1   **Phylogeography of infectious disease: genetic diversity and connectivity of the Ostreid**

2   **herpesvirus 1 population in France**

3

4   Jean Delmotte[1], Camille Pelletier[2], Benjamin Morga[2], Richard Galinier[3], Bruno Petton[4], Jean-

5   Baptiste Lamy[2], Oliver Kaltz[5], Jean-Christophe Avarre[5], Caroline Montagnani[1]*, Jean-

6   Michel Escoubas[1]*

7

8   [1]IHPE, Univ. Montpellier, CNRS, Ifremer, UPVD, F-34095 Montpellier, France

9   [2]Ifremer, RBE-SG2M-LGPMM, Station La Tremblade, F-17390 La Tremblade, France

10   [3]IHPE, Univ. Montpellier, CNRS, Ifremer, UPVD, F-66000 Perpignan, France

11   [4]Ifremer, LEMAR UMR 6539 (Université de Bretagne Occidentale, CNRS, IRD, Ifremer), F-

12   29840 Argenton-en-Landunvez, France

13   [5]ISEM, IRD, CNRS, EPHE, Univ. Montpellier, F-34095 Montpellier, France

14

15   Running Head: Phylogeography of *Ostreid herpesvirus 1*

16

17   *Corresponding   authors: Caroline.Montagnani@ifremer.fr;

18   Jean.Michel.Escoubas@ifremer.fr

19

20

21 **Abstract**

22   The genetic diversity of viral populations is a key to understanding ther

23 phylogeographic and dissemination history of viruses, but studying the diversity of whole

24 genomes from natural populations remains a challenge. Molecular ecology approaches are

25 commonly used for RNA viruses harboring small genomes, but have only rarely been applied

26 to DNA viruses with large genomes. Here, we used the Pacific oyster mortality syndrome

27 (POMS, a disease that affects oyster farms around the world) as a model to study the genetic

28 diversity of its causative agent, the Ostreid herpesvirus 1 (OsHV-1) in the three main French

29 oyster-farming areas. Using ultra-deep sequencing on individual moribund oysters and new

30 bioinformatics methodology, we *de novo* assembled 21 OsHV-1 genomes. Combining whole-

31 genome comparisons with phylogenetic analysis and quantification of major and minor

32 variants, we assessed the connectivity of OsHV-1 viral populations between the three oyster-

33 farming areas. Our results suggest that the Marennes-Oléron Bay represents the main source

34 of OsHV-1 diversity, from where the virus has dispersed to other farming areas, a scenario

35 consistent with current practices of oyster transfers in France. Here, we demonstrate that

36 molecular ecology approaches can be applied to large-genome viruses to determine the extent

37 of their genetic diversity and better understand the spread of viral populations in natural

38 environments.

39

40 **Importance**

41 Phylogeography is a field of research that attempts to reconstruct the relationships between

42 individual genotypes within a species and then correlate these genealogical relationships with

43 their geographic origin. This field of research has become an essential step in the

44 understanding of pandemics, in particular to determine the origin, spread and evolution of a

45 pathogen as currently illustrated in studies on viral pandemics. However, because

46   phylogeographic analyses are based on genome variation, stable genomes yield less

47   information than labile genomes. Accordingly, viruses with double-stranded DNA (dsDNA)

48   genomes generally have lower nucleotide diversity than RNA viruses. In this study, by

49   combining the use of both major and minor variants with phylogeographic analyses of the

50   oyster herpesvirus OsHV-1, we highlight genealogical relationships that have not been

51   depicted in phylogenetic trees based on consensus viral genomes only. These data offer a

52   plausible scenario reflecting the origin and spread of OsHV-1 populations between oyster-

53   farming sites.

54

55   **Keywords** Herpesvirus; OsHV-1; cupped oyster; phylogeography; non-redundant genomes; genetic

56   diversity; viral spread; minor variant

Introduction

58   Viruses are disease agents that have high levels of genetic diversity. This high

59   diversity often means that they lack shared genetic markers, such as ribosomal DNA

60   sequences that are common to all prokaryotes and eukaryotes (1), making it difficult to

61   characterize viruses genetically. Furthermore, many viruses — especially those with RNA

62   genomes — are not stable genetic entities, but exist as clouds of many phylogenetically

63   related genetic variants, known as viral quasispecies. This genetic organization currently

64   encumbers our understanding of viral diseases and their evolution and impedes the

65   straightforward characterization of virus population structure (2, 3). Studies have shown that

66   the level of genetic diversity within these viral populations likely influences viral

67   pathogenicity, dissemination and host immune evasion, and full-length genome analyses are

68   required to identify intra-host viral population structure, reveal molecular traits with

69   epidemiological significance, or detect low-frequency, but nonetheless relevant viral variants.

3

70   These difficulties apply to RNA viruses, but also certain large DNA viruses, such as

71   herpesviruses, whose genomic variability rivals that of many RNA viruses (4-6). For

72   example, human cytomegalovirus shows considerable inter-host and intra-host genetic

73   divergence across tissue compartments and times of infection (7, 8). In addition, the evolution

74   of a disease is sometimes not fully explained by intrinsic host factors, but by the genotypic

75   diversity of herpesviruses (9).

76   Aquaculture is one of the fastest growing food-producing sectors, representing an

77   important animal protein supply for human consumption, with an expanding role in global

78   food security. Today, the biggest threat arising due to the intensification and globalization of

79   aquaculture is infectious diseases. The management and mitigation of the emergence and

80   spread of these infectious diseases are key issues to address to ensure the sustainability of this

81   industry (10-12). One illustration is the Pacific oyster mortality syndrome (POMS), which

82   threatens global *Crassostrea gigas* oyster production, a main sector in aquaculture worldwide

83   (reviewed in (13)). Since 2008, this syndrome has caused mass mortality in cultivated oysters

84   around the world, from Europe to America and Asia (14-22). In 2010, the causative agent of

85   these massive mortalities was identified: it is an emerging genotype of a herpes-like virus

86   named Ostreid herpesvirus 1 (OsHV-1) (23-27). Two major genetic factors seem to affect the

87   severity of POMS: the genetic background of the oysters (28-30) and OsHV-1 genetic

88   diversity (31, 32). Since the characterization of the emergent genotype OsHV-1 μVar,

89   associated with the 2008 high mortality events (27), several variants of the μVar genotype

90   have been described (reviewed in (16)). Interestingly, in 2016, a survey of OsHV-1 genetic

91   diversity carried out on wild *C. gigas* populations along the coasts of Italy demonstrated the

92   high diversity of this virus in natural oyster populations (33). However, most studies on

93   OsHV-1 genetic diversity have been based on PCR molecular markers focusing on a few

94   variable regions of the viral genome, which may not reflect the whole genomic diversity. For

4

95   instance, in 2017, sequencing of the whole genome of OsHV-1 µVar and its comparison with

96   the reference genome published in 2005 (34) showed that the two genomes also differed by

97   the loss or addition of several open reading frames (ORFs), indicating that whole-genome

98   sequencing is necessary to fully reveal viral diversity and to better understand the virus'

99   origin and evolution (35). To date, no study has specifically looked at the links between the

100   geographic distribution of OsHV-1 and its genetic diversity at the level of the whole genome.

101   To date, the entire genomes of only four OsHV-1 infecting *C. gigas* oysters are available,

102   namely the reference genome (AY509253), two OsHV-1 µVar (µVar A KY242785.1 and

103   µVar B KY271630.1) and more recently OsHV-1-PT (MG561751.2) (14, 34, 35). These four

104   assemblages show significant genomic diversity; however, they tell us very little about the

105   geographic distribution and dissemination of the virus.

106   As shown with the H1N1 influenza virus (36), the dengue virus (37), the Zika virus

107   (38), the AIDS virus (39) and more recently SARS-CoV-2 (40), molecular epidemiology

108   based on full genomes can be the key to understanding the epidemics of emerging viruses.

109   Therefore, the purpose of the present study was to characterize the genomic diversity of

110   OsHV-1 µVar in natural viral populations both within and between host individuals, to better

111   understand the dissemination and evolution of OsHV-1 µVar populations. Using a deep-

112   sequencing approach conducted at the individual level and novel bioinformatics analyses, we

113   identified 21 new complete genome sequences of OsHV-1 µVar from the three most

114   important oyster-farming areas in France. Phylogenetic analyses combined with comparative

115   genomics and inspection of variant frequencies showed that viral genetic diversity differs

116   greatly not only between areas, but also between individuals within areas. This

117   phylogeographic approach led us to propose a scenario explaining the phylogenetic

118   relationships between viral populations encountered in the three farming areas. This analysis

119   strategy, combined with epidemiological information, holds promise for understanding

5

120   OsHV-1 evolution, for revealing how the genetic relationships shape epidemic dynamics and

121   for identifying transmission patterns between oyster-producing regions. These data are

122   ultimately useful for developing novel disease management strategies.

123   **Results**

124   **1) Variability in sequencing depth does not impair the analysis of OsHV-1**

125   **diversity.**

126   We used specific-pathogen-free (SPF) *C. gigas* juveniles to sample OsHV-1 diversity

127   during natural infections in the three farming areas. The cumulative moralities for the three

128   batches of oysters were 85%, 81% and 70% for Thau (Th), Marennes-Oléron (Mo) and Brest

129   (Br), respectively (Fig. 1A). No mortality was observed in the control group.

130   For each individual oyster, deep sequencing produced from 86.2 to 145.9 M reads per

131   sample (average 109.7 M reads ±16.7 SD, Fig.1B. and Table S1). To evaluate OsHV-1

132   sequencing depth in each sample, sequencing reads were mapped on the OsHV-1 μVar A

133   genome (KY242785.1) (35). The number of reads mapping to the OsHV-1 μVar A genome

134   greatly varied from one individual to the other, independently of the number of reads

135   obtained for each sample: from 0.65 M reads (0.66%) for Mo ind3 to 11.94 M reads (13.21%)

136   for Th ind10 (Fig.1B, Table S1). Therefore, OsHV-1 genome coverage varied from 480X to

137   8745X for Mo ind3 and Th ind10, respectively.

138   A rarefaction analysis was run to verify that OsHV-1 coverage was sufficient to

139   accurately quantify viral genomic diversity in each sample. For single nucleotide

140   polymorphisms (SNPs) and for insertions-deletions (InDels), rarefaction curves reached a

141   plateau around 300,000 reads, corresponding to an OsHV-1 coverage of 220 X (Fig. 1C).

142   Since the number of mapped reads of each sample was clearly above this threshold,

143   sequencing depth was considered sufficient to capture OsHV-1 genomic diversity.

