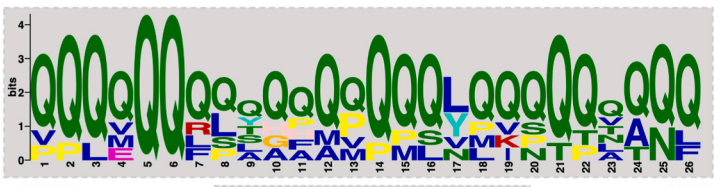
ORF24 in the YC-01 genome is also encoded in two parts (at 16,286–17,272 and 22,354–23,079 nt, respectively, named *ORF24* and *ORF24_2*) and they have opposite transcriptional orientations in different strands, as depicted in Fig. 1B and D, and Table S1. In addition, there are 12 core ORFs (*ORF33*, *ORF46*, *ORF47*, *ORF61*, *ORF71*, *ORF72*, *ORF78*, *ORF79*, *ORF80*, *ORF90*, *ORF92*, and *ORF107*) in the YC-01 genome (Fig. 1B and Table S1) that are significantly conserved among alloherpesviruses.

3.2. Features of the predicted functional proteins encoded by CyHV-2 YC-01

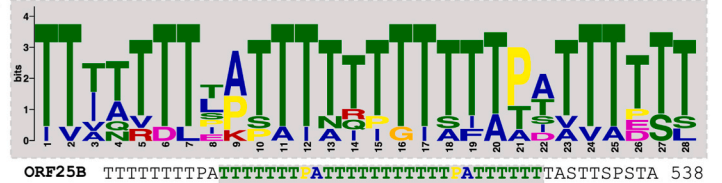
Among the 151 potential proteins ~~of~~ encoded by YC-01 using bioinformatics analysis, 27 membrane proteins were predicted, among which 15 are classified as type I membrane proteins, ten are type III membrane proteins, and the remaining two, ORF25C and ORF132, should be considered as type II membrane proteins. Four ORFs (*ORF66*, *ORF72*, *ORF78*, and *ORF92*) were predicted to encode nucleocapsid proteins. Meanwhile, 27 proteins were predicted to possess a SP sequence, and 40 putative proteins were predicted to possess one or more TMDs, as shown in Table S1. Notably, ORF64, ORF114, ORF152A, and ORF16 have 10, 8, 7, and 7 TMDs, respectively, and were predicted to be important membrane proteins (type III) of CyHV-2. Interestingly, the ORF25 family (ORF25, ORF25B, and ORF25D as type I membrane proteins; ORF25C as type II membrane protein) and the ORF153 family (ORF153 and ORF153A–B as type III membrane proteins) all have one or more TMDs. Meanwhile, the functional features and conserved domains of all 151 potential proteins were predicted using the available databases, as shown in Table 1, Table S1, Table S2, and Table S3. Among them, 43 putative YC-01 proteins were predicted to be related to enzymes, including some functional proteases or conserved domains, as well as functional proteins, which are mainly divided into five categories (Table 1 shows more details). Seven putative proteins (category 1) encoded by YC-01 were predicted to be mainly related to nucleotide synthesis, metabolism, or transport. Six putative proteins (category 2) encoded by YC-01 were predicted to be primarily associated with DNA replication or damage repair. Four putative proteins (category 3) encoded by YC-01 were predicted to be predominantly related to viral capsid morphogenesis. Five putative proteins (category 4) were predicted to be homologs of other viral proteins. 22 putative proteins (category 5) encoded by YC-01 were predicted to be similar to cellular functional proteins.

In addition, as shown in Fig. 1B and Table S1, like ST-J1, six gene families were identified in YC-01 genome, including ORF2 (members: ORF2, ORF2A, ORF3, ORF129, ORF130, and ORF135), ORF22 (members: ORF22, ORF24, and ORF24_2), ORF25 (members: ORF25, ORF25B, ORF25C, and ORF25D), TNFR (members: ORF4 and ORF151A), RING (members: ORF41, ORF54, ORF128, ORF144, and ORF150), and ORF153 (members: ORF153, ORF153A, and ORF153B) families, respectively. They probably have the same functions or share some similar sequences and motifs. The multiple sequence alignment results showed that the CDSs and amino acids of members within each gene family share low sequence identities (Table S4). The TNFR family has two members, ORF4 and ORF151A, named for the similarity of their encoded proteins to TNFR, and ORF4 was predicted to possess two putative TNFR domains located at the middle and C-terminus of the protein, respectively. Although ORF151A has no predicted TNFR domain, it shares several conserved residues with ORF4 (Fig. 2A). The ORF153 family (exclusive to CyHV-2) members share several conserved residues (Fig. 2B), and they are all predicted to be type III membrane proteins with four TMDs, thus they are likely to be important membrane proteins of CyHV-2 (Fig. 2C). The ORF25 family members encode highly divergent predicted type I membrane proteins, except for ORF25C, which is considered as a type II membrane protein; however, they all have one conserved TMD located at the C-terminus (Fig. 2D and Fig. S1A), which in ORF25 and ORF25B exhibit weak similarities to Ig domains (Table 1). Notably, these four proteins all have multiple continuous threonine (T) tandem repeats (the number is not less than 4) with a scattered distribution (Fig. S1B). As shown in Fig. 2E, the RING family members all possess a consensus C3HC4 type RING-finger (RING-HC) domain, i.e., CX2CX(9–39)CX(1–3)HX(2–3)C/HX2CX(4–48)CX2C with the cysteine (Cys) and histidine (His) representing zinc binding residues (a Cys occupying the fifth coordination site is defined as RING-HC). The ORF2 family members are numerous and highly divergent, with only some short regions containing the conserved motif QWXXG shared by all (E-value: 4.4e-019) whose function is unknown (Fig. 2F and Fig. S1C). The members of the ORF22 family are highly divergent and hardly share similarity, in which ORF24 and ORF24_2 are much shorter than ORF22 in length (Fig. S1D).





```
ORF68   EHALQQELNRQQQQQQQQQQQQQQQQLQQQQQLQQQQRDTFATEKA 1632
ORF10   LGSLFKKPSVPQQEQQQQYQQQQQQQYQQQQQQQQQYHSYNEQETF 285
ORF6    KTEPQSQVQVQQQVQQRPAAAQPQQQVQVQQQAQQQAAPAAAQPQQ 495
ORF113  SGQPKQQLLLQPQQQQLLLQPQQQQLLLQPQPQQQQLLQPQQFVTF 207
ORF48   PRFTILKRPTQQQQQQQQQQ QQPPPSYQKNQTTANFISAALEEHGP 109
ORF155  SRSKDIAEGAQQLMQQFLSGFMMQMQNMQQQQNTQQSNNNQ      1414
ORF78   QAAVLPAPAPVQQQQQQSTQQAVQQQL ISTQVANLAAATGRTPSA 638
```



```
ORF25B  TTTTTTTTPATTTTTTTPATTTTTTTTTTTTPATTTTTTTTASTTSPSTA 538
ORF25   TPSTVPTTPKTTTQRTTTPTTTTTTTTTTTTPATTTTTTTIAATPPPTT 556
ORF25D  TLPTTTTTTTTTTTTTTTTTPTTTTTRPTTTTTPTTTTTTTTPATTTTPTP 556
ORF25C  GPTTVGPSTTTTITTTTLASTTTTTGTTTTTSTTTTSTLANTTLLPTT 291
ORF64   TSGHNNGSTSTTVATTTSATATTTTTTSTTPAATTESSTTGASSASTS 143
ORF139  LIQSVIRNATTVATTTTEPPTTNIITTTTTTTDTTTDTTASTMNSTTNE 390
ORF131  PPSPVTAAPIITINVDLLKTTTTQTTTTFAPITTAPTTPSPQQLAASR 144
ORF147B EPVTVETFVTTTIATTTIATTIATTTIAIAATVVTTTLANILTGTTVV 373
```

```
ORF68   QQQQQRQRQQQQQQQQQQQQRQRERERQQR 64
ORF10   LPLQHVDVVQQQQQQQQQQQTTPPIRLTG 656
ORF77   AAATTSPPPMQQQQQQQQQKTAADSQHQRN 42
ORF17   VSAQCTYTDLDSQQQQQQQPDAEPAQQTKR 137
ORF73   DYEPAASDPSRPQQQQQQQIYTLSQSVTVH 49
ORF113  PVAASTPKPQQTQQTQQQPPVSFQPPPQY 264
ORF3    SSDDEDFEDEQPQQEQQPPVVSEKSYQSD 94
ORF108  HRQQQPSSSSNSQQQQQQPPNHIPPPPPVG 157
ORF75   NGGHFTVTETQRQQQQQQAGANGGSQLCRL 96
ORF155  NTMARRSEPPQQQQEQEQQQAAPIVAPVV 541
ORF62   VDSDGGGGGGISLQQQQQQQDDDVVVDGAEI 78
ORF6    QQAQQQAAPAAAQPQQQQPAAAVQVPQVQ 508
ORF150  SSRHHYQPPRQRPQQQQNPFDNNNMATFR 223
ORF78   GVPVAQASTLVEAQQQQQPTLVSVSAGLT 695
ORF97   QFNLPPGVELTGTQQQQQQPLVPDLSFKI 678
ORF154  ELHVYRRRLQQPQLQQLQPDIIPLMRQGS 93
ORF143  LVVVAEPEPEQQ QQPEQPDNDQAEQQTI 36
ORF88   AGEEAAGGSKVRQQQQPPEKKQRKSAAD 868
ORF51   EEEGGEEGEEQPTQSNQQPRKRLRKPSSS 1431
ORF142A STEATAVTTMQPTTIQQQPTTQPTQPQTR 55
ORF133  QLKDDYLNQNQLKQQSQQLSKSSGCLNRR 412
ORF8    QEAAREAERERAAQQEQQRAEEQEQAEPPA 306
ORF145  QPQPSVPSPQRRQSSSQQPKPAPVHHQEP 214
```

(caption on next page)

Fig. 3. The conserved motifs of ORFs and predicted proteins encoded by YC-01. (A–E) The five conserved motifs of the 151 ORFs encoded by YC-01 were scanned and constructed using Multiple Em for Motif Elicitation (MEME) Suite v5.5.2. The logos represent each column of the alignment by a stack of bases, with the height of each base being proportional to the observed frequency of the corresponding nucleotide, and the overall height of each stack being proportional to the sequence conservation at that position, measured in bits. The bases of each stack are ordered from most to least frequent. The width of the motif is indicated in the x-axis, which each motif describes a pattern with a fixed width. (F–H) The prediction of three conserved motifs in all 151 potential proteins using MEME Suite v5.5.2 (E-value <0.05 as a cut-off). The logos represent each column of the alignment by a stack of amino acids, with the height of each amino acid being proportional to its observed frequency, and the overall height of each stack is proportional to the sequence conservation at that position, measured in bits, and the width of the motif is indicated in the x-axis. The corresponding conserved amino acid sequences are listed below the motifs.

