# Phylogeography of infectious disease: genetic diversity and connectivity of the Ostreid herpesvirus 1 populations in France

Running title: Phylogeography of *Ostreid herpesvirus 1*

Jean Delmotte[1], Camille Pelletier[2], Benjamin Morga[2], Richard Galinier[3], Maude Jacquot[2], Bruno Petton[4], Jean-Baptiste Lamy[2], Oliver Kaltz[5], Jean-Christophe Avarre[5], Caroline Montagnani[1]#, Jean-Michel Escoubas[1]#

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

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

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

[4]Ifremer, LEMAR UMR 6539 (Université de Bretagne Occidentale, CNRS, IRD, Ifremer), F-29840 Argenton-en-Landunvez, France

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

[#]Corresponding authors: Caroline.Montagnani@ifremer.fr; Jean.Michel.Escoubas@ifremer.fr

Abstract word count: 211 Importance word count : 152 Text word count: 4104

**Abstract**

The genetic diversity of viral populations is a key to understanding the spatial and temporal diffusion history of viruses, but studying the diversity of whole genomes from natural populations remains a challenge. Phylodynamic approaches are commonly used for RNA viruses harboring small genomes, but have only rarely been applied to DNA viruses with larger genomes. Here, we used the Pacific oyster mortality syndrome (POMS, a disease that affects oyster farms around the world) as a model to study the genetic diversity of its causative agent, the Ostreid herpesvirus 1 (OsHV-1) in the three main French oyster-farming areas. Using ultra-deep sequencing on individual moribund oysters and new bioinformatics methodology, we *de novo* assembled 21 OsHV-1 new genomes. Combining quantification of major and minor variants, phylogenetic analysis and ancestral state reconstruction of discrete traits approaches; we assessed the connectivity of OsHV-1 viral populations between the three oyster-farming areas. Our results suggest that the Marennes-Oléron Bay represents the main source of OsHV-1 diversity, from where the virus has dispersed to other farming areas, a scenario consistent with current practices of oyster transfers in France. Here, we demonstrate that phylodynamic approaches can be applied to marine DNA viruses to determine how epidemiological, immunological, and evolutionary processes act and potentially interact to shape their diversity patterns.

**Importance**

Phylogeography is a field of research that attempts to reconstruct the relationships between individual genotypes within a species and then correlate these genealogical relationships with their geographic and temporal origin. This field of research has become an essential step in the understanding of pandemics, in particular to determine the origin, spread and evolution of

a pathogen as currently illustrated in studies on viral pandemics. However, because phylogeographic analyses are based on genome variation, stable genomes yield less information than labile genomes. Accordingly, viruses with double-stranded DNA (dsDNA) genomes generally have lower nucleotide diversity than RNA viruses. In this study, by combining the use of both major and minor variants with phylogeographic analyses of the oyster herpesvirus OsHV-1, we highlight genealogical relationships that have not been depicted in phylogenetic trees based on consensus viral genomes only. These data offer a plausible scenario reflecting the origin and spread of OsHV-1 populations between oyster-farming sites.

**Keywords:** Herpesvirus; OsHV-1; cupped oyster; phylogeography; non-redundant genomes; genetic diversity; viral spread; minor variants

## Introduction

Viruses are disease agents that have high levels of genetic diversity. This high diversity often means that they lack shared genetic markers, such as ribosomal DNA sequences that are common to all prokaryotes and eukaryotes (Nkili-Meyong et al., 2016), making it difficult to characterize viruses genetically. Furthermore, many viruses — especially those with RNA genomes — are not stable genetic entities, but exist as clouds of many phylogenetically related genetic variants, known as viral quasispecies. This genetic organization currently encumbers our understanding of viral diseases and their evolution and impedes the straightforward characterization of virus population structure (Lauring and Andino, 2010; Vignuzzi et al., 2006). Studies have shown that the level of genetic diversity within these viral populations likely influences viral pathogenicity, dissemination and host immune evasion, and full-length genome analyses are required to identify intra-host viral population structure, reveal molecular traits with epidemiological significance, or detect low-

frequency, but nonetheless relevant viral variants. These difficulties apply to RNA viruses, but also certain large DNA viruses, such as herpesviruses, whose genomic variability rivals that of many RNA viruses (Hammoumi et al., 2016; Renner and Szpara, 2018; Renzette et al., 2013). For example, human cytomegalovirus shows considerable inter-host and intra-host genetic divergence across tissue compartments and times of infection (Renzette et al., 2014; Renzette et al., 2015). In addition, the evolution of a disease is sometimes not fully explained by intrinsic host factors, but by the genotypic diversity of herpesviruses (Akhtar et al., 2019).

Aquaculture is one of the fastest growing food-producing sectors, representing an important animal protein supply for human consumption, with an expanding role in global food security. Today, the biggest threat arising due to the intensification and globalization of aquaculture is infectious diseases. The management and mitigation of the emergence and spread of these infectious diseases are key issues to address to ensure the sustainability of this industry (Burge et al., 2017; Pernet et al., 2016; Stentiford et al., 2012). One illustration is the Pacific oyster mortality syndrome (POMS), which threatens global *Crassostrea gigas* oyster production, a main sector in aquaculture worldwide (reviewed in (Petton et al., 2021)). Since 2008, this syndrome has caused mass mortality in cultivated oysters around the world, from Europe to America and Asia (Abbadi et al., 2018; Barbieri et al., 2019; Barbosa Solomieu et al., 2015; Friedman et al., 2005; Gittenberger et al., 2016; Jenkins et al., 2013; Moss et al., 2007; Peeler et al., 2012; Vasquez-Yeomans et al., 2010). In 2010, the causative agent of these massive mortalities was identified: it is an emerging genotype of a herpes-like virus named Ostreid herpesvirus 1 (OsHV-1) (Paul-Pont et al., 2013; Pernet et al., 2012; Petton et al., 2013; Renault et al., 2012; Segarra et al., 2010). Two major genetic factors seem to affect the severity of POMS: the genetic background of the oysters (de Lorgeril et al., 2018; de Lorgeril et al., 2020; Dégremont et al., 2015) and OsHV-1 genetic diversity (Delmotte et al., 2020; Martenot et al., 2011). Since the characterization of the emergent genotype OsHV-1

µVar, associated with the 2008 high mortality events (Segarra et al., 2010), several variants of the µVar genotype have been described (reviewed in (Barbosa Solomieu et al., 2015)). Interestingly, in 2016, a survey of OsHV-1 genetic diversity carried out on wild *C. gigas* populations along the coasts of Italy demonstrated the high diversity of this virus in natural oyster populations (Burioli et al., 2016). However, most studies on OsHV-1 genetic diversity have been based on PCR molecular markers focusing on a few variable regions of the viral genome, which may not reflect the whole genomic diversity. For instance, in 2017, sequencing of the whole genome of OsHV-1 µVar and its comparison with the reference genome published in 2005 (Davison et al., 2005) showed that the two genomes also differed by the loss or addition of several open reading frames (ORFs), indicating that whole-genome sequencing is necessary to fully reveal viral diversity and to better understand the virus' origin and evolution (Burioli et al., 2017). To date, no study has specifically looked at the links between the geographic distribution of OsHV-1 and its genetic diversity at the level of the whole genome. As far as we know, the entire assembled genomes of only four OsHV-1 infecting *C. gigas* oysters are available, namely the reference genome (AY509253), two OsHV-1 µVar (µVar A KY242785.1 and µVar B KY271630.1) and more recently OsHV-1-PT (MG561751.2) (Abbadi et al., 2018; Burioli et al., 2017; Davison et al., 2005). These four assemblages show significant genomic diversity; however, they tell us very little about the geographic distribution and dissemination of the virus.

As shown, for instance, with the H1N1 influenza virus (Rambaut et al., 2008), the dengue virus (Allicock et al., 2012), the Zika virus (Theze et al., 2018), the AIDS virus (Faria et al., 2014) and more recently SARS-CoV-2 (Martin et al., 2021), molecular epidemiology based on full genomes can be the key to understanding the epidemics of emerging viruses. Therefore, the purpose of the present study was to characterize the genomic diversity of OsHV-1 µVar in natural viral populations both within and between host individuals, to better

understand the dissemination and evolution of OsHV-1 µVar populations and to assess the feasibility of applying phylodynamic and phylogeographic models to OsHV-1 whole genome sequences. Using a deep-sequencing approach conducted at the individual level and novel bioinformatics analyses, we identified 21 new complete genome sequences of OsHV-1 µVar from the three most important oyster-farming areas in France. Phylogenetic analyses combined with comparative genomics and inspection of variant frequencies showed that viral genetic diversity differs greatly not only between areas, but also between individuals within areas. An ancestral state reconstruction of discrete traits approach led us to propose a scenario explaining the phylogenetic relationships between viral populations encountered in the three farming areas. This analysis strategy, combined with epidemiological information, holds promise for understanding OsHV-1 evolution and epidemic dynamics and for identifying transmission patterns between oyster-producing regions. These data are ultimately useful for developing novel disease management strategies.

## Results

### 1) Variability in sequencing depth does not impair the analysis of OsHV-1 diversity.

We used specific-pathogen-free (SPF) *C. gigas* juveniles to sample OsHV-1 diversity during natural infections in the three farming areas (Fig. 1A). The cumulative mortalities for the three batches of oysters were 85%, 81% and 70% for Thau (Th), Marennes-Oléron (Mo) and Brest (Br), respectively. No mortality was observed in the control group.

