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

Bai Chang-Ming [1, 2, 3, 4], Morga Benjamin [5], Rosani Umberto [6], Shi Jie [1, 2, 3, 4], Li Chen [1, 2, 3, 4], Xin Lu-Sheng [1, 2, 3, 4], Wang Chong-Ming [1, 2, 3, 4, *]

[1] Chinese Acad Fishery Sci, Key Lab Maricultural Organism Dis Control, Minist Agr, Qingdao, Shandong, Peoples R China.
[2] Chinese Acad Fishery Sci, Lab Marine Fisheries Sci & Food Prod Proc, Qingdao Natl Lab Marine Sci & Technol, Qingdao, Shandong, Peoples R China.
[3] Chinese Acad Fishery Sci, Qingdao Key Lab Mariculture Epidemiol & Biosecur, Qingdao, Shandong, Peoples R China.
[4] Chinese Acad Fishery Sci, Yellow Sea Fisheries Res Inst, Qingdao, Shandong, Peoples R China.
[5] IFREMER, LGP, Ave Mus Loup, F-17390 La Tremblade, France.
[6] Univ Padua, Dept Biol, Padua, Italy.

* Corresponding author : Chong-Ming Wang, email address : wangcm@ysfri.ac.cn

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.

Highlights

► We aimed to understand genomic diversity and evolution of different *Ostreid herpesvirus* 1 (OsHV-1) variants. ► A long-range PCR based method was developed for the enrichment of OsHV-1 DNA used for high-throughput sequencing. ► The developed method was successfully used for genome sequencing of six OsHV-1 variants. ► Genomic comparison of the six OsHV-1 variants and those from public database confirmed high genetic diversity and revealed complex evolution process of OsHV-1.

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

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; Bai et al., 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).

58 PCR and Sanger sequencing of several molecular markers of OsHV-1 variants also

59 indicated low genetic similarity (Renault et al., 2012). However, these studies always

60 focused on limited and arbitrarily selected genomic regions, which were not sufficient

61 for a complete understanding of genome-wide nucleotide variations. In order to

62 resolve the problem, we need to characterize variations at the genome level, which

63 requires many genome sequences of different OsHV-1 variants. These demands make

64 Sanger sequencing extremely challenging and technically impractical. SGS and

65 Third-Generation Sequencing (TGS) have many advantages compared to Sanger

66 sequencing, which have been widely used for genome sequencing of vertebrate

67 herpesviruses (Depledge et al., 2011; Kolb et al., 2011; Spatz and Rue, 2008).

68 However, SGS and TGS require the preparation of considerable amounts of viral

69 DNA, which is difficult for OsHV-1 due to the lack of permissive cell line for virus

70 purification. Moreover, the short read lengths generated by SGS make them poorly

71 suited for *de novo* assembly of complex genomic regions typical of herpesviruses and

72 to study the diversity of the virus (Kolb et al., 2011).

73 Long-range PCR (LR-PCR) can amplify candidate genomic regions with high

74 sensitivity and specificity (Jia et al., 2014). Combined with SGS, LR-PCR has been

75 used for viral genome sequencing, genetic variation detecting and systematic studies

76 (Chan et al., 2012; Kvisgaard et al., 2013; Morrison et al., 2018; Ozcelik et al., 2012;

77 Uribe-Convers et al., 2014). The genome of OsHV-1 contains two unique regions each

78 flanked by inverted repeats longer than 7500 bp (Davison et al., 2005), which may

79 lead to fragmented assemblies due to the short read lengths generated by SGS

80 sequencing (Quail et al., 2012). The PacBio platform, by virtue of its long read

81 lengths, was found useful for *de novo* assemblies of complex repeat regions (Quail et

82 al., 2012). We describe here the use of 23 primer pairs for LR-PCR to amplify

83 OsHV-1 genomic DNA. Then, we quantified and successfully pooled the PCR

84 products of each sample in equimolecular proportions. The amplicon mixtures were

85 used for genome sequencing with PacBio RS II and/or Illumina HiSeq 4000 platforms.

86 The resolved genome sequences (one complete and five partial genomes) were

87 applied for further genomic comparison and phylogenetic analysis with the other five

88 published OsHV-1 variants.

89

2. Materials and methods

91 *2.1. Sample Preparation*

92 Ten OsHV-1 positive bivalve samples determined by qPCR (Martenot et al., 2010)

93 and collected during mortality events in Shandong Province (China) were employed

94 in the present study (Table 1, Supplementary figure 1). The samples were stored at

95 -40 °C until use. For adult tissues, a homogenate preparation protocol as described by

96 (Schikorski et al., 2011) was used before DNA extraction. Briefly, about 1g of mantles

97 and gills were dissected from a specimen and crushed with tissue homogenizer in 9

98 mL of 0.22 μm filtered seawater. Then the tissue homogenate was centrifuged at 1000

99 *g* for 5 min at 4 °C. Finally, the supernatant was collected and used for DNA

100 extraction. For Pacific oyster larvae, specimens were washed in double distilled water

101 and grinded with liquid nitrogen. Corresponding healthy and OsHV-1 negative sample

102 was determined by qPCR.

103 *2.2. DNA Extraction and OsHV-1 DNA Quantification*

104 Total DNA extraction was performed from 180 µL tissue supernatant or 50 mg

105 grinded larvae using the Qiagen DNeasy Blood and Tissue kit according to the

106 manufacturer's protocol. Elution was performed in 60 µL of AE buffer provided in the

107 kit.

108 The detection and quantification of OsHV-1 DNA were carried out by qPCR adapted

109 from previously published protocol (Martenot et al., 2010) and fully described in Bai

110 et al. (2016). Each sample was tested in duplicate. The quantification of OsHV-1 DNA

111 was estimated as the mean OsHV-1 DNA copies per µL of tissue supernatant or per

112 mg of grinded larvae for the two replicates.

113 *2.3. Primer Design and Long-range PCR*

114 Primer pairs were initially designed with the online version of GenoFrag

115 (http://genoweb1.irisa.fr/Serveur-GPO/outils/generationAmorces/GENOFRAG/index

116 en.php) (Ben Zakour et al., 2004) using the genome sequence of OsHV-1 reference

117 variant (GenBank accession no. AY509253). Potential primers were selected based on

118 default values of a series of filters defining the primer length (25 bp), GC content

119 (48%), hairpin, stability, auto-complementarity and so on. Then the software picked

120 out a list of primer pairs covering the viral genome from the potential primers

121 previously generated. The amplicon length was set in a range of 9-11 kb, and

122 overlapping the adjacent amplicons by 500-1500 bp (Table 2). When the

123 automatically selected primers failed to amplify, new primer pairs were manually

124 selected for testing from the potential primers around its original binding site. Primers

125 of three amplicon sites at the termini of OsHV-1 genome (Table 2) were designed with

126 Primer Premier 5 (Premier Biosoft International, Paolo Alto, CA, USA) according to

127 general primer selection criteria.

128 All LR-PCR reactions were performed using the TaKaRa PrimeSTAR GXL DNA

129 polymerase, which showed high fidelity and stability in amplifying long PCR

130 products as reported previously (Jia et al., 2014). The PCR reaction was performed in

131 a 50 μl reaction volume that included 10 μL 5×PrimeSTAR GXL Buffer, 4 μL dNTP

132 Mixture (2.5 mM each), 1 μL TaKaRa PrimeSTAR GXL DNA Polymerase, 1 μL

133 each of forward and reverse primers (10 μM), 2 μL DNA template and 31 μL of

134 PCR-grade H_2O. LR-PCR were performed using Veriti Thermal Cycler (Applied

135 Biosystems) under the following conditions: 94 ℃ for 1 minutes, followed by 35

136 cycles of amplification (denaturation 98 ℃, 10 s; annealing 50 ℃, 15 s, extension

137 68 ℃, 10 minutes) and hold at 4 ℃. PCR product sizes were detected on 0.8% 1×

138 TAE agarose gels stained with GeneFinder™ (Zeesan Biotechnology Inc.).

139 *2.4. Genome Sequencing and Assembly*

140 PCR amplicons of OsHV-1 infected samples (Table 1) were purified with QIAquick

141 PCR Purification Kit (Qiagen) and quantified using Qubit Fluorometer using the

142 dsDNA Assay Kit (Life Technologies). Then the quantified amplicons for each

143 sample were mixed in equimolecular proportions and send for genome sequencing

144 and assembly at Guangzhou Gene Denovo Biotechnology Co., Ltd. ZK0118 was

145 sequenced with both PacBio and Illumina platforms, the other five samples were

146 sequenced with Illumina platform only.