3.3. The conserved motifs of predicted ORFs, proteins, promoters, and cis-acting elements of CyHV-2 YC-01

All CDSs or amino acid sequences of the 151 ORFs or potential proteins encoded by YC-01 were aligned and scanned using the MEME Suite v5.5.2. As shown in Fig. 3A–E, there are five predicted conserved motifs among the CDSs of the 151 ORFs encoded by YC-01, including CMWCHACAACAACMWCADCMACDDSHACW (shared by 110 ORFs; width: 29; E-value: 9.1e-022), SAWGAVGAVGAVGMNGACGAMRAVGAMGA (shared by 146 ORFs; width: 29; E-value: 2.5e-021), GHTGWKGHYGHTGATGWTGA (shared by 148 ORFs; width: 20; E-value: 4.1e-039), CCAMVRBCAMCAMCA (shared by 145 ORFs; width: 15; E-value: 9.5e-003), and WGWCTCTGDWGRKDNTGNYSR (shared by 135 ORFs; width: 21; E-value: 2.5e-014), which the second and third motifs show GA-rich features and the forth motif shows CA-rich feature. Similar cases were found in other CyHVs (data not shown), suggesting that GA- or CA-rich features are common among the majority of ORFs in CyHVs. As shown in Fig. 3F–H and Table S5, interestingly, there are three predicted conserved motifs containing continuous glutamine (Q)-rich or threonine (T)-rich tandem repeats, including QQQQQQQQQQQQQQQLQQQQQQQQQ (motif 1; shared by seven ORFs; width: 26; E-value: 8.4e-034), QQQQQQQQP (motif 2; shared by 23 ORFs; width: 9; E-value: 1.6e-019), and TTITTTTTPTTTTTTTTTTTPATTTTTT (motif 3; shared by eight ORFs; width: 28; E-value: 3.9e-022). Coincidentally, the eight ORFs sharing motif 3 were all predicted to be membrane proteins and possess one or more TMDs, especially the ORF25 family (Table S1 and Fig. S1B).

Furthermore, a conserved motif named BTGWKGYTGBTGBTG (width: 15; E-value: 5.8e-019) containing continuous "TTG" tandem repeats is shared by all 151 promoters of the YC-01 ORFs (Fig. 4A). We scanned and analyzed the *cis*-acting elements in the promoter regions of the 151 ORFs using the Nsite program in RegsiteAN DB database (human or animals). The results revealed 671 *cis*-acting elements belonging to 215 categories. As shown in Fig. 4B and Tables S5 and 22 categories of *cis*-acting elements are present in more than five promoters of ORFs (some promoter regions might contain more than one of the same elements in different sites). Then, the conserved sequences of the top six elements, including AT-rich (appears 65 times; shared by 23 promotors), specificity protein 1 (Sp1; appears 30 times; shared by 19 promotors), glucose transporter type 2 TAAT motif (GLUT2TAAT; appears 19 times; shared by 16 promotors), interferon-stimulated response element (ISRE; appears 19 times; shared by 15 promotors), AC-box (appears 11 times; shared by 11 promotors), and E-box (appears 11 times; shared by 11 promotors) were visualized (Fig. 4C). Meanwhile, as shown in Table S6, one promoter (of *ORF52*) has a typical TATA-box, four promoters (of *ORF6*, *ORF90*, *ORF152*, and *ORF152A*) have GC-boxes, and one promoter (of *ORF63*) has a CCAAT-box (also known as a CAAT-box). Additionally, an element named G-string recurs many times, and was identified in only five ORFs' promoters (Table S5). Moreover, the *cis*-acting elements in the promoter regions of the 151 ORFs were also scanned in the PLantCARE database (plants), and 16,678 elements belonging to 111 categories were predicted (Fig. 4D and Table S7). The results showed the presence of some typical *cis*-acting elements, such as the CAAT-box (possessed by all 151 promotors) and the TATA-box (possessed by 143 promotors), in the promoters of YC-01, which are similar to those of plants, as well as Sp1, AT-rich, and GATA, which also matched human or animal *cis*-acting elements. Then the putative 151 ORFs with their promoters (2000 bp upstream of the start codon) were analyzed. The results showed that 62 translation initiation sites (TISs) of the genes have a RNNATGG (R represents purine) sequence, considered as a Kozak motif (Fig. 4E).

3.4. Comparison of the genome structures, ORF features, and evolutionary relationships

The genomes of three available reference CyHV-2 isolates that have been sequenced and annotated, including ST-J1, SY-C1, and SY, comprising 290,304, 289,365, and 290,455 bp with 150, 140, and 150 unique ORFs, respectively, and all possess unique (U) and terminal repeat (TR) regions at each end. Our results showed that the genome of YC-01 is 275,367 bp in length, without U and TR features, encoding 151 unique ORFs, and sharing 98.24 %, 97.42 %, and 97.31 % sequence identities with ST-J1, SY-C1, and SY, respectively (Table 2). Considering the absence of TR regions in the YC-01 genome, four ORFs (*ORF5*, *ORF6*, *ORF7*, and *ORF8*) are not repeatedly present compared with those in the other three CyHV-2 isolates, among which *ORF8* is located upstream (the first gene) and the *ORF5–7* segment is located downstream (the last three genes) in the YC-01 genome (Fig. 5). Meanwhile, YC-01 encodes *ORF10A*, which is absent in the CyHV-1 genome, CyHV-3, and AngHV-1, like other CyHV-2 isolates. *ORF33* is also repeatedly present in the YC-01 genome similar to other CyHVs. Notably, different from the other seven strains (Fig. 5), *ORF24* in the YC-01 genome is encoded in two parts, named as *ORF24* (987 bp) and *ORF24_2* (726 bp) for better differentiation, respectively, which have opposite transcriptional orientations (Fig. 1B and Table S1). Its specific functions need to be verified experimentally. Particularly, compared with ST-J1, SY-C1, and SY, the orientations of *ORF25* (like that of CyHV-1) and *ORF25B* are opposite, in which the positions of the two ORFs are reversed in YC-01 (Fig. 5). Meanwhile, the YC-01 genome shares the greatest sequence identity with ST-J1, and SY-C1 lacks 10 unique ORFs compared with YC-01 (Fig. 5). In addition, 116 out of the 150 unique ORFs encoding proteins (excluding ORF24_2) are identical (100 %) between YC-01 and ST-J1, and only 52 ORFs in YC-01 are identical (100 %) with their counterparts in ST-J1, SY, and SY-C1,

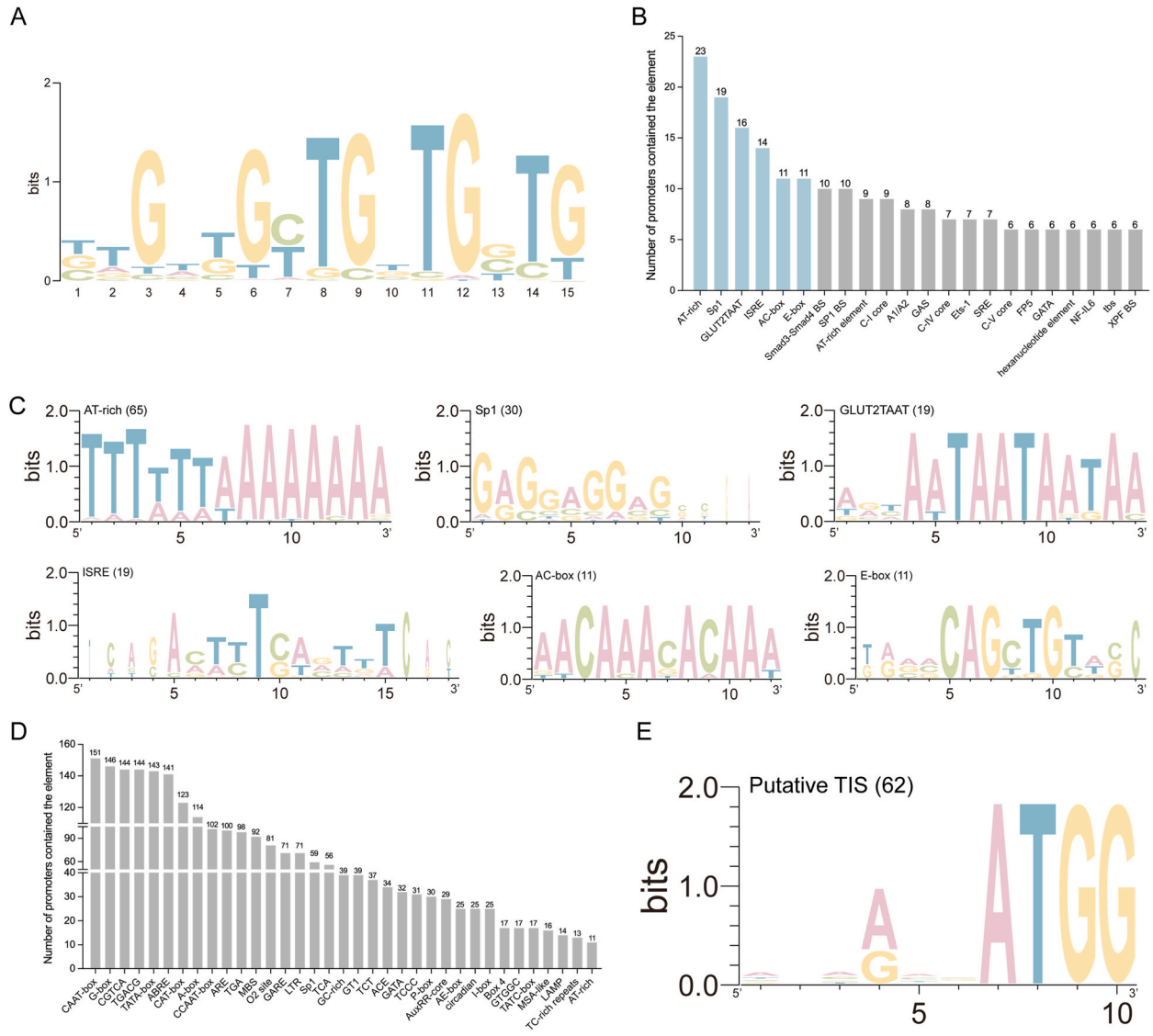


Fig. 4. The *cis*-acting elements in the promoter regions and putative translation initiation site motifs of YC-01. (A) The conserved motifs of 151 promoters of ORFs scanned and constructed by MEME Suite v5.5.2 (E-value <0.05 as a cut-off). (B) The number of promoters contained the *cis*-acting elements searched in RegsiteAN DB database (human or animals) through the Nsite program. The quantity of each element is marked on the corresponding bar (only elements with a quantity >5 are shown), which refers to the number of promoters containing the element. If the promoter region of an ORF contains more than one of the same element (i.e., one element appears more than once in one promoter of the ORF), this element of the same promoter will be counted only once instead of repeatedly. The columns of the top six elements are filled in cyan and the other are grey. (C) Visualization of consensus sequences of the top six *cis*-acting elements by number in WebLogo 3. The number in brackets indicates the number of that element appearing in the YC-01 genome, i.e., all the same elements together to construct the visual sequence logos. (D) The number of promoters containing the *cis*-acting elements searched in the PlantCARE database. The quantity of each element is marked on the corresponding bar (only elements that with a quantity >10 and having predicted functions are shown). The counting method is the same as described in (B). (E) The putative translation initiation site (TIS) region was visualized using WebLogo 3. The number in brackets indicates the quantity of genes possessing a TIS in the YC-01 genome.

including the homologs from CaHV (Table S1).

The CyHV-3 genome remains the largest overall and CyHV-1 contains the fewest ORFs among CyHVs; however, like the three CyHV-2 isolates, the YC-01 genome has more complexity in terms of copy size and arrangement. The absence of the TR regions in the YC-01 genome means that the counterparts of some ORFs in the TR regions of CyHV-3 (*ORF1–8*) and CyHV-1 (*ORF2*, *ORF3*, *ORF5*, and *ORF8*) related to flanking genes are not repeatedly present. Moreover, except for *ORF8*, which is in the upstream region, the other ORFs are located closely downstream in the YC-01 genome (Fig. 5). Unlike the CyHV-1 and CyHV-3 genomes, *ORF2A* is inserted closely downstream (on the negative strand and before *ORF2*) in the YC-01 genome, similar to the other CyHV-2 isolates. Compare with closely

Table 2
Genome features of CyHVs, CaHV, and AngHV-1.