For each individual oyster, deep sequencing produced from 86.2 to 145.9 M reads per sample (average 109.7 M reads ±16.7 SD, Fig.1B. and Table S1). To evaluate OsHV-1 sequencing depth in each sample, sequencing reads were mapped on the OsHV-1 µVar A genome (KY242785.1) (Burioli et al., 2017). The number of reads mapping to the OsHV-1 µVar A genome greatly varied from one individual to the other, independently of the number

of reads obtained for each sample: from 0.65 M reads (0.66%) for Mo ind3 to 11.94 M reads (13.21%) for Th ind10 (Fig.1B, Table S1). Therefore, OsHV-1 genome coverage varied from 480X to 8745X for Mo ind3 and Th ind10, respectively.

A rarefaction analysis was run to verify that OsHV-1 coverage was sufficient to accurately quantify viral genomic diversity in each sample. For single nucleotide polymorphisms (SNPs) and for insertions-deletions (InDels), rarefaction curves reached a plateau around 300,000 reads, corresponding to an OsHV-1 coverage of 220 X (Fig. 1C). Since the number of mapped reads of each sample was clearly above this threshold, sequencing depth was considered sufficient to capture OsHV-1 genomic diversity. Rarefaction curves indicate that viral diversity was higher in infected Mo oysters than in the other two areas. Comparison of the average number of individual polymorphisms per environment showed that i) the number of SNPs in viruses from Mo (170 ± 27.7) was significantly higher than that of viruses from Br (136.4 ± 5.9; $P \leq 0.001$) and Th (143.8 ± 5.8; $P \leq 0.05$) and ii) the number of InDels in viruses from Mo (35.9 ±3.1 SD) was also significantly higher than that of viruses from Th (25.7 ± 0.9; $P \leq 0.0001$) (Fig. 1D, Table S2). In addition, in Mo, viral diversity greatly varied from one individual to the other; for example Mo ind3 contained 222 SNPs and 42 InDels, whereas Mo ind4 contained only 144 SNPs and 33 InDels.

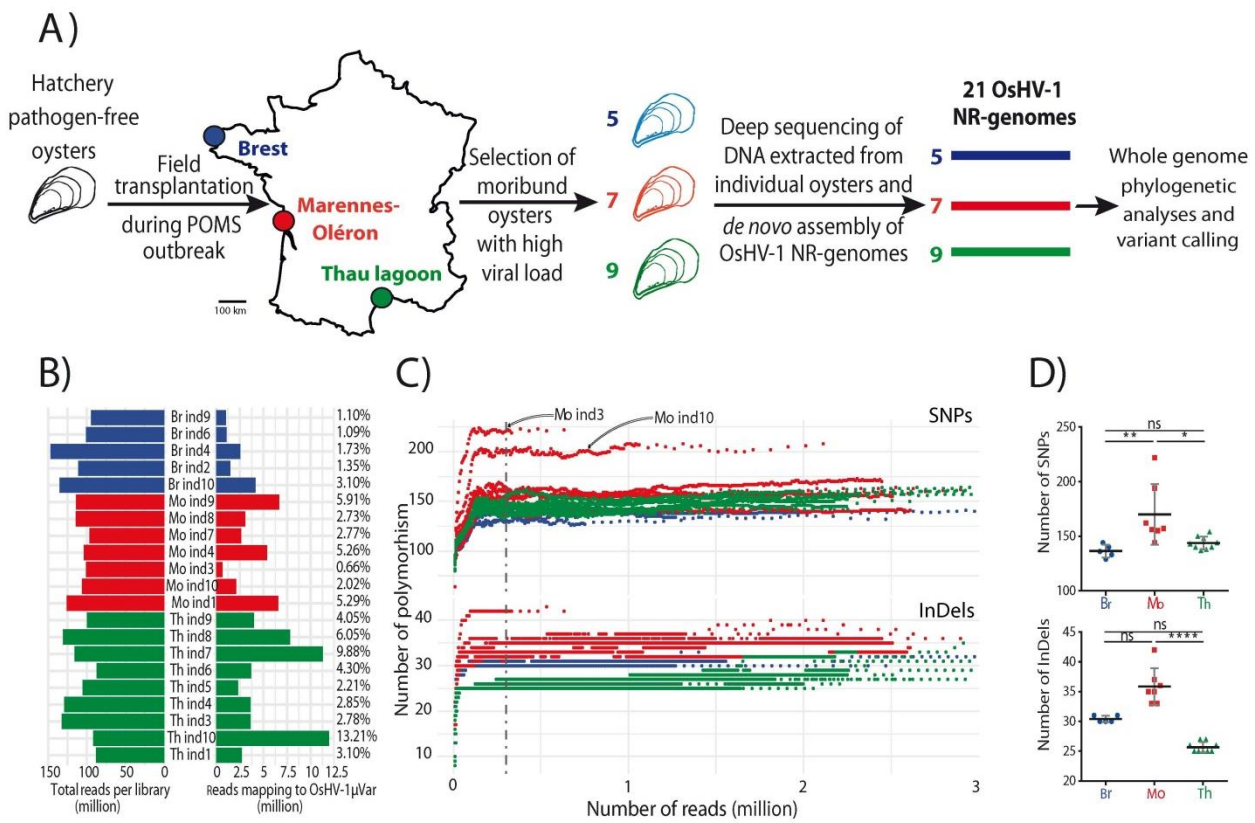


**Fig. 1: Variability in sequencing depth does not impair the analysis of OsHV-1 diversity**. (**A**) Schematic representation of the sampling of OsHV-1 genetic diversity in the three main French oyster farming areas during the Pacific oyster mortality syndrome (POMS) event in 2018 (See Material and Method for more details). Each color represents a farming area, with Brest (Br) in blue (5 individual samples), Marennes-Oléron (Mo) in red (7 individual samples) and Thau Lagoon (Th) in green (9 individual samples). (**B**) Quantification of the number of reads in the library and the number of reads mapping to the OsHV-1 µVar A genome. Numbers on the right side of the figure indicate the percentages of reads mapping to the OsHV-1 µVar A genome. (**C**) Rarefaction curves using the OsHV-1 µVar A genome as the reference sequence. The top panel shows the number of single nucleotide polymorphisms (SNPs) and the bottom panel, the number of insertion-deletions (InDels). The dashed vertical line indicates 0.3 M reads, corresponding to the threshold at which the rarefaction curves reach a plateau. (**D**) Comparison of the number of SNPs (top) and InDels (bottom) between

the three farming areas. Asterisks indicate the *P* value according to Kruskal-Wallis and Dunn's multiple comparison tests (ns: non-significant, $*P \leq 0.05$, $**P \leq 0.01$ $**** P \leq 0.0001$).

**2) Phylogenetic analysis of non-redundant OsHV-1 genomes reveals that all three OsHV-1 populations belong to the µVar genotype.**

We assembled non-redundant (NR) genomes so as to reduce the complexity of the OsHV-1 genome and keep only one copy of each repeated region in the virus sequence. The OsHV-1 µVar A genome contains two unique regions and three repeats: the unique long region (UL, 164,268 base pair (bp)), flanked by two inverted repeats (repeat long [RL], 7338 bp), and the unique short region (US, 3370 bp), flanked by two inverted repeats (repeat short [RS], 9777 bp), and two X regions (1510 bp) (Fig 2A, upper panel). Considering the high sequence identity within each set of repeats (between 99.9 and 100%), removing one of the repeats makes it possible to reliably map the reads from these repeated regions without affecting the overall genome composition. Removing copy of each repeated region led to a NR OsHV-1 µVar A genome (NR-genome) of 186,262 bp, which represents 91% of the full-length genome and contains all the coding sequences (Fig 2A, lower panel). Using this approach, 21 NR-genomes were assembled from the sequencing data generated during this study, as well as 6 other *Ostreid herpesvirus 1* (belonging to the *Malacoherpesviridae* family and *Ostreavirus* genus) published genomes.

A phylogenetic analysis of these 27 NR-genomes revealed that the viruses isolated from *C. gigas*, the 21 new OsHV-1 isolates were phylogenetically closer to OsHV-1 µVar A and B isolates (Fig. 2B) than to the OsHV-1 reference genome (AY509253.2), suggesting that they belong to the µVar genotype. Th and Br isolates grouped into two distinct clusters reflecting their geographic origin (Fig. 2B, clusters 1 and 2). Conversely, Mo OsHV-1 isolates were distributed in these two clusters: five in cluster 1 and two in cluster 2. The

phylogenetic distance within the Th or Br isolates (average 5.33 $10^{-6}$ ± 1.99 $10^{-6}$ SD and 2.16 $10^{-5}$ ± 1.43 $10^{-5}$ SD, respectively) was lower than that observed within the Mo isolates (average 6.16 $10^{-4}$ ± 5.29 $10^{-4}$), indicating greater viral diversity in Mo than in the other two farming areas.