144   Rarefaction curves indicate that viral diversity was higher in infected Mo oysters than in the

145   other two areas. Comparison of the average number of individual polymorphisms per

146   environment showed that i) the number of SNPs in viruses from Mo (170 ± 27.7) was

147   significantly higher than that of viruses from Br (136.4 ± 5.9; $P \leq 0.001$) and Th (143.8 ± 5.8;

148    $P \leq 0.05$) and ii) the number of InDels in viruses from Mo (35.9 ±3.1 SD) was also

149   significantly higher than that of viruses from Th (25.7 ± 0.9; $P \leq 0.0001$) (Fig. 1D, Table S2).

150   In addition, in Mo, viral diversity greatly varied from one individual to the other; for example

151   Mo ind3 contained 222 SNPs and 42 InDels, whereas Mo ind4 contained only 144 SNPs and

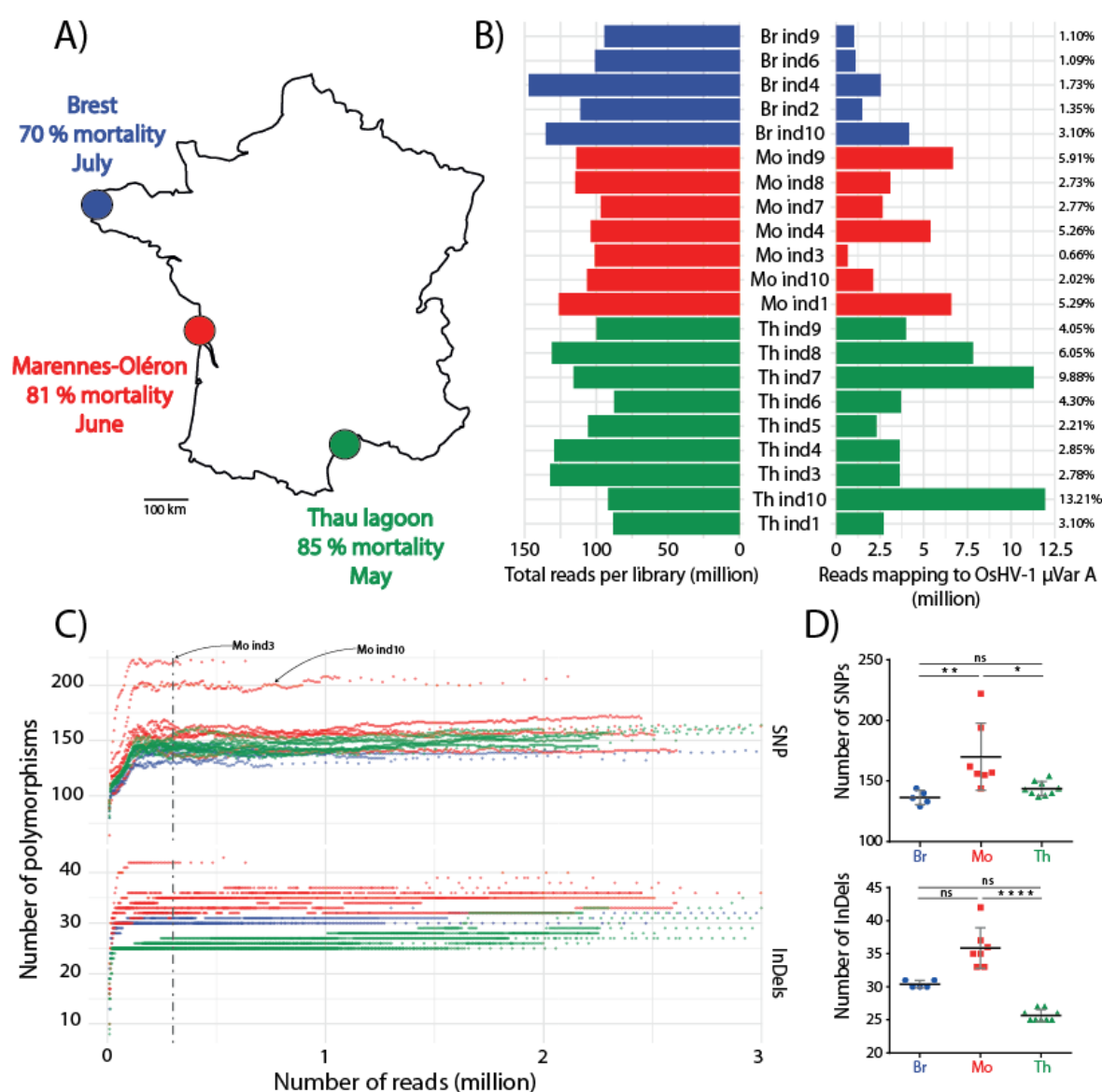152   33 InDels.

153

154 **Fig. 1: Variability in sequencing depth does not impair the analysis of OsHV-1 diversity**.

155 (**A**) Sampling locations of the Ostreid herpesvirus-1 (OsHV-1) infected oysters during the

156 Pacific oyster mortality syndrome (POMS) event in 2018. Each color represents a farming

157 area, with Brest (Br) in blue (5 individual samples), Marennes-Oléron (Mo) in red (7

158 individual samples) and Thau Lagoon (Th) in green (9 individual samples). Below each

159 farming area, the percentage of mortality within the site is given. (**B**) Quantification of the

160 number of reads in the library and the number of reads mapping to the OsHV-1 μVar A

161 genome. Numbers on the right side of the figure indicate the percentages of reads mapping to

162 the OsHV-1 μVar A genome. (**C**) Rarefaction curves using the OsHV-1 μVar A genome as

163 the reference sequence. The top panel shows the number of single nucleotide polymorphisms

164 (SNPs) and the bottom panel, the number of insertion-deletions (InDels). The dashed vertical

165 line indicates 0.3 M reads, corresponding to the threshold at which the rarefaction curves

166 reach a plateau. (**D**) Comparison of the number of SNPs (top) and InDels (bottom) between

167 the three farming areas. Asterisks indicate the *P* value according to Kruskal-Wallis and

168 Dunn's multiple comparison tests (ns: non-significant, $*P \leq 0.05$, $**P \leq 0.01$ $****$ $P \leq$

169 0.0001).

170 **2) Phylogenetic analysis of non-redundant OsHV-1 genomes reveals that all three**

171 **OsHV-1 populations belong to the μVar genotype.**

172 We assembled non-redundant genomes so as to reduce the complexity of the OsHV-1

173 genome and keep only one copy of each repeated region in the virus sequence. The OsHV-1

174 μVar A genome contains two unique regions and three repeats: the unique long region (UL,

175 164,268 bp), flanked by two inverted repeats (repeat long [RL], 7338 bp), and the unique

176 short region (US, 3370 bp), flanked by two inverted repeats (repeat short [RS], 9777 bp), and

177 two X regions (1510 bp) (Fig 2A, upper panel). Considering the high sequence identity

178 within each set of repeats (between 99.9 and 100%), removing one of the repeats makes it

179 possible to reliably map the reads from these repeated regions without affecting the overall

180 genome composition. Removing copy of each repeated region led to a non-redundant OsHV-

181 1 µVar A genome (NR-genome) of 186,262 bp, which represents 91% of the full-length

182 genome and contains all the coding sequences (Fig 2A, lower panel). Using this approach, 21

183 NR-genomes were assembled from the sequencing data generated during this study, as well

184 as 7 other published genomes belonging to the *Malacoherpesviridae* family: six *Ostreid*

185 *herpesvirus 1* (genus *Ostreavirus*) and one *Haliotid herpesvirus 1* (abalone herpesvirus

186 (AbHV-1), genus *Aurivirus*).

187 A phylogenetic analysis of these 28 NR-genomes revealed three main polyphyletic

188 clusters (Fig. 2B). AbHV-1, the sole herpesvirus isolated from a gastropod, was identified as

189 an outgroup, whereas the two herpesviruses isolated from bivalves other than oysters (OsHV-

190 1-SB isolated from the clam *Scapharca broughtonii* and OsHV-1 ZK0118 isolated from the

191 scallop *Azumapecten farreri*) formed two separate branches. Among the viruses isolated from

192 *C. gigas*, the 21 new OsHV-1 isolates were phylogenetically closer to OsHV-1 µVar A and B

193 isolates (Fig. 2B cluster 3) than to the OsHV-1 reference genome (AY509253.2), suggesting

194 that they belong to the µVar genotype. Th and Br isolates grouped into two distinct clusters

195 reflecting their geographic origin (Fig. 2B, cluster 1 and 2). Conversely, Mo OsHV-1 isolates

196 were distributed in these two clusters: five in cluster 1 and two in cluster 2. The phylogenetic

197 distance within the Th or Br isolates (average $5.33 \ 10^{-6} \pm 1.99 \ 10^{-6}$ SD and $2.16 \ 10^{-5} \pm 1.43$

198 10^{-5} SD, respectively) was lower than that observed within the Mo isolates (average $6.16 \ 10^{-4}$

199 $\pm 5.29 \ 10^{-4}$), indicating greater viral diversity in Mo than in the other two farming areas.
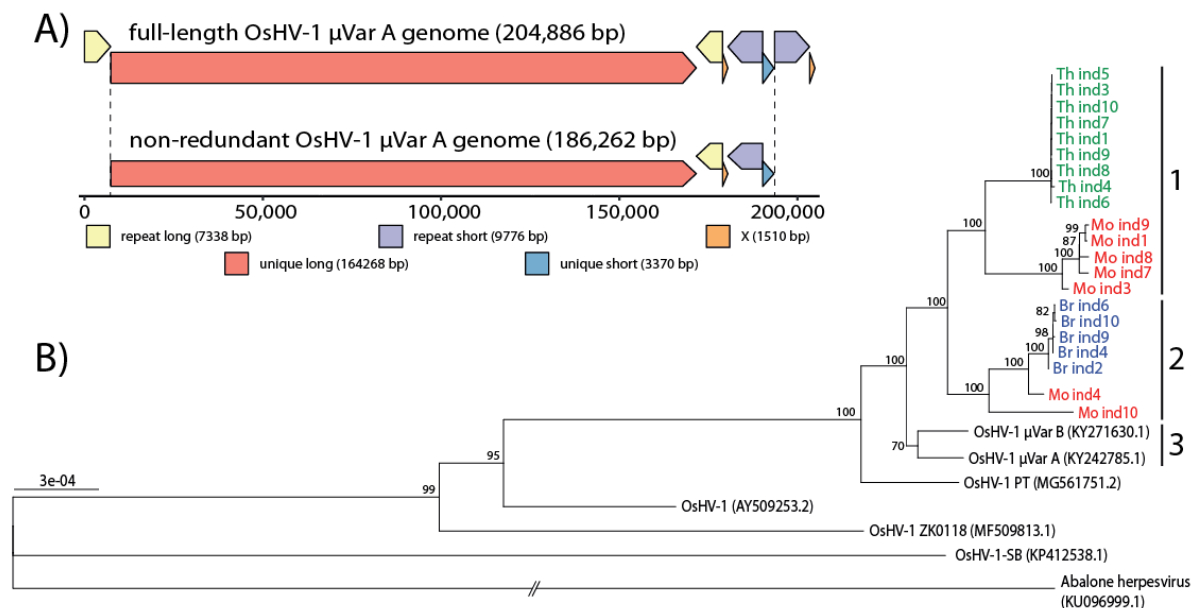
200