147 For PacBio sequencing, 10 μg of ZK0118 amplicon mixtures were directly

148 end-repaired for library preparation (9~13 kb SMRTBell library) according to the

149 manufacturer's specification (Pacific Biosciences, Menlo Park, CA). Sequencing was

150 performed on the PacBio RS II sequencer according to standard protocols (MagBead

151 Standard Seq v2 loading, 1×180 min movie) using the P4-C2 chemistry. Reads

152 longer than 500 bp and with a quality value over 0.75 were merged together into a

153 single dataset. Next, the hierarchical genome-assembly process (HGAP) pipeline

154 (Chin et al., 2013) was used to correct for random errors in the long seed reads (seed

155 length threshold 6 kb) by aligning the shorter reads against them. The resulting

156 corrected and preassembled reads were used for the final *de novo* assembly using

157 Celera Assembler with an Overlap-Layout-Consensus (OLC) strategy (Myers et al.,

158 2000). Since SMRT sequencing features very little variations of the quality

159 throughout the reads, no quality values were used during the assembly. To validate the

160 quality of the assembly and determine the final genome sequence, the Quiver

161 consensus algorithm was used (Chin et al., 2013).

162 For Illumina sequencing, 2 μg amplicon mixtures of each sample were firstly

163 sonicated randomly, and then end-repaired, A-tailed, and adaptor ligated according to

164 the Illumina TruSeq DNA preparation protocol. Genome sequencing was performed

165 on the Illumina HiSeq 4000 sequencer using the pair-end technology (PE 150). Raw

166 reads were firstly processed using an in-house perl script. In this step, clean reads

167 were obtained by removing adaptor sequences and low-quality reads with >40% bases

168 having a PHRED quality scores of $\leq$ 20, with $\geq$ 10 % unidentified nucleotides (N).

169 The resulting clean reads were *de novo* assembled using the SOAPdenovo (ver. 2.04)

170 (Luo et al., 2012), followed by modification using gapclose (version 1.12). The clean

171 reads of each sample were also used for a reference genome mapping on its'

172 phylogenetically closest reference genome using the CLC mapper tool (CLC Genomic

173 Workbench, Qiagen) applying 0.8 and 0.5 of similarity and length parameters,

174 respectively. The number of covered reference positions (with at least 50x of

175 sequencing depth) were counted to calculate the percentage of the reference genome

176 covered by each read dataset.

177 *2.5. Gap Closure, Verification of Ambiguous Regions of ZK0118*

178 No sequence read of GF16 amplicons generated in ZK0118 and the other three

179 samples (ZK2002, ZK2008 and KH2015) with either PacBio or Illumina, although the

180 amplicons were detected by electrophoresis, while GF16-1 amplicons targeting the

181 same genome area were sequenced successfully in the other two samples (ZK2003

182 and ZK2004). In order to obtain the complete genome of ZK0118, GF16-1 primer pair

183 was employed for amplification of the ZK0118 DNA fragment. The amplified

184 fragments were then sequenced with Illumina, and the generated sequence reads were

185 used for filling the gap. The other two PCR amplicons (GF22 and GF23) with a low

186 coverage of PacBio reads were validated with Illumina sequencing. One region with

187 82 ambiguous bases (nt: 48196– 48227) in the assembly of ZK0118 were determined

188 by Sanger sequencing of PCR product. The genome termini of ZK0118 were

189 identified by a PCR method for amplifying termini using the Marathon cDNA

190   amplification kit (Clontech Laboratories) as previously described (Davison et al.,

191   2003).

2.6. Gene Annotation of ZK0118 and Genome Comparisons

193   The potential open reading frames (ORFs) were predicted using the GeneMarkS

194   (Besemer et al., 2001) and GATU (Tcherepanov et al., 2006) relative to OsHV-1 and

195   adjusted manually.

196   The pairwise and multiple alignments of ZK0118, OsHV-1 reference variant, AVNV

197   (GenBank accession no. GQ153938), OsHV-1-SB (No. KP412538), OsHV-1 µVar

198   variant A (No. KY242785) and variant B (No. KY271630) were performed with the

199   mauveAligner tool (Darling et al., 2004) implanted in Geneious (ver. 4.8.5) and

200   adjusted manually. The indels larger than 10 bases and Single-Nucleotide Variations

201   (SNV) among these genomes were identified with Mega v. 7.0.14 (Kumar et al., 2016)

202   based on the multiple alignment.

2.7. Phylogenetic Analysis

204   Since the complete genome sequences were not resolved for the five samples

205   sequenced with only Illumina platform, the scaffolds of each sample were

206   concatenated and aligned with the complete genome sequence of ZK0118 sequenced

207   in the present study and that of the other 5 available OsHV-1 variants in public

208   database. All positions containing aligned gaps and missing data were eliminated in

209   the following analysis. Genomic sequences of the 11 variants were aligned with

210   MAFFT version 7 with the default settings (Katoh and Standley, 2013). The best-fit

211   nucleotide substitution model was determined using the Akaike Information in MEGA

212 v. 7.0.14 (Kumar et al., 2016). Phylogenetic relationship was inferred by using the

213 Neighbor-Joining method implemented in Geneious (ver. 4.8.5). Branch support was

214 estimated with 100 bootstrap replicates. Since there is no suitable outgroup available,

215 the NJ tree was displayed as unrooted.

216

3. Results

217

218 *3.1. Specificity and Sensitivity of OsHV-1 Primers*

219 A total of 23 primer pairs (labeled as GF1-23F/R respectively) were designed for

220 long-range PCR, which span 99.3% of the OsHV-1 reference genome (Table 2). Due

221 to the amplification failure of 7 primer pairs when applied to some specimens, they

222 were redesigned manually (Table 2, Supplementary figure 2). All 23 amplicons were

223 amplified with long-range PCR using as templates supernatants with viral DNA loads

224 $\geq 8.8 \times 10^3$ copies per μL (Figure 1. A). While for specimens with viral DNA loads $\leq$

225 4.8×10^3 per μL, less than 13 amplicons were amplified, despite some optimizations.

226 No amplicon was obtained for negative controls, which indicated that all primer pairs

227 were specific for OsHV-1 DNA (Figure 1. B).

228 *3.2. Genome Sequencing and Assembly*

229 A total of 150, 292 polymerase reads with an average length of 6, 700 bp and an N50

230 of 20, 079 bp were generated from PacBio RS II platform (Table 3). After read quality

231 trimming, 60, 514 reads with an average length of 14, 099 bp and an N50 of 21, 771

232 bp were obtained. *De novo* assembly results in 2 scaffolds with lengths of 136, 911 bp

233 and 59, 926 bp, respectively. The average sequencing depth was 3, 218x across the

234 whole genome, with a range between 0 (GF16) and 4, 807x (GF3) (Figure 2). Regions

235 characterized by overlapping amplicons showed higher coverage values, while a

236 limited coverage bias was observed among different amplicons. Moreover, we found

237 lower sequencing depths in the 3' - terminal and internal repeat regions (IR_S and TR_S).

238 Illumina sequencing generated a total of 41, 738, 566 reads. After trimming, 35, 515,

239 518 clean reads were obtained. *De novo* genome assembly resulted in 9 to 68 scaffolds

240 for each sample, covering 88.09 - 95.23 % of the reference sequence with average

241 sequencing depths ranging from 4, 139x to 7, 319x. The sequencing depths along the

242 whole genome were unbiased; except for the terminal repeat regions. A summary of

243 sequencing data and assemblies generated from the PacBio RS II in comparison to

244 that generated from the Illumina HiSeq 4000 platform was listed in Table 3. Since the

245 assembly of five samples (ZK2002, ZK2003, ZK2004, ZK2008 and KH2015)

246 sequenced with only Illumina platform generated multiple scaffolds, their complete

247 genome sequences were not determined. Reference mapping of the Illumina clean

248 reads of ZK0118.2, ZK2002, ZK2003, ZK2004 and ZK2008 resulted in 96.0 %,

249 95.6 %, 99.0 %, 98.8 % and 95.6 % of the AVNV genome covered with a minimal

250 sequencing depth of 50x, whereas KH2012 reads covered 96.6 % of the OsHV-1-SB

251 genome. The clean Illumina reads and raw PacBio reads of the whole project has been

252 submitted to the Sequence Read Archive database with BioProject ID: PRJNA448032.

