Virology 526 (2019) 81–90

Contents lists available at ScienceDirect

Virology

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

Long-range PCR and high-throughput sequencing of *Ostreid herpesvirus 1* indicate high genetic diversity and complex evolution process

Chang-Ming Bai[a], Benjamin Morga[b], Umberto Rosani[c], Jie Shi[a], Chen Li[a], Lu-Sheng Xin[a], Chong-Ming Wang[a,*]

[a] *Key Laboratory of Maricultural Organism Disease Control, Ministry of Agriculture; Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology; Qingdao Key Laboratory of Mariculture Epidemiology and Biosecurity; Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, Shandong, China*

[b] *Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Laboratoire de Génétique et Pathologie (LGP), Avenue de Mus de Loup, 17390 La Tremblade, France*

[c] *Department of Biology, University of Padua, Padua, Italy*

ARTICLE INFO

Keywords:
Long-range PCR
OsHV-1
High-throughput sequencing
Evolution
Diversity

ABSTRACT

Ostreid herpesvirus 1 (OsHV-1) is an important pathogen associated with mass mortalities of cultivated marine mollusks worldwide. Since no cell line allows OsHV-1 replication *in vitro*, it is difficult to isolate enough high-purity viral DNA for High-Throughput Sequencing (HTS). We developed an efficient approach for the enrichment of OsHV-1 DNA for HTS with long-range PCR. Twenty-three primer pairs were designed to cover 99.3% of the reference genome, and their performances were examined on ten OsHV-1 infected samples. Amplicon mixtures from six successfully amplified samples were sequenced with Illumina platform, and one of them (ZK0118) was also sequenced with the PacBio platform. PacBio reads were assembled into 2 scaffolds compared to 9 − 68 scaffolds assembled from the Illumina reads. Genomic comparison confirmed high genetic diversity among OsHV-1 variants. Phylogenetic analysis revealed that OsHV-1 evolution was mainly impacted by its host species rather than spatial segregation.

1. Introduction

Ostreid herpesvirus 1 (OsHV-1), as the unique member of the genus *Ostreavirus* (family Malacoherpesviridae, order Herpesvirales), is the first herpesvirus isolated from invertebrates in the early 1990s (Hine et al., 1992; Nicolas et al., 1992). OsHV-1 is able to infect different marine mollusk species with a wide geographic distribution (Arzul et al., 2017). Several OsHV-1 genotypes exhibit difference in host species and virulence have been identified around the world (Bai et al., 2015; Renault et al., 2012). For example, OsHV-1 µVar and related variants has been associated with increased mortality of Pacific oysters, *Crassostrea gigas*, in Europe and Australia since 2008 (Jenkins et al., 2013; Segarra et al., 2010). OsHV-1-SB and AVNV has been associated with mass mortalities of *Scapharca broughtonii* and *Chlamys farreri* respectively in China (Bai et al., 2016, 2015). Genome sequences of three OsHV-1 variants have been obtained with bacteriophage lambda libraries or genome walking procedures (Davison et al., 2005; Ren et al., 2013; Xia et al., 2015). More recently, two OsHV-1 µVar genomes were resolved with Second-Generation Sequencing (SGS) (Burioli et al., 2017). Since no cell line is available for the propagation of OsHV-1, it is difficult to get enough OsHV-1 DNA with high purity for SGS (Burioli et al., 2017). Although OsHV-1 particles could be purified with density gradient centrifugation, the purification rates were low and thus always require more than 10 g tissues (Burioli et al., 2017).

It is inferred that OsHV-1 variants have a higher level of genetic differentiation compared to that of vertebrate herpesviruses (Burioli et al., 2017; Xia et al., 2015). PCR and Sanger sequencing of several molecular markers of OsHV-1 variants also indicated low genetic similarity (Renault et al., 2012). However, these studies always focused on limited and arbitrarily selected genomic regions, which were not sufficient for a complete understanding of genome-wide nucleotide variations. In order to resolve the problem, we need to characterize variations at the genome level, which requires many genome sequences of different OsHV-1 variants. These demands make Sanger sequencing extremely challenging and technically impractical. SGS and Third-Generation Sequencing (TGS) have many advantages compared to Sanger sequencing, which have been widely used for genome sequencing of vertebrate herpesviruses (Depledge et al., 2011; Kolb et al.,

* Corresponding author.
E-mail address: wangcm@ysfri.ac.cn (C.-M. Wang).

https://doi.org/10.1016/j.virol.2018.09.026
Received 19 July 2018; Received in revised form 29 September 2018; Accepted 29 September 2018
Available online 24 October 2018
0042-6822/

2011; Spatz and Rue, 2008). However, SGS and TGS require the preparation of considerable amounts of viral DNA, which is difficult for OsHV-1 due to the lack of permissive cell line for virus purification. Moreover, the short read lengths generated by SGS make them poorly suited for *de novo* assembly of complex genomic regions typical of herpesviruses and to study the diversity of the virus (Kolb et al., 2011).

Long-range PCR (LR-PCR) can amplify candidate genomic regions with high sensitivity and specificity (Jia et al., 2014). Combined with SGS, LR-PCR has been used for viral genome sequencing, genetic variation detecting and systematic studies (Chan et al., 2012; Kvisgaard et al., 2013; Morrison et al., 2018; Ozcelik et al., 2012; Uribe-Convers et al., 2014). The genome of OsHV-1 contains two unique regions each flanked by inverted repeats longer than 7500 bp (Davison et al., 2005), which may lead to fragmented assemblies due to the short read lengths generated by SGS sequencing (Quail et al., 2012). The PacBio platform, by virtue of its long read lengths, was found useful for *de novo* assemblies of complex repeat regions (Quail et al., 2012). We describe here the use of 23 primer pairs for LR-PCR to amplify OsHV-1 genomic DNA. Then, we quantified and successfully pooled the PCR products of each sample in equimolecular proportions. The amplicon mixtures were used for genome sequencing with PacBio RS II and/or Illumina HiSeq. 4000 platforms. The resolved genome sequences (one complete and five partial genomes) were applied for further genomic comparison and phylogenetic analysis with the other five published OsHV-1 variants.

2. Materials and methods

2.1. Sample preparation

Ten OsHV-1 positive bivalve samples determined by qPCR (Martenot et al., 2010) and collected during mortality events in Shandong Province (China) were employed in the present study (Table 1, Supplementary figure 1). The samples were stored at −40 °C until use. For adult tissues, a homogenate preparation protocol as described by (Schikorski et al., 2011) was used before DNA extraction. Briefly, about 1 g of mantles and gills were dissected from a specimen and crushed with tissue homogenizer in 9 mL of 0.22 μm filtered seawater. Then the tissue homogenate was centrifuged at 1000*g* for 5 min at 4 °C. Finally, the supernatant was collected and used for DNA extraction. For Pacific oyster larvae, specimens were washed in double distilled water and grinded with liquid nitrogen. Corresponding healthy and OsHV-1 negative sample was determined by qPCR.

2.2. DNA extraction and OsHV-1 DNA quantification

Total DNA extraction was performed from 180 μL tissue supernatant or 50 mg grinded larvae using the Qiagen DNeasy Blood and Tissue kit according to the manufacturer's protocol. Elution was performed in 60 μL of AE buffer provided in the kit.

The detection and quantification of OsHV-1 DNA were carried out by qPCR adapted from previously published protocol (Martenot et al., 2010) and fully described in Bai et al. (2016). Each sample was tested in duplicate. The quantification of OsHV-1 DNA was estimated as the mean OsHV-1 DNA copies per μL of tissue supernatant or per mg of grinded larvae for the two replicates.

2.3. Primer design and long-range PCR

Primer pairs were initially designed with the online version of GenoFrag (http://genoweb1.irisa.fr/Serveur-GPO/outils/generationAmorces/GENOFRAG/indexen.php) (Ben Zakour et al., 2004) using the genome sequence of OsHV-1 reference variant (GenBank accession no. AY509253). Potential primers were selected based on default values of a series of filters defining the primer length (25 bp), GC content (48%), hairpin, stability, auto-complementarity and so on. Then the software picked out a list of primer pairs covering the viral genome from the potential primers previously generated. The amplicon length was set in a range of 9–11 kb, and overlapping the adjacent amplicons by 500–1500 bp (Table 2). When the automatically selected primers failed to amplify, new primer pairs were manually selected for testing from the potential primers around its original binding site. Primers of three amplicon sites at the termini of OsHV-1 genome (Table 2) were designed with Primer Premier 5 (Premier Biosoft International, Paolo Alto, CA, USA) according to general primer selection criteria.