Virus		Size (bp)			Nucleotide composition (%)	No. of ORFs				Identity (%) [j]
		Genome	U [f]	TR [g]	G + C	Genome [h]	Unique [i]	U [g]	TR [h]	
CyHV-2	YC-01	275,367	–	–	51.80	151	–	–	–	***
	ST-J1 [a]	290,304	260,238	15,033	51.70	154	150	146	4	98.24
	SY-C1 [b]	289,365	259,555	14,905	51.60	143	140	137	3	97.42
	SY [c]	290,455	259,749	15,353	51.60	154	150	146	4	97.31
CaHV [d]		275,348	–	–	51.73	150	–	–	–	91.23
CyHV-1 [a]		291,144	224,784	33,180	51.30	143	137	131	6	41.46
CyHV-3 [a]		295,146	250,208	22,469	59.20	163	155	147	8	44.36
AngHV-1 [e]		248,526	227,258	10,634	53.00	134	129	124	5	37.23

Note.
[a] Adapted with permission from Ref. [19].
[b] Adapted with permission from Ref. [22].
[c] Adapted with permission from Ref. [64].
[d] Adapted with permission from Ref. [17].
[e] Adapted with permission from Ref. [15].
[f] U represents the unique region in the genome (except in that of YC-01 and CaHV).
[g] TR represents the terminal repeat in the genome (except in that of YC-01 and CaHV).
[h] Number of ORFs in U plus two copies of TR (except in that of YC-01 and CaHV).
[i] Number of ORFs in U plus one copy of TR (except in that of YC-01 and CaHV).
[j] The genome identities of eight strains were aligned through MAFFT by Geneious Prime v2022.0.2. CyHV-2 YC-01 (GenBank accession number: MN593216); CyHV-2 ST-J1 (NC_019495.1); CyHV-2 SY-C1 (KM200722.1); CyHV-2 SY (KT387800.1); CaHV (KU199244.1); CyHV-1 (NC_019491.1); CyHV-3 (NC_009127.1); AngHV-1 (NC_013668.3).

downstream in the YC-01 genome, CyHV-1 lacks *ORF1*, *ORF4*, and *ORF7* but possesses *ORF1A*; two genes, *ORF4* and *ORF140*, are apparently rearranged among the three strains. In detail, *ORF4* is located downstream in YC-01, like the other CyHV-2 isolates (*ORF4* is located at the TR regions and duplicate present in CyHV-3; for other three CyHV-2 isolates that contained TRs, it has transferred from the TR to the U region; and the homolog *ORF144* in CaHV is also located downstream), while it is deleted in CyHV-1; *ORF140* has undergone a large translocation in CyHV-1 to become located upstream of the U region, while in CyHV-2 and CyHV-3 it is located downstream (the homologs *ORF121* and *ORF77*, respectively, in CaHV and AngHV-1, are also located downstream). Compared with YC-01, there are 28 inserted (25 unique ORFs; three ORFs, i.e., *ORF1A*, *ORF2*, and *ORF3* are duplicate in the TRs) and 29 deleted (*ORF7* is in the TR) ORFs in CyHV-1, as well as 30 inserted (26 unique ORFs; four ORFs, i.e., *ORF1–4* segment is duplicate in the TRs) and 15 deleted ORFs in CyHV-3. In addition, there are many rearrangements and inversions of the ORFs in YC-01 (Fig. 5). For example, the orientation of *ORF21–23*, *ORF114–115*, *ORF120–121*, *ORF123–124*, and *ORF126–127* in YC-01 are opposite to those in CyHV-1; the orientation of *ORF4*, *ORF25*, *ORF138*, and *ORF140* are opposite to those in CyHV-3; and the orientation of *ORF128–133* and *ORF135–136* are all opposite compared with those in CyHV-1 and CyHV-3 (Table S1). Moreover, YC-01 and CaHV both have no U and TR features and are with similar sequence lengths, although the genomes of two strains share 91.23 % sequence identity only. Compared with YC-01, 12 homologs are evidently translocated in CaHV, i.e., *ORF1–12* upstream in the CaHV genome correspond to *ORF153B–C*, *ORF154–156*, *ORF1–2*, *ORF2A*, *ORF3*, and *ORF5–7* downstream in the YC-01 genome (Table 2 and Fig. 5). Additionally, among the genus *Cyvirus*, AngHV-1 is distantly related to YC-01, with only 55 homologous ORFs (Fig. 5 and Table S1).

Moreover, we aligned the amino acid sequences of some potential proteins among YC-01, ST-J1, SY-C1, SY, and CaHV, and some typical peptide repeats are shared by all five strains, but exhibit some differences (Fig. 6). Some peptide repeats are specific to CyHV-2 but absent in other CyHVs. There are two Q-rich tandem repeats QQQYQQQQ containing a tyrosine (Y) in ORF10 of YC-01, ST-J1, and SY-C1, while SY and CaHV have one more tandem repeat, but there is an insertion of QQQH containing a histidine (H) between the first and second Q-rich tandem repeat (the sixth Q in the first Q-rich tandem is replaced with a lysine (K), in CaHV). Meanwhile, like ST-J1 and SY-C1, there is a continuous Q tandem repeat (11 Qs) in ORF10 of YC-01, while one Q of this continuous tandem is missing in SY and CaHV. In ORF25D, like ST-J1, YC-01 possesses two threonine (T)-rich tandem repeats, TTTPTTTTP, containing two prolines (P), while there is only one tandem repeat in the other three isolates. Interestingly, a tandem repeat, LDD(E/N), is shared by ORF107, encoded by one of the core genes that conserved among alloherpesviruses in all four CyHV-2 isolates and CaHV; the tandem repeat is quadruplicate in ST-J1 and SY-C1, but just triplicate in SY and CaHV, while it is surprisingly repeated five times in YC-01 (in which the fourth site glutamic acid (E), in the last tandem repeat is replaced with asparagine (N) in all five isolates). The tandem repeat PTVTGI (T/D)QQS is repeated seven times in ORF156 encoded by YC-01, whereas it is repeated six times in ST-J1 and SY-C1, and it just triplicated in SY and CaHV (the seventh site threonine (T) in the last tandem is replaced with aspartic acid (D) in all five isolates). Notably, the two different amino acid sequences encoded by *ORF107* and *ORF156* could possibly be used as genetic markers of the YC-01 isolate to distinguish it from ST-J1, SY-C1, SY, and CaHV. Furthermore, based on the amino acid sequences alignment of the DNA polymerase catalytic subunit from 19 representative strains in the family *Alloherpesviridae*, including YC-01 (Fig. 7), phylogenetic analysis shows that they are divided into four groups corresponding to the genera *Cyvirus*, *Batravirus*, *Ictavirus*, and *Salmovirus*. YC-01 clusters with the members of the genus *Cyvirus* (together with the other six CyHV-2 isolates), and it is more closely related to CyHV-3 than CyHV-1 and is most distant from AngHV-1.

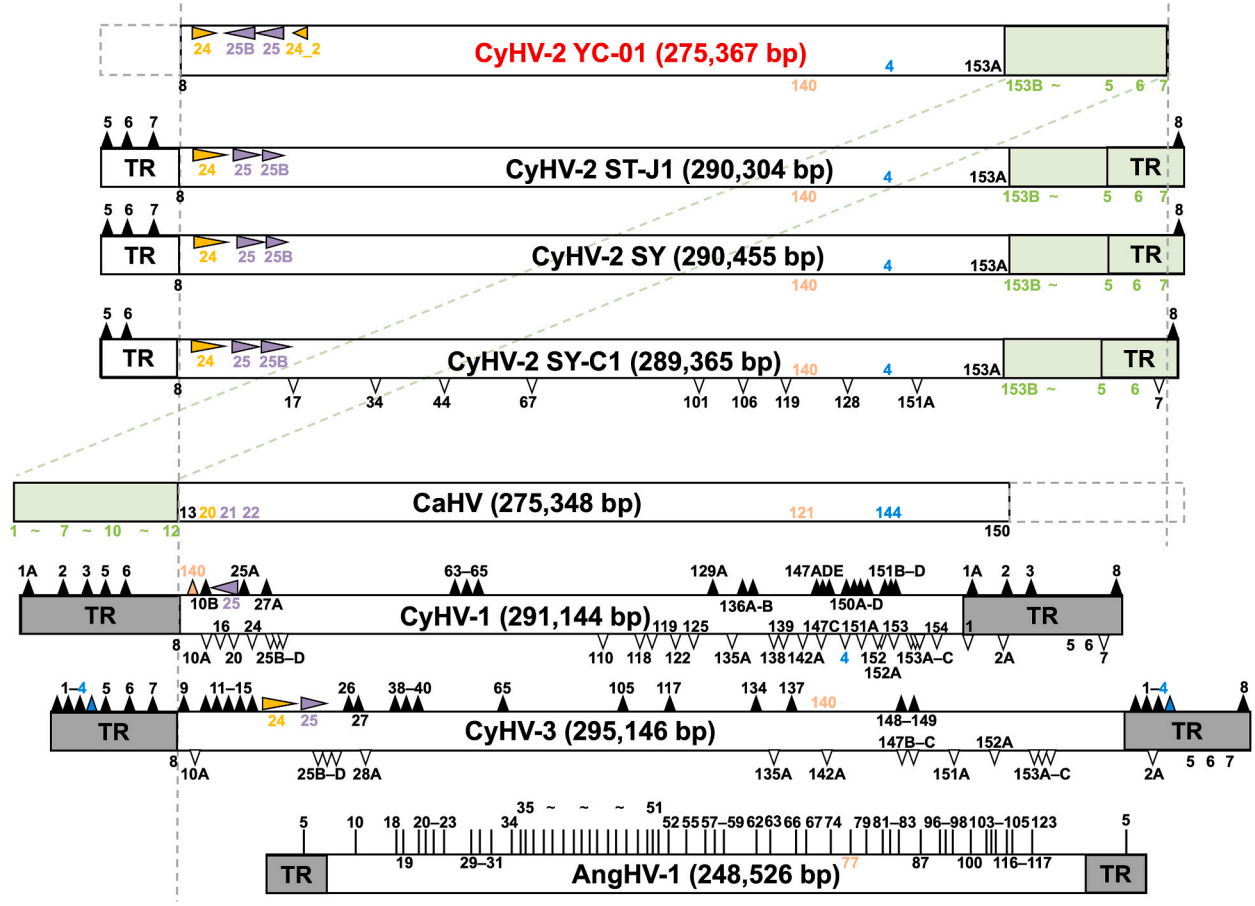


Fig. 5. Comparison of genome structures among YC-01 and other seven strains. Long frames and short grey frames indicate the sizes of the genomes and terminal repeat (TR) regions (TRs in the three strains of CyHV-2 are marked with transparent frames), respectively. Gene names are shown lacking the "ORF" prefix and "~" indicates the omitted ORFs. Left or right long purple triangles show that the positions of *ORF25* or *ORF25B* in YC-01 are opposite relative to those in other CyHV-2 isolates, CaHV, and CyHV-3 (the orientation of *ORF25* in YC-01 is the same as that in CyHV-1). Left or right long orange triangles show that *ORF24* of YC-01 is encoded in two parts (*ORF24* and *ORF24_2*). Short white triangles in ST-J1, SY, SY-C1, CaHV, CyHV-1 and CyHV-3 indicate deletions relative to YC-01. Short black triangles in ST-J1, SY, SY-C1, CyHV-1 and CyHV-3 indicate insertions relative to YC-01. *ORF140* (including homolog *ORF121* and *ORF77* in CaHV and AngHV-1, respectively) are marked in light pink. *ORF4* (including homolog *ORF144* in CaHV and deletion in CyHV-1) are marked in light blue. The dashed frame at the upstream end of the CaHV genome indicates nucleotides deleted relative to YC-01. The long light green diagonal dashed lines between YC-01 and CaHV indicates that the 5′-terminal *ORF1–12* of CaHV (light green frame) corresponding to the 3′-terminal *ORF153B–ORF7* of YC-01 (light green frame), similar to ST-J1, SY and SY-C1 (absence of *ORF7*) marked in light green frames. The short vertical lines in AngHV-1 indicate ORFs homologous to YC-01.