253 Since no read generated from GF16 amplicons of ZK0118, Illumina reads of GF16-1

254 amplicons were used for filling the gap between the 2 Pacbio scaffolds. After

255 verification of ambiguous regions and determination of genomic terminals, the

256   ZK0118 genome was finally resolved to be 204, 652 bp in length. The G/C content of

257   ZK0118 genome is 38.6 %, which is in agree with the previously published OsHV-1

258   variants. The presence of a new large deletion (> than 5, 000 bp) was verified by PCR,

259   cloning and Sanger sequencing (data not shown). The genome structure of ZK0118 is

260   similar to that of OsHV-1 reference, which contains two unique regions each flanked

261   by two inverted repeats (TRL/IRL and TRS/IRS), with a third unique region situated

262   between IRL and IRS. The genome sequences reported in the present study has been

263   submitted to GenBank Nucleotide Sequence Database with Accession No. MF509813.

264   *3.3. Gene Annotation of ZK0118 and Genome Comparisons*

265   Computer-assisted analysis revealed 126 distinct ORFs in ZK0118 genome, ranging

266   from 42 to 1879 amino acids (aa) in length. For the 123 ORFs with a counterpart in

267   the OsHV-1 reference, we assigned the same number as those in OsHV-1 reference

268   (ORFs 1–122 and 124). Three ORFs, absent in OsHV-1 reference, were named as

269   ORFs 125–127, respectively. Due to the large deletion (more than 5000 bp) occurred

270   at the IR_S-U_S region of ZK0118, 4 OsHV-1 ORFs (ORF 120, 121 and 122 in the IR_S

271   region and ORF 123 in the U_S region) were deleted in ZK0118.

272   The pairwise comparison of ZK0118 with OsHV-1 reference, AVNV, OsHV-1-SB,

273   OsHV-1 μVar variant A and variant B showed 94.7%, 96.8%, 92.5%, 92.7% and 92.7%

274   of similarity, respectively (Table 4). Multiple sequence alignments of the six OsHV-1

275   genomes resulted in a 216, 079 bp aligned positions, because of the large indel

276   (Supplementary Data 1). We identified 1,583 SNVs over the whole alignment of 6

277   OsHV-1 variants, which occurred at a rate of 7.33 per kb (Table 5). We identified 496

278 SNVs among the 3 European variants from the same species (*C. gigas*), which were

279 lower than 1, 027 SNVs identified among the 3 Chinese variants from 2 different host

280 species (*C. farreri* and *S. broughtonii*). While the lowest number of SNVs (274) and

281 frequency (1.31 per kb) were identified when only the 2 variants from the same

282 species (*C. farreri*) in China were analyzed. The SNVs distributed unevenly across

283 OsHV-1 genome within all groups. A higher frequency of SNVs was always detected

284 in the inverted repeats than the unique regions.

285 Further indel polymorphism analysis of the multiple alignment results revealed a total

286 of 40 indels (>10 bp) characteristic of a specific variant or shared by a subgroup of

287 variants (Table 6). There were 6 indels occurred only in OsHV-1 variants found in *C.*

288 *farreri* in China, 2 identical indels in OsHV-1 variants found in *C. farreri* and *S.*

289 *broughtonii* in China. Interestingly, there was also 1 deletion of 600 bp occurred only

290 in *S. broughtonii* in China and *C. gigas* in Europe (OsHV-1 μVar). The indels

291 occurred most frequently in IR_S and TR_S regions (19 indels), which involved the

292 alteration of ORFs 115 - 117 and 120 - 123. While only 7 indels were found in the IR_L

293 and TR_L regions, 6 of which were 3 pairs of inverted repeats found in the non-coding

294 regions of OsHV-1 reference variant. The other one was a large deletion occurred in

295 the IR_L regions of OsHV-1 reference variant, which involved the deletion of ORFs

296 114 and 4. Thirteen indels were found in the U_L region, 11 of them were associated

297 with variation of the alteration of 15 ORFs, which include ORFs 11, 32, 36 - 38,

298 IN.1-IN.4, 48, 50, 62, 63, 106, 114. Additionally, a deletion of X region in the genome

299 of OsHV-1-SB led to the complete loss of ORF 115. The putative functions of 11

300 ORFs from the 23 affected ORFs described above were identified. Three ORFs were

301 predicted to encode membrane proteins (ORFs 32, 36 and IN.4), two of which were

302 putative membrane glycoproteins (ORFs 32 and IN.4). Four ORFs included RING

303 finger domain (ORFs 38, 106, 117 and 123), three ORFs were predicted to encode

304 secreted proteins (ORFs IN.1, 50 and 120), and one ORF encode Replication

305 Origin-binding Protein (ORF 115).

306 *3.4. Phylogenetic Analysis*

307 Alignment of 6 OsHV-1 genomes sequenced in the present study with that of 5

308 available variants in public database generated 216,626 aligned positions

309 (Supplementary Data 2). The nucleotide positions containing gaps and missing data

310 were discharged and resulted in 139, 205 informative positions. The evolutionary

311 distances were computed using the Tajima-Nei model. The estimated phylogenetic

312 tree is drawn to scale, with branch lengths measured in the number of base

313 substitutions per site (Figure 3). The tree divided the 11 OsHV-1 variants into three

314 main groups with 100% of bootstrap support, corresponding to the hosting species.

315 The evolution relationship of variants identified from each host were also resolved

316 with high bootstrap support values (>90%). The distribution pattern of the six variants

317 identified from *C. farreri* was slightly correlated to their sampling time except

318 ZK2003.

319

4. Discussion

321 OsHV-1 and variants characterized by broad genetic, host, and virulence diversities

322   have been identified in China and the other countries (Bai et al., 2015; Barbosa

323   Solomieu et al., 2015). However, genome sequence data of these variants was scarce

324   due to technical constrains, mainly linked to the difficulty of obtaining enough

325   purified viral particles for high-throughput sequencing (Burioli et al., 2017). One of

326   the focus of the present study was to develop a fast and robust method for enrichment

327   of viral DNA prior to high-throughput sequencing. To date, two methods have been

328   used for the enrichment of herpesvirus DNA from tissue samples (Depledge et al.,

329   2011; Donaldson et al., 2013; Hammoumi et al., 2016; Kwok et al., 2014; Olp et al.,

330   2015; Tweedy et al., 2015). The first and commonly used method is a

331   hybridization-based approach, which captures viral DNA by hybridization with RNA

332   baits designed across the genome (Depledge et al., 2011). The second one is a

333   LR-PCR based approach, which amplifies viral DNA using primers designed covering

334   the complete genomic sequence (Jia et al., 2014). Application of these two approaches

335   for Human Herpesvirus 6A (HHV-6A) genome sequencing generated identical

336   consensus sequences, with similar variant-calling efficacy (Tweedy et al., 2015).

337   However, there is no commercial RNA baits available for OsHV-1 and custom design

338   and production of RNA baits is expensive. Additionally, hybridization-based

339   approaches typically require more DNA than LR-PCR based approach. In addition,

340   PCR technique requires very low amounts of DNA samples, inexpensive regents and

341   common equipments. And a long-range DNA polymerase with higher fidelity and

342   better performance has also been picked out of 6 commercially available polymerases

343   (Jia et al., 2014).

344   In the present study, a protocol was developed for enrichment of OsHV-1 DNA using

345   LR-PCR, followed by high throughput sequencing in samples from a collection of

346   tissues collected during mass mortalities. The protocol was applied successfully on

347   freshly collected and long term frozen tissue samples. Initially, DNA extracted

348   directly from tissue samples could not be stably amplified by most of the 23 primer

349   pairs even after extensive optimization of PCR conditions (data not shown). We

350   inferred the amplification failure was due to large amount of host DNA. To minimize

351   the contamination of host DNA, tissue samples were treated with a simple

352   homogenate preparation protocol as described by Schikorski et al. (2011). Then

353   OsHV-1 DNA extracted from supernatants was amplified successfully for all the 23

354   primer pairs. Further results indicated that the estimated viral DNA loads influenced

355   the amplification success rates. The sensitivity of LR-PCR was relatively low

356   compared to common PCR, supernatants with viral DNA load $\leq 4.8 \times 10^3$ per μL will

357   not be completely amplified by all 23 primer pairs.