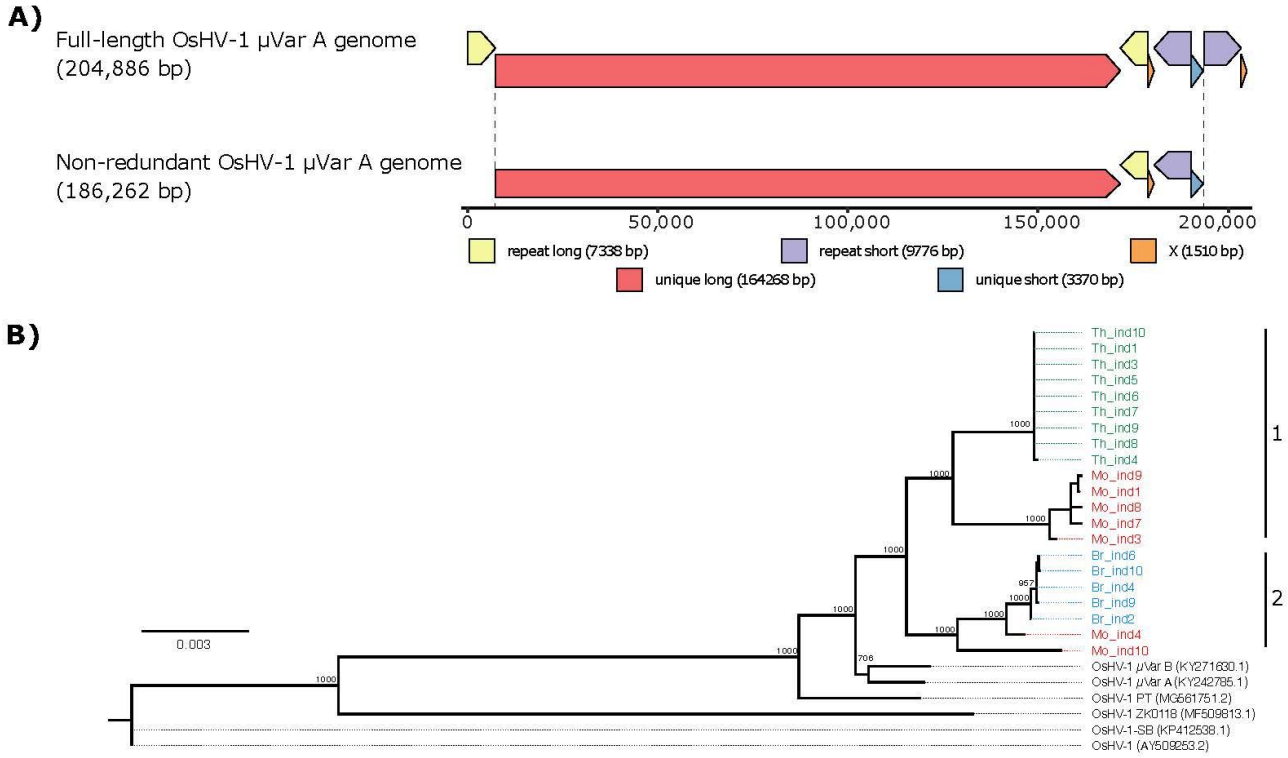


**Fig. 2: Phylogenetic relationship between the 21 *de novo* assembled and other OsHV-1 non-redundant genomes**. (**A**) Full-length (upper panel) and non-redundant (NR) (lower panel) genomes of OsHV-1 µVar A (KY242785). The OsHV-1 µVar A genome consists of two unique regions, referred to as unique long (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 27 OsHV-1 NR-genomes. Bootstrap values were obtained from 1000 replicates. Numbers at branch nodes indicate branch confidence (number 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. The 21 isolates from this study are colored according to their geographic origin, Brest (Br) in blue, Marennes-Oléron (Mo) in red, and Thau Lagoon (Th) in green. Within those, two clusters were defined based on genetic distances (numbered on the right).

### 3) Comparative genomics of the 21 non-redundant genomes confirms that OsHV-1 diversity is higher in Marennes-Oléron than in the other two farming areas.

To compare the genetic diversity at the whole genome scale between the 21 samples, we used the 21 NR-genomes to construct a consensus non-redundant genome (C-NR-genome). Comparison of the 21 NR-genomes to the C-NR-genome identified a total of 399 variable positions (SNPs and InDels), hereafter called variants. The average number of variants per genome was significantly higher in samples from Br (159.8 ± 4.9) than from Th (125 ± 5.6; $P < 0.01$), whereas the average number of variants in Mo (139.4 ± 20.8) was not significantly different from the other two farming areas (Fig. 3A, Table S3).

However, viruses from Mo showed the highest diversity, with 321 variants accounting for 80.5% of the total diversity. Among the variants, 151 (47%) were specific to Mo, and 87 were observed in one sample only (singletons) (Fig. 3B, Table S3). None of these 151 variants were shared by all the samples from this area. The majority of Mo singleton variants were observed in two samples: ind10 (49) and ind4 (13) (Fig. 3C). For the Th and Br farming areas, 134 and 168 variants were identified, among which 51 and 24 were area-specific, respectively (Fig. 3B, Table S3). In contrast to Mo, 70.8% (17/24) and 94.1% (48/51) of the variants were common to all the samples of Br and Th, respectively (Fig. 3B, Table S3).

The three variants common to Br and Th were present in almost all individuals of the two farming areas (Fig. 3C, green arrows). Namely, among the 90 variants shared by Br and Mo, 78 (86.7%) were shared by two Mo individuals (ind10 and ind4) and all Br individuals (Fig. 3C, purple arrow). These two Mo individuals also shared six variants with all Th individuals (Fig. 3C, red arrows).

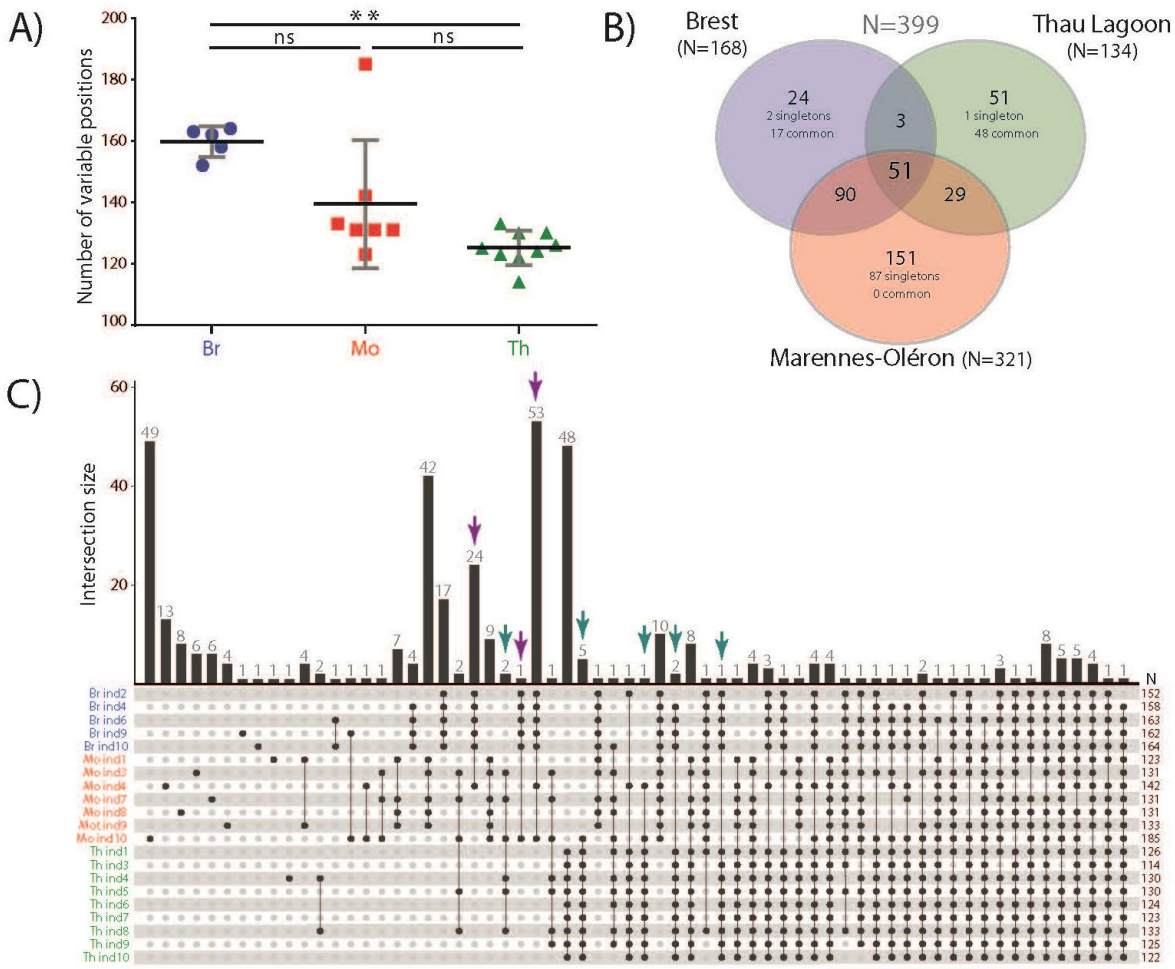


**Fig. 3: Comparative genomics of the 21 non-redundant genomes of OsHV-1 isolated from three oyster farming areas**. (**A**) Distribution of the 399 variable positions (hereafter named variants) between the 21 individuals and the three farming areas. Significance was calculated according to Kruskal-Wallis 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-) genomes. The bottom panel shows the specific combinations (or intersections), and the vertical bars indicate the number of variants within these combinations. Colored arrows on top indicate variants shared by almost all individuals from Br and Th (green arrows) and variants shared by two individuals from Mo (ind4 and ind10) and all individuals from Br (purple arrows) or all individuals from Th (red arrows). Numbers (N) on the right side indicate the number of variants per sample.

### 4) Diversity of minor OsHV-1 variants is higher in Marennes-Oléron than in the other two farming areas.

To fully characterize OsHV-1 genetic diversity, we performed variant calling analyses, using the C-NR-genome as the reference, to quantify both major (frequency >50%) and minor (frequency <50%) variants in the 21 individual samples. Among the 399 variants identified as major variants, 117 had variable frequencies (above or below 50%) across samples; moreover, 129 additional minor variants were identified (Fig. 4A).

Frequencies of the 129 strictly minor variants ranged from 5.1% to 45.9% (mean = 8.75% ± 5.75% SD). The number of minor variants was lower in oysters from Br (39) than in oysters from Th (71) or Mo (75) (Fig. 4B, Table S4). Twenty-five of these 129 minority variants (19.3%) were found in all three farming areas. Moreover, the majority of site-specific variations were found in only one individual oyster (singletons).