3.5. Potential virulence factors and B cell epitopes prediction of CyHV-2

Based on virulence factors identified by experimental validation on CyHV-3, in which deletion of single or multiple genes resulted in a significant reduction in viral proliferation or virulence [56–62], we referenced and compared the sequence identities of these eight genes and their encoded proteins between CyHV-2 and CyHV-3, as well as different CyHV-2 isolates (Table S8). Among them, ORF121 was not described in CyHV-3, but it is considered as a potential virulence factor for CyHV-2 and was identified as an immediate-early (IE) gene of CyHV-2 in a previous work [63]. The results showed that these genes are similar and conserved between CyHV-2 and CyHV-3, and highly consistent among CyHV-2 isolates. Notably, ORF150 in CyHV-2 and CyHV-3 share a highly conserved C3HC4 type RING-finger domain (with E-values of 4.05e-05 and 1.43e-05, respectively), although the amino acid sequence identity between the two strains is only 27.87 %, presumably the protein has the same function in the two closely related strains (Fig. S2A and Table S8). Furthermore, we aligned the nucleotide and amino acid sequences of the eight potential virulence factors described above among YC-01 and four other isolates, and several mutations, deletions, and insertions were identified (Figs. S2B–C). For ORF25, the initiation codon (ATG) of YC-01 and SY-C1 appears later (one codon) than in the other three isolates, and a similar situation occurs in ORF123 (dUTPase) of YC-01, ST-J1, and SY (delayed by four codons) compared with that in SY-C1 and CaHV (Fig. S2C). Meanwhile, a mutation from threonine (T) to isoleucine (I) (caused by a point mutation from "C" of "ACA" to "T" at position 542) was observed at aa 182 of ORF25 only in YC-01. For the ORF55 (TK), notably, compared with other three isolates, nine bases "GAGGACGAG" have been

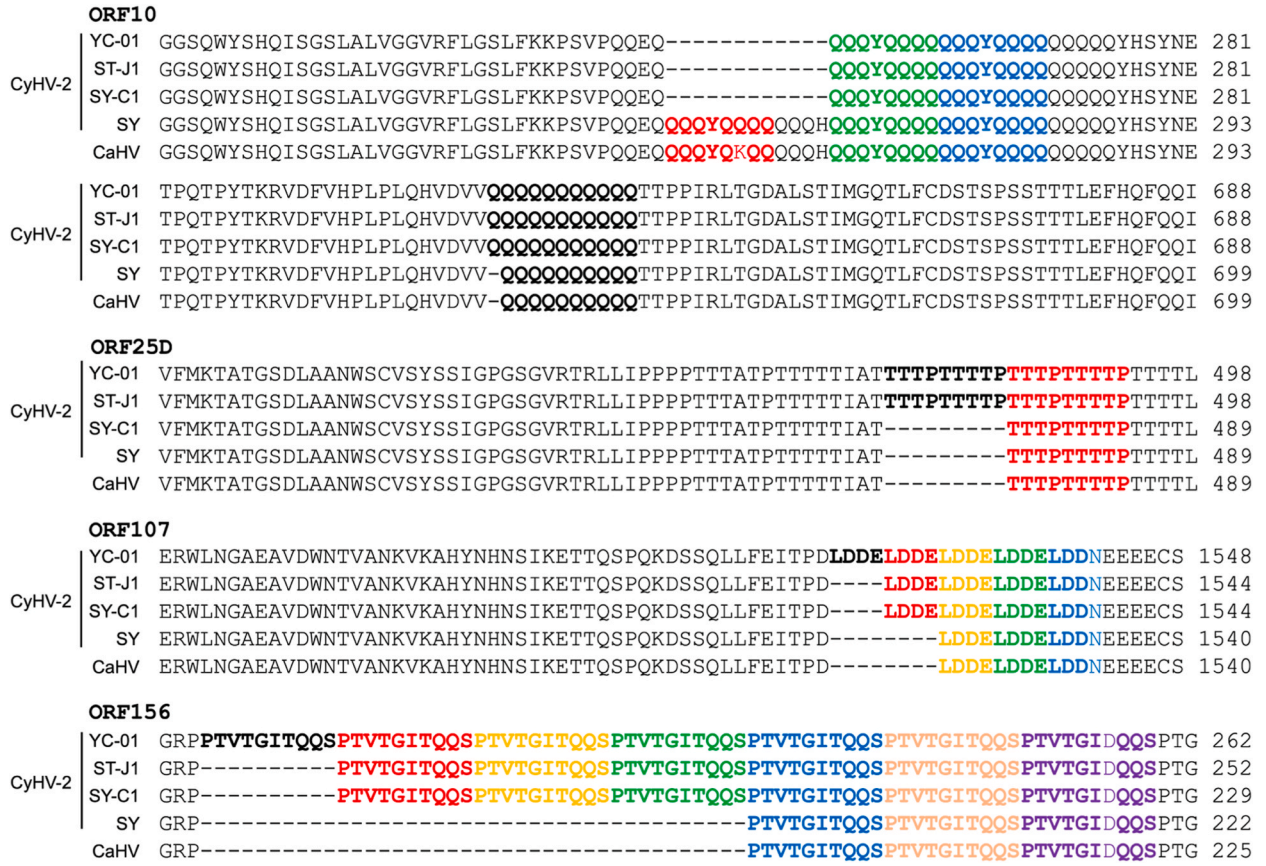


Fig. 6. Amino acid sequence alignment among CyHV-2 isolates and CaHV. Amino acid sequence alignment of ORF10, ORF25D, ORF107, and ORF156 homologs of CyHV-2 isolates (YC-01, ST-J1, SY-C1, and SY) and CaHV. The tandem repeats are in bold, and the same multiple tandem repeats of an amino acid sequence are marked in different colors to highlight differences among the five isolates. The following amino acids are not in bold but are marked with corresponding colors: the sixth glutamine (Q) in the first Q-rich tandem is replaced with lysine (K), in CaHV; the fourth site glutamic acid (E), in the last tandem is replaced with asparagine (N), in all five isolates; and the seventh site threonine (T), in the last tandem is replaced with aspartic acid (D) in all five isolates.

duplicately inserted at position 580–588 only in YC-01 and ST-J1, in which "EDE" is expressed repeatedly, and thus might be used as a molecular marker for YC-01 and ST-J1 to distinguish them from SY-C1, SY, and CaHV. For ORF56, there are two mutations located at position 1192 (from "G" to "A") and 1594 (from first "C" to "A") only in YC-01 and ST-J1, respectively, resulting in two mutations from alanine (A) to threonine (T) (aa 398) and from P to threonine (T) (aa 532) compared with SY-C1, SY, and CaHV. For ORF57, two mutations from leucine (L) to methionine (M) and from N to D (by a point mutation from "C" of CTG to "A" and from first "A" of AAT to "G", respectively, located at the positions 952 and 955) occur at aa 318 and aa 319 only in YC-01 and ST-J1. Moreover, compared with the four other isolates, 33 nucleotides have been deleted only in CaHV (Fig. S2B). For ORF121, there are three obvious tandem repeats in CyHV-2, while some differences remain among the five isolates. "KKSEKS" is duplicate in SY and CaHV, "KK" is inserted repeatedly (caused by the duplication of "AAGAAA") only in SY-C1, and a lack of "A" encoded by "GCA" occurs in SY, while it is quadruplicate in the other four isolates. For ORF123, except for the delayed appearance of the initiation codon mentioned above, L has been inserted only in YC-01 and ST-J1, located at position 18. For ORF141 (RNR α subunit), only in YC-01, there is a base substitution (the first "G" of "GCG" is replaced by an "A" resulting in a translation from the original alanine (A) to threonine (T) located at position 341) and a deletion (an alanine (A) encoded by "GCA") located at positions 1021 and 2415–2417, respectively. Moreover, for ORF150, only in YC-01 and ST-J1, there are four base mutations (substitutions from first "C" of "CCT" to "T" at position 268, from first "G" of "GTG" to "A" at position 559, from second "A" of "AAC" to "C" at position 1295, and from "G" of "AGC" to "A" at position 1694, resulting in translations from P to serine (S), from valine (V) to M, from N to threonine (T), and from S to N, respectively, located at positions 90, 187, 432, and 565) and an insertion (threonine (T) and E encoded by "ACT" and "GAA", respectively, located at positions 189 and 190) compared with the other three isolates (Fig. S2C).

To predict potential B cell epitopes of CyHV-2, we used the IEDB and ABCpred server. B cell epitope predictions were carried out using the eight immunogenic proteins of CyHV-2, including ORF25 (predicted as a type I membrane protein), ORF57 (unknown function), ORF66 (predicted as putative capsid triplex subunit 1), ORF72 (predicted as putative capsid triplex subunit 2), ORF92 (predicted as a major capsid protein), ORF115 (predicted as a type I membrane protein), ORF131 (predicted as a type I membrane