358   In addition, 7 primer pairs failed to amplify PCR products of some specimens.

359   Although perfectly matching the ZK0118 genome, two primer pairs (GF 17 and GF

360   18) failed to amplify. These results indicated that the amplification failures might not

361   be owing to primer mismatches but might result from the intricate primer binding

362   energies influenced by template concentrations and viral-to-host genomic DNA ratios

363   (Sipos et al., 2007). More intricately, although GF16 amplicons were amplified

364   successfully from 4 specimens (ZK0118, ZK2002, ZK2008 and KH2015), no

365   sequencing read was detected. Because of the sequencing failure, the assembled

366   genome from PacBio data was broken into two scaffolds. However, the normal

367   amounts of sequence reads were obtained when PCR products of GF16-1 (ZK2002

368   and ZK2003) were used for amplification and sequencing. To fill the gap between the

369   two scaffolds, we amplified ZK0118 genome with GF16-1 primers, and the PCR

370   products were sequenced successfully with Illumina platform. Similar sequencing

371   failure has not been reported by previous studies associated with LR-PCR (Hagberg et

372   al., 2016; Hernan et al., 2012; Kvisgaard et al., 2013). Since we have successfully

373   obtained the GF16 amplicons, there should be no problem in DNA quality and PCR

374   process. We inferred that GF16 amplicons might form some kind of complex

375   construction, which influenced the downstream HTS process.

376   The number of sequenced herpesvirus species and variants increased rapidly as the

377   cost of SGS fell gradually. The genome structures of herpesvirus are particularly

378   complex, which make it difficult to determine using short reads (Newman et al., 2015).

379   To overcame the limitations of SGS, TGS platform which is capable of generating

380   longer reads have been employed to sequence Pseudorabies virus (an

381   *Alphaherpesvirus*) and Human herpesvirus type 1 (HHV-1) (Karamitros et al., 2016;

382   Mathijs et al., 2016; Tombacz et al., 2014). Karamitros et al. (2016) reported that

383   Oxford Nanopore MinION sequencer, a sequencing platform producing long reads,

384   could improve the *de novo* assembly of complex genomic regions of HHV-1. In the

385   present study, all PacBio reads were assembled into two scaffolds compared to 9 - 68

386   scaffolds assembled using Illumina reads. The gap between the two PacBio scaffolds

387   was because of the sequencing failure of the GF16 amplicon. We could expect that a

388   single contig covering the whole OsHV-1 genome would be obtained from PacBio

389   reads if GF16-1 were used in the LR-PCR amplification. These results indicated that

390   PacBio sequencing can pass through repetitive genome regions and achieve complete

391   *de novo* assembly of OsHV-1 genome. However, more variable coverages were

392   obtained among PacBio amplicons compared to that of Illumina. Thus, both two

393   platforms have their own merits and drawbacks in sequencing OsHV-1 DNA enriched

394   with LR-PCR. Considering the lower expenditure of Illumina sequencing, it is still a

395   prior choice for detecting OsHV-1 genomic variations by reference-dependent

396   assembly.

397   Pairwise genomic alignment of ZK0118 with other five published OsHV-1 variants

398   revealed that ZK0118 and AVNV displayed the highest similarity (96.8%).

399   Accordingly, only 3 large indels were found between their genome sequences, while

400   more than 10 large indels were found between ZK0118 and the other four OsHV-1

401   genomes. Moreover, OsHV-1-SB showed a higher similarity with ZK0118 and AVNV

402   (about 92 %) compared to variants collected from Europe (about 89 %). OsHV-1-SB

403   was collected from *S. broughtonii*, ZK0118 and AVNV were collected from *C. farreri*

404   from the same areas (Qingdao, China). These results indicated that both geographic

405   origin (Qingdao, China) and host species (Zhikong scallops) played a role in the

406   genetic differentiation process of OsHV-1 variants. However, ZK0118 and AVNV

407   showed higher genetic similarity with variants collected from Europe than with

408   OsHV-1-SB. Although both OsHV-1 reference variant and OsHV-1 μVars were

409   collected from Europe, OsHV-1 reference variant showed higher genetic similarity

410 with ZK0118 and AVNV than with OsHV-1 μVars. These results indicated that the

411 evolutionary process of OsHV-1 is complex, which involved many factors such as

412 host species, temporal and geographic origins.

413 The mean pairwise identity among the considered OsHV-1 variants was 93.5%, which

414 is lower than CyHV-3 (99.5%) (Hammoumi et al., 2016), Herpes simplex virus 2

415 (99.6%) and Herpes simplex virus 1 (96.8%) (Norberg et al., 2007; Szpara et al.,

416 2014). Correspondingly, a lower amino acid identity (about 93% on average) between

417 the reference and μVar variants compared to that of HHV-6A and HHV-6B (about 94%

418 on average) has also been reported (Burioli et al., 2017). Pairwise genetic identity

419 between two known variants of *Haliotid herpesvirus* 1 (HaHV-1), the other member

420 in the family Malacoherpesviridae, order Herpesvirales, was also very low (90.1%)

421 (Savin et al., 2010). The large number of indel and SNV polymorphisms found in

422 OsHV-1 genomes should be responsible for their low genetic identity. Our results also

423 showed that some indels in the coding regions were shared by variants identified from

424 the same geographic origin and/or host species (Table 6). We inferred that these indels

425 should be partially responsible for the differences related to host tropism and

426 virulence of these OsHV-1 variants. However, due to the high divergence of OsHV-1

427 and the well-studied vertebrate groups of herpesvirus, it is impossible to infer the

428 specific functions of most ORFs based on bioinformatics alone. We proposed that

429 correlation analysis based on more genomic sequence and associated epidemiology

430 data will be helpful to our better understanding of the problem mentioned above.

431 Higher SNV frequency was detected among three variants identified from two

432 different host species in northern China compared to that from the same species in

433 Europe. While the SNV frequency detected from two variants identified from the

434 same host species in China was compatible to that from the same species in Europe.

435 These results indicated that host species play an import role in the genetic divergence

436 of the OsHV-1 variants. This inference was further supported by phylogenetic analysis,

437 in which OsHV-1 variants clustered together according to their host species.

438 Different to the narrow host range and specificity of vertebrate herpesviruses (Tischer

439 and Osterrieder, 2010), OsHV-1 is capable of infecting multiple bivalve species across

440 a wide geographical range (Arzul et al., 2001; Bai et al., 2016). Phylogenetic analysis

441 based on nucleotide sequences of 32 ORFs indicated conflicting results about the

442 relationship of geographically separated variants (closer relationship between

443 OsHV-1-SB and AVNV and distant relationship between OsHV-1 reference type and

444 µVars) (Xia et al., 2015). The present results based on a higher number of longer

445 sequences revealed a robust correlation between OsHV-1 variants and the host. These

446 results were readily compatible with the fact that herpesvirus generally evolved

447 synchronously with the host (Tischer and Osterrieder, 2010). However, the three clade

448 pattern denied the spatial segregation effects on OsHV-1 evolution, because the

449 variants identified in *C. farreri* and *S. broughtonii* from a local region in China

450 allocated to different clades. Although OsHV-1 variants identified from Europe were

451 clustered together, the spatial segregation effects were still uncertain because they

452 were from the same host species (*C. gigas*) and different time. Phylogenetic analysis

453 of Herpes simplex virus type 1 (HSV-1) led to geographically separated clades (Kolb

454 et al., 2013; Szpara et al., 2014). HSV-1 infects only one host, human, while OsHV-1

455 has multiple host species. For OsHV-1 clade with the same *C. farreri* host species, our

456 results showed a few impacts of the temporal appearance of these variants on the

457 subclade pattern. Temporal separation of OsHV-1 variants has also been reported in

458 previous phylogenetic studies based on hypervariable markers (Bai et al., 2015).

459

5. Conclusions

461 The present study provided an effective method for rapid enrichment of DNA of

462 multiple OsHV-1 variants, even from long-term frozen samples. Genome sequences

463 obtained from the enriched DNA with HTS confirmed the high genetic diversity and

464 complex evolution process of OsHV-1. Subsequent phylogenetic analysis revealed

465 that the evolution of OsHV-1 was impacted by host species rather than spatial

466 segregation. Genome sequences of OsHV-1 collected from a wide geographic range of

467 the same host species will be helpful to our better understanding of the spatial

468 segregation effects on OsHV-1 evolution.

469

475

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).