Fig. 2: Phylogenetic relationship between the 21 *de novo* assembled non-redundant genomes and other *Malacoherpesviridae* viruses. (**A**) Full-length (upper panel) and non-redundant (NR) (lower panel) genomes of OsHV-1 µVar A (KY242785). The OsHV-1 µVar A genome consists of two unique regions, referred to as unique long (UL; in red) and unique short (US, in blue), and three repeat regions, referred to as repeat long (RL, in yellow), repeat short (RS, in purple) and X region (X, in orange). The first two repeat regions each have copies at a terminal locus (TRL, TRS) and an internal locus (IRL, IRS) within the genome. The NR-genome consists of the linear fragment starting at the beginning of the UL and ending at the end of US region. This strategy includes all the genetic information without duplication. All the regions within the genome and the NR genome are to scale. (**B**) Maximum-likelihood phylogenetic tree of the 28 OsHV-1 NR-genomes. Bootstrap values were obtained from 1000 replicates. Numbers at branch nodes indicate branch confidence (percentage of bootstrap replicates for that branch). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree, except for the outgroup, AbHV-1. The 21 isolates from this study are colored according to their geographic origin, Brest (Br) in blue, Marennes-Oléron (Mo) in red, and Thau

218   Lagoon (Th) in green. Three clusters were defined based on the clustering of genetic

219   distances (numbered on the right from 1 to 3).

220   **3) Comparative genomics of the 21 non-redundant genomes confirms that OsHV-1**

221   **diversity is higher in Marennes-Oléron than in the other two farming areas.**

222   To compare the genetic diversity at the whole genome scale between the 21 samples, we

223   used the 21 NR-genomes to construct a consensus non-redundant genome (C-NR-genome).

224   Comparison of the 21 NR-genomes to the C-NR-genome identified a total of 399 variable

225   positions (SNPs and InDels), hereafter called variants. The average number of variants per

226   genome was significantly higher in samples from Br (159.8 ± 4.9) than from Th (125 ± 5.6; P

227   < 0.01), whereas the average number of variants in Mo (139.4 ± 20.8) was not significantly

228   different from the other two farming areas (Fig. 3A, Table S3).

229   However, viruses from Mo showed the highest diversity, with 321 variants accounting for

230   80.5% of the total diversity. Among the variants, 151 (47%) were specific to Mo, and 87

231   were observed in one sample only (singletons) (Fig. 3B, Table S3). None of these 151

232   variants were shared by all the samples from this area. The majority of Mo singleton variants

233   were observed in two samples: ind10 (49) and ind4 (13) (Fig. 3C). For the Th and Br farming

234   areas, 134 and 168 variants were identified, among which 51 and 24 were area-specific,

235   respectively (Fig. 3B, Table S3). In contrast to Mo, 70.8% (17/24) and 94.1% (48/51) of the

236   variants were common to all the samples of Br and Th, respectively (Fig. 3B, Table S3).

237   The three variants common to Br and Th were present in almost all individuals of the two

238   farming areas (Fig. 3C, green arrows). Namely, among the 90 variants shared by Br and Mo,

239   78 (86.7%) were shared by two Mo individuals (ind10 and ind4) and all Br individuals (Fig.

240   3C, purple arrow). These two Mo individuals also shared six variants with all Th individuals
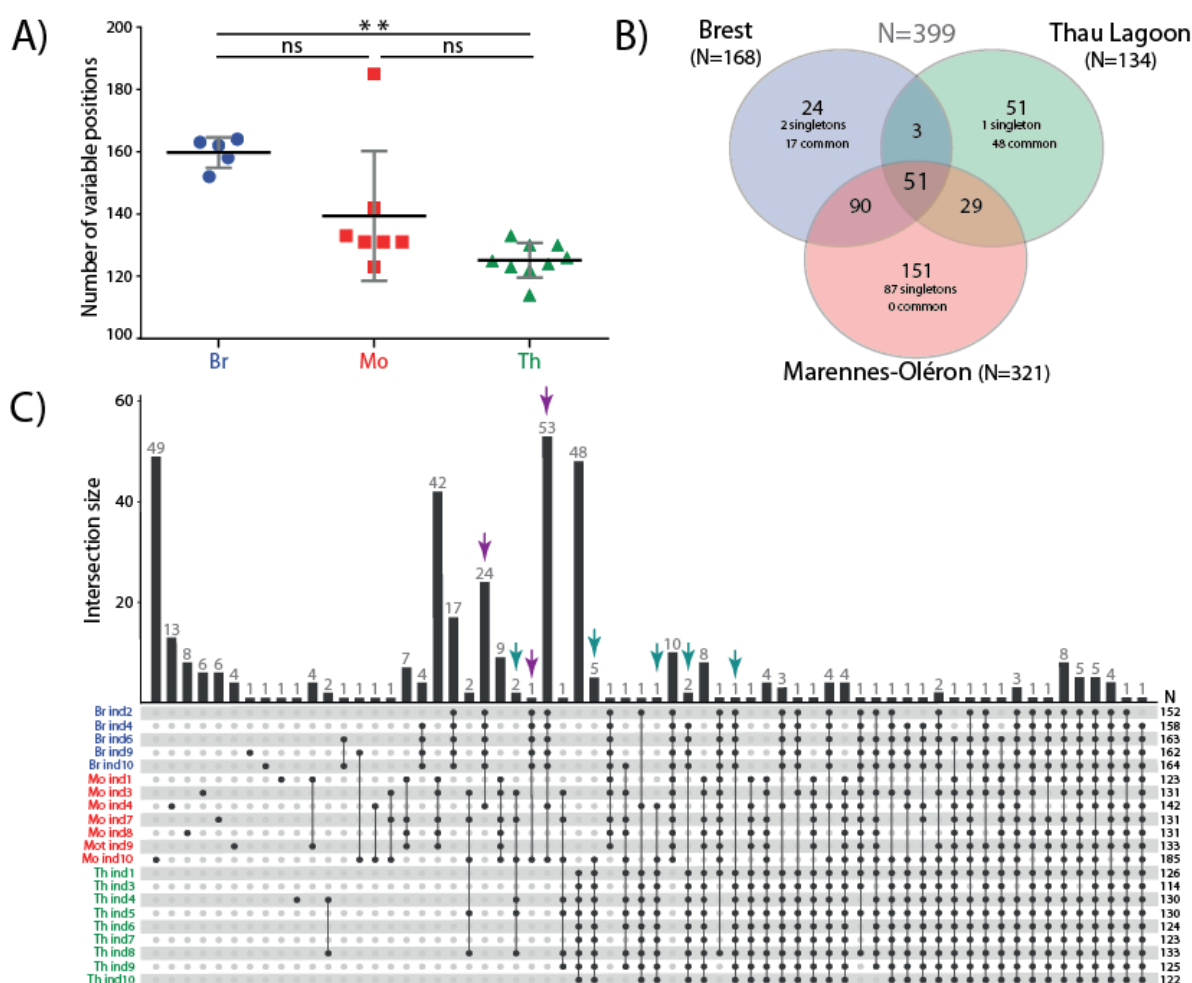
241   (Fig. 3C, red arrows).

11

242

243

Fig. 3: Comparative genomics of the 21 non-redundant genomes of OsHV-1 isolated from three oyster farming areas. (**A**) Distribution of the 399 variable positions (hereafter named variants) between the 21 individuals and the three farming areas. Significance was calculated according to Kruskal-Wallis and Dunn's multiple comparison tests (ns: non-significant, ∗∗ P < 0.01). **B)** Venn diagram summarizing the distribution of variants between the three farming areas. Color depends on the farming area with Brest (Br) in blue, Marennes-Oléron (Mo) in red and Thau Lagoon (Th) in green. The numbers of variants per farming area are indicated in parentheses. Singletons are variants that occur in only one sample, and the term "common" indicates variants present in all the sampled individuals at a given location. (**C**) UpSet plot of variant distribution among the 21 non-redundant (NR-)

254   genomes. The bottom panel shows the specific combinations (or intersections), and the

255   vertical bars indicate the number of variants within these combinations. Colored arrows on

256   top indicate variants shared by almost all individuals from Br and Th (green arrows) and

257   variants shared by two individuals from Mo (ind4 and ind10) and all individuals from Br

258   (purple arrows) or all individuals from Th (red arrows). Numbers (N) on the right side

259   indicate the number of variants per sample.

260   **4) Diversity of minor OsHV-1 variants is higher in Marennes-Oléron than in the**

261   **other two farming areas.**

262   To fully characterize OsHV-1 genetic diversity, we performed variant calling analyses,

263   using the C-NR-genome as the reference, to quantify both major (frequency >50%) and

264   minor (frequency <50%) variants in the 21 individual samples. Among the 399 variants

265   identified as major variants, 117 had variable frequencies (above or below 50%) across

266   samples; moreover, 129 additional minor variants were identified (Fig. 4A).

267   Frequencies of the 129 strictly minor variants ranged from 5.1% to 45.9% (mean =

268   8.75% ± 5.75% SD). The number of minor variants was lower in oysters from Br (39) than in

269   oysters from Th (71) or Mo (75) (Fig. 4B, Table S4). Twenty-five of these 129 minority

270   variants (19.3%) were found in all three farming areas. Moreover, the majority of site-

271   specific variations were found in only one individual oyster (singletons).