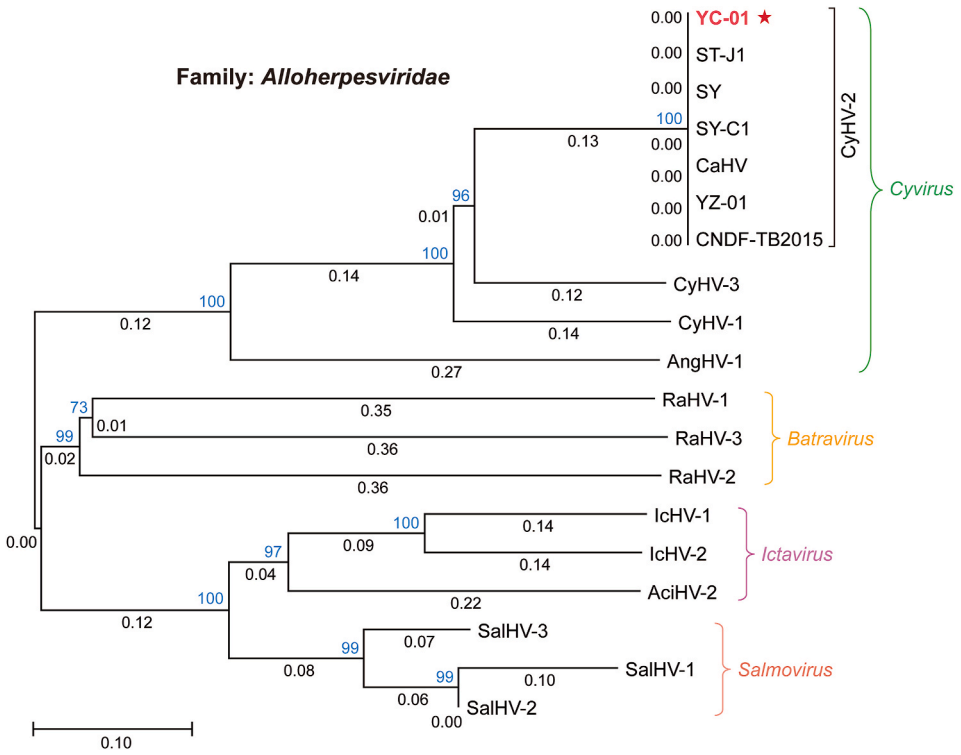


Fig. 7. Phylogenetic analysis. Phylogenetic tree constructed based on the amino acid sequences of the DNA polymerase catalytic subunit among YC-01 and other 18 members in the family *Alloherpesviridae* using the neighbor-joining method in MEGA v11 with bootstrap values of 1000 replications, which were divided into four genera, including *Cyvirus*, *Batravirus*, *Ictavirus*, and *Salmovirus*, distinguished by different colors. The consensus bootstrap confidence values of 1000 replications are shown at nodes of the branches and marked in blue. The degree of genetic variation is labeled on the branch line, representing the degree of evolutionary branch change. The scale bar represents 0.10 fixed mutation per amino acid position. CyHV-2 YC-01 (MN593216.1); CyHV-2 ST-J1 (NC_019495.1); CyHV-2 SY (KT387800.1); CyHV-2 SY-C1 (KM200722.1); CyHV-2 YZ-01 (MK260012.1); CyHV-2 CNDF-TB2015 (MN201961.1); CaHV (KU199244.1); CyHV-1 (NC_019491.1); CyHV-3 (NC_009127.1); AngHV-1 (NC_013668.3); ranid herpesvirus 1 (RaHV-1; NC_008211.1); RaHV-2 (NC_008210.1); RaHV-3 (NC_034618.1); ictalurid herpesvirus 1 (IcHV-1; NC_001493.2); IcHV-2 (NC_036579.1); acipenserid herpesvirus 2 (AciHV-2; NC_043042.1); salmonid herpesvirus 1 (SalHV-1; OK337613.1); SalHV-2 (NC_075802.1; partial sequence); SalHV-3 (NC_043468.1; partial sequence).

protein), and ORF132 (predicted as a type I membrane protein). In parallel, we performed predictions for linear B cell epitopes using the BepiPred-2.0 algorithm embedded in IEDB and ABCpred server. As shown in Table S9, for each protein, the epitope probability score for each amino acid residue was retrieved by the BepiPred-2.0 algorithm, in which the residues with scores above the threshold (default value is 0.500) were predicted to be part of an epitope and marked with "E". Meanwhile, using BepiPred-2.0 with a threshold of 0.550 (corresponding to a specificity cutoff of 80 %) and ABCpred server with a cutoff of 0.50 (showed 65.93 % accuracy with equal sensitivity and specificity using default window length of 16), a list of predicted dominant B cell epitopes is provided in Table 3, among which the short peptides containing helix and strand structures have been removed (the secondary structure prediction results of the eight proteins described above are shown in Fig. S3). Using PSIPRED server prediction, the results showed the secondary structures of the eight immunogenic proteins all contain rich α-helix, β-sheet (strand), and coils, as well as many potential B cell epitopes. We deleted the short peptide sequences containing α-helix and β-sheet (strand) structures because they would find it difficult to form suitable epitopes. Among them, thirteen B cell epitopes were derived from three capsid proteins (ORF66, ORF72, and ORF92), another 13 from four membrane proteins (ORF25, ORF115, ORF131, and ORF132), and eight from ORF57, as shown in Table 3. The MCP encoded by *ORF92* has the most dominant B cell epitopes, followed by ORF57, thus the two proteins might be used as dominant CyHV-2 antigens to rationally design subunit vaccines and vaccination strategies designed to target the immune responses to these conserved epitopes, ultimately producing immunity. Moreover, the eight immunogenic proteins in the four CyHV-2 isolates and CaHV are highly conserved and share high sequence identities (nearly more than 99 %) and are highly divergent compared with those of CyHV-1 and CyHV-3 (Table S10).

4. Discussion

CyHV-2 infections have resulting in severe economic losses in all farmed fishes of *Carassius* species including gibel carp suffered from the HVHN with different levels of susceptibility and mortality in China [2–4,7–10]. However, not much has been reported on the isolation, identification, and characterization of CyHV-2 isolates in the past decades, and only seven isolates such as ST-J1 [19], SY-C1

Table 3
B cell epitopes prediction of CyHV-2 available in IEDB and ABCpred server.

Peptide sequence [a]	Protein	Protein ID	Start (aa) [b]	End (aa) [b]	Length (aa) [b]	Database/server
VMQINDDVP	ORF25	QIV66841.1	405	413	9	IEDB
TTTPSTVPTTPKTTTQRTTTPTT	ORF25	QIV66841.1	507	529	23	IEDB
PTTTAEPTAGD	ORF25	QIV66841.1	554	564	11	IEDB
TPPPTTTAEPTAGDTS	ORF25	QIV66841.1	551	566	16	ABCpred
EDEEYVP	ORF57	QIV66875.1	24	30	7	IEDB
NADGGGGKIPGLDD	ORF57	QIV66875.1	205	218	14	IEDB
FPTAKSVTLD	ORF57	QIV66875.1	345	354	10	IEDB
RPSTSSSASSTDE	ORF57	QIV66875.1	481	493	13	IEDB
KAKKPVKPKRLTAGATGTSRATGAAGT SRATGATGTSRSYGEMGAKPKAP	ORF57	QIV66875.1	526	575	50	IEDB
RGRSSKQPECGKGCAC	ORF57	QIV66875.1	458	473	16	ABCpred
CGKGCACGECPKCPRP	ORF57	QIV66875.1	467	482	16	ABCpred
TGTSRSYGEMGAKPKA	ORF57	QIV66875.1	559	574	16	ABCpred
QQATGDGLPA	ORF66	QIV66882.1	5	14	10	IEDB
SSSSSSRGSL	ORF66	QIV66882.1	226	235	10	IEDB
IESQLEPDTGASSF	ORF66	QIV66882.1	342	355	14	IEDB
PDPGVAH	ORF72	QIV66888.1	63	69	7	IEDB
MTMALPG	ORF92	QIV66908.1	147	153	7	IEDB
ICTEGPSMGRPFVS	ORF92	QIV66908.1	400	413	14	IEDB
IASPSSTNRDPNH	ORF92	QIV66908.1	801	813	13	IEDB
DLAQEIGLA	ORF92	QIV66908.1	959	967	9	IEDB
GLQAKPRYSMNNYDH	ORF92	QIV66908.1	1010	1024	15	IEDB
STHPPIQTAATG	ORF92	QIV66908.1	1039	1050	12	IEDB
FNPLLGEVTSSDYFRPDAYH	ORF92	QIV66908.1	1061	1080	20	IEDB
GDPNADMFGAG	ORF92	QIV66908.1	1110	1120	11	IEDB
LSTPTDPYVYFGDNKP	ORF92	QIV66908.1	756	771	16	ABCpred
SVPDVPSKCS	ORF115	QIV66930.1	44	53	10	IEDB
PGSTDAN	ORF115	QIV66930.1	197	203	7	IEDB
PLELEDQDPYMNDDAFP	ORF115	QIV66930.1	366	382	17	IEDB
RSGGTHQVR	ORF115	QIV66930.1	426	434	9	IEDB
ARCSLTWSDFLGKCIP	ORF115	QIV66930.1	545	560	16	ABCpred
LELEDQDPYMNDDAFP	ORF115	QIV66930.1	367	382	16	ABCpred
VSPAPIPQPMIARSVP	ORF131	QIV66949.1	201	216	16	ABCpred
TPVPTVPTVTTASTTP	ORF131	QIV66949.1	161	176	16	ABCpred
SGKNKRSGGYDNTAFM	ORF132	QIV66948.1	140	155	16	ABCpred

Note.
[a] The potential B cell epitopes were predicted using a cutoff of 0.550 (corresponding to specificity is 0.817 and sensitivity is 0.292, approximately) through the BepiPred-2.0 algorithm embedded in the B cell epitope prediction analysis tool available in Immune Epitope Database (IEDB), as well as using a threshold of 0.50 showed 65.93 % accuracy with equal sensitivity and specificity by window length of 16 through ABCpred server (the peptide sequences contained helix and strand structures were removed).
[b] aa, amino acid.

[22], SY [64] have been whole-genome sequenced. In this paper, we present the complete genome structure and potential molecular characterization analysis of a CyHV-2 isolate YC-01 and a comparison with other representative published strains of the genus *Cyvirus*. Identification of new isolates will assist studies to elucidate the diversity of CyHV-2.

Phylogenetically, CyHV-2 is closely related to CyHV-3 and CyHV-1, especially the former (Fig. 7). The three viruses possess several similar biological characteristics and pathogenic properties, hence the differential diagnosis among these three diseases is vital. The three viruses are host-selective, mainly confined to common carp, koi, goldfish, and gibel carp [7,13,14,65]. There is some evidence suggesting that the three viruses can remain latent in clinically healthy fish, and that the infection can be reactivated by some stress [66–68]. The diseases are seasonal and temperature-sensitive, which lesions caused by CyHV-1 usually develop when water temperature is lower than 15 °C; disease outbreaks induced by CyHV-2 mainly occur between 15 and 25 °C; outbreaks infected by CyHV-3 appear when the water temperature is 18–28 °C [65,69,70]. Another differential diagnosis is based on clinical symptoms and histopathological changes. CyHV-1 causes benign papillomatous lesions with mucoid to smooth waxy multifocal-tocoalescing, slightly-raised-to-nodular, milky-white, and cutaneous growths in common carp and koi (the lesions are mainly concentrated in the skin) [71,72]. CyHV-2 induces gill bleeding, body swelling, abdominal hemorrhages, and generalized edema in goldfish and gibel carp (the lesions focus on gills and hematopoietic organs) [8]. Affected koi induced by CyHV-3 exhibit apathy, gill epithelium necrosis, pale patches on the skin, pale and irregularly colored gills, increased mucus production, and enophthalmia (the lesions are mainly on the gills and eyeballs) [73,74]. Additionally, based on whole-genome sequencing, many nucleic acid and immunological assays have been developed to specifically detect these viruses [75–77].

For the genome structure, unlike the other strains in the genus *Cyvirus*, YC-01 lacks TRs at the ends of the U region (i.e., no U and TR features), similar to that of CaHV, which two strains share only 91.23 % sequence identity, although the two genomes have similar lengths (Table 2 and Fig. 5). The genome organization in the majority of the genus *Cyvirus* consists of one long U region flanked by two short TRs at the ends, which do not seem to be general features of alloherpesviruses, since the SalHV-1 genome is known to have a long

noninverting region (U_L) linked to a short inverting region (U_S) flanked by an inverted repeat (R_S) similar to the characteristic of mammalian alphaherpesviruses in the genus *Varicellovirus* [78]. For CyHVs, Davison et al. (2013) confirmed that the tandem reiterations are most common in TRs, and the most prominent are the telomere-like repeats based on the element TTAGGG, which are located very close to the ends of the TR and, hence, to the genome ends in CyHVs [19]. There is evidence that such reiterations might mediate the integration of herpesvirus genomes into the telomeric regions of host chromosomes, such as in Marek's disease virus (MDV) and human herpesvirus 6 (HHV-6) [79]. The presence of TRs and the duplications of *ORF5–8* of CyHV-2 seem to be associated with restriction of flanking genes (i.e., telomere-like repeats). In both TRs of each CyHV genome, for the right end, telomere-like repeats are located downstream of *ORF8*; for the left end, in CyHV-1 and CyHV-3, they are both located downstream of *ORF156*, however, in CyHV-2 they situate between *ORF3* and *ORF5*, which the TR contains only *ORF5–8* in CyHV-2 [19]. However, for the YC-01 isolate, although the TRs are deleted, we found that some telomere-like repeats, specified as TTAGGG element incompletely remain located downstream of *ORF8* (data not shown) and do not appear at other positions such as upstream of *ORF5*, like that of ST-J1. We speculated that this is the reason for the deleted TR in YC-01, i.e., the lack of telomere-like repeats.