A hierarchical clustering analysis applied to the minor variants showed that individuals did not cluster according to their geographic origin, except for individuals 4, 6 and 10 from Th (Fig. 4C). In some cases, we even observed two or three individuals from different

farming areas clustering together (cluster 1, 2 and 3 in Fig. 4C), suggesting that some OsHV-1 found in these individuals had a common origin. Figure 4C also revealed that most of the 25 variants shared by three farming areas were found in all sampled individuals, suggesting that they had not arisen within individual oysters, but had been transmitted horizontally to other oysters.

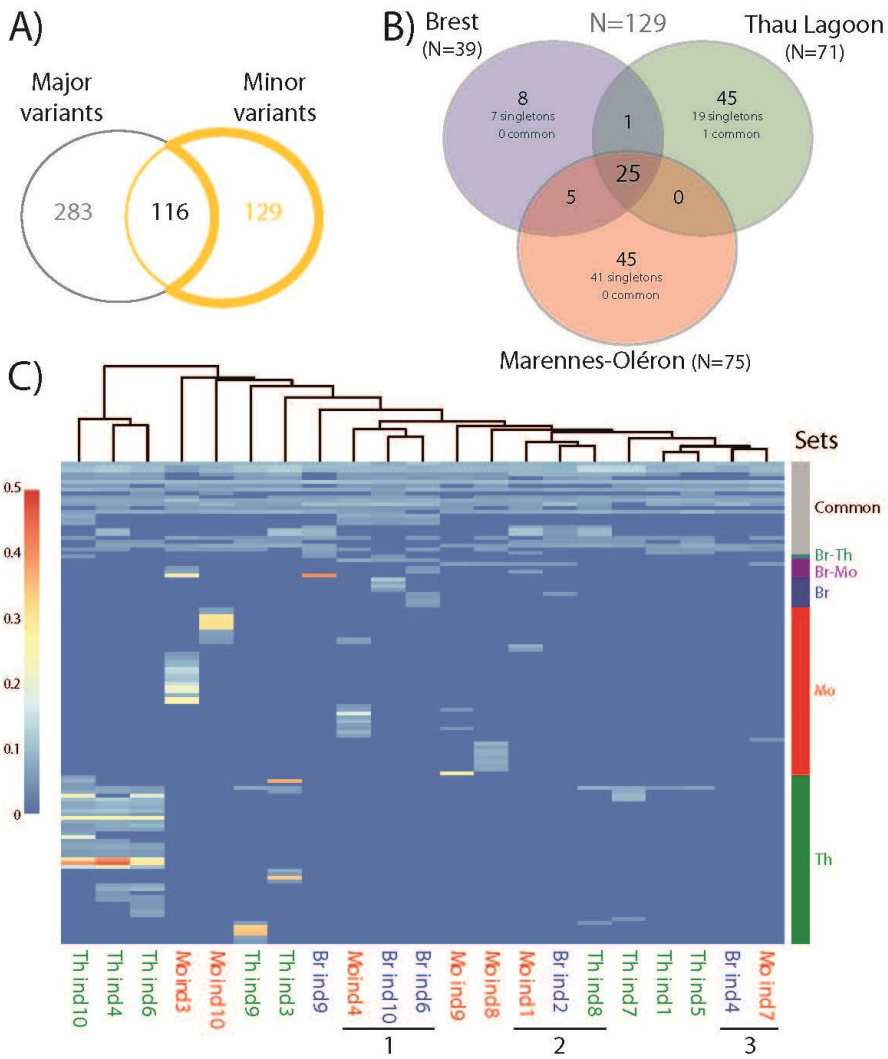


**Fig. 4: Variant calling analysis and minor variant distribution in the three farming areas.** (**A**) Quantification of both major and minor variants using a variant calling analysis (**B**) Venn diagram summarizing the distribution of the 129 minor variants (frequency <50%) between the three farming areas (Brest (Br) in blue, Marennes-Oléron (Mo) in red, and Thau

Lagoon (Th) in green). The number of minor variants within each farming area is indicated in parentheses. Singletons correspond to variants that occur in only one oyster sample, whereas the term "common" corresponds to variants present in all the oyster samples within the dataset. (**C**) Heatmap of the allelic frequency of minor variants across the 21 samples with a gradient from 0 (blue) to 0.5 (red). Hierarchical clustering using the Euclidean distance of the allele frequency is displayed on top. Each line of the heatmap corresponds to a minority variant. Variant sets on the right are derived from the Venn diagram. Samples are indicated on the bottom, and colored according to their geographic origin.

**5) Some minor OsHV-1 variants from the Marennes-Oléron farming area are major variants in the other two farming areas.**

The 116 variants that had variable frequencies across samples (Fig. 4A) were present as minor variants in only four samples: three from Mo (ind3, ind4, and ind10) and one from Br (ind2) (Fig. 5A). Among them, samples Mo ind3 and Mo ind10 contained 114 of these minor variants. Sample Mo ind10 contained 29 minor variants (with a frequency ranging from 5 to 10%; Fig. 5B) that were major variants in samples Mo ind1/3/7/8/9. For instance, the variant at position 65,016 (T>A) of the C-NR-genome had a frequency of 7% in Mo ind10, whereas its frequency varied from 76% to 99% in Mo ind1/3/7/8/9 (Fig. 5C). Similar fluctuations in variant frequency were observed for the other 28 variants (Table S5). These 29 variants highlighted a clear genetic link between the Mo ind10 sample and the five Mo samples (ind1/3/7/8/9) located in the other phylogenetic clade (Fig. 5D).

Furthermore, Mo ind3 contained 85 minor variants that corresponded to major variants in Mo ind10, 38 that corresponded to major variants in Mo ind4 and all Br samples, and 4 that corresponded to major variants in all Th samples (Fig. 5 A and E). For instance, the variant at position 57,091 (T>C) in the C-NR-genome had a frequency of 24% in Mo ind3, whereas its frequency was 93% in Mo ind10 and 99% in all Th samples (Fig. 5F, Table S5).

Another example is the variant at position 65,624 (C>T) in the C-NR-genome: its frequency grew from 25% in Mo ind3 to 93% in Mo ind10, and 99% in Mo ind4 and all Br samples (Fig. 5F, Table S5). Again, minor variants highlighted genetic links between samples from different farming areas and from phylogenetic clusters 1 and 2 (Fig. 5G, Table S5).