272   A hierarchical clustering analysis applied to the minor variants showed that individuals

273   did not cluster according to their geographic origin, except for individuals 4, 6 and 10 from

274   Th (Fig. 4C). In some cases, we even observed two or three individuals from different

275   farming areas clustering together (cluster 1, 2 and 3 in Fig. 4C), suggesting that some OsHV-

276   1 found in these individuals had a common origin. Figure 4C also revealed that most of the

277   25 variants shared by three farming areas were found in all sampled individuals, suggesting

278   that they had not arisen within individual oysters, but had been transmitted horizontally to
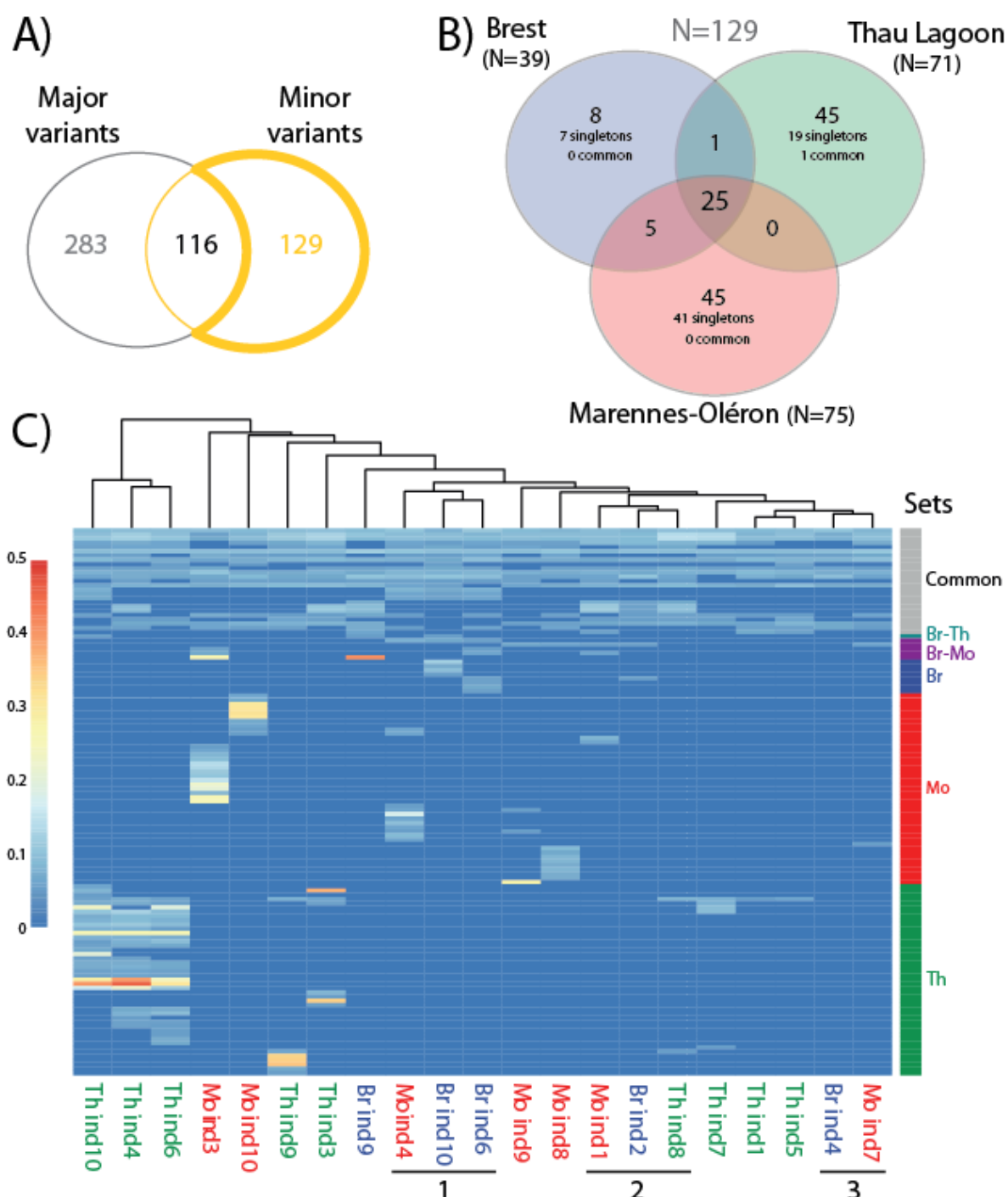
279   other oysters.

280

281

Fig. 4: Variant calling analysis and minor variant distribution in the three farming

areas. (**A**) Quantification of both major and minor variants using a variant calling analysis

14

284 (**B**) Venn diagram summarizing the distribution of the 129 minor variants (frequency <50%)

285 between the three farming areas (Brest (Br) in blue, Marennes-Oléron (Mo) in red, and Thau

286 Lagoon (Th) in green). The number of minor variants within each farming area is indicated in

287 parentheses. Singletons correspond to variants that occur in only one oyster sample, whereas

288 the term "common" corresponds to variants present in all the oyster samples within the

289 dataset. (**C**) Heatmap of the allelic frequency of minor variants across the 21 samples with a

290 gradient from 0 (blue) to 0.5 (red). Hierarchical clustering using the Euclidean distance of the

291 allele frequency is displayed on top. Each line of the heatmap corresponds to a minority

292 variant. Variant sets on the right are derived from the Venn diagram. Samples are indicated

293 on the bottom, and colored according to their geographic origin.

294 **5) Some minor OsHV-1 variants from the Marennes-Oléron farming area are**

295 **major variants in the other two farming areas.**

296 The 116 variants that had variable frequencies across samples (Fig. 4A) were present

297 as minor variants in only four samples: three from Mo (ind3, ind4, and ind10) and one from

298 Br (ind2) (Fig. 5A). Among them, samples Mo ind3 and Mo ind10 contained 114 of these

299 minor variants. Sample Mo ind10 contained 29 minor variants (with a frequency ranging

300 from 5 to 10%; Fig. 5B) that were major variants in samples Mo ind1/3/7/8/9. For instance,

301 the variant at position 65,016 (T>A) of the C-NR-genome had a frequency of 7% in Mo

302 ind10, whereas its frequency varied from 76% to 99% in Mo ind1/3/7/8/9 (Fig. 5C). Similar

303 fluctuations in variant frequency were observed for the other 28 variants (Table S5). These 29

304 variants highlighted a clear genetic link between the Mo ind10 sample and the five Mo

305 samples (ind1/3/7/8/9) located in the other phylogenetic clade (Fig. 5D).

306 Furthermore, Mo ind3 contained 85 minor variants that corresponded to major variants

307 in Mo ind10, 38 that corresponded to major variants in Mo ind4 and all Br samples, and 4

15

308 that corresponded to major variants in all Th samples (Fig. 5 A and E). For instance, the

309 variant at position 57,091 (T>C) in the C-NR-genome had a frequency of 24% in Mo ind3,

310 whereas its frequency was 93% in Mo ind10 and 99% in all Th samples (Fig. 5F, Table S5).

311 Another example is the variant at position 65,624 (C>T) in the C-NR-genome: its frequency

312 grew from 25% in Mo ind3 to 93% in Mo ind10, and 99% in Mo ind4 and all Br samples

313 (Fig. 5F, Table S5). Again, minor variants highlighted genetic links between samples from

314 different farming areas and from phylogenetic clusters 1 and 2 (Fig. 5G, Table S5).

315 Altogether, the analysis of minor variants highlighted links not only between Mo

316 samples and samples from the other two farming areas, but also between Mo samples, which

317 segregated into two different clusters according to whole-genome phylogenetic analyses.

318 These results strongly suggest that all the OsHV-1 variants characterized in this study

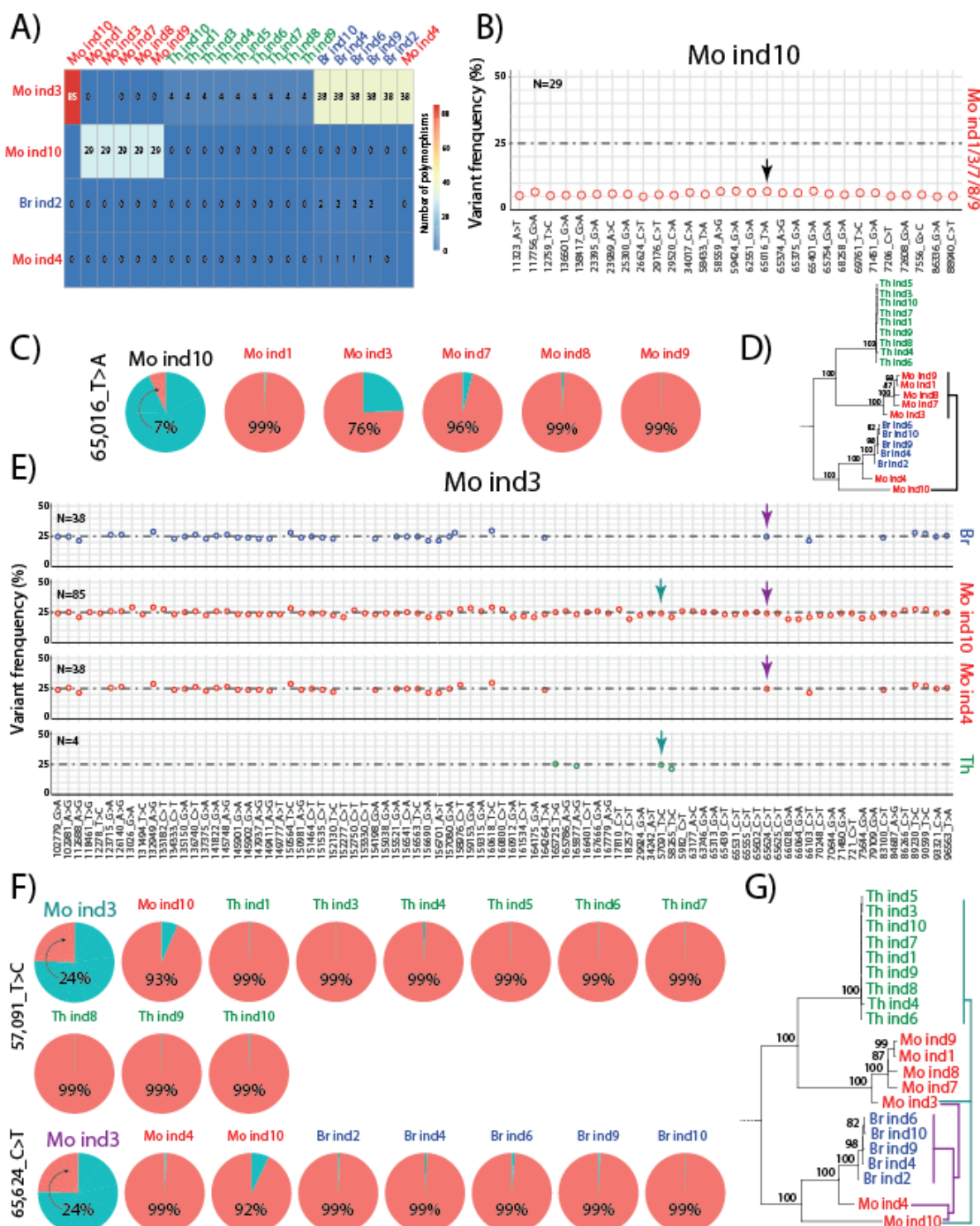319 originated from the Mo farming area.