Further genomic comparisons demonstrated several sequence variations in YC-01 comparing with CyHV-2 isolates, CaHV, and another three strains in the genus *Cyvirus*. The variations comprise gene recombination, such as deletions, insertions, duplications, rearrangements, and horizontal transfers (compared with CyHV-1 and CyHV-3), and/or nucleotide mutations, such as base substitutions, insertions, deletions, and duplications (Fig. 5), which have resulted in the high complexity and diversity of these genomes through evolution. Compared with CyHV-2 isolates, because of the deletion of the TRs, *ORF5–8* segment is not duplicate in the YC-01 genome. The orientations of the *ORF25* and *ORF25B* are opposite, and their positions are reserved compared with the other three CyHV-2 isolates. Except for *ORF25B* (which is absent in CyHV-1 and CyHV-3), the orientation of *ORF25* in YC-01 is consistent with that in CyHV-1, but opposite to that in CyHV-3 (Fig. 5). In addition, the evolutionarily dynamic nature of the ORF25 family seems to be illustrated in CyHV-2, particularly in YC-01. In the case of *ORF25* and *ORF25B*, the two have migrated from the positive to the negative strand and the orientation of *ORF25* is consistent with that of CyHV-1. *ORF24_2* is inserted between *ORF25* and *ORF25C*, which is not found in other CyHVs. It seems that ORF25 family members might have come and gone frequently or "jumped" or captured the expression patterns of other viral genes during CyHV-2 evolution [19]. Compared with CyHV-1 and CyHV-3, like the other three CyHV-2 isolates, there is more complexity in the arrangement of the ORFs in the YC-01 genome. First, the genome size, the differences in the TRs, the rearrangement of *ORF4* and *ORF140*, and the cases of the inserted and deleted ORFs among the three strains have been mentioned above. These non-conserved ORFs presumably provide functions that fit the individual viruses to their biological niches [19]. Second, among the three CyHVs, including YC-01, 119 unique ORFs are orthologously arranged with a large degree of colinearity. Each CyHV contains a large central section (*ORF28–113*) in which conserved genes are ordered equivalently. However, compared with CyHV-3, like ST-J1, evident rearrangements to the left and right of the large and central section are found in YC-01. The gene rearrangement or lateral transfer reveals that the viruses might have undergone events involving genes jumping under long-term host selection pressure, allowing them to occupy more diverse niches [80].

Interestingly, in the YC-01 genome, we found two potential spliced genes, *ORF10A* and *ORF33*, respectively (Fig. 1B and D). Like other CyHVs, we found that *ORF33* (in negative strand) consists of one long part and a cluster of another three short parts, i.e., assuming a description like this, it has three introns, and there have 12 other genes (nine are on the positive strand and three are on the negative strand) inserting into the first intron similar to that of CyHV-2 isolates, while in CyHV-1 and CyHV-3, it contains one long part and a cluster of another two short parts (consisting of two introns) [19,81], yet five parts with more complexity are shown in the orthologous *ORF10* of AngHV-1 [15]. Simlarly, *ORF10* has two exons separated by introns of 1001 bp and *ORF10A* (on the negative strand) is inserted into it. This arrangement is conserved in CyHV-2 isolates (in CaHV, *ORF14*, the *ORF10* homolog, is not spliced, but it swaps position with *ORF15*, the *ORF10A* homolog), but not in CyHV-1 and CyHV-3, because *ORF10A* is unique to CyHV-2 (Fig. 5). Additionally, *ORF24* is encoded in two parts (*ORF24* and *ORF24_2*) with opposite transcriptional orientations on the positive and negative strands in YC-01, respectively, but we don't think it is spliced. Through amino acid sequence comparison, we found that the ORF24 and ORF24_2 of YC-01 coincide with the anterior (317 aa) and posterior (241 aa) segments of ORF24 encoded by other CyHV-2 isolates, respectively (data not shown). We speculated that this situation arose via gene duplication events, in which the viral gene was captured several times in different situations and broken to encode two parts and inserted into a location by chance, but is likely to be meaningless. Thus, it is expected that future structural analysis will further elucidate these relationships and provide more research ideas to further our understanding of the structure and functions of CyHV-2 proteins.

Based on bioinformatic analysis, we predicted 27 putative membrane proteins of YC-01, among which 15 proteins are classified as type I membrane proteins, ten are type III, and the remaining two, ORF25C and ORF132 are considered as type II, as well as four ORFs (*ORF66*, *ORF72*, *ORF78*, and *ORF92*) predicted to be related to nucleocapsid proteins (Table 1 and Table S1). In a previous work, Gao et al. (2020) identified 18 membrane proteins (12 type I and five type III) and three capsid proteins using liquid chromatography-tandem mass spectrometry (LC-MS/MS), in which ten membrane proteins and one capsid protein (ORF78) were not identified experimentally, possibly because of the influence of proteins with relatively low abundance in CyHV-2 [27]. In a study of CyHV-3 encoded proteins, 16 virion transmembrane proteins (VTPs) that affect the virus infectious cycle were identified by LC-MS/MS [57], among them, nine proteins (ORF25, ORF64, ORF81, ORF83, ORF99, ORF115, ORF131, ORF132, and ORF136) are well conserved in CyHV-2 and CyHV-3. Among them, type III membrane proteins are considered as multipass transmembrane proteins in which the polypeptide crosses the lipid bilayer multiple times [82], i.e., this type of membrane proteins possesses multiple TMDs. In CyHV-2, it was predicted that ORF64, ORF114, ORF152A, and ORF16 have 10, 8, 7, and 7 TMDs, respectively, and ORF81–83 and the ORF153 family all have four TMDs. As shown in Table S1, there are 10 putative TMDs in ORF64, similar to the 12 TMDs in ORF64 of CyHV-3 [19], indicating a nucleoside transporter domain or similar.

Furthermore, the promoters contain multiple *cis*-acting elements that are specific binding sites for proteins involved in the initiation

and regulation of transcription [83]. We identified 671 *cis*-acting elements belonging to 215 categories, among which 22 categories are present in more than five gene promoters and top six include AT-rich, Sp1, GLUT2TAAT, ISRE, AC-box, and E-box (Fig. 4B and C). Most of them are involved in the regulation of host cell physiological activities, such as Sp1, GLUT2TAAT, and AC-box [84–86]; and some research have confirmed that these elements also regulate the transcriptional expression of viral genes, such as Sp1, IRSE, and E-box [87–89]. Meanwhile, some elements have been found in the viral genome and are involved in the anti-host immune response or the transcription of viral genes, such as AT-rich, IRSE and E-box [90–92]. In addition, an element named G-string recurs many times, and the copy number and precise positions of these G-strings might determine the extent to which promoter regions respond to different activation modes [93]. Moreover, we found 62 TISs of genes that follow the Kozak's rules (Fig. 4E) encoded by YC-01. The nucleotide sequence surrounding the initiation codon has been shown to affect the efficiency of translation initiation, in which the conserved (A/G)NNATGG, named the Kozak sequence, is characterized as the most efficient TIS in eukaryotes [94]. Sequences around the initiation codon of some viral genes have been also found to follow the Kozak's rules [95,96].

Advances in molecular virology and the advent of recombinant-virus systems have led to the identification of many viral genes associated with virulence and immunogenicity [97]. In CyHV-3, different genes thought to encode virulence factors, such as *ORF55*, *ORF123*, and *ORF141* [58], were deleted singly or together, leading to a significant reduction in viral proliferation or virulence. However, related reports in CyHV-2 are scarce. Based on virulence factors identified by experimental validation in CyHV-3, we aligned the mutations of nucleotide and amino acid sequences of eight potential virulence factors among CyHV-2 isolates. These genes are conserved between the two viruses, and highly consistent among CyHV-2 isolates (Table S8). The results provide a reference for the screening of virulence factors and the construction of attenuated mutants for CyHV-2 using recombinant-virus systems. For the biological implications induced by variations of YC-01 genome, we make the following speculations. First, TRs are absent in YC-01 similar to that in CaHV, suggesting that CyHV-2 might have lost part of its genome sequence or discarded redundant regions, such as repeated TRs (the lack of telomere-like repeats) during evolution with different hosts to allow the virus to better adapt to its host or be more stable potentially. Second, the mutations are found in several genes particularly potential virulence factors by comparing genomes of these different CyHV-2 isolates, implying they could influence viral infectivity and host adaptation. Third, in terms of virus traceability, CyHV-2 infection was originally identified in goldfish in Japan [6] and subsequently cases of CyHV-2 infection have been reported from many countries and regions worldwide [7]. Some isolates were classified into J or C genotypes, based on differences in their genomes and isolation sites, which there have ten deleted genes in SY-C1 (C genotype) comparing with ST-J1 (J genotype) [22]. Epidemic CyHV-2 isolates in China have been reported mostly in gibel carp, including YC-01 was also isolated from it [3,4,17,22,25,64]. Hence, different CyHV-2 isolates may co-evolve with their hosts to better adapt them. Fourth, YC-01 is well adaptable in cell infection [26] and may induced some variations in its genome, which they could be related to viral virulence attenuation by serial subculture in cell lines. Notably, ORF150 in CyHV-2 and CyHV-3 share a highly conserved C3HC4 type RING-finger domain, although with a low amino acid sequence identity (Fig. S2A). The protein presumably has the same function in the two closely related strains. Klafack et al. (2019) confirmed the partial sequence deletion occurs only in *ORF150* of CyHV-3 by serial subculture in cell [98]. Then they demonstrated the relevance of ORF150 as a virulence factor by targeted deletion and indicated that ORF150-deleted CyHV-3 can be used as an efficacious and safe vaccine in carp production [56]. The case may also occur in CyHV-2 and further experimental verification is needed.

5. Conclusions

In this study, we analyzed the complete genome structure and potential molecular characterization, then compared with those of the genus *Cyvirus*, and several evident variations were found in CyHV-2 YC-01. Two potential molecular genetic markers were confirmed to distinguish YC-01 and other CyHV-2 isolates. Moreover, the potential virulence factors and linear B cell epitopes of CyHV-2 were predicted to provide possibilities for rational design of novel vaccines. The identification and analysis of new isolates will assist studies to elucidate the diversity of CyHV-2. This paper provides more insights for better understanding genomic structure, genetic evolution, and potential molecular mechanisms of CyHV-2 in future research.

Data availability statement

Data is contained within the article and Supplementary data.

CRediT authorship contribution statement

Jia Yang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Simin Xiao:** Validation, Software, Investigation. **Liqun Lu:** Supervision, Project administration, Funding acquisition. **Hao Wang:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Funding acquisition, Data curation, Conceptualization. **Yousheng Jiang:** Validation, Investigation.

Declaration of competing interest

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 32102842) and the Earmarked Fund for China Agriculture Research System (Grant No. CARS-45-16).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e32811.