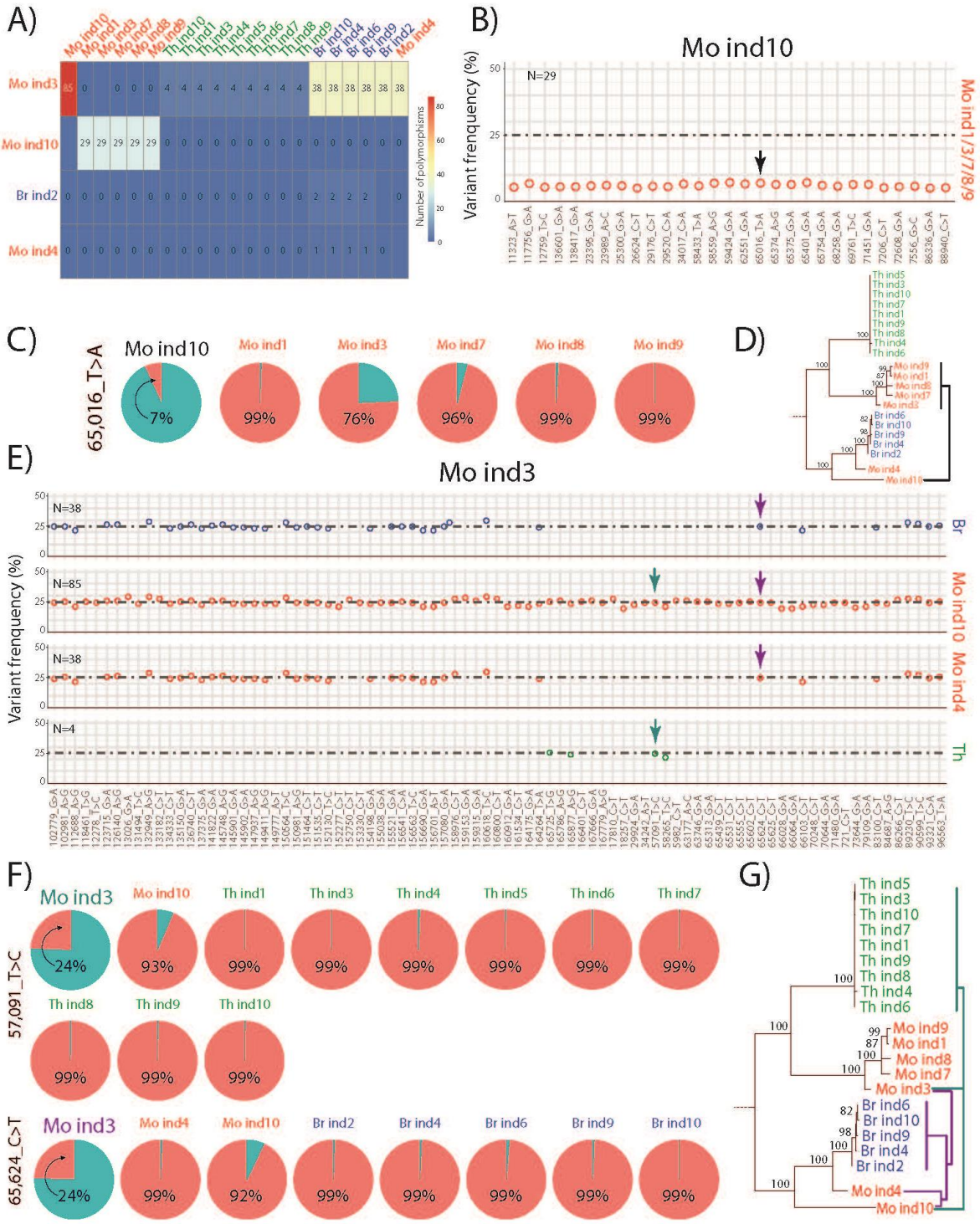

**Fig. 5: Some minor OsHV-1 variants found in the Marennes-Oléron farming area are major variants in the other two farming areas.** (**A**) Matrix of the 116 variants which were major or minor, according to the sample. Rows display the four samples in which the 116 variants were minor, and columns show all 21 samples. Intersections give the number of shared variants with minor (row) and major (column) frequency. For self-comparisons, no number is given. (**B**) Dot plot of the 29 minor variants of Mo ind10 which were major variants in Mo ind1/3/7/8/9. Positions on the C-NR-genome and the type of SNPs are indicated on the x-axis. Variant frequency (in %) is indicated in ordinate. Black arrow indicates the variant that is detailed in (C). (**C**) Pie charts of the frequency of the minority variant found at position 65,016 of the NR-genome in Mo ind1/3/7/8/9/10. (**D**) Phylogenetic tree of the 21 *de novo* assembled NR-genomes. The black line shows the link established between Mo ind10 and Mo ind3/4/7/8/9 from variant frequency at position 65,016. (**E**) Dot plot of the 85 minor variants of Mo ind3 that were major variants in Mo ind4/10 and in all samples from Br and Th. Green and turquoise arrows indicate the two variants (57,091_T>C and 65,264_C>T respectively) that are detailed in (F). (**F**) The minor variant found in Mo ind3 at position 57,091 of the C-NR-genome corresponds to major variants in Mo ind10 and all Th samples (upper panel). The minor variant found in Mo ind3 at position 65,624 of the NR-genome corresponds to major variants in Mo ind4/10 and all Br samples (lower panel). (**G**) Phylogenetic tree of the 21 *de novo* assembled NR-genomes. The turquoise lines show the link established between Mo ind3 and Mo ind10 and all Th samples based on the study of variant frequency at position 57,091. The purple lines show the link established between Mo ind3 and Mo ind4/10 and all Br samples based on the study of variant frequency at position 65,624.

### 6) Ancestral state reconstruction of discrete trait

The analysis of minor variants highlighted links not only between Mo samples and samples from the other two farming areas, but also between Mo samples, which segregated into two different clusters according to whole-genome phylogenetic analyses. These results strongly suggest that all the OsHV-1 variants characterized in this study originated from the Mo farming area. This is congruent with ancestral state reconstruction of the discrete locations (Fig. 6A) that showed evidence for Mo being the ancestral location of the 21 OsHV-1 isolates sequenced. Moreover, the Bayesian stochastic search variable selection approach suggested significant transition rates of OsHV-1 between the three sampling sites. Indeed, two rates were inferred with decisive supports: from Mo to Th (BF=920) and from Mo to Br (BF=11,051) (Fig. 6B).

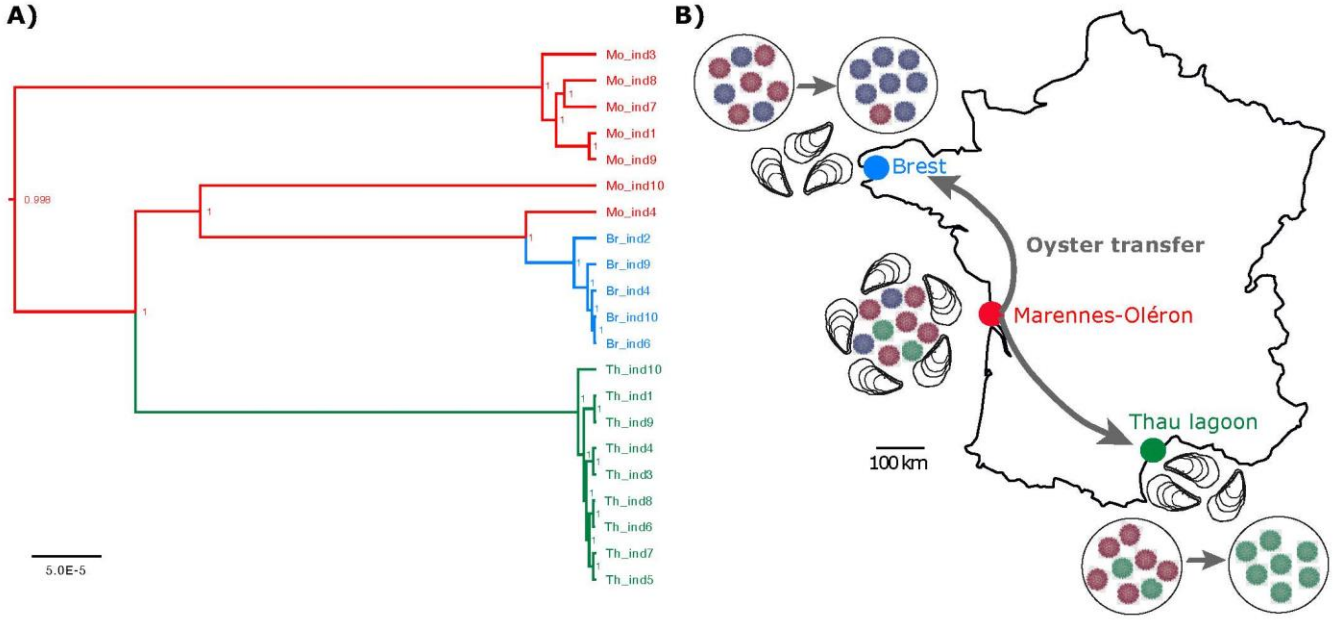


**Fig. 6: Graphical representation of OsHV-1 population connectivity in France.**

**(A)** Maximum clade credibility (MCC) phylogeny of NR-genomes of OsVH-1 isolates and ancestral state reconstruction of the discrete locations: Brest (Br), Marennes-Oléron (Mo) and Thau lagoon (Th). Color of the branches represents the most probable location state of their descendent nodes. The color-coding is consistent with other figures: Br in blue, Mo in red, and Th in green. Numbers at branch nodes indicate the posterior location probabilities. Given

that we did not incorporate any timing information into our analysis (no calibration prior), the scale beneath the tree is expressed in average number of substitutions per time unit. (**B)** Bayesian stochastic search variable selection (BSSVS) was used to infer dispersal routes with none-zero rates using Bayes factors. Routes that were inferred to have decisive support (Bayes Factor > 100) are plotted on map (grey arrows).

## Discussion

The accurate description of viral genetic diversity (major and minor variants) depends on several parameters, such as the quantity and quality of viral genetic material, sequencing quality and depth, as well as the bioinformatics tools used for data analyses (Sanjuán and Domingo-Calap, 2019). In the present study, we characterized OsHV-1 genetic diversity in three oyster-farming areas during POMS outbreaks. To do so, oysters from the same cohort were exposed to POMS outbreaks in the field and OsHV-1 genetic diversity was investigated in individual moribund oysters. To optimize the characterization of OsHV-1 variants, we selected moribund oysters with high viral loads. Then, DNA extracted from these oysters was used to prepare sequencing libraries using a PCR-free kit to remove genomic coverage biases associated with PCR amplification. Using this approach, we obtained an average OsHV-1 genome coverage of ≈3000X. However, because OsHV-1 coverage varied among samples (ranging from 480X to 8745X), it was necessary to find a way to verify that coverage was sufficient to accurately quantify viral genomic diversity in each sample. Using a rarefaction analysis, we determined the threshold at which sequencing depth was sufficient to capture all of the genetic diversity present in the studied samples. Although this analysis is commonly used in ecology to assess the species richness in a community (Chase et al., 2018), it is, to our knowledge, rarely applied to study viral genetic diversity. This information can be crucial when studying viral diseases in which the high diversity of the virus may potentially increase

the fitness of the viral population, making it hard to eradicate (Al Khatib et al., 2020; Domingo et al., 2012).

Having a great sequencing depth was an advantage for OsHV-1 genome assembly; however, the task was complicated by the presence of large repeated sequences, corresponding to 9.1% of the genome of OsHV-1 µVar A. To solve this problem, we used the approach proposed by Morse *et al.* (Morse et al., 2017): contigs generated by *de novo* assembly were ordered using a reference-based approach to ultimately assemble non-redundant genomes in which only one copy of each repeated element was kept. These non-redundant genomes constituted the keystone of our bioinformatics analyses. Because each of them contained the entire genomic sequence of the virus, we were able to carry out an exhaustive search for intra-individual (sample) genetic variants and characterize the viral population structure. These genomes were also aligned with each other to perform whole-genome phylogeny and genome-wide comparative genomics.

Our results revealed that viral populations of OsHV-1 are a heterogeneous set of genomes rather than a single dominating genome, as revealed for another aquatic herpesviruses, e.g. the Cyprinid herpesvirus 3 (Hammoumi et al., 2016). They also suggest that all OsHV-1 characterized during this study share a common ancestor with OsHV-1 µVar characterized by Burioli *et al.* in Normandy (France), reinforcing the idea that genotypes closely related to µVar have replaced the reference genotype and are predominant in oyster-farming areas along the French coasts (Burioli et al., 2017; Burioli et al., 2018). Our findings also showed that OsHV-1 µVar populations are much more diverse than previously thought. This is congruent with results from Delmotte *et al.* showing that OsHV-1 µVar from the Bay of Brest and Thau Lagoon constitute two distinct viral populations (Delmotte et al., 2020) as well as with the recent relatively high estimate of OsHV-1 evolutionary rate (Morga et al., 2021). We therefore argue it is now necessary to combine common approaches relying on the

generation and the analysis of consensus genomes to intra host diversity characterization in order to accurately study the epidemiology of DNA viruses (Renner and Szpara, 2018).