Fig. 5: Some minor OsHV-1 variants found in the Marennes-Oléron farming area are major variants in the other two farming areas. (**A**) Matrix of the 116 variants which were major or minor, according to the sample. Rows display the four samples in which the 116 variants were minor, and columns show all 21 samples. Intersections give the number of

325 shared variants with minor (row) and major (column) frequency. For self-comparisons, no

326 number is given. (**B**) Dot plot of the 29 minor variants of Mo ind10 which were major

327 variants in Mo ind1/3/7/8/9. Positions on the C-NR-genome and the type of SNPs are

328 indicated on the x-axis. Variant frequency (in %) is indicated in ordinate. Black arrow

329 indicates the variant that is detailed in (C). (**C**) Pie charts of the frequency of the minority

330 variant found at position 65,016 of the NR-genome in Mo ind1/3/7/8/9/10. (**D**) Phylogenetic

331 tree of the 21 *de novo* assembled NR-genomes. The black line shows the link established

332 between Mo ind10 and Mo ind3/4/7/8/9 from variant frequency at position 65,016. (**E**) Dot

333 plot of the 85 minor variants of Mo ind3 that were major variants in Mo ind4/10 and in all

334 samples from Br and Th. Green and turquoise arrows indicate the two variants (57,091_T>C

335 and 65,264_C>T respectively) that are detailed in (F). (**F**) The minor variant found in Mo

336 ind3 at position 57,091 of the C-NR-genome corresponds to major variants in Mo ind10 and

337 all Th samples (upper panel). The minor variant found in Mo ind3 at position 65,624 of the

338 NR-genome corresponds to major variants in Mo ind4/10 and all Br samples (lower panel).

339 (**G**) Phylogenetic tree of the 21 *de novo* assembled NR-genomes. The turquoise lines show

340 the link established between Mo ind3 and Mo ind10 and all Th samples based on the study of

341 variant frequency at position 57,091. The purple lines show the link established between Mo

342 ind3 and Mo ind4/10 and all Br samples based on the study of variant frequency at position

343 65,624.

Discussion

345   The accurate description of viral genetic diversity (major and minor variants) depends

346 on several parameters, such as the quantity and quality of viral genetic material, sequencing

347 quality and depth, as well as the bioinformatics tools used for data analyses (41). In the

348 present study, we characterized OsHV-1 genetic diversity in three oyster-farming areas

18

349   during POMS outbreaks. To do so, oysters from the same cohort were exposed to POMS

350   outbreaks in the field and OsHV-1 genetic diversity was investigated in individual moribund

351   oysters. To optimize the characterization of OsHV-1 variants, we selected moribund oysters

352   with high viral loads. Then, DNA extracted from these oysters was used to prepare

353   sequencing libraries using a PCR-free kit to remove genomic coverage biases associated with

354   PCR amplification. Using this approach, we obtained an average OsHV-1 genome coverage

355   of ≈3000X. However, because OsHV-1 coverage varied among samples (ranging from 480X

356   to 8745X), it was necessary to find a way to verify that coverage was sufficient to accurately

357   quantify viral genomic diversity in each sample. Using a rarefaction analysis, we determined

358   the threshold at which sequencing depth was sufficient to capture all of the genetic diversity

359   present in the studied samples. Although this analysis is commonly used in ecology to assess

360   the species richness in a community (42), it is, to our knowledge, rarely applied to study viral

361   genetic diversity. This information can be crucial when studying viral diseases in which the

362   high diversity of the virus may potentially increase the fitness of the viral population, making

363   it hard to eradicate (43, 44).

364   Having a great sequencing depth was an advantage for OsHV-1 genome assembly;

365   however, the task was complicated by the presence of large repeated sequences,

366   corresponding to 9.1% of the genome of OsHV-1 µVar A. To solve this problem, we used the

367   approach proposed by Morse *et al.* (45): contigs generated by *de novo* assembly were ordered

368   using a reference-based approach to ultimately assemble non-redundant genomes in which

369   only one copy of each repeated element was kept. These non-redundant genomes constituted

370   the keystone of our bioinformatics analyses. Because each of them contained the entire

371   genomic sequence of the virus, we were able to carry out an exhaustive search for intra-

372   individual (sample) genetic variants and characterize the viral population structure. These

373   genomes were also aligned with each other to perform whole-genome phylogeny and

374   genome-wide comparative genomics.

375   Our results revealed that viral populations of OsHV-1 are a heterogeneous set of

376   genomes rather than a single dominating genome, as revealed for another aquatic

377   herpesviruses, e.g. the Cyprinid herpesvirus 3 (4). They also suggest that all OsHV-1

378   characterized during this study share a common ancestor with OsHV-1 µVar characterized by

379   Burioli *et al.* in Normandy (France), reinforcing the idea that genotypes closely related to

380   µVar have replaced the reference genotype and are predominant in oyster-farming areas

381   along the French coasts (35, 46). Our findings also showed that OsHV-1 µVar populations

382   are much more diverse than previously thought. Phylogenetic analyses corroborated results

383   from Delmotte *et al.* showing that OsHV-1 µVar from the Bay of Brest and Thau Lagoon

384   constitute two distinct viral populations (31). Clearly, it is now necessary to change the

385   common practice of relying on the generation of consensus genomes to study the

386   epidemiology of DNA viruses, because they do not accurately depict the actual genetic

387   diversity (5). For this reason, we developed this novel approach, applicable to any other DNA

388   virus model.

389   Our approach demonstrated that OsHV-1 diversity in Marennes-Oléron Bay was

390   higher than in the Bay of Brest and Thau Lagoon. This variability may be related to the fact

391   that the Marennes-Oléron Bay has become one of the largest French oyster-farming areas,

392   with chronic overstocking (47). Moreover, a network analysis of oyster movements revealed

393   that oysters are moved up to 9 times during their production cycle, with peaks of transfers in

394   spring and autumn. These aquacultural practices likely contribute to the spread of OsHV-1

395   infection, and in turn increase its genetic diversity (48).

396   Our approach also made it possible to identify major and minor variants, and variants

397   with frequencies varying dramatically among sampled oyster individuals. Thus, ultra-deep

398   sequencing can support the reconstruction of the true structure of a viral population, and can

399   even highlight potential genetic links between individuals from the same farming area or

400   between individuals from the three farming areas (Fig. 5 D and G, respectively). These

401   genetic links, combined with whole-genome phylogeny, can be used to infer the dynamics of

402   dispersal of OsHV-1 populations across the three farming areas (Fig. 6). Samples from

403   Marennes-Oléron have the most heterogeneous viral populations with the fastest evolution

404   rate. Moreover, major variants characterized in the other two farming areas were also present

405   in Marennes-Oléron, as minor variants. This suggests that Marennes-Oléron farming area is

406   the main source of OsHV-1 viral diversity, from where variants are dispersed to the other

407   farming areas. This hypothesis is in agreement with what is known about oyster transfers.

408   Indeed, Marennes-Oléron is the main area of *C. gigas* spat collection, and juvenile oysters are

409   then transferred to growing sites along the coasts of Brittany, Normandy and also along the

410   Mediterranean (47). Importantly, the three studied sites are separated by long distances (>500

411   km) and therefore are not directly connected through ocean currents. OsHV-1 thus likely

412   disseminates via "travel during transportation of infected oysters, similar to the spread of

413   other diseases during livestock transfer" (*e.g.*, (49)). Viral populations originating from the

414   Marennes-Oléron Bay are then subsequently shaped by local evolutionary and ecological

415   processes in the different farming areas. The next challenge is to characterize these

416   evolutionary and ecological processes that affect OsHV-1 genetic diversity and, in turn, their

417   impact on OsHV-1 virulence in the different farming areas. Given that the fate of new genetic

418   variants is largely determined by host selection, one hypothesis is that oyster populations

419   located in the Bay of Brest and Thau Lagoon exert a bottleneck on viral diversity imported

420   from Marennes-Oléron. However, several studies conducted in Europe to document oyster

421   genetic variability and population structure have revealed a high level of genetic diversity,

422   but no genetic differentiation between French populations (50-52). This lack of

423   differentiation suggests that selection pressure exerted by oysters is not sufficient to explain

424   differences in viral genetic diversity between farming areas. Genetic drift also influences the

425   probability and rate by which alleles increase or decrease in frequency in a viral population;

426   drift results in the loss of genetic diversity, because only a subset of the population

427   contributes to the next generation (41). Furthermore, the relationship between environmental

428   factors and mortality events in Pacific oysters has been well documented (reviewed in (53)).

429   Environmental factors such as temperature, food availability, water quality and salinity or UV

430   radiation have been identified as risk factors that directly or indirectly affect POMS

431   dynamics. Given the variation in these factors among the studied oyster farming areas,

432   identifying the effects of these factors on the structure of OsHV-1 µVar populations may

433   shed new light on the processes of variant selection.