All LR-PCR reactions were performed using the TaKaRa PrimeSTAR GXL DNA polymerase, which showed high fidelity and stability in amplifying long PCR products as reported previously (Jia et al., 2014). The PCR reaction was performed in a 50 μL reaction volume that included 10 μL 5 × PrimeSTAR GXL Buffer, 4 μL dNTP Mixture (2.5 mM each), 1 μL TaKaRa PrimeSTAR GXL DNA Polymerase, 1 μL each of forward and reverse primers (10 μM), 2 μL DNA template and 31 μL of PCR-grade H_2O. LR-PCR were performed using Veriti Thermal Cycler (Applied Biosystems) under the following conditions: 94 °C for 1 min, followed by 35 cycles of amplification (denaturation 98 °C, 10 s; annealing 50 °C, 15 s, extension 68 °C, 10 min) and hold at 4 °C. PCR product sizes were detected on 0.8% 1 × TAE agarose gels stained with GeneFinder™ (Zeesan Biotechnology Inc.).

2.4. Genome sequencing and assembly

PCR amplicons of OsHV-1 infected samples (Table 1) were purified with QIAquick PCR Purification Kit (Qiagen) and quantified using Qubit Fluorometer using the dsDNA Assay Kit (Life Technologies). Then the quantified amplicons for each sample were mixed in equimolecular proportions and send for genome sequencing and assembly at Guangzhou Gene Denovo Biotechnology Co., Ltd. ZK0118 was

Table 1
Ostreid herpesvirus 1 (OsHV-1) infected samples used for long-range PCR.

Sample ID	Date of sampling	Origin	Host species	Age[a]	Viral loads
ZK0118	**August 2001**	**Qingdao**	***Chlamys farreri***	**Adult**	**3.4×10^4**
ZK2002	**August 2002**	**Qingdao**	***Chlamys farreri***	**Adult**	**2.1×10^4**
ZK2003	**July 2003**	**Qingdao**	***Chlamys farreri***	**Adult**	**1.6×10^4**
ZK2004	**August 2004**	**Qingdao**	***Chlamys farreri***	**Adult**	**3.2×10^4**
ZK2006	July 2006	Qingdao	*Chlamys farreri*	Adult	8.5×10^2
ZK2007	September 2007	Rongcheng	*Chlamys farreri*	Adult	4.8×10^3
ZK2008	**August 2008**	**Rongcheng**	***Chlamys farreri***	**Juvenile**	**8.8×10^3**
ZK2011	August 2011	Changdao	*Chlamys farreri*	Adult	7.3×10^2
KH2015	**June 2015**	**Rizhao**	***Scapharca broughtonii***	**Adult**	**4.6×10^4**
ML2015	May 2015	Qingdao	*Crassostrea gigas*	larva	3.0×10^3

The six samples printed in bold were successfully amplified and sequenced.

[a] Developmental stages corresponding to juvenile (6–12 months old) or adult (over 12 months old for *Chlamys farreri* and over 24 months old for *Scapharca broughtonii*).

Table 2
Primer sequences and primary characteristics for long-range PCR of *Ostreid herpesvirus 1*.

Primer ID	Primer sequences (5'-3')	Size (bp)	Starting coordinates	Stopping coordinates	*Tm*[*] (°C)	Amplicon size (bp)	Amplicon overlap (bp)	Genomic location	Variation[†]
GF1F[#]	TCCCGCCAATACCCATAATGCA	22	67	88	61	11,102	None	TR_L-U_L	None
GF1R	GGTTTCCAGTAGGGTGTTTAAGAGC	25	11168	11144	69				None
GF2F	GCAATCCAGTTCCCAAACCAATAGG	25	9991	10015	69	9530	1178	U_L	None
GF2R	ATCGCTTCCTATCACCTTGTGGTCT	25	19520	19496	69				None
GF3F	CCGTGAAATATCTGCCAAGGTGTTG	25	18895	18919	69	9570	626	U_L	Absent in KY
GF3R	CGACCAGGAGAACATGAACGACTTT	25	28464	28440	69				None
GF4F	GCCCTCCTATTGGTACAAGATTGCT	25	27425	24449	69	9571	1040	U_L	1 SNP in KY
GF4R	GAGCACAAACACTACCGCATACATG	25	36995	36971	69				None
GF5F	GCTTGTTTCTTGGTGTCTGAGGTCA	25	36429	36453	69	9660	567	U_L	None
GF5R	ATGGCAGAAATAGAAACCCGAGGTC	25	46088	46064	69				None
GF6F	ATCCCAGTCTGTCAAATGCTCTCTC	25	45470	45494	69	9166	619	U_L	None
GF6R	CCAGATATGAAGAGGAAGGGATGTC	25	54635	54611	69				None
GF7F	ATGCCTGGGCGTAATTGTCTCTTGA	25	54371	54395	69	9418	265	U_L	None
GF7R	CCAACTCTTCATCGTCACTCATCTC	25	63788	63764	69				None
GF8F	GGGCGTTCACTTTAGACTTCCAATC	25	63129	63153	69	9564	660	U_L	None
GF8R	TTCCCTGGCGATACTCTCATAGACA	25	72692	72668	69				1 SNP in KP
GF9F	GGTCCGTCAACATCGAGAAAGAGAA	25	71994	72018	69	9509	699	U_L	1 SNP in KP
GF9R	CACGATAAATATGCTGCCTGGGTCA	25	81502	81478	69				None
GF10F	AGGGCGAGCATGGTCACATTTCAAA	25	80922	80946	69	9671	581	U_L	1 SNP in KP
GF10R	GGGATATTCTGAGGGTTGTTGTGGA	25	90592	90568	69				None
GF11F	GCCCAATAAACCTACAGAGGATGAG	25	89659	89683	69	9573	934	U_L	None
GF11R	ATGGCAGATTCAGGAGAGGGTTGTA	25	99231	99207	69				None
GF12F	TGGCTTCTGTGGTGGTAGTTGTTGT	25	98631	98655	69	9344	601	U_L	None
GF12R	CGGCATTACCAAATATAGGCACACG	25	107974	107950	69				None
GF13F	GGCACGCTCATTCTTACAACTCTTG	25	106810	106834	69	9418	1165	U_L	None
GF13R	CAACCAATCAGATCGACGAGACTCA	25	116227	116203	69				Absent in KP
GF14F	GCGATGCCTTAATTGTTGCCAGAGT	25	115058	115082	69	10,145	1170	U_L	None
GF14R	TCCTGTGGAATGGTTGTTGGTGATG	25	125202	125178	69				None
GF15F	GCAAACGAAAGAGCGGCTATAACAG	25	124673	124697	69	9755	530	U_L	None
GF15R	CTCCGTCATCGGTGTTATTACTAGG	25	134427	134403	69				3 SNP in KY
GF15-1F	***GAGCGGCTATAACAGTGGTACAAAG***	***25***	***124683***	***124707***	***69***	***9739***	***520***	U_L	None
GF15–1R	***CATCGGTGTTATTACTAGGCGTCCA***	***25***	***134421***	***134397***	***69***				2 SNP in KY
GF16F	GTCGGTTGTGGGTTTGGAAATGTAG	25	133704	133728	69	9302	724	U_L	None
GF16R	CTGGAAGCGAGTGTCAAGGTTAAAC	25	143005	142981	69				None
GF16-1F	***GCTTTATGAATGGGAGGCTGGTGAT***	***25***	***133068***	***133092***	***69***	***9618***	***1354***	U_L	None
GF16–1R	***GGAAGAGGTTGGGAAGACATACATG***	***25***	***142685***	***142661***	***69***				None
GF17F	ACCAAAGACCATCACTGCCAACAAC	25	142044	142068	69	9438	642	U_L	None
GF17R	ACTGCCGATTTACTTGCCTCTTCTG	25	151481	151457	69				None
GF17-1F	***GCAGTTTGATTCATGTGTGGCAGAG***	***25***	***141165***	***141189***	***69***	***10,662***	***1841***	U_L	None
GF17–1R	***GCCATCCACCTCATATCCATTTCTC***	***25***	***151826***	***151802***	***69***				None
GF18F	GGATGATGGATTGTTGGACGAGAGA	25	149769	149793	69	9993	2058	U_L	None
GF18R	GCTGCGGTCAGTACATGGTCATTTA	25	159761	159737	69				None
GF19F	CGTGAAGACGCCATGAAGAGAAGTT	25	159169	159193	69	9758	593	U_L	None
GF19R	TCTGCCAGCCTCTGTGAACTTGTAA	25	168926	168902	69				None
GF20F	TTGGCAGATGAGGACACCTTATACC	25	167525	167549	69	9861	1402	U_L-IR_L	None
GF20R	TTCCTGATTCCTCCACGCCATAACA	25	177385	177361	69				Absent in KP
GF20-1F	***GCTGCTGCTCCAACTCAAGAATACT***	***25***	***166772***	***166796***	***69***	***11,558***	***2155***	U_L-IR_L	None
GF20–1R	***CCATTCATCTTGCCGCACATCACAT***	***25***	***178329***	***178305***	***69***				Absent in KP
GF21F	GGCAGCTAGTAAGGTCAATCTCAAC	25	175889	175913	69	9273	1497	IR_L-X-IR_S	Absent in KP
GF21R	TGGTTCCCTGGCGACGTTTACATAA	25	185161	185137	69				None
GF21-1F	***GCGAGAAGACGGAATTGGAAATCAC***	***25***	***176005***	***176029***	***69***	***9157***	***2325***	IR_L-X-IR_S	Absent in KP
GF21R	TGGTTCCCTGGCGACGTTTACATAA	25	185161	185137	69				None
GF22F[#]	CGAAACGACAGGTTGAAGTGAGG	23	183809	183831	65	13,029	1353	X-IR_S-U_S	Absent in KP 2 SNP in KY
GF22R[#]	CAATGAATCGCCAATTAAGGAGG	23	196837	196815	61				None
GF23F[#]	CCATTTGTCAATCTCGGTTCTGC	23	196385	196407	63	9903	453	U_S-TR_S	None
GF23R[#]	GGAGGTGGGGTTTGAATACGAAG	23	206287	206265	65				None
GF23-1F[#]	***GGCAGTCTGGTAGCAATG***	***18***	***195266***	***195283***	***55***	***11,037***	***1572***	U_S-TR_S	Absent in KP
GF23–1R[#]	***AGATGACGAATCGGAGGA***	***18***	***206302***	***206285***	***51***				1 SNP in GQ