491 **References**

492 Arzul, I., Corbeil, S., Morga, B., Renault, T., 2017. Viruses infecting marine molluscs. J Invertebr

493 Pathol 147, 118-135.

494 Arzul, I., Nicolas, J.L., Davison, A.J., Renault, T., 2001. French scallops: A new host for Ostreid

495 herpesvirus-1. Virology 290, 342-349.

496 Bai, C., Gao, W., Wang, C., Yu, T., Zhang, T., Qiu, Z., Wang, Q., Huang, J., 2016. Identification

497 and characterization of ostreid herpesvirus 1 associated with massive mortalities of

498 *Scapharca broughtonii* broodstocks in China. Dis Aquat Organ 118, 65-75.

499 Bai, C., Wang, C., Xia, J., Sun, H., Zhang, S., Huang, J., 2015. Emerging and endemic types of

500 *Ostreid herpesvirus 1* were detected in bivalves in China. J Invertebr Pathol 124, 98-106.

501 Barbosa Solomieu, V., Renault, T., Travers, M.A., 2015. Mass mortality in bivalves and the

502 intricate case of the Pacific oyster, *Crassostrea gigas*. J Invertebr Pathol 131, 2-10.

503 Ben Zakour, N., Gautier, M., Andonov, R., Lavenier, D., Cochet, M.F., Veber, P., Sorokin, A., Le

504 Loir, Y., 2004. GenoFrag: software to design primers optimized for whole genome scanning

505 by long-range PCR amplification. Nucleic Acids Res 32, 17-24.

506 Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for

507 prediction of gene starts in microbial genomes. Implications for finding sequence motifs in

508 regulatory regions. Nucleic Acids Res 29, 2607-2618.

509 Burioli, E.A.V., Prearo, M., Houssin, M., 2017. Complete genome sequence of Ostreid herpesvirus

510 type 1 microVar isolated during mortality events in the Pacific oyster *Crassostrea gigas* in

511 France and Ireland. Virology 509, 239-251.

512 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.,

513 Lee, A.S.G., 2012. Development of a Next-Generation Sequencing Method for BRCA

514   Mutation Screening A Comparison between a High-Throughput and a Benchtop Platform. J

515   Mol Diagn 14, 602-612.

516   Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland,

517   A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished

518   microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10,

519   563-569.

520   Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved

521   genomic sequence with rearrangements. Genome Res 14, 1394-1403.

522   Davison, A.J., Dolan, A., Akter, P., Addison, C., Dargan, D.J., Alcendor, D.J., McGeoch, D.J.,

523   Hayward, G.S., 2003. The human cytomegalovirus genome revisited: comparison with the

524   chimpanzee cytomegalovirus genome. J Gen Virol 84, 17-28.

525   Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C., Le Deuff,

526   R.M., Renault, T., 2005. A novel class of herpesvirus with bivalve hosts. J Gen Virol 86,

527   41-53.

528   Depledge, D.P., Palser, A.L., Watson, S.J., Lai, I.Y.C., Gray, E.R., Grant, P., Kanda, R.K., Leproust,

529   E., Kellam, P., Breuer, J., 2011. Specific Capture and Whole-Genome Sequencing of Viruses

530   from Clinical Samples. Plos One 6, e27805.

531   Donaldson, C.D., Clark, D.A., Kidd, I.M., Breuer, J., Depledge, D.D., 2013. Genome Sequence of

532   Human Herpesvirus 7 Strain UCL-1. Genome Announc 1, e00830-00813.

533   Hagberg, E.E., Krarup, A., Fahnoe, U., Larsen, L.E., Dam-Tuxen, R., Pedersen, A.G., 2016. A fast

534   and robust method for whole genome sequencing of the Aleutian Mink Disease Virus

535   (AMDV) genome. J. Virol. Methods 234, 43-51.

536   Hammoumi, S., Vallaeys, T., Santika, A., Leleux, P., Borzym, E., Klopp, C., Avarre, J.C., 2016.

537   Targeted genomic enrichment and sequencing of CyHV-3 from carp tissues confirms low

538   nucleotide diversity and mixed genotype infections. PeerJ 4, e2516.

539   Hernan, I., Borras, E., de Sousa Dias, M., Gamundi, M.J., Mane, B., Llort, G., Agundez, J.A.,

540   Blanca, M., Carballo, M., 2012. Detection of genomic variations in BRCA1 and BRCA2

541   genes by long-range PCR and next-generation sequencing. J Mol Diagn 14, 286-293.

542   Hine, P., Wesney, B., Hay, B., 1992. Herpesviruses associated with mortalities among

543   hatchery-reared larval Pacific oysters, *Crassostrea-gigas*. Dis Aquat Organ 12, 135-142.

544   Jenkins, C., Hick, P., Gabor, M., Spiers, Z., Fell, S.A., Gu, X., Read, A., Go, J., Dove, M.,

545   O'Connor, W., Kirkland, P.D., Frances, J., 2013. Identification and characterisation of an

546   ostreid herpesvirus-1 microvariant (OsHV-1 μvar) in *Crassostrea gigas* (Pacific oysters) in

547   Australia. Dis Aquat Organ 105, 109-126.

548   Jia, H.Y., Guo, Y.F., Zhao, W.W., Wang, K., 2014. Long-range PCR in next-generation sequencing:

549   comparison of six enzymes and evaluation on the MiSeq sequencer. Sci Rep-Uk 4, 5737.

550   Karamitros, T., Harrison, I., Piorkowska, R., Katzourakis, A., Magiorkinis, G., Mbisa, J.L., 2016.

551   *De Novo* Assembly of Human Herpes Virus Type 1 (HHV-1) Genome, Mining of

552   Non-Canonical Structures and Detection of Novel Drug-Resistance Mutations Using Short-

553   and Long-Read Next Generation Sequencing Technologies. Plos One 11, e0157600.

554   Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7:

555   Improvements in Performance and Usability. Mol Biol Evol 30, 772-780.

556   Kolb, A.W., Adams, M., Cabot, E.L., Craven, M., Brandt, C.R., 2011. Multiplex sequencing of

557   seven ocular herpes simplex virus type-1 genomes: phylogeny, sequence variability, and SNP

558   distribution. Invest Ophth Vis Sci 52, 9061-9073.

559   Kolb, A.W., Ane, C., Brandt, C.R., 2013. Using HSV-1 Genome Phylogenetics to Track Past

560   Human Migrations. Plos One 8, e76267.

561   Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis

562   Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870-1874.

563   Kvisgaard, L.K., Hjulsager, C.K., Fahnoe, U., Breum, S.O., Ait-Ali, T., Larsen, L.E., 2013. A fast

564   and robust method for full genome sequencing of Porcine Reproductive and Respiratory

565   Syndrome Virus (PRRSV) Type 1 and Type 2. J. Virol. Methods 193, 697-705.

566   Kwok, H., Wu, C.W., Palser, A.L., Kellam, P., Sham, P.C., Kwong, D.L.W., Chiang, A.K.S., 2014.

567   Genomic Diversity of Epstein-Barr Virus Genomes Isolated from Primary Nasopharyngeal

568   Carcinoma Biopsy Samples. J Virol 88, 10662-10672.

569   Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J.,

570   Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu,

571   S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., Wang,

572   J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo

573   assembler. Gigascience 1, 18.

574   Martenot, C., Oden, E., Travaille, E., Malas, J.P., Houssin, M., 2010. Comparison of two real-time

575   PCR methods for detection of ostreid herpesvirus 1 in the Pacific oyster *Crassostrea gigas*. J.

576   Virol. Methods 170, 86-89.

577   Mathijs, E., Vandenbussche, F., Verpoest, S., De Regge, N., Van Borm, S., 2016. Complete

578   Genome Sequence of Pseudorabies Virus Reference Strain NIA3 Using Single-Molecule

579   Real-Time Sequencing. Genome Announc 4, e00440-00416.

580 Morrison, C.L., Iwanowicz, L., Work, T.M., Fahsbender, E., Breitbart, M., Adams, C., Iwanowicz,

581 D., Sanders, L., Ackermann, M., Cornman, R.S., 2018. Genomic evolution, recombination,

582 and inter-strain diversity of chelonid alphaherpesvirus 5 from Florida and Hawaii green sea

583 turtles with fibropapillomatosis. PeerJ 6, e4386.