434   Combining data on OsHV-1 viral population structure using our new genomics

435   pipeline with robust epidemiological information from the different farming areas can offer

436   valuable insight into the dynamics of POMS infection (routes of transmission, potential

437   reservoirs) and thus ultimately help implement effective integrated viral disease management
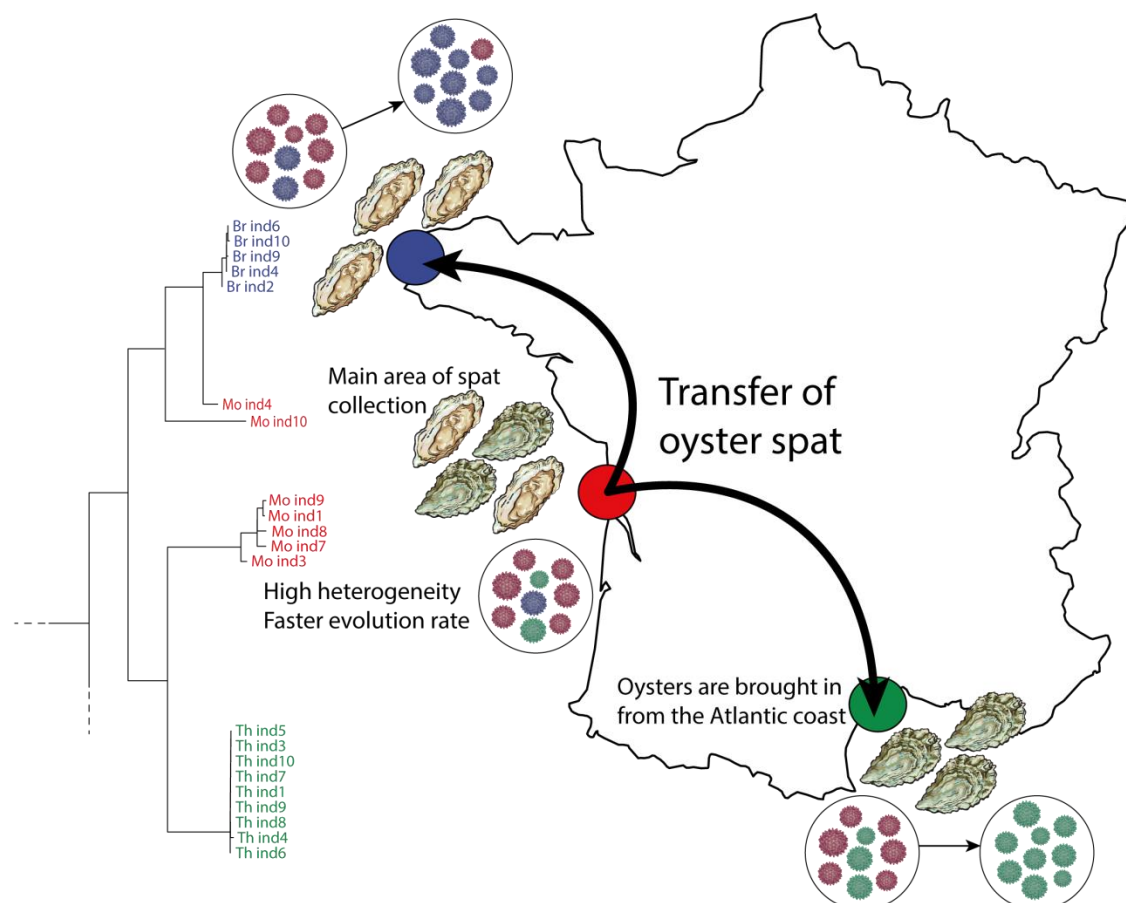
438   strategies (54).

Fig. 6. Graphical representation of OsHV-1 population connectivity in France.

Materials and Methods

OsHV-1 field infection and oyster sampling

All *Crassostrea gigas* oysters used in the present study were produced in August 2017 in the Ifremer experimental facilities located in Argenton (Brittany,France), as specific-pathogen-free (SPF) juveniles (55). To sample OsHV-1 diversity, SPF oysters originating from the same batch were transplanted into three oyster farming areas (~1000 individuals per farm; average weight 1.5 g/oyster) during a disease outbreak when seawater temperature was above 16°C: Thau Lagoon (Th) (Mèze, Lat. 43.378888 long. 3.571111) in May 2018, Marennes-Oléron Bay (Mo) (La Floride lat.: 45.802968 and long.: -1.153421) in June 2018

23

451   and the Bay of Brest (Br) (Logonna-Daoulas, lat.: 48.335263 long.: -4.317922) in July 2018

452   (Fig. 1A). It has been established that massive mortalities can occur as early as 5 days after

453   oyster transplantation in the natural environment (55, 56). For this reason, oysters were

454   transferred back to laboratory facilities 5 days after transplantation. They were then

455   maintained in seawater tanks at 20°C, and moribund oysters were collected daily and stored

456   at -80°C until analysis. A control group consisted of oysters that had not been transplanted in

457   natural environment. Monitoring was stopped after three consecutive days without oyster

458   mortality. During transport, oysters were packed in polystyrene containers and kept moist by

459   covering them with a damp cloth.

460   **DNA extraction, viral load quantification and sequencing**

461   DNA was extracted from moribund oysters using the MagAttract® HMW DNA kit

462   (Qiagen) according to the manufacturer's protocol. DNA purity and concentration were

463   checked using a Nano-Drop ND-1000 spectrometer (Thermo Scientific) and Qubit® dsDNA

464   HS assay kits (Molecular Probes Life Technologies), respectively. Quantification of OsHV-1

465   was carried out using quantitative PCR (qPCR). Amplification reactions were performed

466   using the Roche LightCycler 480 Real-Time thermocycler on three technical replicates

467   (qPHD-Montpellier GenomiX platform, Montpellier University, France). The total qPCR

468   reaction volume was 1.5 µL, consisting of 0.5 µL of DNA (40 ng/µL) and 1 µL of

469   LightCycler 480 SYBR Green I Master mix (Roche) containing 0.5 µM PCR primers

470   (Eurogenetec SA). The primers used were virus-specific and targeted the region of the

471   OsHV-1 genome predicted to encode a catalytic subunit of DNA polymerase (ORF100,

472   AY509253):   Fw-5'-ATTGATGATGTGGATAATCTGTG-3'   and   Rev-5'-

473   GGTAAATACCATTGGTCTTGTTCC-3' (57). The following program was applied: enzyme

474   activation at 95° C for 10 min followed by 40 cycles of denaturation (95°C, 10 s), annealing

475   (60°C, 20 s) and elongation (72°C, 25 s). To check the specificity of the amplification, a

476   subsequent melting step was applied. Twenty-one samples with a viral load greater than or

477   equal to 10^5 genomic units (GU)/ng of DNA were selected for sequencing (5 from Br, 7 from

478   Mo and 9 from Th). DNA-Seq library preparation and sequencing were performed by the

479   Genome Quebec Company (Genome Quebec Innovation Center, McGill University,

480   Montreal, Canada) using the Shotgun PCR-free library preparation kit (Lucigen) and the

481   NovaSeq™ 6000 Sequencing system (Illumina®) (paired ends, 150 bp).

Raw read QC and selection of viral reads

483   Read quality was evaluated using FastQC v0.11.8 (58), and sequence adapters were

484   removed using trimmomatic 0.39 (59). Reads shorter than 50 bp were discarded, and bases at

485   the start and the end of a read were trimmed when their PHRED quality score was below 30.

486   Additionally, reads were clipped if the average quality within a 4-bp sliding window fell

487   below 15. PCR duplicates were removed using Picard v2.22.4 ("Picard toolkit," 2019

488   http://broadinstitute.github.io/picard/). Then, viral reads were extracted using KrakenUniq

489   0.5.8 (60) and Seqkit v0.12.0 (61) with the seq --name --only-id and grep --pattern-file. To

490   avoid any host contamination, extracted reads were aligned to a *C. gigas* reference genome

491   (assembly version V9 (62)) using Bowtie2 4.8.2 (63), and non-aligned viral reads were kept

492   for subsequent analyses.

De novo **assembly of non-redundant OsHV-1 genomes**

494   Viral reads were used for *de novo* assembly using SPAdes with the --meta and --only-

495   assembler options (64). Assembled scaffolds were aligned to the OsHV-1 µVar A genome

496   (GenBank no. KY242785.1) using BLAST 2.9.0 (65) with an e-value of 0.00001. Scaffolds

497   were then extended using SSPACE v3.0 (66). Considering that the OsHV-1 genome is

498   composed of a combination of unique (U) and repeated (R) regions (TRL–UL–IRL–X–IRS–

25

499   US–TRS-X, (34)) and that the latter have a size varying from 160 to 9777 bp, most of these

500   repeat sequences could not be resolved using a kmer approach.

501   Scaffolds were grouped according to their size using Seqkit v0.12.0 (61). The first

502   group of scaffolds had a size between 2 kbp and 4 kbp and made it possible to select the US

503   region. The second group was composed of sequences with a size between 4 kbp and 20 kbp,

504   which made it possible to select the scaffold constituting IRL-X-IRS. Finally, the last group

505   contained sequences greater than 20 kbp, to select the UL region. We used these three groups

506   of scaffolds to construct non-redundant genomes (NR-genomes) that contain only one copy

507   of each repeated region (UL-IRL-X-IRS-US) (45). Because the repeat regions are inverted

508   complements of each other, several adjustments were made to avoid assembling sequences in

509   the wrong orientation. Using this strategy, 21 OsHV-1 NR-genomes were assembled from the

510   21 individual oyster samples.

511   **Whole-genome comparisons and phylogenetic analyses**

512   To assess the relationships between OsHV-1 genomes, the 21 NR-genome assemblies

513   were aligned using MAFFT v7.455 with default parameters (67). A consensus non-redundant

514   genome (C-NR-genome) was subsequently built using the "cons" argument from the

515   EMBOSS suite v6.6.0.0 with the default settings (68). Then the 21 NR-genomes were

516   compared to the C-NR-genome using MUMer4 4.0.0beta2 (69) with the option "nucmer -c

517   100 -l 15 –f" to identify all the variable positions between the 21 NR-genomes. A maximum-

518   likelihood phylogenetic inference was conducted on the 21 NR-genomes and 7 NR-genomes

519   derived from previously published OsHV-1 genomes (AY509253, KU096999, KP412538,

520   KY242785, KY271630, MG561751, MF509813) using IqTree 1.6.7 (70), with 1000

521   bootstrap replications. The NR genome of the abalone herpesvirus (AbHV-1, KU096999)

522   was used as an outgroup. The ModelFinder implementation (71) resulted in the following

523   best-fit model according to Bayesian information criterion: Transversion model + empirical

524   base frequencies + gamma model with 4 rate categories. Results were locally visualized using

525   the R package ggTree v2.0.4 (72).

Variant calling analysis

527   To assess the genomic variability between the farming areas, both minor and major variants

528   of each sequencing library (21) were called on the C-NR-genome previously generated using

529   Freebayes v1.3.2-dirty (73), with the following settings: --use-mapping-quality, --min-repeat-

530   entropy 1, --haplotype-length 0, --min-alternate-count 5, --pooled-continuous, --hwe-priors-

531   off, --allele-balance-priors-off. The resulting variant calling outputs were normalized using

532   BCFtools v- (74), decomposed with vt v1.0.0 (75) and split with vcflib v1.0.0.. Considering

533   the very low occurrence of multi-nucleotide polymorphisms, these latter were counted as

534   InDels. Variable positions with a frequency of >50% were compared to those obtained with

535   MUMer4 for comparative genomics and all were validated. They were subsequently

536   subtracted from the variant calling files to retain only minority variants (with a frequency of

537   <50%). A unique identifier composed of the position and sequence information (e.g.