The primer pairs printed in italicized and bold were used for the amplifications of only some specimens listed below. GF15-1F/R used for ZK2003, ZK2004 and KH2015; GF16-1F/R and GF20-1F/R used for ZK2003 and ZK2004; GF17-1F/R used for ZK2001; GF21-1F used for ZK2004; GF23-1F/R used for KH2015.

[*] Tm value (°C) = [(nA + nT) × 2] + [(nC + nG) × 4] - 5, n represents the number of nucleotides.

[#] represents primers designed by Primer Premier 5.

[†] Primer sequence variations against the other available OsHV-1 genome. GQ represents AVNV, KP represents OsHV-1-SB, KY represents OsHV-1 μVar variant A and B.

sequenced with both PacBio and Illumina platforms, the other five samples were sequenced with Illumina platform only.

For PacBio sequencing, 10 μg of ZK0118 amplicon mixtures were directly end-repaired for library preparation (9–13 kb SMRTBell library) according to the manufacturer's specification (Pacific Biosciences, Menlo Park, CA). Sequencing was performed on the PacBio RS II sequencer according to standard protocols (MagBead Standard Seq v2 loading, 1 × 180 min movie) using the P4-C2 chemistry. Reads longer than 500 bp and with a quality value over 0.75 were merged together into a single dataset. Next, the hierarchical genome-assembly

process (HGAP) pipeline (Chin et al., 2013) was used to correct for random errors in the long seed reads (seed length threshold 6 kb) by aligning the shorter reads against them. The resulting corrected and preassembled reads were used for the final *de novo* assembly using Celera Assembler with an Overlap-Layout-Consensus (OLC) strategy (Myers et al., 2000). Since SMRT sequencing features very little variations of the quality throughout the reads, no quality values were used during the assembly. To validate the quality of the assembly and determine the final genome sequence, the Quiver consensus algorithm was used (Chin et al., 2013).

For Illumina sequencing, 2 μg amplicon mixtures of each sample were firstly sonicated randomly, and then end-repaired, A-tailed, and adaptor ligated according to the Illumina TruSeq DNA preparation protocol. Genome sequencing was performed on the Illumina HiSeq. 4000 sequencer using the pair-end technology (PE 150). Raw reads were firstly processed using an in-house perl script. In this step, clean reads were obtained by removing adaptor sequences and low-quality reads with $>$ 40% bases having a PHRED quality scores of $\leq$ 20, with $\geq$ 10% unidentified nucleotides (N). The resulting clean reads were *de novo* assembled using the SOAPdenovo (ver. 2.04) (Luo et al., 2012), followed by modification using gapclose (version 1.12). The clean reads of each sample were also used for a reference genome mapping on its' phylogenetically closest reference genome using the CLC mapper tool (CLC Genomic Workbench, Qiagen) applying 0.8 and 0.5 of similarity and length parameters, respectively. The number of covered reference positions (with at least 50x of sequencing depth) were counted to calculate the percentage of the reference genome covered by each read dataset.

2.5. Gap closure, verification of ambiguous regions of ZK0118

No sequence read of GF16 amplicons generated in ZK0118 and the other three samples (ZK2002, ZK2008 and KH2015) with either PacBio or Illumina, although the amplicons were detected by electrophoresis, while GF16-1 amplicons targeting the same genome area were sequenced successfully in the other two samples (ZK2003 and ZK2004). In order to obtain the complete genome of ZK0118, GF16-1 primer pair was employed for amplification of the ZK0118 DNA fragment. The amplified fragments were then sequenced with Illumina, and the generated sequence reads were used for filling the gap. The other two PCR amplicons (GF22 and GF23) with a low coverage of PacBio reads were validated with Illumina sequencing. One region with 82 ambiguous bases (nt: 48196–48227) in the assembly of ZK0118 were determined by Sanger sequencing of PCR product. The genome termini of ZK0118 were identified by a PCR method for amplifying termini using the Marathon cDNA amplification kit (Clontech Laboratories) as previously described (Davison et al., 2003).

2.6. Gene annotation of ZK0118 and genome comparisons

The potential open reading frames (ORFs) were predicted using the GeneMarkS (Besemer et al., 2001) and GATU (Tcherepanov et al., 2006) relative to OsHV-1 and adjusted manually.

The pairwise and multiple alignments of ZK0118, OsHV-1 reference variant, AVNV (GenBank accession no. GQ153938), OsHV-1-SB (No. KP412538), OsHV-1 μVar variant A (No. KY242785) and variant B (No. KY271630) were performed with the mauveAligner tool (Darling et al., 2004) implanted in Geneious (ver. 4.8.5) and adjusted manually. The indels larger than 10 bases and Single-Nucleotide Variations (SNV) among these genomes were identified with Mega v. 7.0.14 (Kumar et al., 2016) based on the multiple alignment.

2.7. Phylogenetic analysis

Since the complete genome sequences were not resolved for the five samples sequenced with only Illumina platform, the scaffolds of each sample were concatenated and aligned with the complete genome sequence of ZK0118 sequenced in the present study and that of the other 5 available OsHV-1 variants in public database. All positions containing aligned gaps and missing data were eliminated in the following analysis. Genomic sequences of the 11 variants were aligned with MAFFT version 7 with the default settings (Katoh and Standley, 2013). The best-fit nucleotide substitution model was determined using the Akaike Information in MEGA v. 7.0.14 (Kumar et al., 2016). Phylogenetic relationship was inferred by using the Neighbor-Joining method implemented in Geneious (ver. 4.8.5). Branch support was estimated with 100 bootstrap replicates. Since there is no suitable outgroup available, the NJ tree was displayed as unrooted.

3. Results

3.1. Specificity and sensitivity of OsHV-1 primers

A total of 23 primer pairs (labeled as GF1–23F/R respectively) were designed for long-range PCR, which span 99.3% of the OsHV-1 reference genome (Table 2). Due to the amplification failure of 7 primer pairs when applied to some specimens, they were redesigned manually (Table 2, Supplementary figure 2). All 23 amplicons were amplified with long-range PCR using as templates supernatants with viral DNA loads $\geq 8.8 \times 10^3$ copies per μL (Fig. 1. A). While for specimens with viral DNA loads $\leq 4.8 \times 10^3$ per μL, less than 13 amplicons were amplified, despite some optimizations. No amplicon was obtained for negative controls, which indicated that all primer pairs were specific for OsHV-1 DNA (Fig. 1. B).

3.2. Genome sequencing and assembly

A total of 150, 292 polymerase reads with an average length of 6700 bp and an N50 of 20, 079 bp were generated from PacBio RS II platform (Table 3). After read quality trimming, 60, 514 reads with an average length of 14, 099 bp and an N50 of 21, 771 bp were obtained. *De novo* assembly results in 2 scaffolds with lengths of 136, 911 bp and 59, 926 bp, respectively. The average sequencing depth was 3, 218x across the whole genome, with a range between 0 (GF16) and 4807 × (GF3) (Fig. 2). Regions characterized by overlapping amplicons showed higher coverage values, while a limited coverage bias was observed among different amplicons. Moreover, we found lower sequencing depths in the 3' - terminal and internal repeat regions (IR_S and TR_S).