References

[1] J. Gui, L. Zhou, Genetic basis and breeding application of clonal diversity and dual reproduction modes in polyploid *Carassius auratus gibelio*, Sci. China Life Sci. 53 (4) (2010) 409–415.
[2] L. Wang, J. He, L. Liang, X. Zheng, P. Jia, X. Shi, et al., Mass mortality caused by cyprinid herpesvirus 2 (CyHV-2) in prussian carp (*Carassius gibelio*) in China, bull, Eur. Assoc. Fish Pathol 32 (5) (2012) 164–173.
[3] J. Xu, L. Zeng, H. Zhang, Y. Zhou, J. Ma, Y. Fan, Cyprinid herpesvirus 2 infection emerged in cultured gibel carp, *Carassius auratus gibelio* in China, Vet. Microbiol. 166 (1–2) (2013) 138–144.
[4] Y.Z. Luo, L. Lin, Y. Liu, Z.X. Wu, Z.M. Gu, L.J. Li, et al., Haematopoietic necrosis of cultured Prussian carp, *Carassius gibelio* (Bloch), associated with Cyprinid herpesvirus 2, J. Fish. Dis. 36 (12) (2013) 1035–1039.
[5] L. Hanson, A. Doszpoly, S.J. Van Beurden, P.H. de Oliveira Viadanna, T. Waltzek, Alloherpesviruses of fish, in: F.S.B. Kibenge, M.G. Godoy (Eds.), Aquaculture Virology, Academic Press, London, 2016, pp. 153–172.
[6] S.J. Jung, T. Miyazaki, Herpesviral haematopoietic necrosis of goldfish, *Carassius auratus* (L.), J. Fish. Dis. 18 (3) (1995) 211–220.
[7] R.S. Thangaraj, S.R. Nithianantham, A. Dharmaratnam, R. Kumar, P.K. Pradhan, S. Thangalazhy Gopakumar, et al., Cyprinid herpesvirus-2 (CyHV-2): a comprehensive review, Rev. Aquacult. 13 (2) (2021) 796–821.
[8] J. Wen, Y. Xu, M. Su, L. Lu, H. Wang, Susceptibility of goldfish to cyprinid herpesvirus 2 (CyHV-2) SH01 isolated from cultured crucian carp, Viruses 13 (9) (2021) 1761.
[9] N. Jiang, D. Yuan, M. Zhang, L. Luo, N. Wang, W. Xing, et al., Diagnostic case report: disease outbreak induced by CyHV-2 in goldfish in China, Aquaculture 523 (2020) 735156.
[10] M. Zhu, B. Liu, G. Cao, X. Hu, Y. Wei, J. Yi, et al., Identification and rapid diagnosis of the pathogen responsible for haemorrhagic disease of the gill of Allogynogenetic crucian carp, J. Virol. Methods 219 (2015) 67–74.
[11] K.R. Jeffery, K. Bateman, A. Bayley, S.W. Feist, J. Hulland, C. Longshaw, et al., Isolation of a cyprinid herpesvirus 2 from goldfish, *Carassius auratus* (L.), in the UK, J. Fish. Dis. 30 (11) (2007) 649–656.
[12] A.E. Goodwin, G.E. Merry, J. Sadler, Detection of the herpesviral hematopoietic necrosis disease agent (Cyprinid herpesvirus 2) in moribund and healthy goldfish: validation of a quantitative PCR diagnostic method, Dis. Aquat. Org. 69 (2–3) (2006) 137–143.
[13] T.B. Waltzek, G.O. Kelley, D.M. Stone, K. Way, L. Hanson, H. Fukuda, et al., Koi herpesvirus represents a third cyprinid herpesvirus (CyHV-3) in the family *Herpesviridae*, J. Gen. Virol. 86 (Pt 6) (2005) 1659–1667.
[14] T.B. Waltzek, G.O. Kelley, M.E. Alfaro, T. Kurobe, A.J. Davison, R.P. Hedrick, Phylogenetic relationships in the family *Alloherpesviridae*, Dis. Aquat. Org. 84 (3) (2009) 179–194.
[15] S.J. van Beurden, A. Bossers, M.H. Voorbergen-Laarman, O.L. Haenen, S. Peters, M.H. Abma-Henkens, et al., Complete genome sequence and taxonomic position of anguillid herpesvirus 1, J. Gen. Virol. 91 (Pt 4) (2010) 880–887.
[16] A. Doszpoly, E.R. Kovács, G. Bovo, S.E. LaPatra, B. Harrach, M. Benko, Molecular confirmation of a new herpesvirus from catfish (*Ameiurus melas*) by testing the performance of a novel PCR method, designed to target the DNA polymerase gene of *alloherpesviruses*, Arch. Virol. 153 (11) (2008) 2123–2127.
[17] X.T. Zeng, Z.Y. Chen, Y.S. Deng, J.F. Gui, Q.Y. Zhang, Complete genome sequence and architecture of crucian carp *Carassius auratus* herpesvirus (CaHV), Arch. Virol. 161 (12) (2016) 3577–3581.
[18] T. Wu, Z.F. Ding, M. Ren, L. An, Z.Y. Xiao, P. Liu, et al., The histo- and ultra-pathological studies on a fatal disease of Prussian carp (*Carassius gibelio*) in mainland China associated with cyprinid herpesvirus 2 (CyHV-2), Aquaculture 412 (2013) 8–13.
[19] A.J. Davison, T. Kurobe, D. Gatherer, C. Cunningham, I. Korf, H. Fukuda, et al., Comparative genomics of carp herpesviruses, J. Virol. 87 (5) (2013) 2908–2922.
[20] X. Huo, C. Fan, T. Ai, J. Su, The combination of molecular adjuvant CCL35.2 and DNA vaccine significantly enhances the immune protection of *Carassius auratus gibelio* against CyHV-2 infection, Vaccines (Basel) 8 (4) (2020) 567.
[21] M. Su, R. Tang, H. Wang, L. Lu, Suppression effect of plant-derived berberine on cyprinid herpesvirus 2 proliferation and its pharmacokinetics in Crucian carp (*Carassius auratus gibelio*), Antivir. Res. 186 (2021) 105000.
[22] L. Li, Y. Luo, Z. Gao, J. Huang, X. Zheng, H. Nie, et al., Molecular characterization and prevalence of a new genotype of Cyprinid herpesvirus 2 in mainland China, Can. J. Microbiol. 61 (6) (2015) 381–387.
[23] F.Z. Wang, Y. Xu, Y. Zhou, C. Ding, H.G. Duan, Isolation and characterization of a Cyprinid herpesvirus strain YZ01 from apparently healthy goldfish after rising water temperature, bioRxiv (2021), https://doi.org/10.1101/2021.07.05.451087 [preprint].
[24] Q. Zhang, J.F. Gui, Virus genomes and virus-host interactions in aquaculture animals, Sci. China Life Sci. 58 (2) (2015) 156–169.
[25] L. Xu, P. Podok, J. Xie, L. Lu, Comparative analysis of differential gene expression in kidney tissues of moribund and surviving crucian carp (*Carassius auratus gibelio*) in response to cyprinid herpesvirus 2 infection, Arch. Virol. 159 (8) (2014) 1961–1974.
[26] J. Lu, D. Xu, L. Lu, A novel cell line established from caudal fin tissue of *Carassius auratus gibelio* is susceptible to cyprinid herpesvirus 2 infection with the induction of apoptosis, Virus Res. 258 (2018) 19–27.
[27] W. Gao, H. Wen, H. Wang, J. Lu, L. Lu, Y. Jiang, Identification of structure proteins of cyprinid herpesvirus 2, Aquaculture 523 (4) (2020) 735184.
[28] B. Ewing, L. Hillier, M.C. Wendl, P. Green, Base-calling of automated sequencer traces using phred. I. Accuracy assessment, Genome Res. 8 (3) (1998) 175–185.
[29] B. Ewing, P. Green, Base-calling of automated sequencer traces using phred. II. Error probabilities, Genome Res. 8 (3) (1998) 186–194.
[30] S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890.
[31] G. de Sena Brandine, A.D. Smith, Falco: high-speed FastQC emulation for quality control of sequencing data, F1000Res 8 (2019) 1874.
[32] S.D. Jackman, B.P. Vandervalk, H. Mohamadi, J. Chu, S. Yeo, S.A. Hammond, et al., ABySS 2.0: resource-efficient assembly of large genomes using a Bloom filter, Genome Res. 27 (5) (2017) 768–777.
[33] R. Luo, B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, et al., SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler, GigaScience 1 (1) (2012) 18.
[34] M. Xu, L. Guo, S. Gu, O. Wang, R. Zhang, B.A. Peters, et al., TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads, GigaScience 9 (9) (2020) giaa094.
[35] J. Besemer, A. Lomsadze, M. Borodovsky, GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions, Nucleic Acids Res. 29 (12) (2001) 2607–2618.