584 Myers, E.W., Sutton, G.G., Delcher, A.L., Dew, I.M., Fasulo, D.P., Flanigan, M.J., Kravitz, S.A.,

585 Mobarry, C.M., Reinert, K.H., Remington, K.A., Anson, E.L., Bolanos, R.A., Chou, H.H.,

586 Jordan, C.M., Halpern, A.L., Lonardi, S., Beasley, E.M., Brandon, R.C., Chen, L., Dunn, P.J.,

587 Lai, Z., Liang, Y., Nusskern, D.R., Zhan, M., Zhang, Q., Zheng, X., Rubin, G.M., Adams,

588 M.D., Venter, J.C., 2000. A whole-genome assembly of Drosophila. Science 287, 2196-2204.

589 Newman, R.M., Lamers, S.L., Weiner, B., Ray, S.C., Colgrove, R.C., Diaz, F., Jing, L., Wang, K.,

590 Saif, S., Young, S., Henn, M., Laeyendecker, O., Tobian, A.A., Cohen, J.I., Koelle, D.M.,

591 Quinn, T.C., Knipe, D.M., 2015. Genome Sequencing and Analysis of Geographically

592 Diverse Clinical Isolates of Herpes Simplex Virus 2. J Virology 89, 8219-8232.

593 Nicolas, J., Comps, M., Cochennec, N., 1992. Herpes-like virus infecting Pacific-oyster larvae,

594 *Crassostrea gigas*. B Eur Assoc Fish Pat 12, 11-13.

595 Norberg, P., Kasubi, M.J., Haarr, L., Bergstrom, T., Liljeqvist, J.A., 2007. Divergence and

596 recombination of clinical herpes simplex virus type 2 isolates. J Virol 81, 13158-13167.

597 Olp, L.N., Jeanniard, A., Marimo, C., West, J.T., Wood, C., 2015. Whole-Genome Sequencing of

598 Kaposi's Sarcoma-Associated Herpesvirus from Zambian Kaposi's Sarcoma Biopsy

599 Specimens Reveals Unique Viral Diversity. J Virol 89, 12299-12308.

600 Ozcelik, H., Shi, X.J., Chang, M.C., Tram, E., Vlasschaert, M., Di Nicola, N., Kiselova, A., Yee,

601 D., Goldman, A., Dowar, M., Sukhu, B., Kandel, R., Siminovitch, K., 2012. Long-Range

602   PCR and Next-Generation Sequencing of BRCA1 and BRCA2 in Breast Cancer. J Mol Diagn

603   14, 467-475.

604   Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A.,

605   Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms:

606   comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC

607   Genomics 13, 341.

608   Ren, W.C., Chen, H.X., Renault, T., Cai, Y.Y., Bai, C.M., Wang, C.M., Huang, J., 2013. Complete

609   genome sequence of acute viral necrosis virus associated with massive mortality outbreaks in

610   the Chinese scallop, *Chlamys farreri*. Virol J 10, 110.

611   Renault, T., Moreau, P., Faury, N., Pepin, J.F., Segarra, A., Webb, S., 2012. Analysis of clinical

612   *ostreid herpesvirus 1* (*Malacoherpesviridae*) specimens by sequencing amplified fragments

613   from three virus genome areas. J Virol 86, 5942-5947.

614   Savin, K.W., Cocks, B.G., Wong, F., Sawbridge, T., Cogan, N., Savage, D., Warner, S., 2010. A

615   neurotropic herpesvirus infecting the gastropod, abalone, shares ancestry with oyster

616   herpesvirus and a herpesvirus associated with the amphioxus genome. Virol J 7, 308.

617   Schikorski, D., Renault, T., Saulnier, D., Faury, N., Moreau, P., Pepin, J.F., 2011. Experimental

618   infection of Pacific oyster *Crassostrea gigas* spat by ostreid herpesvirus 1: demonstration of

619   oyster spat susceptibility. Vet Res 42, 27.

620   Segarra, A., Pepin, J.F., Arzul, I., Morga, B., Faury, N., Renault, T., 2010. Detection and

621   description of a particular *Ostreid herpesvirus 1* genotype associated with massive mortality

622   outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. Virus Res 153, 92-99.

623   Sipos, R., Szekely, A.J., Palatinszky, M., Revesz, S., Marialigeti, K., Nikolausz, M., 2007. Effect

624   of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA

625   gene-targetting bacterial community analysis. FEMS Microbiol Ecol 60, 341-350.

626   Spatz, S.J., Rue, C.A., 2008. Sequence determination of a mildly virulent strain (CU-2) of Gallid

627   herpesvirus type 2 using 454 pyrosequencing. Virus Genes 36, 479-489.

628   Szpara, M.L., Gatherer, D., Ochoa, A., Greenbaum, B., Dolan, A., Bowden, R.J., Enquist, L.W.,

629   Legendre, M., Davison, A.J., 2014. Evolution and Diversity in Human Herpes Simplex Virus

630   Genomes. J Virol 88, 1209-1227.

631   Tcherepanov, V., Ehlers, A., Upton, C., 2006. Genome Annotation Transfer Utility (GATU): rapid

632   annotation of viral genomes using a closely related reference genome. BMC Genomics 7,

633   150.

634   Tischer, B.K., Osterrieder, N., 2010. Herpesviruses--a zoonotic threat? Vet Microbiol 140,

635   266-270.

636   Tombacz, D., Sharon, D., Olah, P., Csabai, Z., Snyder, M., Boldogkoi, Z., 2014. Strain Kaplan of

637   Pseudorabies Virus Genome Sequenced by PacBio Single-Molecule Real-Time Sequencing

638   Technology. Genome Announc 2, e00628-00614.

639   Tweedy, J., Spyrou, M.A., Donaldson, C.D., Depledge, D., Breuer, J., Gompels, U.A., 2015.

640   Complete Genome Sequence of the Human Herpesvirus 6A Strain AJ from Africa Resembles

641   Strain GS from North America. Genome Announc 3, e01498-01414.

642   Uribe-Convers, S., Duke, J.R., Moore, M.J., Tank, D.C., 2014. A Long Pcr Based Approach for

643   DNA Enrichment Prior to Next-Generation Sequencing for Systematic Studies. Appl Plant

644   Sci 2, 1300063.

645   Xia, J., Bai, C., Wang, C., Song, X., Huang, J., 2015. Complete genome sequence of Ostreid

646   herpesvirus-1 associated with mortalities of *Scapharca broughtonii* broodstocks. Virol J 12,

647   110.

648

649 **Figure legends**

650 **Figure 1. Agarose gel electrophoreses (0.8 %) analysis of long range PCR**

651 **amplicons.**

652 (A) PCR amplicons for ZK0118, (B) PCR amplicons for negative control, lanes 1-23

653 corresponds to 23 PCR amplicons respectively.

654

655 **Figure 2. Coverage and GC content across the assembled genomes.**

656 OsHV-1 reference (AY509253) was used as a reference of nucleotide positions.

657 ZK0118.2: sequence data of Illumina HiSeq 4000 platform, ZK0118.3: sequence data

658 of PacBio RS II platform. GC content of OsHV-1 reference was displayed.

659

660 **Figure 3. Phylogenetic tree of 11 OsHV-1 variants inferred using the**

661 **Neighbor-Joining method.**

662 Bootstrap values were obtained from 100 resampled data sets. Numbers at the

663 branches indicate bootstrap support value > 50%. The clade colored in red indicates

664 variants identified in *Chlamys farreri* from China, the clade colored in green indicates

665 variants identified in *Crassostrea gigas* from Europe, the clade colored in black

666 indicates variants identified in *Scapharca broughtonii* from China.