Illumina sequencing generated a total of 41, 738, 566 reads. After trimming, 35, 515, 518 clean reads were obtained. *De novo* genome assembly resulted in 9–68 scaffolds for each sample, covering 88.09–95.23% of the reference sequence with average sequencing depths ranging from 4, 139x to 7, 319 ×. The sequencing depths along the whole genome were unbiased; except for the terminal repeat regions. A summary of sequencing data and assemblies generated from the PacBio RS II in comparison to that generated from the Illumina HiSeq. 4000 platform was listed in Table 3. Since the assembly of five samples (ZK2002, ZK2003, ZK2004, ZK2008 and KH2015) sequenced with only Illumina platform generated multiple scaffolds, their complete genome sequences were not determined. Reference mapping of the Illumina clean reads of ZK0118.2, ZK2002, ZK2003, ZK2004 and ZK2008 resulted in 96.0%, 95.6%, 99.0%, 98.8% and 95.6% of the AVNV genome covered with a minimal sequencing depth of 50 ×, whereas KH2012 reads covered 96.6% of the OsHV-1-SB genome. The clean Illumina reads and raw PacBio reads of the whole project has been submitted to the Sequence Read Archive database with BioProject ID: PRJNA448032.

Since no read generated from GF16 amplicons of ZK0118, Illumina reads of GF16-1 amplicons were used for filling the gap between the 2 Pacbio scaffolds. After verification of ambiguous regions and determination of genomic terminals, the ZK0118 genome was finally resolved

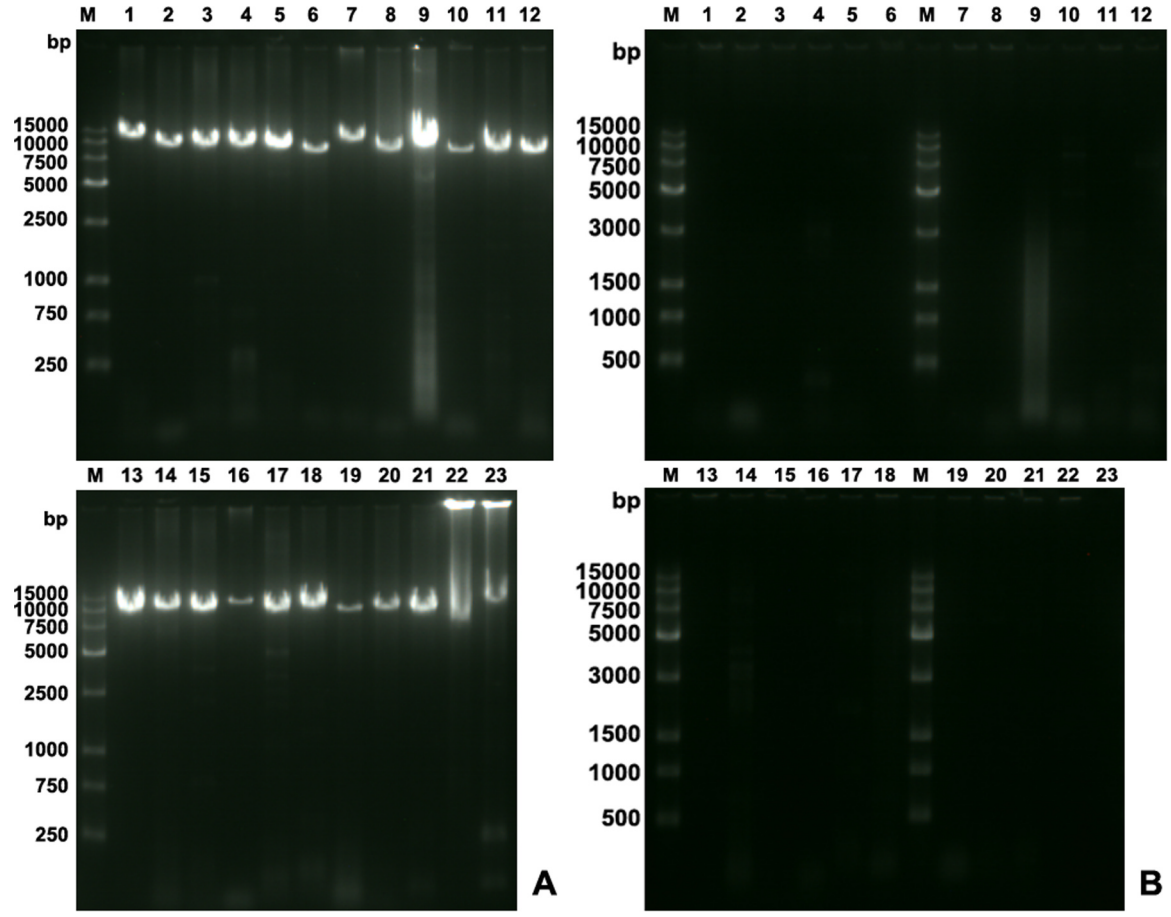

Fig. 1. Agarose gel electrophoreses (0.8%) analysis of long range PCR amplicons. (A) PCR amplicons for ZK0118, (B) PCR amplicons for negative control, lanes 1–23 corresponds to 23 PCR amplicons respectively.

to be 204, 652 bp in length. The G/C content of ZK0118 genome is 38.6%, which is in agree with the previously published OsHV-1 variants. The presence of a new large deletion (> than 5000 bp) was verified by PCR, cloning and Sanger sequencing (data not shown). The genome structure of ZK0118 is similar to that of OsHV-1 reference, which contains two unique regions each flanked by two inverted repeats (TRL/IRL and TRS/IRS), with a third unique region situated between IRL and IRS. The genome sequences reported in the present study has been submitted to GenBank Nucleotide Sequence Database with Accession No. MF509813.

3.3. Gene annotation of ZK0118 and genome comparisons

Computer-assisted analysis revealed 126 distinct ORFs in ZK0118 genome, ranging from 42 to 1879 amino acids (aa) in length. For the 123 ORFs with a counterpart in the OsHV-1 reference, we assigned the same number as those in OsHV-1 reference (ORFs 1–122 and 124). Three ORFs, absent in OsHV-1 reference, were named as ORFs 125–127, respectively. Due to the large deletion (more than 5000 bp) occurred at the IR_S-U_S region of ZK0118, 4 OsHV-1 ORFs (ORF 120, 121 and 122 in the IR_S region and ORF 123 in the U_S region) were deleted in ZK0118.

The pairwise comparison of ZK0118 with OsHV-1 reference, AVNV, OsHV-1-SB, OsHV-1 µVar variant A and variant B showed 94.7%, 96.8%, 92.5%, 92.7% and 92.7% of similarity, respectively (Table 4). Multiple sequence alignments of the six OsHV-1 genomes resulted in a 216, 079 bp aligned positions, because of the large indel (Supplementary Data 1). We identified 1583 SNVs over the whole alignment of 6 OsHV-1 variants, which occurred at a rate of 7.33 per kb (Table 5). We identified 496 SNVs among the 3 European variants from the same species (*C. gigas*), which were lower than 1, 027 SNVs identified among the 3 Chinese variants from 2 different host species (*C. farreri* and *S. broughtonii*). While the lowest number of SNVs (274) and frequency (1.31 per kb) were identified when only the 2 variants from the same species (*C. farreri*) in China were analyzed. The SNVs distributed unevenly across OsHV-1 genome within all groups. A higher

Table 3
Summary of high-throughput sequencing results.

Sample ID	No. of Raw reads	No. of Clean reads	No. of Scaffolds	Mean length of Scaffolds	N50 of Scaffolds	Coverage on OsHV-1 (%)	Estimated viral loads
ZK0118.3	150,292	60514	2	984,19	136,911	95.50	3.4×10^4
ZK0118.2	15,014,566	14,575,518	9	21,046	129,144	93.64	3.4×10^4
ZK2002	5,304,000	4236,000	36	5019	9888	92.71	2.1×10^4
ZK2003	5,336,000	4228,000	42	4353	8028	91.46	1.6×10^4
ZK2004	5,784,000	4360,000	17	1,1200	14,746	95.23	3.2×10^4
ZK2008	5,412,000	4248,000	68	2547	3813	88.09	8.8×10^3
KH2015	4,888,000	3868,000	16	1,2042	28,233	91.35	4.6×10^4

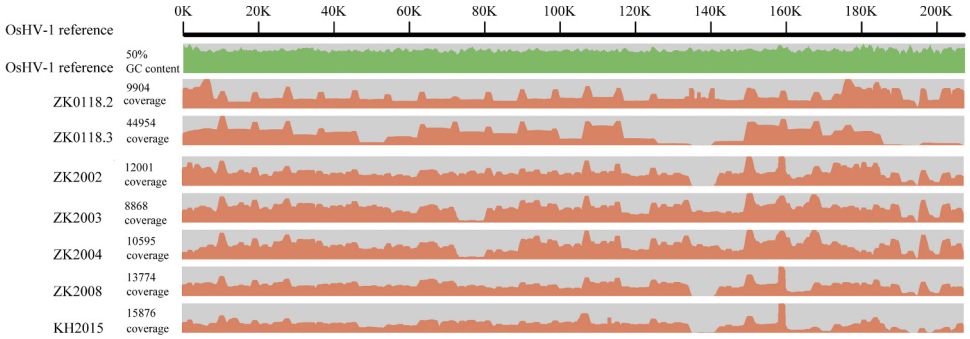

Fig. 2. Coverage and GC content across the assembled genomes. OsHV-1 reference (AY509253) was used as a reference of nucleotide positions. ZK0118.2: sequence data of Illumina HiSeq. 4000 platform, ZK0118.3: sequence data of PacBio RS II platform. GC content of OsHV-1 reference was displayed.

frequency of SNVs was always detected in the inverted repeats than the unique regions.