538   128737_A>G) of each minority variant was created. Set analyses of nucleotide variations

539   were performed with a multiplication matrix using R (76) operation and visualized with the

540   Venn Diagram package (77), UpSetR v1.4.0 (78) or pheatmap v1.0.12.

Downstream analyses

542   All downstream analyses (tables, graphs, plot creation and edition) were performed locally in

543   R 3.6 (76) on R studio IDE (79) with an extensive use of Dplyr v1.0.0 (80), and ggplot2 (81)

544   from the tidyverse v1.3.0 package (82). Genome visualization was carried out using gggenes

545   v0.4.0.

Statistical Analyses

For statistical analyses, the samples were treated as three groups, representing the different farming areas (Br, Mo and Th. Kruskal-Wallis and Dunn's multiple comparison tests were used locally (GraphPad Prism 8.4.2) to compare the number of SNPs or InDels.

Data availability

The datasets generated from this study can be found in the SRA database BioProject accession number PRJNA681599 with submission ID SUB8642385. All the scripts generated in this study are freely available at https://github.com/propan2one/OshV-1-molepidemio. Complementary information is available from the corresponding authors upon reasonable request.

Acknowledgments

JCA, OK, BM, JD and JME were involved in the study conception and design. BM, BP, CP, JD, and JME were involved in the collection of samples and in the experimental

28

577   **Reference**

578   1.   Nkili-Meyong AA, Bigarre L, Labouba I, Vallaeys T, Avarre JC, Berthet N. 2016.

579   Contribution of Next-Generation Sequencing to Aquatic and Fish Virology.

580   Intervirology 59:285-300.

581   2.   Lauring AS, Andino R. 2010. Quasispecies Theory and the Behavior of RNA Viruses.

582   Plos Pathogens 6.

583   3.   Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. 2006. Quasispecies

584   diversity determines pathogenesis through cooperative interactions in a viral

585   population. Nature 439:344-348.

586   4.   Hammoumi S, Vallaeys T, Santika A, Leleux P, Borzym E, Klopp C, Avarre JC.

587   2016. Targeted genomic enrichment and sequencing of CyHV-3 from carp tissues

588   confirms low nucleotide diversity and mixed genotype infections. PeerJ 4:e2516.

589   5.   Renner DW, Szpara ML. 2018. Impacts of Genome-Wide Analyses on Our

590   Understanding of Human Herpesvirus Diversity and Evolution. J Virol 92.

591   6.   Renzette N, Gibson L, Bhattacharjee B, Fisher D, Schleiss MR, Jensen JD, Kowalik

592   TF. 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by

593   demography and positive selection. PLoS Genet 9:e1003735.

594   7.   Renzette N, Gibson L, Jensen JD, Kowalik TF. 2014. Human cytomegalovirus

595   intrahost evolution-a new avenue for understanding and controlling herpesvirus

596   infections. Curr Opin Virol 8:109-15.

597   8.   Renzette N, Pokalyuk C, Gibson L, Bhattacharjee B, Schleiss MR, Hamprecht K,

598   Yamamoto AY, Mussi-Pinhata MM, Britt WJ, Jensen JD, Kowalik TF. 2015. Limits

599   and patterns of cytomegalovirus genomic diversity in humans. Proc Natl Acad Sci U

600   S A 112:E4120-8.

601   9.   Akhtar LN, Bowen CD, Renner DW, Pandey U, Della Fera AN, Kimberlin DW,

602   Prichard MN, Whitley RJ, Weitzman MD, Szpara ML. 2019. Genotypic and

603   Phenotypic Diversity of Herpes Simplex Virus 2 within the Infected Neonatal

604   Population. mSphere 4.

605 10.   Pernet F, Lupo C, Bacher C, Whittington RJ. 2016. Infectious diseases in oyster

606   aquaculture require a new integrated approach. Philos Trans R Soc Lond B Biol Sci

607   371.

608 11.   Stentiford GD, Neil DM, Peeler EJ, Shields JD, Small HJ, Flegel TW, Vlak JM, Jones

609   B, Morado F, Moss S, Lotz J, Bartholomay L, Behringer DC, Hauton C, Lightner DV.

610   2012. Disease will limit future food supply from the global crustacean fishery and

611   aquaculture sectors. J Invertebr Pathol 110:141-57.

612 12.   Burge CA, Shore-Maggio A, Rivlin ND. 2017. Ecology of Emerging Infectious

613   Diseases of Invertebrates. *In* Hajek AE (ed), Ecology of Invertebrate Diseases

614   doi:doi.org/10.1002/9781119256106.ch16. John Wiley & Sons Ltd.

615　13.　Petton B, Destoumieux-Garzon D, Pernet F, Toulza E, de Lorgeril J, Degremont L,

616　　　Mitta G. 2021. The Pacific Oyster Mortality Syndrome, a Polymicrobial and

617　　　Multifactorial Disease: State of Knowledge and Future Directions. Front Immunol

618　　　12:630343.

619　14.　Abbadi M, Zamperin G, Gastaldelli M, Pascoli F, Rosani U, Milani A, Schivo A,

620　　　Rossetti E, Turolla E, Gennari L, Toffan A, Arcangeli G, Venier P. 2018.

621　　　Identification of a newly described OsHV-1 microvar from the North Adriatic Sea

622　　　(Italy). J Gen Virol 99:693-703.

623　15.　Barbieri ES, Medina CD, Vazquez N, Fiorito C, Martelli A, Wigdorovitz A, Schwindt

624　　　E, Morga B, Renault T, Parreno V, Baron PJ. 2019. First detection of Ostreid

625　　　herpesvirus 1 in wild Crassostrea gigas in Argentina. J Invertebr Pathol 166:107222.

626　16.　Barbosa Solomieu V, Renault T, Travers MA. 2015. Mass mortality in bivalves and

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

628　17.　Friedman CS, Estes RM, Stokes NA, Burge CA, Hargove JS, Barber BJ, Elston RA,

629　　　Burreson EM, Reece KS. 2005. Herpes virus in juvenile Pacific oysters Crassostrea

630　　　gigas from Tomales Bay, California, coincides with summer mortality episodes. Dis

631　　　Aquat Organ 63:33-41.

632　18.　Gittenberger A, Voorbergen-Laarman MA, Engelsma MY. 2016. Ostreid herpesvirus

633　　　OsHV-1 mu Var in Pacific oysters Crassostrea gigas (Thunberg 1793) of the Wadden

634　　　Sea, a UNESCO world heritage site. Journal of Fish Diseases 39:105-109.

635　19.　Jenkins C, Hick P, Gabor M, Spiers Z, Fell SA, Gu X, Read A, Go J, Dove M,

636　　　O'Connor W, Kirkland PD, Frances J. 2013. Identification and characterisation of an

637　　　ostreid herpesvirus-1 microvariant (OsHV-1 micro-var) in Crassostrea gigas (Pacific

638　　　oysters) in Australia. Dis Aquat Organ 105:109-26.

639   20.   Moss JA, Burreson EM, Cordes JF, Dungan CF, Brown GD, Wang A, Wu X, Reece
640   KS. 2007. Pathogens in Crassostrea ariakensis and other Asian oyster species:
641   implications for non-native oyster introduction to Chesapeake Bay. Dis Aquat Organ
642   77:207-23.

643   21.   Peeler EJ, Reese RA, Cheslett DL, Geoghegan F, Power A, Thrush MA. 2012.
644   Investigation of mortality in Pacific oysters associated with Ostreid herpesvirus-1 mu
645   Var in the Republic of Ireland in 2009. Preventive Veterinary Medicine 105:136-143.

646   22.   Vasquez-Yeomans R, Garcia-Ortega M, Caceres-Martinez J. 2010. Gill erosion and
647   herpesvirus in Crassostrea gigas cultured in Baja California, Mexico. Dis Aquat
648   Organ 89:137-44.

649   23.   Paul-Pont I, Dhand NK, Whittington RJ. 2013. Spatial distribution of mortality in
650   Pacific oysters Crassostrea gigas: reflection on mechanisms of OsHV-1 transmission.
651   Diseases of Aquatic Organisms 105:127-138.

652   24.   Pernet F, Barret J, Le Gall P, Corporeau C, Dégremont L, Lagarde F, Pépin J-F, Keck
653   N. 2012. Mass mortalities of Pacific oysters Crassostrea gigas reflect infectious
654   diseases and vary with farming practices in the Mediterranean Thau lagoon, France.
655   Acquaculture environmental interactions 2:215-237.

656   25.   Petton B, Pernet F, Robert R, Boudry P. 2013. Temperature influence on pathogen
657   transmission and subsequent mortalities in juvenile Pacific oysters Crassostrea gigas.
658   Aquaculture Environment Interactions 3:257-273.

659   26.   Renault T, Moreau P, Faury N, Pepin JF, Segarra A, Webb S. 2012. Analysis of
660   Clinical Ostreid Herpesvirus 1 (Malacoherpesviridae) Specimens by Sequencing
661   Amplified Fragments from Three Virus Genome Areas. Journal of Virology 86:5942-
662   5947.

663 27. Segarra A, Pepin JF, Arzul I, Morga B, Faury N, Renault T. 2010. Detection and
664   description of a particular Ostreid herpesvirus 1 genotype associated with massive
665   mortality outbreaks of Pacific oysters, Crassostrea gigas, in France in 2008. Virus Res
666   153:92-9.

667 28. de Lorgeril J, Lucasson A, Petton B, Toulza E, Montagnani C, Clerissi C, Vidal-
668   Dupiol J, Chaparro C, Galinier R, Escoubas J-M, Haffner P, Dégremont L, Charrière
669   GM, Lafont M, Delort A, Vergnes A, Chiarello M, Faury N, Rubio TP, Leroy MA,
670   Pérignon A, Régler D, Morga B, Alunno-Bruscia M, Boudry P, Le Roux F,
671   Destoumieux-Garzón D, Gueguen Y, Mitta G. 2018. Immune-suppression by OsHV-1
672   viral infection causes fatal bacteraemia in Pacific oysters. Nature Communications
673   9:4215.

674 29. de Lorgeril J, Petton B, Lucasson A, Perez V, Stenger PL, Degremont L, Montagnani
675   C, Escoubas JM, Haffner P, Allienne JF, Leroy M, Lagarde F, Vidal-Dupiol J,
676   Gueguen Y, Mitta G. 2020. Differential basal expression of immune genes confers
677   Crassostrea gigas resistance to Pacific oyster mortality syndrome. BMC Genomics
678   21:63.

679 30. Dégremont L, Nourry M, Maurouard E. 2015. Mass selection for survival and
680   resistance to OsHV-1 infection in Crassostrea gigas spat in field conditions: response
681   to selection after four generations. Aquaculture 446:111-121.