Here, we identified major and minor variants within viral populations. Variants frequencies vary dramatically among sampled oyster individuals. Thus, ultra-deep sequencing can support the reconstruction of the true structure of a viral population, and can inform on relationships between individuals from the same farming area or between individuals from the three farming areas (Fig. 5 D and G, respectively). Ancestral state reconstruction, used to infer the dynamics of dispersal of OsHV-1 populations across the three farming areas (Fig. 6 A), along with the distribution of major and minor variants, suggest that Marennes-Oléron farming area is the main source of OsHV-1 viral diversity, from where variants are dispersed to the other farming areas. This hypothesis is in agreement with what is known about oyster transfers. Indeed, Marennes-Oléron is the main area of *C. gigas* spat collection, and juvenile oysters are then transferred to growing sites along the coasts of Brittany, Normandy and also along the Mediterranean (Buestel et al., 2009) (Fig. 6 B). Importantly, the three studied sites are separated by long distances (>500 km) and therefore are not directly connected through ocean currents. OsHV-1 thus likely disseminates via “travel during transportation of infected oysters, similar to the spread of other diseases during livestock transfer” (*e.g.*, (Keeling et al., 2001)).

OsHV-1 diversity found in Marennes-Oléron Bay was higher than in the Bay of Brest and Thau Lagoon. This may be related to the fact that the Marennes-Oléron Bay has become one of the largest French oyster-farming areas, with intensive shellfish cultivation and chronic overstocking (Buestel et al., 2009). While we acknowledged the potential bias introduced here by low sampling size, this is coherent with large viral population size and increased opportunities for recombination. Mutation and recombination rates being good prognostic markers for possible emergence of phenotypically different viruses, studying them

at the different studied areas would help to better prevent future new OsHV-1 variant outbreaks. Moreover, a network analysis of oyster movements revealed that oysters are moved up to 9 times during their production cycle, with peaks of transfers in spring and autumn. These aquacultural practices likely contribute to the spread of OsHV-1, and in turn increase its genetic diversity (Lupo et al., 2016).

Given that the fate of new genetic variants is largely determined by host selection, one hypothesis to explain the lower diversity in the Bay of Brest and Thau Lagoon is that oyster populations there exert a bottleneck on viral diversity imported from Marennes-Oléron. However, several studies conducted in Europe to document oyster genetic variability and population structure have revealed a high level of genetic diversity, but no genetic differentiation between French populations (Lapegue et al., 2020; Rohfritsch et al., 2013; Vendrami et al., 2019). This lack of differentiation suggests that selection pressure exerted by oysters is not sufficient to explain differences in viral genetic diversity between farming areas. Genetic drift also influences the probability and rate by which alleles increase or decrease in frequency in a viral population; drift results in the loss of genetic diversity, because only a subset of the population contributes to the next generation (Sanjuán and Domingo-Calap, 2019). Furthermore, the relationship between environmental factors and mortality events in Pacific oysters has been well documented (reviewed in (Alfaro et al., 2018)). Environmental factors such as temperature, food availability, water quality and salinity or UV radiation have been identified as risk factors that directly or indirectly affect POMS dynamics. Given the variation in these factors among the studied oyster farming areas, identifying the effects of these factors on the structure of OsHV-1 µVar populations may shed new light on the processes of variant selection. Therefore, a larger and well-designed OsHV-1 study should focus on characterizing the epidemiological and evolutionary processes that shape OsHV-1 genetic diversity patterns and on identifying the ecological,

environmental and anthropogenic factors influencing OsHV-1 dispersal using a combination of phylogeography and spatial epidemiology approaches (Dellicour et al., 2016; Pybus et al., 2012).

Combining data on OsHV-1 viral population structure using our new genomics pipeline with robust epidemiological information from the different farming areas can offer valuable insight into the dynamics of POMS infection (routes of transmission, potential reservoirs) and thus ultimately help implement effective integrated viral disease management strategies (Avarre, 2017).

## 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 (Petton et al., 2015). 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: 1) the Thau Lagoon (Th) (Mèze, lat. 43.379 long. 3.571) in May 2018, 2) the Marennes-Oléron Bay (Mo) (La Floride, lat.: 45.803 and long.: -1.153) in June 2018 and 3) the Bay of Brest (Br) (Logonna-Daoulas, lat.: 48.335 long.: -4.318) in July 2018 (Fig. 1A). It has been established that massive mortalities can occur as early as 5 days after oyster transplantation in the natural environment (Dupont et al., 2020; Petton et al., 2015). For this reason, oysters were transferred back to laboratory facilities 5 days after transplantation. They were then maintained in seawater tanks at 20°C, and moribund oysters were collected daily and stored at -80°C until analysis. A control group was composed of oysters that had not been transplanted in natural environment. Monitoring was stopped after

three consecutive days without any oyster mortality. During transport, oysters were packed in polystyrene containers and kept moist by covering them with a damp cloth.

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

DNA was extracted from individual moribund oysters using the MagAttract® HMW DNA kit (Qiagen) according to the manufacturer's protocol. DNA purity and concentration were checked using a Nano-Drop ND-1000 spectrometer (Thermo Scientific) and Qubit® dsDNA HS assay kits (Molecular Probes Life Technologies), respectively. Quantification of OsHV-1 was carried out using quantitative PCR (qPCR). Amplification reactions were performed using the Roche LightCycler 480 Real-Time thermocycler on three technical replicates (qPHD-Montpellier GenomiX platform, Montpellier University, France). The total qPCR reaction volume was 1.5 µL, comprising 0.5 µL of DNA (40 ng/µL) and 1 µL of LightCycler 480 SYBR Green I Master mix (Roche) containing 0.5 µM PCR primers (Eurogenetec SA). The primers used were virus-specific and targeted the region of the OsHV-1 genome predicted to encode a catalytic subunit of DNA polymerase (ORF100, AY509253): Forward: 5'-ATTGATGATGTGGATAATCTGTG-3' and Reverse: 5'-GGTAAATACCATTGGTCTTGTTCC-3' (Webb et al., 2007). The following program was applied: enzyme activation at 95° C for 10 min followed by 40 cycles of denaturation (95°C, 10 s), annealing (60°C, 20 s) and elongation (72°C, 25 s). To check the specificity of the amplification, a subsequent melting step was applied. Twenty-one samples with a viral load greater than or equal to $10^5$ genomic units (GU)/ng of DNA were selected for sequencing (5 from Br, 7 from Mo and 9 from Th). DNA-Seq library preparation and sequencing were performed by the Genome Quebec Company (Genome Quebec Innovation Center, McGill University, Montreal, Canada) using the Shotgun PCR-free library preparation kit (Lucigen) and the NovaSeq™ 6000 Sequencing system (Illumina®) (paired ends, 150 bp).

**Raw read QC and selection of viral reads**

Read quality was assessed using FastQC v0.11.8 (Andrews, 2010), and sequence adapters were removed using trimmomatic 0.39 (Bolger et al., 2014). Reads shorter than 50 bp were discarded, and bases at the start and the end of a read were trimmed when their PHRED quality score was below 30. Additionally, reads were clipped if the average quality within a 4-bp sliding window fell below 15. PCR duplicates were removed using Picard v2.22.4 ("Picard toolkit," 2019 http://broadinstitute.github.io/picard/). Then, viral reads were extracted using KrakenUniq 0.5.8 (Breitwieser et al., 2018) and Seqkit v0.12.0 (Shen et al., 2016) with the seq --name --only-id and grep --pattern-file. To avoid any host contamination, extracted reads were aligned to a *C. gigas* reference genome (assembly version V9 (Zhang et al., 2012)) using Bowtie2 4.8.2 (Langmead and Salzberg, 2012), and non-aligned viral reads were kept for subsequent analyses.

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

Viral reads were used for *de novo* assembly using SPAdes with the --meta and --only-assembler options (Nurk et al., 2017). Assembled scaffolds were aligned to the OsHV-1 µVar A genome (GenBank no. KY242785.1) using BLAST 2.9.0 (Altschul et al., 1990) with an e-value of 0.00001. Scaffolds were then extended using SSPACE v3.0 (Boetzer et al., 2011). Considering that the OsHV-1 genome is composed of a combination of unique (U) and repeated (R) regions (TRL–UL–IRL–X–IRS–US–TRS-X, (Davison et al., 2005)) and that the latter have a size varying from 160 to 9777 bp, most of these repeat sequences could not be resolved using a kmer approach.

Scaffolds were grouped according to their size using Seqkit v0.12.0 (Shen et al., 2016). The first group of scaffolds had a size between 2 kbp and 4 kbp and made it possible to select the US region. The second group was composed of sequences with a size between 4 kbp and 20 kbp, which made it possible to select the scaffold constituting IRL-X-IRS.

Finally, the last group contained sequences greater than 20 kbp, to select the UL region. We used these three groups of scaffolds to construct non-redundant genomes (NR-genomes) that contain only one copy of each repeated region (UL-IRL-X-IRS-US) (Morse et al., 2017). Because the repeat regions are inverted complements of each other, several adjustments were made to avoid assembling sequences in the wrong orientation. Using this strategy, 21 OsHV-1 NR-genomes were assembled from the 21 individual oyster samples.

**Whole-genome comparisons and phylogenetic analyses**

The 21 NR-genomes were then aligned with the 6 NR-genomes derived from previously published OsHV-1 genomes (AY509253, KU096999, KP412538, KY242785, KY271630, MG561751, MF509813) using MAFFT v7.475 with default parameters (Katoh et al., 2002). A maximum-likelihood phylogenetic inference was conducted on the 27 NR-genomes using PhyML v3.1 (Guindon et al., 2009) with 1000 bootstrap replications. We used a Hasegawa-Kishino-Yano nucleotide substitution model with invariant sites and gamma distributed categories of rate variation (HKY+G+I), which was considered as the most appropriate model by jModelTest v2.1.10 (Darriba et al., 2012)based on Akaike information criterion. Results were locally visualized using the R package ggTree v2.0.4 (Yu et al., 2016).