Further indel polymorphism analysis of the multiple alignment results revealed a total of 40 indels (> 10 bp) characteristic of a specific variant or shared by a subgroup of variants (Table 6). There were 6 indels occurred only in OsHV-1 variants found in *C. farreri* in China, 2 identical indels in OsHV-1 variants found in *C. farreri* and *S. broughtonii* in China. Interestingly, there was also 1 deletion of 600 bp occurred only in *S. broughtonii* in China and *C. gigas* in Europe (OsHV-1 µVar). The indels occurred most frequently in IR_S and TR_S regions (19 indels), which involved the alteration of ORFs 115–117 and 120–123. While only 7 indels were found in the IR_L and TR_L regions, 6 of which were 3 pairs of inverted repeats found in the non-coding regions of OsHV-1 reference variant. The other one was a large deletion occurred in the IR_L regions of OsHV-1 reference variant, which involved the deletion of ORFs 114 and 4. Thirteen indels were found in the U_L region, 11 of them were associated with variation of the alteration of 15 ORFs, which include ORFs 11, 32, 36–38, IN.1-IN.4, 48, 50, 62, 63, 106, 114. Additionally, a deletion of X region in the genome of OsHV-1-SB led to the complete loss of ORF 115. The putative functions of 11 ORFs from the 23 affected ORFs described above were identified. Three ORFs were predicted to encode membrane proteins (ORFs 32, 36 and IN.4), two of which were putative membrane glycoproteins (ORFs 32 and IN.4). Four ORFs included RING finger domain (ORFs 38, 106, 117 and 123), three ORFs were predicted to encode secreted proteins (ORFs IN.1, 50 and 120), and one ORF encode Replication Origin-binding Protein (ORF 115).

3.4. Phylogenetic analysis

Alignment of 6 OsHV-1 genomes sequenced in the present study with that of 5 available variants in public database generated 216,626 aligned positions (Supplementary Data 2). The nucleotide positions containing gaps and missing data were discharged and resulted in 139, 205 informative positions. The evolutionary distances were computed using the Tajima-Nei model. The estimated phylogenetic tree is drawn to scale, with branch lengths measured in the number of base substitutions per site (Fig. 3). The tree divided the 11 OsHV-1 variants into three main groups with 100% of bootstrap support, corresponding to the hosting species. The evolution relationship of variants identified from each host were also resolved with high bootstrap support values (> 90%). The distribution pattern of the six variants identified from *C. farreri* was slightly correlated to their sampling time except ZK2003.

4. Discussion

OsHV-1 and variants characterized by broad genetic, host, and virulence diversities have been identified in China and the other countries (Bai et al., 2015; Barbosa Solomieu et al., 2015). However, genome sequence data of these variants was scarce due to technical constrains, mainly linked to the difficulty of obtaining enough purified viral particles for high-throughput sequencing (Burioli et al., 2017). One of the focus of the present study was to develop a fast and robust method for enrichment of viral DNA prior to high-throughput sequencing. To date, two methods have been used for the enrichment of herpesvirus DNA from tissue samples (Depledge et al., 2011; Donaldson et al., 2013; Hammoumi et al., 2016; Kwok et al., 2014; Olp et al., 2015; Tweedy et al., 2015). The first and commonly used method is a hybridization-based approach, which captures viral DNA by hybridization with RNA baits designed across the genome (Depledge et al., 2011). The second one is a LR-PCR based approach, which amplifies viral DNA using primers designed covering the complete genomic sequence (Jia et al., 2014). Application of these two approaches for Human Herpesvirus 6A (HHV-6A) genome sequencing generated identical consensus sequences, with similar variant-calling efficacy (Tweedy et al., 2015). However, there is no commercial RNA baits available for OsHV-1 and custom design and production of RNA baits is expensive. Additionally, hybridization-based approaches typically require more DNA than LR-PCR based approach. In addition, PCR technique requires very low amounts of DNA samples, inexpensive regents and common equipments. And a long-range DNA polymerase with higher fidelity and better performance has also been picked out of 6 commercially available polymerases (Jia et al., 2014).

In the present study, a protocol was developed for enrichment of OsHV-1 DNA using LR-PCR, followed by high throughput sequencing in

Table 4
Pairwise genome comparison of published OsHV-1 genomes and ZK0118.

	OsHV-1 (AY509253)[a]	AVNV (GQ153938)	OsHV-1-SB (KP412538)	OsHV-1 µVar variant A (KY242785)	OsHV-1 µVar variant B (KY271630)	ZK0118 (MF509813)
OsHV-1		96.7	89.7	94.4	94.4	94.7
AVNV	96.7		92.6	94.0	94.0	96.8
OsHV-1-SB	89.7	92.6		88.3	88.1	92.5
OsHV-1 µVar variant A	94.4	94.0	88.3		99.96	92.7
OsHV-1 µVar variant B	94.4	94.0	88.1	99.96		92.7
ZK0118	94.7	96.8	92.5	92.7	92.7	

[a] GenBank Accession numbers of each variant was provided in parentheses.

Table 5
Characterization of Single-Nucleotide Variations (SNVs) in all six and different groups of OsHV-1 variants.

Regions	All six variants			Three variants from oysters in Europe			Three variants from scallops and clams in China			Two variants from scallops in China		
	Alignment length (bp)	No. of SNV	Frequency (bp/Kb)	Alignment length (bp)	No. of SNV	Frequency (bp/Kb)	Alignment length (bp)	No. of SNVs	Frequency (bp/Kb)	Alignment length (bp)	No. of SNV	Frequency (bp/Kb)
Genome	216,079	1583	7.33	211,882	496	2.34	213,330	1027	4.81	211,074	274	1.30
TR_L	7705	120	15.57	7693	40	5.20	7682	81	10.54	7646	12	1.57
U_L	170,656	1042	6.11	170,540	354	2.08	170,607	691	4.05	170,436	154	0.90
IR_L	8613	83	9.64	7689	40	5.20	8572	37	4.32	7646	12	1.57
X	1511	10	6.62	1510	2	1.32	1510	3	1.99	1510	3	1.99
IR_S	10,963	160	14.59	9785	27	2.76	10,402	106	10.19	10,231	48	4.69
U_S	3372	41	12.16	3370	5	1.48	3372	34	10.08	3372	7	2.08
TR_S	13,259	127	9.58	11,295	28	2.48	11,185	75	6.71	10,233	38	3.71

Table 6
Large insertions and deletions (> 10 bp) identified in ZK0118 and the other five published OsHV-1 variants.