682 31. Delmotte J, Chaparro C, Galinier R, de Lorgeril J, Petton B, Stenger PL, Vidal-Dupiol
683   J, Destoumieux-Garzon D, Gueguen Y, Montagnani C, Escoubas JM, Mitta G. 2020.
684   Contribution of Viral Genomic Diversity to Oyster Susceptibility in the Pacific Oyster
685   Mortality Syndrome. Frontiers in Microbiology 11:1579.

686   32.   Martenot C, Oden E, Travaille E, Malas JP, Houssin M. 2011. Detection of different
687   variants of Ostreid Herpesvirus 1 in the Pacific oyster, Crassostrea gigas between
688   2008 and 2010. Virus Research 160:25-31.

689   33.   Burioli EA, Prearo M, Riina MV, Bona MC, Fioravanti ML, Arcangeli G, Houssin M.
690   2016. Ostreid herpesvirus type 1 genomic diversity in wild populations of Pacific
691   oyster Crassostrea gigas from Italian coasts. J Invertebr Pathol 137:71-83.

692   34.   Davison AJ, Trus BL, Cheng N, Steven AC, Watson MS, Cunningham C, Le Deuff
693   RM, Renault T. 2005. A novel class of herpesvirus with bivalve hosts. J Gen Virol
694   86:41-53.

695   35.   Burioli EAV, Prearo M, Houssin M. 2017. Complete genome sequence of Ostreid
696   herpesvirus type 1 microVar isolated during mortality events in the Pacific oyster
697   Crassostrea gigas in France and Ireland. Virology 509:239-251.

698   36.   Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. 2008.
699   The genomic and epidemiological dynamics of human influenza A virus. Nature
700   453:615-9.

701   37.   Allicock OM, Lemey P, Tatem AJ, Pybus OG, Bennett SN, Mueller BA, Suchard
702   MA, Foster JE, Rambaut A, Carrington CV. 2012. Phylogeography and population
703   dynamics of dengue viruses in the Americas. Mol Biol Evol 29:1533-43.

704   38.   Theze J, Li T, du Plessis L, Bouquet J, Kraemer MUG, Somasekar S, Yu G, de Cesare
705   M, Balmaseda A, Kuan G, Harris E, Wu CH, Ansari MA, Bowden R, Faria NR, Yagi
706   S, Messenger S, Brooks T, Stone M, Bloch EM, Busch M, Munoz-Medina JE,
707   Gonzalez-Bonilla CR, Wolinsky S, Lopez S, Arias CF, Bonsall D, Chiu CY, Pybus
708   OG. 2018. Genomic Epidemiology Reconstructs the Introduction and Spread of Zika
709   Virus in Central America and Mexico. Cell Host Microbe 23:855-864 e7.

710   39.   Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa

711   JD, Arinaminpathy N, Pepin J, Posada D, Peeters M, Pybus OG, Lemey P. 2014. HIV

712   epidemiology. The early spread and epidemic ignition of HIV-1 in human

713   populations. Science 346:56-61.

714   40.   Martin MA, VanInsberghe D, Koelle K. 2021. Insights from SARS-CoV-2 sequences.

715   Science 371:466-467.

716   41.   Sanjuán R, Domingo-Calap P. 2019. Genetic Diversity and Evolution of Viral

717   Populations. doi:10.1016/b978-0-12-809633-8.20958-8.

718   42.   Chase JM, McGill BJ, McGlinn DJ, May F, Blowes SA, Xiao X, Knight TM,

719   Purschke O, Gotelli NJ. 2018. Embracing scale-dependence to achieve a deeper

720   understanding of biodiversity and its change across communities. Ecol Lett 21:1737-

721   1751.

722   43.   Al Khatib HA, Benslimane FM, Elbashir IE, Coyle PV, Al Maslamani MA, Al-Khal

723   A, Al Thani AA, Yassine HM. 2020. Within-Host Diversity of SARS-CoV-2 in

724   COVID-19 Patients With Variable Disease Severities. Front Cell Infect Microbiol

725   10:575613.

726   44.   Domingo E, Sheldon J, Perales C. 2012. Viral Quasispecies Evolution. Microbiology

727   and Molecular Biology Reviews 76:159-216.

728   45.   Morse AM, Calabro KR, Fear JM, Bloom DC, McIntyre LM. 2017. Reliable

729   Detection of Herpes Simplex Virus Sequence Variation by High-Throughput

730   Resequencing. Viruses 9.

731   46.   Burioli EAV, Varello K, Lavazza A, Bozzetta E, Prearo M, Houssin M. 2018. A

732   novel divergent group of Ostreid herpesvirus 1 muVar variants associated with a

733   mortality event in Pacific oyster spat in Normandy (France) in 2016. J Fish Dis

734   doi:10.1111/jfd.12883.

735   47.   Buestel D, Ropert M, Prou J, Goulletquer P. 2009. HISTORY, STATUS, AND
736   FUTURE OF OYSTER CULTURE IN FRANCE. Journal of Shellfish Research
737   28:813-820.

738   48.   Lupo C, Ezanno P, Arzul I, Garcia C, Jadot C, Joly J-P, Renault T, Bareille N. How
739   network analysis of oyster movements can improve surveillance and control programs
740   of infectious diseases?, p. *In* (ed), Frontiers,

741   49.   Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT,
742   Cornell SJ, Kappey J, Wilesmith J, Grenfell BT. 2001. Dynamics of the 2001 UK foot
743   and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science
744   294:813-7.

745   50.   Lapegue S, Heurtebise S, Cornette F, Guichoux E, Gagnaire PA. 2020. Genetic
746   Characterization of Cupped Oyster Resources in Europe Using Informative Single
747   Nucleotide Polymorphism (SNP) Panels. Genes (Basel) 11.

748   51.   Rohfritsch A, Bierne N, Boudry P, Heurtebise S, Cornette F, Lapegue S. 2013.
749   Population genomics shed light on the demographic and adaptive histories of
750   European invasion in the Pacific oyster, Crassostrea gigas. Evol Appl 6:1064-78.

751   52.   Vendrami DLJ, Houston RD, Gharbi K, Telesca L, Gutierrez AP, Gurney-Smith H,
752   Hasegawa N, Boudry P, Hoffman JI. 2019. Detailed insights into pan-European
753   population structure and inbreeding in wild and hatchery Pacific oysters (Crassostrea
754   gigas) revealed by genome-wide SNP data. Evolutionary Applications 12:519-534.

755   53.   Alfaro AC, Nguyen TV, Merien F. 2018. The complex interactions of Ostreid
756   herpesvirus 1, Vibrio bacteria, environment and host factors in mass mortality
757   outbreaks of Crassostrea gigas. Reviews in Aquaculture 11:1148-1168.

758   54.   Avarre JC. 2017. Editorial: Molecular Tracing of Aquatic Viruses: Where
759   Epidemiology Needs to Meet Genomics. Front Microbiol 8:1498.

760 55. Petton B, Bruto M, James A, Labreuche Y, Alunno-Bruscia M, Le Roux F. 2015.

761   Crassostrea gigas mortality in France: the usual suspect, a herpes virus, may not be

762   the killer in this polymicrobial opportunistic disease. Front Microbiol 6:686.

763 56. Dupont S, Lokmer A, Corre E, Auguet JC, Petton B, Toulza E, Montagnani C,

764   Tanguy G, Pecqueur D, Salmeron C, Guillou L, Desnues C, La Scola B, Bou Khalil J,

765   de Lorgeril J, Mitta G, Gueguen Y, Escoubas JM. 2020. Oyster hemolymph is a

766   complex and dynamic ecosystem hosting bacteria, protists and viruses. Animal

767   Microbiome 2.

768 57. Webb SC, Fidler A, Renault T. 2007. Primers for PCR-based detection of ostreid

769   herpes virus-1 (OsHV-1): Application in a survey of New Zealand molluscs.

770   Aquaculture 272:126-139.

771 58. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.

772   *Available online at:* http://wwwbioinformaticsbabrahamacuk/projects/fastqc.

773 59. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina

774   sequence data. Bioinformatics 30:2114-20.

775 60. Breitwieser FP, Baker DN, Salzberg SL. 2018. KrakenUniq: confident and fast

776   metagenomics classification using unique k-mer counts. Genome Biology 19.

777 61. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for

778   FASTA/Q File Manipulation. PLoS One 11:e0163962.

779 62. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H,

780   Xiong Z, Que H, Xie Y, Holland PW, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C,

781   Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A,

782   Hedgecock D, Xu Z, Liu Y, Domazet-Loso T, Du Y, Sun X, Zhang S, Liu B, Cheng

783   P, Jiang X, Li J, Fan D, Wang W, Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li

784   N, Wang J, et al. 2012. The oyster genome reveals stress adaptation and complexity of

785   shell formation. Nature 490:49-54.

786 63. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat

787   Methods 9:357-9.

788 64. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new

789   versatile metagenomic assembler. Genome Res 5:824-834.

790 65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment

791   search tool. J Mol Biol 215:403-10.

792 66. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-

793   assembled contigs using SSPACE. Bioinformatics 27:578-9.

794 67. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid

795   multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res

796   30:3059-66.

797 68. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology

798   Open Software Suite. Trends Genet 16:276-7.

799 69. Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018.

800   MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol

801   14:e1005944.

802 70. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and

803   effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol

804   Biol Evol 32:268-74.

805 71. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.

806   ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods

807   14:587-589.

808   72.   Yu G, Smith DK, Zhu H, Guan YY, Lam TTY, McInerny G. 2016. GGTREE : an R
809   package for visualization and annotation of phylogenetic trees with their covariates
810   and other associated data. Methods in Ecology and Evolution 8:28-36.

811   73.   Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read
812   sequencing. arXivorg.

813   74.   Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016.
814   BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from
815   next-generation sequencing data. Bioinformatics 32:1749-51.

816   75.   Tan A, Abecasis GR, Kang HM. 2015. Unified representation of genetic variants.
817   Bioinformatics 31:2202-4.

818   76.   R Core Team. 2018. R: A language and environment for statistical computing, R
819   Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

820   77.   Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of highly-
821   customizable Venn and Euler diagrams in R. BMC Bioinformatics 12:35.

822   78.   Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: Visualization
823   of Intersecting Sets. IEEE Trans Vis Comput Graph 20:1983-92.

824   79.   R-Studio-Team. 2015. RStudio: Integrated Development Environment for R., RStudio
825   Inc., Boston, MA.

826   80.   Wickham H, François R, Henry L, Müller K. 2018. dplyr: A Grammar of Data
827   Manipulation, RStudio, https://dplyr.tidyverse.org/reference/dplyr-package.html.

828   81.   Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag,
829   New York,

830   82.   Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G,
831   Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K,

832   Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K,

833   Yutani H. 2019. Welcome to the Tidyverse. Journal of Open Source Software 4:1686.

834