Figure 1
Click here to download high resolution image
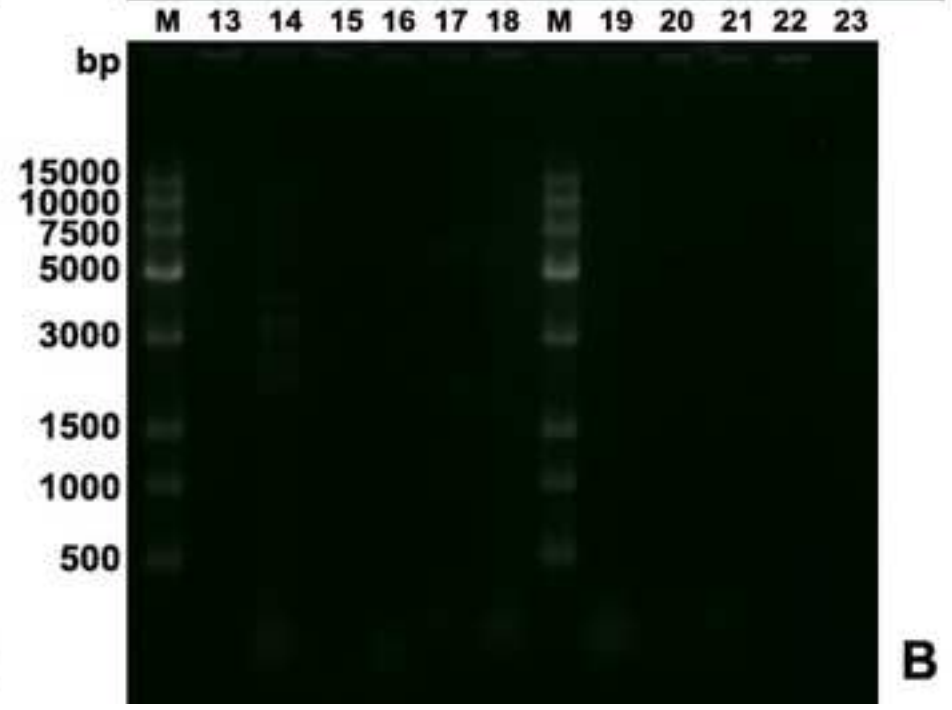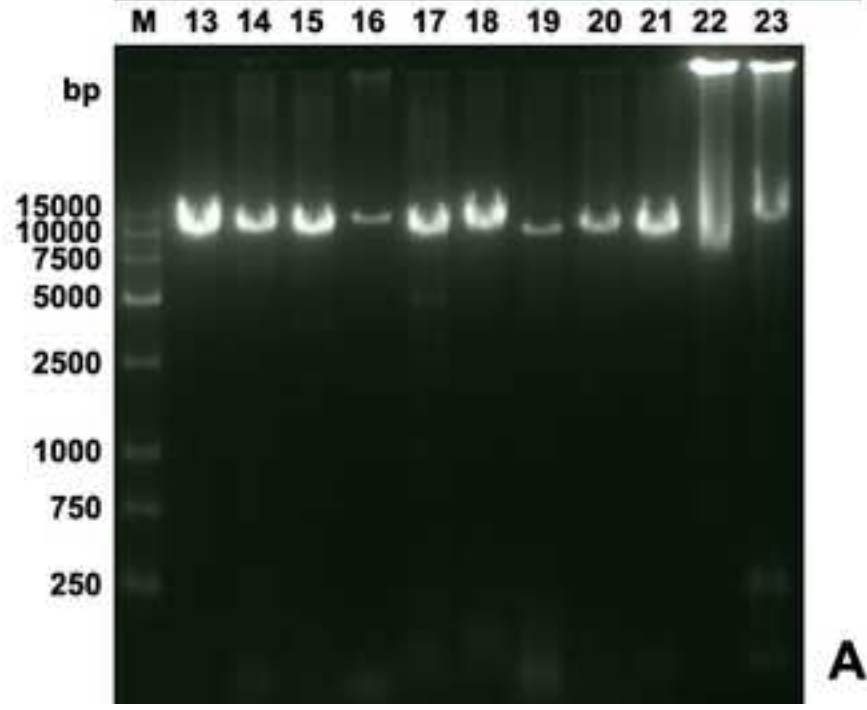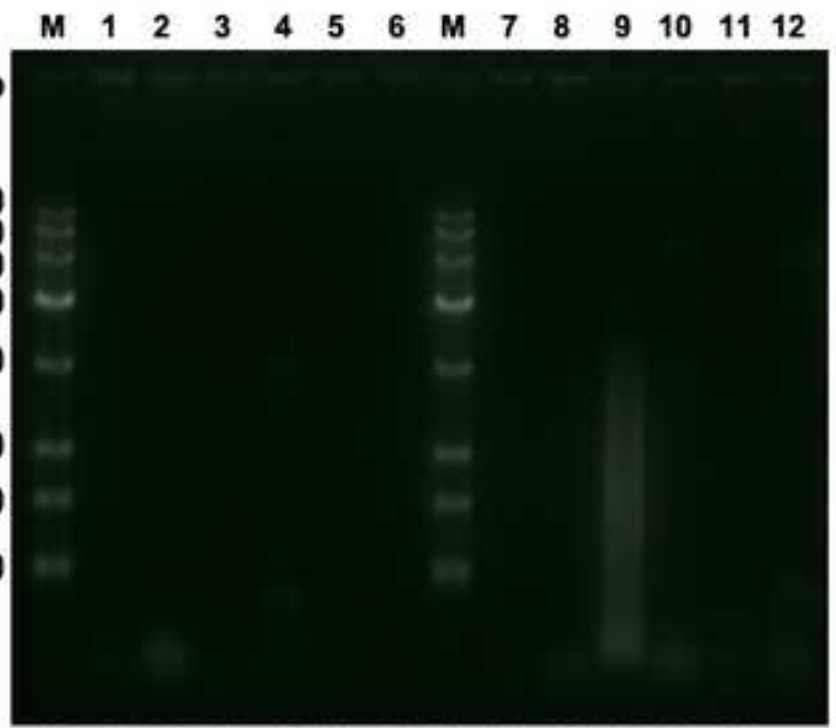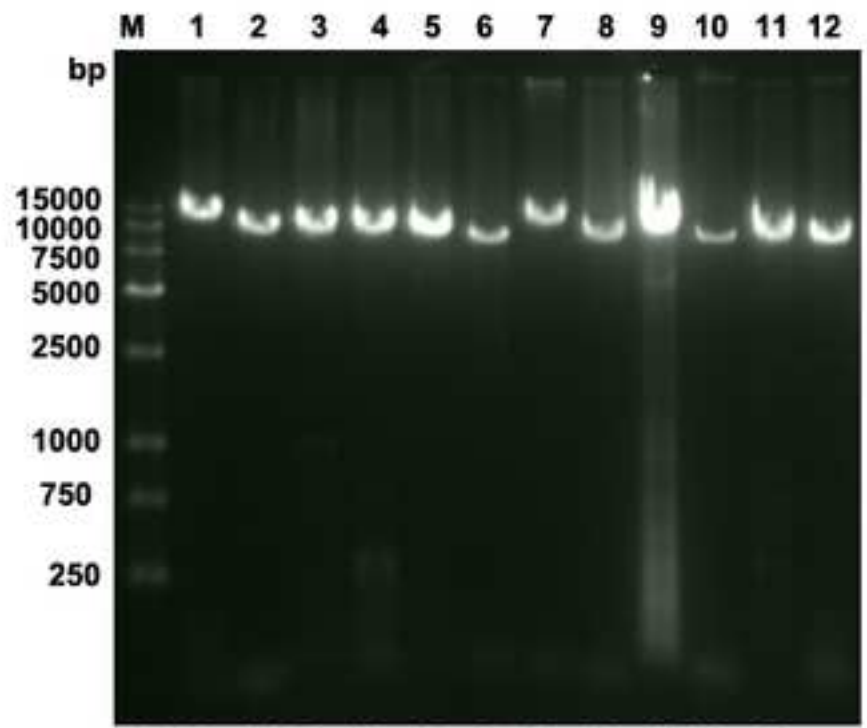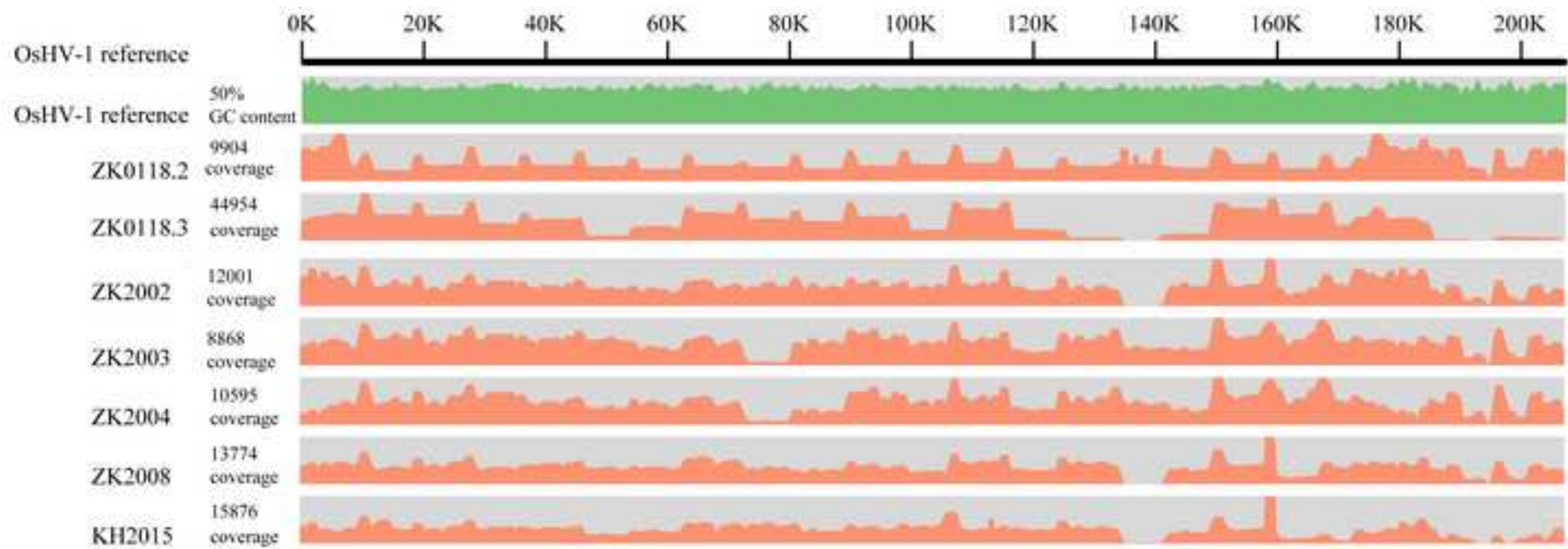
Figure3
Click here to download high resolution image