Positions Index*	Size (bp)	Variations#	Genomic regions*	Affected ORFs*	Annotation of the involved ORF(s)†
1654–1655	86	Insertion in AY	TR_L	None	
2761–2776	4–16	Insertion in AY	TR_L	None	
4446–4460	3–15	Insertion in AY	TR_L	None	
17707–19092	1386	Deletion in KY	U_L	ORF 11	
49823–49855	63	Deletion in GQ and MF	U_L	ORF 32	Transmembrane glycoprotein
52251–52856	606	Deletion in KY	U_L	ORF 36–38	Membrane protein (ORF36), Zinc-finger, Ring type (ORF38)
60741–60742	2659	Deletion in AY	U_L	***ORF IN.1-IN.4***	Secreted (ORF IN.1) Transmembrane glycoprotein (ORF IN.4)
67973–68572	600	Deletion in KP and KY	U_L	ORF 48	
68608–68609	11	Insertion in GQ	U_L	None	
73395–75243	1849	Deletion in KP	U_L	ORF 50	Secreted protein
93120–96669	3550	Deletion in KY	U_L	ORF 62 and 63	
95979–96000	22	Deletion in GQ and MF	U_L	ORF 63	
116089–116447	359	Deletion in KP	U_L	None	
163514–163516	9/12	Insert 9 bp in KP Delete 3 bp in GQ and MF	U_L	ORF106	Zinc-finger, Ring type, BIR domain
175019–175743	725	Deletion in KY	U_L	ORF114	
175428–179554	4127	Deletion in KP	IR_L	ORF114 , ORF4	
178547–178561	12/15	Insertion in AY	IR_L	None	
179554–179555	906	Insertion in KP	***U_L***	ORF114	
180241–180251	3/11	Insertion in AY	IR_L	None	
181356–181357	86	Deletion in AY	IR_L	None	
183013–184520	1510	Deletion in KP	X	ORF115	Replication origin-binding protein
184880–184942	63	Deletion in GQ and MF	IR_S	None	
187220–188407	1188	Deletion in KP	IR_S	ORF117	Zinc-finger, Ring type
187713–188152	440	Deletion in GQ and MF	IR_S	ORF117	Zinc-finger, Ring type
190459–195843	5385	Deletion in MF	IR_S/U_S	ORF120–123	Secreted (ORF120), Zinc-finger, Ring type (ORF123)
190823–190824	1045	Insertion in GQ and KP	IR_S	None	
191784–191898	115	Deletion in GQ and KP	IR_S	None	
192060–192294	235	Deletion in KP	IR_S	None	
192472–192473	99	Insertion in KP	IR_S	None	
192473–195843	3371	Deletion in KP	IR_S/U_S	122–123	Zinc-finger, Ring type (ORF123)
197902–197903	884	Insertion in KP	IR_S	None	
199645–199879	235	Deletion in KP	TR_S	None	
200066–200810	115	Deletion in GQ, KP and MF	TR_S	None	
201134–201135	1045	Insertion in GQ, KP and MF	TR_S	None	
203554–204741	1188	Deletion in KP	TR_S	ORF117	Zinc-finger, Ring type
203807–204246	440	Deletion in GQ and MF	TR_S	ORF117	Zinc-finger, Ring type
205194–205195	15	Insertion in MF	TR_S	None	
205608–205626	19	Deletion in MF	TR_S	ORF116	
207014–207076	63	Deletion in GQ and MF	TR_S	None	
207439-	1510	Insertion in KY	TR_S	ORF115	Replication origin-binding protein

* Referenced to OsHV-1 reference type, except two items printed in bold and italic.
\# AY represents OsHV-1 reference type, GQ represents AVNV, KP represents OsHV-1-SB, KY represents OsHV-1 μVar variant A and B, MF represents ZK0118.
† Putative ORF annotations were retrieved from Burioli et al. (2017).

samples from a collection of tissues collected during mass mortalities. The protocol was applied successfully on freshly collected and long term frozen tissue samples. Initially, DNA extracted directly from tissue samples could not be stably amplified by most of the 23 primer pairs even after extensive optimization of PCR conditions (data not shown). We inferred the amplification failure was due to large amount of host DNA. To minimize the contamination of host DNA, tissue samples were treated with a simple homogenate preparation protocol as described by Schikorski et al. (2011). Then OsHV-1 DNA extracted from supernatants was amplified successfully for all the 23 primer pairs. Further results indicated that the estimated viral DNA loads influenced the amplification success rates. The sensitivity of LR-PCR was relatively low compared to common PCR, supernatants with viral DNA load $\leq 4.8 \times 10^3$ per μL will not be completely amplified by all 23 primer pairs.

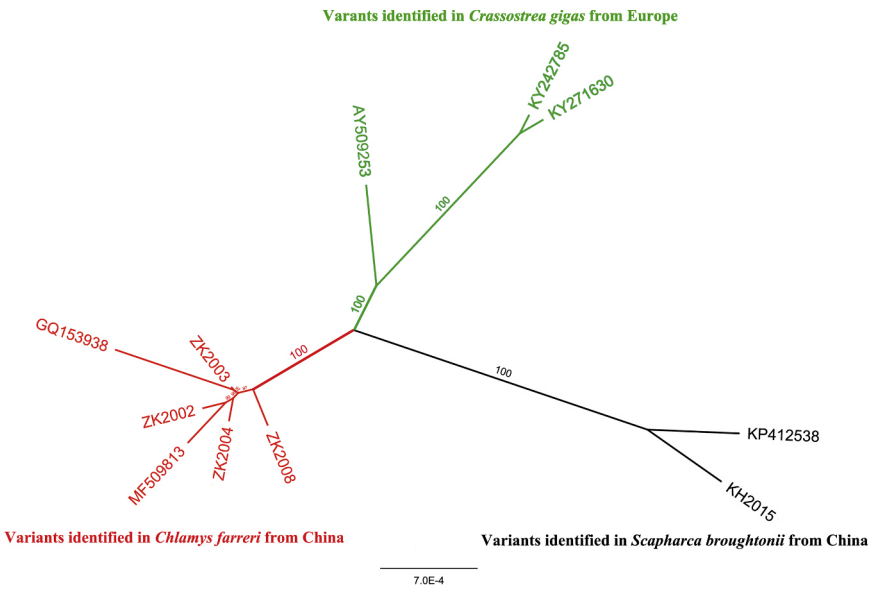

Fig. 3. Phylogenetic tree of 11 OsHV-1 variants inferred using the Neighbor-Joining method. Bootstrap values were obtained from 100 resampled data sets. Numbers at the branches indicate bootstrap support value > 50%. The clade colored in red indicates variants identified in *Chlamys farreri* from China, the clade colored in green indicates variants identified in *Crassostrea gigas* from Europe, the clade colored in black indicates variants identified in *Scapharca broughtonii* from China.

In addition, 7 primer pairs failed to amplify PCR products of some specimens. Although perfectly matching the ZK0118 genome, two primer pairs (GF 17 and GF 18) failed to amplify. These results indicated that the amplification failures might not be owing to primer mismatches but might result from the intricate primer binding energies influenced by template concentrations and viral-to-host genomic DNA ratios (Sipos et al., 2007). More intricately, although GF16 amplicons were amplified successfully from 4 specimens (ZK0118, ZK2002, ZK2008 and KH2015), no sequencing read was detected. Because of the sequencing failure, the assembled genome from PacBio data was broken into two scaffolds. However, the normal amounts of sequence reads were obtained when PCR products of GF16-1 (ZK2002 and ZK2003) were used for amplification and sequencing. To fill the gap between the two scaffolds, we amplified ZK0118 genome with GF16-1 primers, and the PCR products were sequenced successfully with Illumina platform. Similar sequencing failure has not been reported by previous studies associated with LR-PCR (Hagberg et al., 2016; Hernan et al., 2012; Kvisgaard et al., 2013). Since we have successfully obtained the GF16 amplicons, there should be no problem in DNA quality and PCR process. We inferred that GF16 amplicons might form some kind of complex construction, which influenced the downstream HTS process.

The number of sequenced herpesvirus species and variants increased rapidly as the cost of SGS fell gradually. The genome structures of herpesvirus are particularly complex, which make it difficult to determine using short reads (Newman et al., 2015). To overcame the limitations of SGS, TGS platform which is capable of generating longer reads have been employed to sequence Pseudorabies virus (an *Alphaherpesvirus*) and Human herpesvirus type 1 (HHV-1) (Karamitros et al., 2016; Mathijs et al., 2016; Tombacz et al., 2014). Karamitros et al. (2016) reported that Oxford Nanopore MinION sequencer, a sequencing platform producing long reads, could improve the *de novo* assembly of complex genomic regions of HHV-1. In the present study, all PacBio reads were assembled into two scaffolds compared to 9–68 scaffolds assembled using Illumina reads. The gap between the two PacBio scaffolds was because of the sequencing failure of the GF16 amplicon. We could expect that a single contig covering the whole OsHV-1 genome would be obtained from PacBio reads if GF16-1 were used in the LR-PCR amplification. These results indicated that PacBio sequencing can pass through repetitive genome regions and achieve complete *de novo* assembly of OsHV-1 genome. However, more variable coverages were obtained among PacBio amplicons compared to that of Illumina. Thus, both two platforms have their own merits and drawbacks in sequencing OsHV-1 DNA enriched with LR-PCR. Considering the lower expenditure of Illumina sequencing, it is still a prior choice for detecting OsHV-1 genomic variations by reference-dependent assembly.

Pairwise genomic alignment of ZK0118 with other five published OsHV-1 variants revealed that ZK0118 and AVNV displayed the highest similarity (96.8%). Accordingly, only 3 large indels were found between their genome sequences, while more than 10 large indels were found between ZK0118 and the other four OsHV-1 genomes. Moreover, OsHV-1-SB showed a higher similarity with ZK0118 and AVNV (about 92%) compared to variants collected from Europe (about 89%). OsHV-1-SB was collected from *S. broughtonii*, ZK0118 and AVNV were collected from *C. farreri* from the same areas (Qingdao, China). These results indicated that both geographic origin (Qingdao, China) and host species (Zhikong scallops) played a role in the genetic differentiation process of OsHV-1 variants. However, ZK0118 and AVNV showed higher genetic similarity with variants collected from Europe than with OsHV-1-SB. Although both OsHV-1 reference variant and OsHV-1 μVars were collected from Europe, OsHV-1 reference variant showed higher genetic similarity with ZK0118 and AVNV than with OsHV-1 μVars. These results indicated that the evolutionary process of OsHV-1 is complex, which involved many factors such as host species, temporal and geographic origins.