**Genomic variability between farming areas**

A consensus non-redundant genome (C-NR-genome) of the 21 newly generated NR-genomes was built using the “cons” argument from the EMBOSS suite v6.6.0.0 with the default settings (Rice et al., 2000). The NR-genomes were compared to the C-NR-genome using MUMer4 4.0.0beta2 (Marcais et al., 2018) with the option “nucmer -c 100 -l 15 –f” to identify all the variable positions between the 21 NR-genomes.

To assess the genomic variability between the farming areas, both minor and major variants of each sequencing library (21) were called on the C-NR-genome previously

generated using Freebayes v1.3.2-dirty (Garrison and Marth, 2012), with the following settings: --use-mapping-quality, --min-repeat-entropy 1, --haplotype-length 0, --min-alternate-count 5, --pooled-continuous, --hwe-priors-off, --allele-balance-priors-off. The resulting variant calling outputs were normalized using BCFtools v- (Narasimhan et al., 2016), decomposed with vt v1.0.0 (Tan et al., 2015) and split with vcflib v1.0.0. Considering the very low occurrence of multi-nucleotide polymorphisms, these latter were counted as InDels. Variable positions with a frequency of >50% were compared to those obtained with MUMer4 for comparative genomics and all were validated. They were subsequently subtracted from the variant calling files to retain only minority variants (with a frequency of <50%). A unique identifier composed of the position and sequence information (e.g. 128737_A>G) of each minority variant was created. Set analyses of nucleotide variations were performed with a multiplication matrix using R (R Core Team, 2018) operation and visualized with the Venn Diagram package (Chen and Boutros, 2011), UpSetR v1.4.0 (Lex et al., 2014) or pheatmap v1.0.12.

All downstream analyses (tables, graphs, plot creation and edition) were performed in R 3.6 (R Core Team, 2018) on R studio IDE (R-Studio-Team, 2015) with an extensive use of Dplyr v1.0.0 (Wickham et al., 2018), and ggplot2 (Wickham, 2016) from the tidyverse v1.3.0 package (Wickham et al., 2019). Genome visualization was carried out using gggenes v0.4.0. 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.

**Ancestral states reconstruction of discrete trait**

To infer the geographic origin of our sampled isolates, we first used Recombination Detection Program (RDP) version 4.97 (Martin et al., 2015), with default settings, to identify any recombination events between the OsHV-1 NR-genome we sequenced. As evidence for

recombination, we took events detected by at least two of seven different recombination detection methods implemented in RDP: RDP, MAXCHI, and GENECONV methods in primary scanning mode and the Bootscan, CHIMAERA, SisScan, and 3SEQ methods in secondary scanning mode, each with a Bonferroni-corrected P-value cut-off of 0.05. Two genomic regions were detected by RDP4 as potentially impacted by recombination events and removed from the alignment. Sampling sites (Th, Mo, Br) were modelled as a discrete trait for each NR-genome sequence over the genealogy by ancestral state inference using a discrete asymmetric phylogenetic diffusion model (Lemey et al., 2009) in BEAST v1.10.4 (Suchard et al., 2018). This approach estimates the probability of the internal nodes and branches being associated with a specific sampling site, based on information about sampling site states of the samples at the branch tips. A Bayesian stochastic search variable selection procedure (Lemey et al., 2009) was employed to allow for diffusion between specific location pairs to be included or excluded from the model. Based on Akaike's information criterion ranking in jModelTest v2.1.10 (Darriba et al., 2012), analyses were performed under a HKY+I model of nucleotide substitution. A strict molecular clock and a constant coalescent model were used and a MCMC chain of 10,000,000 steps was run and sub-sampled every 1,000 generations. Convergence to the stationary distribution and sufficient mixing (effective sample size >200) for all parameter estimates were checked in Tracer after removing the initial 10% of the samples as burn-in. The BEAGLE library was used to increase computational speed (Ayres et al., 2012). Bayes Factors (BF) were computed in SpreaD3 (Bielejec et al., 2016) and used as a measure of support for identifying frequently invoked transition rates to explain the diffusion process. Support for a rate was considered substantial when BF > 3, strong if BF > 10, and decisive if >100 (Jeffreys, 1961). A maximum clade credibility (MCC) tree was generated from tree output files from BEAST in

TreeAnnotator v1.10.4, and annotated in the FigTree v1.4.4 graphical user interface (available at http://tree.bio.ed.ac.uk/software/figtree/).

**Figure legends**

**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**

The present study was supported by the EU project VIVALDI (H2020 program, no. 678589) led by Ifremer; the CNRS and the University of Montpellier and the University of Perpignan Via Domitia. JD was supported by grant from the University of Montpellier. This work also benefitted from support from the “Laboratoire d’excellence” (LabEx) CeMEB, through the exploratory research project HaploFit and the use of the “environmental genomics” facility (http://www.labex-cemeb.org/fr/genomique-environnementale-1). This work was also supported by the Ifremer Scientific Board, through the HemoVir and GT-oyster projects. This study is set within the framework of the "Laboratoires d’Excellences (LABEX)" TULIP (ANR-10-LABX-41). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

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

work. JD, MJ and JME were involved in bioinformatics and statistical analyses. JD, RG, CM and JME drafted the manuscript, and all authors revised and approved the final manuscript.

We thank the staff of the Ifremer stations at Argenton (LPI, PFOM) and La Tremblade (SGMM) and the Comité Régional Conchylicole de Méditerranée (CRCM) for technical support in the production of standard (NSI) oysters and transplantation experiments. We also thank Nicole Faury (Ifremer, SGMM) and Marc Leroy for technical assistance. We are also grateful to Eric Rivals (LIRMM) for fruitful discussions.

We declare that we have no competing interests.

**Reference**

Abbadi, M., Zamperin, G., Gastaldelli, M., Pascoli, F., Rosani, U., Milani, A., Schivo, A., Rossetti, E., Turolla, E., Gennari, L., *et al.* (2018). Identification of a newly described OsHV-1 microvar from the North Adriatic Sea (Italy). J Gen Virol *99*, 693-703.

Akhtar, L.N., Bowen, C.D., Renner, D.W., Pandey, U., Della Fera, A.N., Kimberlin, D.W., Prichard, M.N., Whitley, R.J., Weitzman, M.D., and Szpara, M.L. (2019). Genotypic and Phenotypic Diversity of Herpes Simplex Virus 2 within the Infected Neonatal Population. mSphere *4*.

Al Khatib, H.A., Benslimane, F.M., Elbashir, I.E., Coyle, P.V., Al Maslamani, M.A., Al-Khal, A., Al Thani, A.A., and Yassine, H.M. (2020). Within-Host Diversity of SARS-CoV-2 in COVID-19 Patients With Variable Disease Severities. Front Cell Infect Microbiol *10*, 575613.

Alfaro, A.C., Nguyen, T.V., and Merien, F. (2018). The complex interactions of Ostreid herpesvirus 1, Vibrio bacteria, environment and host factors in mass mortality outbreaks of Crassostrea gigas. Reviews in Aquaculture *11*, 1148-1168.

Allicock, O.M., Lemey, P., Tatem, A.J., Pybus, O.G., Bennett, S.N., Mueller, B.A., Suchard, M.A., Foster, J.E., Rambaut, A., and Carrington, C.V. (2012). Phylogeography and

population dynamics of dengue viruses in the Americas. Molecular biology and evolution *29*, 1533-1543.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. *Available online at:* http://wwwbioinformaticsbabrahamacuk/projects/fastqc.

Avarre, J.C. (2017). Editorial: Molecular Tracing of Aquatic Viruses: Where Epidemiology Needs to Meet Genomics. Front Microbiol *8*, 1498.

Ayres, D.L., Darling, A., Zwickl, D.J., Beerli, P., Holder, M.T., Lewis, P.O., Huelsenbeck, J.P., Ronquist, F., Swofford, D.L., Cummings, M.P., *et al.* (2012). BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. Systematic biology *61*, 170-173.

Barbieri, E.S., Medina, C.D., Vazquez, N., Fiorito, C., Martelli, A., Wigdorovitz, A., Schwindt, E., Morga, B., Renault, T., Parreno, V., *et al.* (2019). First detection of Ostreid herpesvirus 1 in wild Crassostrea gigas in Argentina. J Invertebr Pathol *166*, 107222.

Barbosa Solomieu, V., Renault, T., and Travers, M.A. (2015). Mass mortality in bivalves and the intricate case of the Pacific oyster, Crassostrea gigas. J Invertebr Pathol *131*, 2-10.

Bielejec, F., Baele, G., Vrancken, B., Suchard, M.A., Rambaut, A., and Lemey, P. (2016). SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes. Molecular biology and evolution *33*, 2167-2169.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics *27*, 578-579.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114-2120.

Breitwieser, F.P., Baker, D.N., and Salzberg, S.L. (2018). KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biology *19*.

Buestel, D., Ropert, M., Prou, J., and Goulletquer, P. (2009). History, status, and future of oyster culture in France. J Shellfish Res *28*, 813-820.

Burge, C.A., Shore-Maggio, A., and Rivlin, N.D. (2017). Ecology of Emerging Infectious Diseases of Invertebrates. In Ecology of Invertebrate Diseases A.E. Hajek, ed. (John Wiley & Sons Ltd).

Burioli, E.A., Prearo, M., Riina, M.V., Bona, M.C., Fioravanti, M.L., Arcangeli, G., and Houssin, M. (2016). Ostreid herpesvirus type 1 genomic diversity in wild populations of Pacific oyster Crassostrea gigas from Italian coasts. J Invertebr Pathol *137*, 71-83.