Varants identified in *Crassostrea gigas* from Europe

KY242785

KY271630

AY509253

100

100

GQ153938

ZK2003

100

100

ZK2002

ZK2004

ZK2008

MF509813

KP412538

KH2015

Variants identified in *Chlamys farreri* from China

Variants identified in *Scapharca broughtonii* from China

7.0E-4
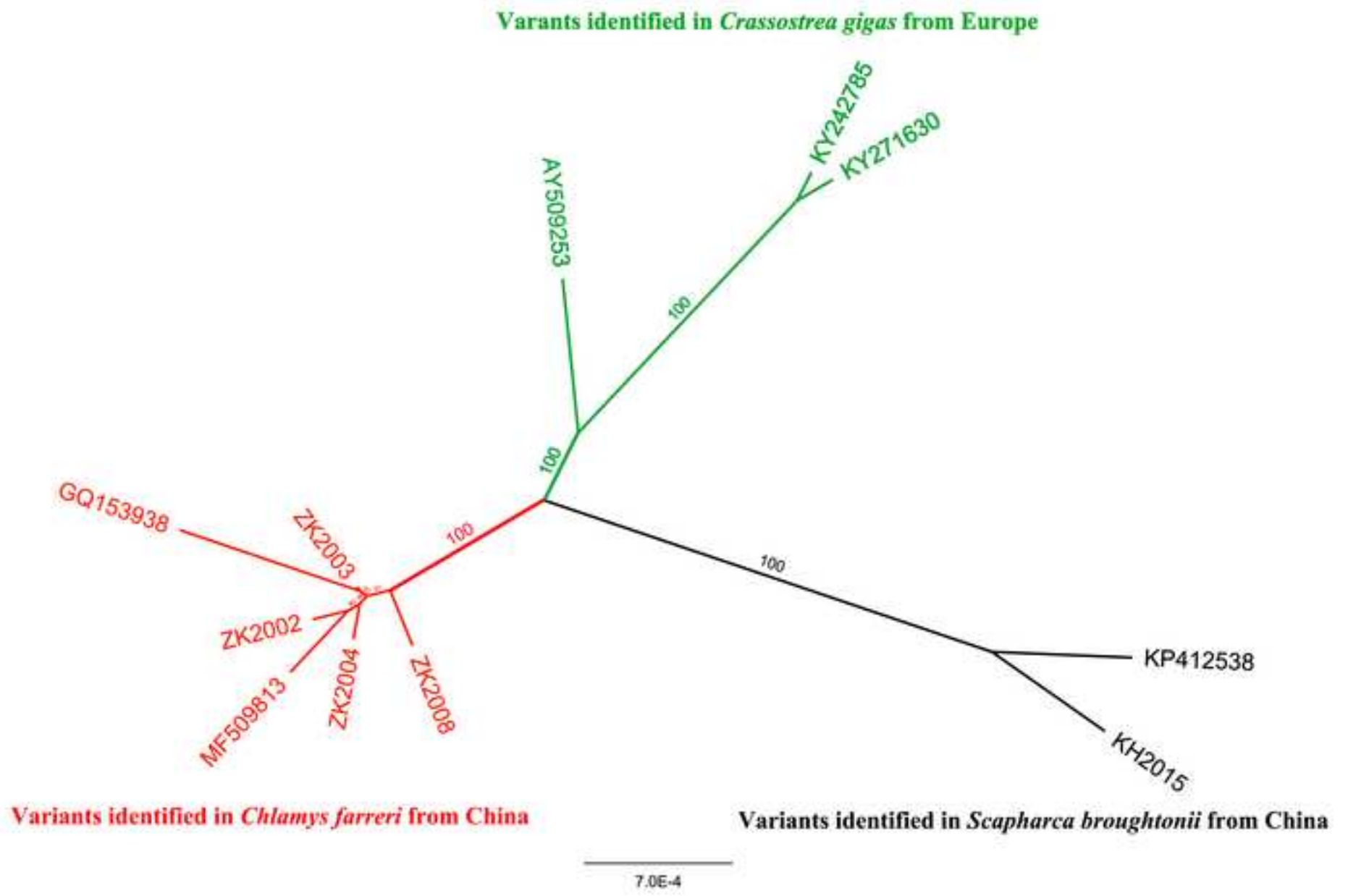
Table 1

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

Sample ID	Date of sampling	Origin	Host species	Age*	Viral loads
ZK0118	**August 2001**	**Qingdao**	*Chlamys farreri*	**Adult**	$\mathbf{3.4 \times 10^4}$
ZK2002	**August 2002**	**Qingdao**	*Chlamys farreri*	**Adult**	$\mathbf{2.1 \times 10^4}$
ZK2003	**July 2003**	**Qingdao**	*Chlamys farreri*	**Adult**	$\mathbf{1.6 \times 10^4}$
ZK2004	**August 2004**	**Qingdao**	*Chlamys farreri*	**Adult**	$\mathbf{3.2 \times 10^4}$
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**	$\mathbf{8.8 \times 10^3}$
ZK2011	August 2011	Changdao	*Chlamys farreri*	Adult	7.3×10^2
KH2015	**June 2015**	**Rizhao**	*Scapharca broughtonii*	**Adult**	$\mathbf{4.6 \times 10^4}$
ML2015	May 2015	Qingdao	*Crassostrea gigas*	larva	2.3×10^3

* 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*).

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

Table2

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	11102	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	10145	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*	*10662*	*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*	*11558*	*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	13029	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*	*11037*	*1572*	U_S-TR_S	Absent in KP
GF23-1R[#]	*AGATGACGAATCGGAGGA*	*18*	*206302*	*206285*	*51*				1 SNP in GQ

*: 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.

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.

Table 3

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	4,236,000	36	5,019	9,888	92.71	2.1×10^4
ZK2003	5,336,000	4,228,000	42	4,353	8,028	91.46	1.6×10^4
ZK2004	5,784,000	4,360,000	17	1,1200	14,746	95.23	3.2×10^4
ZK2008	5,412,000	4,248,000	68	2,547	3,813	88.09	8.8×10^3
KH2015	4,888,000	3,868,000	16	1,2042	28,233	91.35	4.6×10^4

Table 4

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

	OsHV-1 (AY509253)*	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	

*: GenBank Accession numbers of each variant was provided in parentheses.

Table5

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	216079	1583	7.33	211882	496	2.34	213330	1027	4.81	211074	274	1.30
TR_L	7705	120	15.57	7693	40	5.20	7682	81	10.54	7646	12	1.57
U_L	170656	1042	6.11	170540	354	2.08	170607	691	4.05	170436	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	10963	160	14.59	9785	27	2.76	10402	106	10.19	10231	48	4.69
U_S	3372	41	12.16	3370	5	1.48	3372	34	10.08	3372	7	2.08
TR_S	13259	127	9.58	11295	28	2.48	11185	75	6.71	10233	38	3.71

Table 6

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.