The mean pairwise identity among the considered OsHV-1 variants was 93.5%, which is lower than CyHV-3 (99.5%) (Hammoumi et al., 2016), Herpes simplex virus 2 (99.6%) and Herpes simplex virus 1 (96.8%) (Norberg et al., 2007; Szpara et al., 2014). Correspondingly, a lower amino acid identity (about 93% on average) between the reference and μVar variants compared to that of HHV-6A and HHV-6B (about 94% on average) has also been reported (Burioli et al., 2017). Pairwise genetic identity between two known variants of *Haliotid herpesvirus* 1 (HaHV-1), the other member in the family Malacoherpesviridae, order Herpesvirales, was also very low (90.1%) (Savin et al., 2010). The large number of indel and SNV polymorphisms found in OsHV-1 genomes should be responsible for their low genetic identity. Our results also showed that some indels in the coding regions were shared by variants identified from the same geographic origin and/or host species (Table 6). We inferred that these indels should be partially responsible for the differences related to host tropism and virulence of these OsHV-1 variants. However, due to the high divergence of OsHV-1 and the well-studied vertebrate groups of herpesvirus, it is impossible to infer the specific functions of most ORFs based on bioinformatics alone. We proposed that correlation analysis based on more genomic sequence and associated epidemiology data will be helpful to our better understanding of the problem mentioned above.

Higher SNV frequency was detected among three variants identified from two different host species in northern China compared to that from the same species in Europe. While the SNV frequency detected from two variants identified from the same host species in China was compatible to that from the same species in Europe. These results indicated that host species play an import role in the genetic divergence of the OsHV-1 variants. This inference was further supported by phylogenetic analysis, in which OsHV-1 variants clustered together according to their host species.

Different to the narrow host range and specificity of vertebrate herpesviruses (Tischer and Osterrieder, 2010), OsHV-1 is capable of infecting multiple bivalve species across a wide geographical range (Arzul et al., 2001; Bai et al., 2016). Phylogenetic analysis based on nucleotide sequences of 32 ORFs indicated conflicting results about the relationship of geographically separated variants (closer relationship between OsHV-1-SB and AVNV and distant relationship between OsHV-1 reference type and μVars) (Xia et al., 2015). The present results based on a higher number of longer sequences revealed a robust correlation between OsHV-1 variants and the host. These results were readily compatible with the fact that herpesvirus generally evolved synchronously with the host (Tischer and Osterrieder, 2010). However, the three clade pattern denied the spatial segregation effects on OsHV-1 evolution, because the variants identified in *C. farreri* and *S. broughtonii* from a local region in China allocated to different clades. Although OsHV-1 variants identified from Europe were clustered together, the spatial segregation effects were still uncertain because they were from the same host species (*C. gigas*) and different time. Phylogenetic analysis of Herpes simplex virus type 1 (HSV-1) led to geographically separated clades (Kolb et al., 2013; Szpara et al., 2014). HSV-1 infects only one host, human, while OsHV-1 has multiple host species. For OsHV-1 clade with the same *C. farreri* host species, our results showed a few impacts of the temporal appearance of these variants on the subclade pattern. Temporal separation of OsHV-1 variants has also been reported in previous phylogenetic studies based on hypervariable markers (Bai et al., 2015).

5. Conclusions

The present study provided an effective method for rapid enrichment of DNA of multiple OsHV-1 variants, even from long-term frozen samples. Genome sequences obtained from the enriched DNA with HTS confirmed the high genetic diversity and complex evolution process of OsHV-1. Subsequent phylogenetic analysis revealed that the evolution of OsHV-1 was impacted by host species rather than spatial segregation. Genome sequences of OsHV-1 collected from a wide geographic range of the same host species will be helpful to our better understanding of the spatial segregation effects on OsHV-1 evolution.

Acknowledgements

We would like to thank Li Li from IOCAS for providing the Pacific oyster samples. Qing-Chen Wang for his valuable advices during the analysis of SGS and TGS data. We are also grateful to Ya-Nan Li for her assistance in genome annotation and submission to Genbank.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 31502208, U1706204 and 31302233]; the Central Public-interest Scientific Institution Basal Research Fund, CAFS [grant number 2017HY-ZD1002]; and the China Agriculture Research System [grant number CARS-49].

Database linking

The genome sequences of ZK0118 can be found in GenBank Nucleotide Sequence Database with Accession No. MF509813 (https://www.ncbi.nlm.nih.gov/nuccore/MF509813). The clean Illumina reads (including ZK0118.2, ZK2002, ZK2003, ZK2004, ZK2008, KH2015 and G16 amplicon of ZK0118) and raw PacBio reads (ZK0118.3) of the whole project can be found in the Sequence Read Archive (SRA) database with BioProject ID: PRJNA448032 (http://www.ncbi.nlm.nih.gov/bioproject/448032).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.virol.2018.09.026.

References

Arzul, I., Corbeil, S., Morga, B., Renault, T., 2017. Viruses infecting marine molluscs. J. Invertebr. Pathol. 147, 118–135.
Arzul, I., Nicolas, J.L., Davison, A.J., Renault, T., 2001. French scallops: a new host for Ostreid herpesvirus-1. Virology 290, 342–349.
Bai, C., Gao, W., Wang, C., Yu, T., Zhang, T., Qiu, Z., Wang, Q., Huang, J., 2016. Identification and characterization of ostreid herpesvirus 1 associated with massive mortalities of *Scapharca broughtonii* broodstocks in China. Dis. Aquat. Organ. 118, 65–75.
Bai, C., Wang, C., Xia, J., Sun, H., Zhang, S., Huang, J., 2015. Emerging and endemic types of *Ostreid herpesvirus 1* were detected in bivalves in China. J. Invertebr. Pathol. 124, 98–106.
Barbosa Solomieu, V., Renault, T., Travers, M.A., 2015. Mass mortality in bivalves and the intricate case of the Pacific oyster, *Crassostrea gigas*. J. Invertebr. Pathol. 131, 2–10.
Ben Zakour, N., Gautier, M., Andonov, R., Lavenier, D., Cochet, M.F., Veber, P., Sorokin, A., Le Loir, Y., 2004. GenoFrag: software to design primers optimized for whole genome scanning by long-range PCR amplification. Nucleic Acids Res. 32, 17–24.
Besemer, J., Lomsadze, A., Borodovsky, M., 2001. 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, 2607–2618.
Burioli, E.A.V., Prearo, M., Houssin, M., 2017. Complete genome sequence of Ostreid herpesvirus type 1 microVar isolated during mortality events in the Pacific oyster *Crassostrea gigas* in France and Ireland. Virology 509, 239–251.
Chan, M., Ji, S.M., Yeo, Z.X., Gan, L.D., Yap, E., Yap, Y.S., Ng, R., Tan, P.H., Ho, G.H., Ang, P., Lee, A.S.G., 2012. Development of a next-generation sequencing method for BRCA mutation screening a comparison between a high-throughput and a Benchtop platform. J. Mol. Diagn. 14, 602–612.
Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10, 563–569.
Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403.
Davison, A.J., Dolan, A., Akter, P., Addison, C., Dargan, D.J., Alcendor, D.J., McGeoch, D.J., Hayward, G.S., 2003. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. J. Gen. Virol. 84, 17–28.
Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C., Le Deuff, R.M., Renault, T., 2005. A novel class of herpesvirus with bivalve hosts. J. Gen. Virol. 86, 41–53.
Depledge, D.P., Palser, A.L., Watson, S.J., Lai, I.Y.C., Gray, E.R., Grant, P., Kanda, R.K., Leproust, E., Kellam, P., Breuer, J., 2011. Specific capture and whole-genome sequencing of viruses from clinical samples. Plos One 6, e27805.
Donaldson, C.D., Clark, D.A., Kidd, I.M., Breuer, J., Depledge, D.D., 2013. Genome sequence of human herpesvirus 7 strain UCL-1. Genome Announc. 1, e00830–00813.
Hagberg, E.E., Krarup, A., Fahnoe, U., Larsen, L.E., Dam-Tuxen, R., Pedersen, A.G., 2016. A fast and robust method for whole genome sequencing of the Aleutian Mink Disease Virus (AMDV) genome. J. Virol. Methods 234, 43–51.
Hammoumi, S., Vallaeys, T., Santika, A., Leleux, P., Borzym, E., Klopp, C., Avarre, J.C., 2016. Targeted genomic enrichment and sequencing of CyHV-3 from carp tissues confirms low nucleotide diversity and mixed genotype infections. PeerJ 4, e2516.
Hernan, I., Borras, E., de Sousa Dias, M., Gamundi, M.J., Mane, B., Llort, G., Agundez, J.A., Blanca, M., Carballo, M., 2012. Detection of genomic variations in BRCA1 and BRCA2 genes by long-range PCR and next-generation sequencing. J. Mol. Diagn. 14, 286–293.
Hine, P., Wesney, B., Hay, B., 1992. Herpesviruses associated with mortalities among hatchery-reared larval Pacific oysters, *Crassostrea-gigas*. Dis. Aquat. Organ 12, 135–142.
Jenkins, C., Hick, P., Gabor, M., Spiers, Z., Fell, S.A., Gu, X., Read, A., Go, J., Dove, M., O'Connor, W., Kirkland, P.D., Frances, J., 2013. Identification and characterisation of an ostreid herpesvirus-1 microvariant (OsHV-1 μvar) in *Crassostrea gigas* (Pacific oysters) in Australia. Dis. Aquat. Organ 105, 109–126.
Jia, H.Y., Guo, Y.F., Zhao, W.W., Wang, K., 2014. Long-range PCR in next-generation sequencing: comparison of six enzymes and evaluation on the MiSeq sequencer. Sci. Rep. 4, 5737.
Karamitros, T., Harrison, I., Piorkowska, R., Katzourakis, A., Magiorkinis, G., Mbisa, J.L., 2016. *De Novo* assembly of Human Herpes Virus Type 1 (HHV-1) genome, mining of non-canonical structures and detection of novel drug-resistance mutations using short- and long-read next generation sequencing technologies. Plos One 11,