[36] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 33 (Database issue) (2005) D501–D504.
[37] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, et al., BLAST+: architecture and applications, BMC Bioinf. 10 (2009) 421.
[38] T.L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proc. Int. Conf. Intell. Syst. Mol. Biol. 2 (1994) 28–36.
[39] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (6) (2004) 1188–1190.
[40] Y. Choi, T. Ryu, A.C. Lee, H. Choi, H. Lee, J. Park, et al., High information capacity DNA-based data storage with augmented encoding characters using degenerate bases, Sci. Rep. 9 (1) (2019) 6582.
[41] J.J. Almagro Armenteros, K.D. Tsirigos, C.K. Sønderby, T.N. Petersen, O. Winther, S. Brunak, et al., SignalP 5.0 improves signal peptide predictions using deep neural networks, Nat. Biotechnol. 37 (4) (2019) 420–423.
[42] UniProt Consortium, UniProt: the universal protein knowledgebase in 2023, Nucleic Acids Res. 51 (D1) (2023) D523–D531.
[43] J. Huerta-Cepas, D. Szklarczyk, K. Forslund, H. Cook, D. Heller, M.C. Walter, et al., eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences, Nucleic Acids Res. 44 (D1) (2016) D286–D293.
[44] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, M. Ishiguro-Watanabe, KEGG for taxonomy-based analysis of pathways and genomes, Nucleic Acids Res. 51 (D1) (2023) D587–D592.
[45] A. Marchler-Bauer, S.H. Bryant, CD-Search: protein domain annotations on the fly, Nucleic Acids Res. 32 (Web Server issue) (2004) W327–W331.
[46] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, et al., CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Res. 48 (D1) (2020) D265–D268.
[47] I. Letunic, S. Khedkar, P. Bork, SMART: recent updates, new developments and status in 2020, Nucleic Acids Res. 49 (D1) (2021) D458–D460.
[48] F. Madeira, M. Pearce, A.R.N. Tivey, P. Basutkar, J. Lee, O. Edbali, et al., Search and sequence analysis tools services from EMBL-EBI in 2022, Nucleic Acids Res. 50 (W1) (2022) W276–W279.
[49] X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server, Nucleic Acids Res. 42 (Web Server issue) (2014) W320–W324.
[50] I.A. Shahmuradov, V.V. Solovyev, Nsite, NsiteH and NsiteM computer tools for studying transcription regulatory elements, Bioinformatics 31 (21) (2015) 3544–3545.
[51] M. Lescot, P. Déhais, G. Thijs, K. Marchal, Y. Moreau, Y. Van de Peer, et al., PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences, Nucleic Acids Res. 30 (1) (2002) 325–327.
[52] M.C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes, Nucleic Acids Res. 45 (W1) (2017) W24–W29.
[53] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, et al., The immune epitope database (IEDB): 2018 update, Nucleic Acids Res. 47 (D1) (2019) D339–D343.
[54] S. Saha, G.P. Raghava, Prediction methods for B-cell epitopes, Methods Mol. Biol. 409 (2007) 387–394.
[55] D.W.A. Buchan, D.T. Jones, The PSIPRED protein analysis workbench: 20 years on, Nucleic Acids Res. 47 (W1) (2019) W402–W407.
[56] S. Klafack, L. Schröder, Y. Jin, M. Lenk, P.Y. Lee, W. Fuchs, et al., Development of an attenuated vaccine against Koi Herpesvirus Disease (KHVD) suitable for oral administration and immersion, NPJ Vaccines 7 (1) (2022) 106.
[57] C. Vancsok, M.M.D. Peñaranda, V.S. Raj, B. Leroy, J. Jazowiecka-Rakus, M. Boutier, et al., Proteomic and functional analyses of the virion transmembrane proteome of cyprinid herpesvirus 3, J. Virol. 91 (21) (2017) e01209–e01217.
[58] W. Fuchs, D. Fichtner, S.M. Bergmann, T.C. Mettenleiter, Generation and characterization of koi herpesvirus recombinants lacking viral enzymes of nucleotide metabolism, Arch. Virol. 156 (6) (2011) 1059–1063.
[59] B. Costes, G. Fournier, B. Michel, C. Delforge, V.S. Raj, B. Dewals, et al., Cloning of the koi herpesvirus genome as an infectious bacterial artificial chromosome demonstrates that disruption of the thymidine kinase locus induces partial attenuation in Cyprinus carpio koi, J. Virol. 82 (10) (2008) 4955–4964.
[60] L. Schröder, S. Klafack, S.M. Bergmann, D. Fichtner, Y. Jin, P.Y. Lee, et al., Generation of a potential koi herpesvirus live vaccine by simultaneous deletion of the viral thymidine kinase and dUTPase genes, J. Gen. Virol. 100 (4) (2019) 642–655.
[61] M. Boutier, M. Ronsmans, P. Ouyang, G. Fournier, A. Reschner, K. Rakus, et al., Rational development of an attenuated recombinant cyprinid herpesvirus 3 vaccine using prokaryotic mutagenesis and in vivo bioluminescent imaging, PLoS Pathog. 11 (2) (2015) e1004690.
[62] M. Boutier, Y. Gao, C. Vancsok, N.M. Suárez, A.J. Davison, A. Vanderplasschen, Identification of an essential virulence gene of cyprinid herpesvirus 3, Antivir. Res. 145 (2017) 60–69.
[63] R. Tang, L. Lu, B. Wang, J. Yu, H. Wang, Identification of the immediate-early genes of cyprinid herpesvirus 2, Viruses 12 (9) (2020) 994.
[64] B. Liu, Y. Zhou, K. Li, X. Hu, C. Wang, G. Cao, et al., The complete genome of Cyprinid herpesvirus 2, a new strain isolated from Allogynogenetic crucian carp, Virus Res. 256 (2018) 6–10.
[65] A. Lepa, A.K. Siwicki, Fish herpesvirus diseases: a short review of current knowledge, Acta Vet. 81 (4) (2013) 383–389.
[66] N. Sano, M. Moriwake, R. Hondo, T. Sano, Herpesvirus cyprini: a search for viral genome in infected fish by infected fish by in situ hybridization, J. Fish. Dis. 16 (5) (1993) 495–499.
[67] W. Chai, L. Qi, Y. Zhang, M. Hong, L. Jin, L. Li, et al., Evaluation of cyprinid herpesvirus 2 latency and reactivation in *Carassius gibel*, Microorganisms 8 (3) (2020) 445.
[68] K.E. Eide, T. Miller-Morgan, J.R. Heidel, M.L. Kent, R.J. Bildfell, S. LaPatra, et al., Investigation of koi herpesvirus latency in koi, J. Virol. 85 (10) (2011) 4954–4962.
[69] J.M. Groff, S.E. LaPatra, R.J. Munn, J.G. Zinkl, A viral epizootic in cultured populations of juvenile goldfish due to a putative herpesvirus etiology, J. Vet. Diagn. Invest. 10 (4) (1998) 375–378.
[70] A. Perelberg, M. Smirnov, M. Hutoran, A. Diamant, Y. Bejerano, M. Kotler, Epidemilogical description of a new viral disease afflicting cultured *Cyprinus carpio* in Israel, Isr. J. Aquac. Bamidgeh 55 (1) (2003) 5–12.
[71] T. Sano, N. Morita, N. Shima, M. Akimoto, Herpesvirus cyprini: lethality and oncogenicity, J. Fish. Dis. 14 (5) (1991) 533–543.
[72] H. Rahmati-Holasoo, S. Ahmadivand, S. Shokrpoor, M. El-Matbouli, Detection of Carp pox virus (CyHV-1) from koi (*Cyprinus carpio* L.) in Iran; clinico-pathological and molecular characterization, Mol. Cell. Probes 54 (2020) 101668.
[73] R.P. Hedrick, O. Gilad, S. Yun, J.V. Spangenberg, G.D. Marty, R.W. Nordhausen, et al., A herpesvirus associated with mass mortality of juvenile and adult koi, a strain of common carp, J. Aquat. Anim. Health 12 (1) (2000) 44–57.
[74] H. Rahmati-Holasoo, A. Zargar, S. Ahmadivand, S. Shokrpoor, S. Ezhari, H.A. Ebrahimzadeh Mousavi, First detection of koi herpesvirus from koi, *Cyprinus carpio* L. experiencing mass mortalities in Iran: clinical, histopathological and molecular study, J. Fish. Dis. 39 (2016).
[75] P.H.O. Viadanna, T. Miller-Morgan, T. Peterson, K. Way, D.M. Stone, G.D. Marty, et al., Development of a PCR assay to detect cyprinid herpesvirus 1 in koi and common carp, Dis. Aquat. Org. 123 (1) (2017) 19–27.
[76] P.G. Preena, T.V.A. Kumar, T.K. Johny, A. Dharmaratnam, T.R. Swaminathan, Quick hassle-free detection of cyprinid herpesvirus 2 (CyHV-2) in goldfish using recombinase polymerase amplification-lateral flow dipstick (RPA-LFD) assay, Aquacult. Int. 30 (3) (2022) 1211–1220.
[77] M. Saleh, M. El-Matbouli, Rapid detection of Cyprinid herpesvirus-3 (CyHV-3) using a gold nanoparticle-based hybridization assay, J. Virol. Methods 217 (2015) 50–54.
[78] A.J. Davison, The genome of salmonid herpesvirus 1, J. Virol. 72 (3) (1998) 1974–1982.
[79] B.B. Kaufer, K.W. Jarosinski, N. Osterrieder, Herpesvirus telomeric repeats facilitate genomic integration into host telomeres and mobilization of viral DNA during reactivation, J. Exp. Med. 208 (3) (2011) 605–615.
[80] J.S.L. Patané, J. Martins Jr., L.T. Rangel, J. Belasque, L.A. Digiampietri, A.P. Facincani, et al., Origin and diversification of *Xanthomonas citri subsp. citri* pathotypes revealed by inclusive phylogenomic, dating, and biogeographic analyses, BMC Genom. 20 (1) (2019) 700.

[81] T. Aoki, I. Hirono, K. Kurokawa, H. Fukuda, R. Nahary, A. Eldar, et al., Genome sequences of three koi herpesvirus isolates representing the expanding distribution of an emerging disease threatening koi and common carp worldwide, J. Virol. 81 (10) (2007) 5058–5065.
[82] K.C. Chou, D.W. Elrod, Prediction of membrane protein types and subcellular locations, Proteins 34 (1) (1999) 137–153.
[83] C.M. Hernandez-Garcia, J.J. Finer, Identification and validation of promoters and cis-acting regulatory elements, Plant Sci. 217–218 (2014) 109–119.
[84] J. Kaczynski, T. Cook, R. Urrutia, Sp1- and Krüppel-like transcription factors, Genome Biol. 4 (2) (2003) 206.
[85] B. Thorens, GLUT2, glucose sensing and glucose homeostasis, Diabetologia 58 (2) (2015) 221–232.
[86] S. Khochbin, Histone H1 diversity: bridging regulatory signals to linker histone function, Gene 271 (1) (2001) 1–12.
[87] J. Li, J.H. Ou, Differential regulation of hepatitis B virus gene expression by the Sp1 transcription factor, J. Virol. 75 (18) (2001) 8400–8406.
[88] A. Rang, T. Heise, H. Will, Lack of a role of the interferon-stimulated response element-like region in interferon alpha-induced suppression of Hepatitis B virus in vitro, J. Biol. Chem. 276 (5) (2001) 3531–3535.
[89] S.H. Ou, L.F. Garcia-Martínez, E.J. Paulssen, R.B. Gaynor, Role of flanking E box motifs in human immunodeficiency virus type 1 TATA element function, J. Virol. 68 (11) (1994) 7188–7199.
[90] F. Yu, S. Li, H. Chen, K. Hao, L. Meng, J. Yang, et al., Multiple AT-rich sequences function as a cis-element in the ORF3 promoter in channel catfish virus (Ictalurid herpesvirus 1), J. Fish. Dis. 44 (10) (2021) 1609–1617.
[91] S. Ramamoorthy, T. Opriessnig, N. Pal, F.F. Huang, X.J. Meng, Effect of an interferon-stimulated response element (ISRE) mutant of porcine circovirus type 2 (PCV2) on PCV2-induced pathological lesions in a porcine reproductive and respiratory syndrome virus (PRRSV) co-infection model, Vet. Microbiol. 147 (1–2) (2011) 49–58.
[92] C. Calomme, A. Dekoninck, S. Nizet, E. Adam, T.L. Nguyên, A. Van den Broeke, et al., Overlapping CRE and E box motifs in the enhancer sequences of the bovine leukemia virus 5' long terminal repeat are critical for basal and acetylation-dependent transcriptional activity of the viral promoter: implications for viral latency, J. Virol. 78 (24) (2004) 13848–13864.
[93] J.L. Whitton, J.B. Clements, Replication origins and a sequence involved in coordinate induction of the immediate-early gene family are conserved in an intergenic region of herpes simplex virus, Nucleic Acids Res. 12 (4) (1984) 2061–2079.
[94] M. Kozak, Interpreting cDNA sequences: some insights from studies on translation, Mamm, Genome 7 (8) (1996) 563–574.
[95] P. Menegazzi, M. Galvan, A. Rotola, T. Ravaioli, A. Gonelli, E. Cassai, et al., Temporal mapping of transcripts in human herpesvirus-7, J. Gen. Virol. 80 (Pt 10) (1999) 2705–2712.
[96] A.A. Cullinane, J. Neilan, L. Wilson, A.J. Davison, G. Allen, The DNA sequence of the equine herpesvirus 4 gene encoding glycoprotein gp17/18, the homologue of herpes simplex virus glycoprotein gD, J. Gen. Virol. 74 (Pt 9) (1993) 1959–1964.
[97] A.S. Lauring, J.O. Jones, R. Andino, Rationalizing the development of live attenuated virus vaccines, Nat. Biotechnol. 28 (6) (2010) 573–579.
[98] S. Klafack, A.-S. Fiston-Lavier, S.M. Bergmann, S. Hammoumi, L. Schröder, W. Fuchs, et al., Cyprinid herpesvirus 3 evolves in vitro through an assemblage of haplotypes that alternatively become dominant or under-represented, Viruses 11 (8) (2019) 754.