e0157600.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
Kolb, A.W., Adams, M., Cabot, E.L., Craven, M., Brandt, C.R., 2011. Multiplex sequencing of seven ocular herpes simplex virus type-1 genomes: phylogeny, sequence variability, and SNP distribution. Investig. Ophthalmol. Vis. Sci. 52, 9061–9073.
Kolb, A.W., Ane, C., Brandt, C.R., 2013. Using HSV-1 genome Phylogenetics to track past human migrations. Plos One 8, e76267.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Kvisgaard, L.K., Hjulsager, C.K., Fahnoe, U., Breum, S.O., Ait-Ali, T., Larsen, L.E., 2013. A fast and robust method for full genome sequencing of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Type 1 and Type 2. J. Virol. Methods 193, 697–705.
Kwok, H., Wu, C.W., Palser, A.L., Kellam, P., Sham, P.C., Kwong, D.L.W., Chiang, A.K.S., 2014. Genomic diversity of epstein-barr virus genomes isolated from primary nasopharyngeal carcinoma biopsy samples. J. Virol. 88, 10662–10672.
Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1, 18.
Martenot, C., Oden, E., Travaille, E., Malas, J.P., Houssin, M., 2010. Comparison of two real-time PCR methods for detection of ostreid herpesvirus 1 in the Pacific oyster *Crassostrea gigas*. J. Virol. Methods 170, 86–89.
Mathijs, E., Vandenbussche, F., Verpoest, S., De Regge, N., Van Borm, S., 2016. Complete genome sequence of pseudorabies virus reference strain NIA3 using single-molecule real-time sequencing. Genome Announc. 4, e00440–00416.
Morrison, C.L., Iwanowicz, L., Work, T.M., Fahsbender, E., Breitbart, M., Adams, C., Iwanowicz, D., Sanders, L., Ackermann, M., Cornman, R.S., 2018. Genomic evolution, recombination, and inter-strain diversity of chelonid alphaherpesvirus 5 from Florida and Hawaii green sea turtles with fibropapillomatosis. PeerJ 6, e4386.
Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A., Mobarry, C.M., Reinert, K.H., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou, H.H., Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J., Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams, M.D., Venter, J.C., 2000. A whole-genome assembly of Drosophila. Science 287, 2196–2204.
Newman, R.M., Lamers, S.L., Weiner, B., Ray, S.C., Colgrove, R.C., Diaz, F., Jing, L., Wang, K., Saif, S., Young, S., Henn, M., Laeyendecker, O., Tobian, A.A., Cohen, J.I., Koelle, D.M., Quinn, T.C., Knipe, D.M., 2015. Genome sequencing and analysis of geographically diverse clinical isolates of Herpes Simplex Virus 2. J. Virol. 89, 8219–8232.
Nicolas, J., Comps, M., Cochennec, N., 1992. Herpes-like virus infecting Pacific-oyster larvae, *Crassostrea gigas B*. Eur. Assoc. Fish. Pat. 12, 11–13.
Norberg, P., Kasubi, M.J., Haarr, L., Bergstrom, T., Liljeqvist, J.A., 2007. Divergence and recombination of clinical herpes simplex virus type 2 isolates. J. Virol. 81, 13158–13167.
Olp, L.N., Jeanniard, A., Marimo, C., West, J.T., Wood, C., 2015. Whole-genome sequencing of Kaposi's Sarcoma-associated Herpesvirus from Zambian Kaposi's Sarcoma biopsy specimens reveals unique viral diversity. J. Virol. 89, 12299–12308.
Ozcelik, H., Shi, X.J., Chang, M.C., Tram, E., Vlasschaert, M., Di Nicola, N., Kiselova, A., Yee, D., Goldman, A., Dowar, M., Sukhu, B., Kandel, R., Siminovitch, K., 2012. Long-range PCR and next-generation sequencing of BRCA1 and BRCA2 in breast cancer. J. Mol. Diagn. 14, 467–475.
Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genom. 13, 341.
Ren, W.C., Chen, H.X., Renault, T., Cai, Y.Y., Bai, C.M., Wang, C.M., Huang, J., 2013. Complete genome sequence of acute viral necrosis virus associated with massive mortality outbreaks in the Chinese scallop, *Chlamys farreri*. Virol. J. 10, 110.
Renault, T., Moreau, P., Faury, N., Pepin, J.F., Segarra, A., Webb, S., 2012. Analysis of clinical *ostreid herpesvirus 1* (*Malacoherpesviridae*) specimens by sequencing amplified fragments from three virus genome areas. J. Virol. 86, 5942–5947.
Savin, K.W., Cocks, B.G., Wong, F., Sawbridge, T., Cogan, N., Savage, D., Warner, S., 2010. A neurotropic herpesvirus infecting the gastropod, abalone, shares ancestry with oyster herpesvirus and a herpesvirus associated with the amphioxus genome. Virol. J. 7, 308.
Schikorski, D., Renault, T., Saulnier, D., Faury, N., Moreau, P., Pepin, J.F., 2011. Experimental infection of Pacific oyster *Crassostrea gigas* spat by ostreid herpesvirus 1: demonstration of oyster spat susceptibility. Vet. Res. 42, 27.
Segarra, A., Pepin, J.F., Arzul, I., Morga, B., Faury, N., Renault, T., 2010. Detection and description of a particular *Ostreid herpesvirus 1* genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. Virus Res. 153, 92–99.
Sipos, R., Szekely, A.J., Palatinszky, M., Revesz, S., Marialigeti, K., Nikolausz, M., 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. FEMS Microbiol. Ecol. 60, 341–350.
Spatz, S.J., Rue, C.A., 2008. Sequence determination of a mildly virulent strain (CU-2) of Gallid herpesvirus type 2 using 454 pyrosequencing. Virus Genes 36, 479–489.
Szpara, M.L., Gatherer, D., Ochoa, A., Greenbaum, B., Dolan, A., Bowden, R.J., Enquist, L.W., Legendre, M., Davison, A.J., 2014. Evolution and diversity in human herpes simplex virus genomes. J. Virol. 88, 1209–1227.
Tcherepanov, V., Ehlers, A., Upton, C., 2006. Genome Annotation Transfer Utility (GATU): rapid annotation of viral genomes using a closely related reference genome. BMC Genom. 7, 150.
Tischer, B.K., Osterrieder, N., 2010. Herpesviruses–a zoonotic threat? Vet. Microbiol. 140, 266–270.
Tombacz, D., Sharon, D., Olah, P., Csabai, Z., Snyder, M., Boldogkoi, Z., 2014. Strain Kaplan of Pseudorabies Virus Genome Sequenced by PacBio single-molecule real-time sequencing technology. Genome Announc. 2, e00628–00614.
Tweedy, J., Spyrou, M.A., Donaldson, C.D., Depledge, D., Breuer, J., Gompels, U.A., 2015. Complete genome sequence of the human herpesvirus 6A strain AJ from Africa resembles strain GS from North America. Genome Announc. 3, e01498–01414.
Uribe-Convers, S., Duke, J.R., Moore, M.J., Tank, D.C., 2014. A long Pcr based approach for DNA enrichment prior to next-generation sequencing for systematic studies. Appl. Plant Sci. 2, 1300063.
Xia, J., Bai, C., Wang, C., Song, X., Huang, J., 2015. Complete genome sequence of Ostreid herpesvirus-1 associated with mortalities of *Scapharca broughtonii* broodstocks. Virol. J. 12, 110.