Burioli, E.A.V., Prearo, M., and Houssin, M. (2017). Complete genome sequence of Ostreid herpesvirus type 1 microVar isolated during mortality events in the Pacific oyster Crassostrea gigas in France and Ireland. Virology *509*, 239-251.

Burioli, E.A.V., Varello, K., Lavazza, A., Bozzetta, E., Prearo, M., and Houssin, M. (2018). A novel divergent group of Ostreid herpesvirus 1 muVar variants associated with a mortality event in Pacific oyster spat in Normandy (France) in 2016. Journal of fish diseases.

Chase, J.M., McGill, B.J., McGlinn, D.J., May, F., Blowes, S.A., Xiao, X., Knight, T.M., Purschke, O., and Gotelli, N.J. (2018). Embracing scale-dependence to achieve a deeper understanding of biodiversity and its change across communities. Ecol Lett *21*, 1737-1751.

Chen, H., and Boutros, P.C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics *12*, 35.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat Methods *9*, 772.

Davison, A.J., Trus, B.L., Cheng, N., Steven, A.C., Watson, M.S., Cunningham, C., Le Deuff, R.M., and Renault, T. (2005). A novel class of herpesvirus with bivalve hosts. J Gen Virol *86*, 41-53.

de Lorgeril, J., Lucasson, A., Petton, B., Toulza, E., Montagnani, C., Clerissi, C., Vidal-Dupiol, J., Chaparro, C., Galinier, R., Escoubas, J.-M., *et al.* (2018). Immune-suppression by OsHV-1 viral infection causes fatal bacteraemia in Pacific oysters. Nature Communications *9*, 4215.

de Lorgeril, J., Petton, B., Lucasson, A., Perez, V., Stenger, P.L., Degremont, L., Montagnani, C., Escoubas, J.M., Haffner, P., Allienne, J.F., *et al.* (2020). Differential basal expression of immune genes confers Crassostrea gigas resistance to Pacific oyster mortality syndrome. BMC Genomics *21*, 63.

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

Dellicour, S., Rose, R., and Pybus, O.G. (2016). Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. BMC bioinformatics *17*, 82.

Delmotte, J., Chaparro, C., Galinier, R., de Lorgeril, J., Petton, B., Stenger, P.L., Vidal-Dupiol, J., Destoumieux-Garzon, D., Gueguen, Y., Montagnani, C., *et al.* (2020). Contribution of Viral Genomic Diversity to Oyster Susceptibility in the Pacific Oyster Mortality Syndrome. Frontiers in Microbiology *11*, 1579.

Domingo, E., Sheldon, J., and Perales, C. (2012). Viral Quasispecies Evolution. Microbiol Mol Biol R *76*, 159-216.

Dupont, S., Lokmer, A., Corre, E., Auguet, J.C., Petton, B., Toulza, E., Montagnani, C., Tanguy, G., Pecqueur, D., Salmeron, C., *et al.* (2020). Oyster hemolymph is a complex and dynamic ecosystem hosting bacteria, protists and viruses. Animal Microbiome *2*.

Faria, N.R., Rambaut, A., Suchard, M.A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pepin, J., *et al.* (2014). HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science *346*, 56-61.

Friedman, C.S., Estes, R.M., Stokes, N.A., Burge, C.A., Hargove, J.S., Barber, B.J., Elston, R.A., Burreson, E.M., and Reece, K.S. (2005). Herpes virus in juvenile Pacific oysters Crassostrea gigas from Tomales Bay, California, coincides with summer mortality episodes. Dis Aquat Organ *63*, 33-41.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. In arXivorg.

Gittenberger, A., Voorbergen-Laarman, M.A., and Engelsma, M.Y. (2016). Ostreid herpesvirus OsHV-1 mu Var in Pacific oysters Crassostrea gigas ( Thunberg 1793) of the Wadden Sea, a UNESCO world heritage site. Journal of fish diseases *39*, 105-109.

Guindon, S., Delsuc, F., Dufayard, J.F., and Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol *537*, 113-137.

Hammoumi, S., Vallaeys, T., Santika, A., Leleux, P., Borzym, E., Klopp, C., and Avarre, J.C. (2016). Targeted genomic enrichment and sequencing of CyHV-3 from carp tissues confirms low nucleotide diversity and mixed genotype infections. PeerJ *4*, e2516.

Jeffreys, H. (1961). Theory of Probability.

Jenkins, C., Hick, P., Gabor, M., Spiers, Z., Fell, S.A., Gu, X., Read, A., Go, J., Dove, M., O'Connor, W., *et al.* (2013). Identification and characterisation of an ostreid herpesvirus-1 microvariant (OsHV-1 micro-var) in Crassostrea gigas (Pacific oysters) in Australia. Dis Aquat Organ *105*, 109-126.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res *30*, 3059-3066.

Keeling, M.J., Woolhouse, M.E., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., and Grenfell, B.T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science *294*, 813-817.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Lapegue, S., Heurtebise, S., Cornette, F., Guichoux, E., and Gagnaire, P.A. (2020). Genetic Characterization of Cupped Oyster Resources in Europe Using Informative Single Nucleotide Polymorphism (SNP) Panels. Genes *11*.

Lauring, A.S., and Andino, R. (2010). Quasispecies Theory and the Behavior of RNA Viruses. Plos Pathogens *6*.

Lemey, P., Rambaut, A., Drummond, A.J., and Suchard, M.A. (2009). Bayesian phylogeography finds its roots. PLoS computational biology *5*, e1000520.

Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. IEEE transactions on visualization and computer graphics *20*, 1983-1992.

Lupo, C., Ezanno, P., Arzul, I., Garcia, C., Jadot, C., Joly, J.-P., Renault, T., and Bareille, N. (2016). How network analysis of oyster movements can improve surveillance and control programs of infectious diseases? Paper presented at: Front Vet Sci Conference Abstract: AquaEpi I (Oslo, Norway: Frontiers).

Marcais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. PLoS computational biology *14*, e1005944.

Martenot, C., Oden, E., Travaille, E., Malas, J.P., and Houssin, M. (2011). Detection of different variants of Ostreid Herpesvirus 1 in the Pacific oyster, Crassostrea gigas between 2008 and 2010. Virus research *160*, 25-31.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol *1*, vev003.

Martin, M.A., VanInsberghe, D., and Koelle, K. (2021). Insights from SARS-CoV-2 sequences. Science *371*, 466-467.

Morga, B., Jacquot, M., Pelletier, C., Chevignon, G., Dégremont, L., Biétry, A., Pepin, J.-F., Heurtebise, S., Escoubas, J.-M., Bean, T.P., *et al.* (2021). Genomic Diversity of the Ostreid Herpesvirus Type 1 Across Time and Location and Among Host Species. Frontiers in Microbiology *12*.

Morse, A.M., Calabro, K.R., Fear, J.M., Bloom, D.C., and McIntyre, L.M. (2017). Reliable Detection of Herpes Simplex Virus Sequence Variation by High-Throughput Resequencing. Viruses *9*.

Moss, J.A., Burreson, E.M., Cordes, J.F., Dungan, C.F., Brown, G.D., Wang, A., Wu, X., and Reece, K.S. (2007). Pathogens in Crassostrea ariakensis and other Asian oyster species: implications for non-native oyster introduction to Chesapeake Bay. Dis Aquat Organ *77*, 207-223.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics *32*, 1749-1751.

Nkili-Meyong, A.A., Bigarre, L., Labouba, I., Vallaeys, T., Avarre, J.C., and Berthet, N. (2016). Contribution of Next-Generation Sequencing to Aquatic and Fish Virology. Intervirology *59*, 285-300.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome research *5*, 824-834.

Paul-Pont, I., Dhand, N.K., and Whittington, R.J. (2013). Spatial distribution of mortality in Pacific oysters Crassostrea gigas: reflection on mechanisms of OsHV-1 transmission. Diseases of Aquatic Organisms *105*, 127-138.

Peeler, E.J., Reese, R.A., Cheslett, D.L., Geoghegan, F., Power, A., and Thrush, M.A. (2012). Investigation of mortality in Pacific oysters associated with Ostreid herpesvirus-1 mu Var in the Republic of Ireland in 2009. Prev Vet Med *105*, 136-143.

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

Pernet, F., Lupo, C., Bacher, C., and Whittington, R.J. (2016). Infectious diseases in oyster aquaculture require a new integrated approach. Philos Trans R Soc Lond B Biol Sci *371*.

Petton, B., Bruto, M., James, A., Labreuche, Y., Alunno-Bruscia, M., and Le Roux, F. (2015). Crassostrea gigas mortality in France: the usual suspect, a herpes virus, may not be the killer in this polymicrobial opportunistic disease. Front Microbiol *6*, 686.

Petton, B., Destoumieux-Garzon, D., Pernet, F., Toulza, E., de Lorgeril, J., Degremont, L., and Mitta, G. (2021). The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions. Front Immunol *12*, 630343.

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

Pybus, O.G., Suchard, M.A., Lemey, P., Bernardin, F.J., Rambaut, A., Crawford, F.W., Gray, R.R., Arinaminpathy, N., Stramer, S.L., Busch, M.P., *et al.* (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. Proc Natl Acad Sci U S A *109*, 15066-15071.

R-Studio-Team (2015). RStudio: Integrated Development Environment for R. (Boston, MA: RStudio Inc.).

R Core Team (2018). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing).

Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., and Holmes, E.C. (2008). The genomic and epidemiological dynamics of human influenza A virus. Nature *453*, 615-619.

Renault, T., Moreau, P., Faury, N., Pepin, J.F., Segarra, A., and Webb, S. (2012). Analysis of Clinical Ostreid Herpesvirus 1 (Malacoherpesviridae) Specimens by Sequencing Amplified Fragments from Three Virus Genome Areas. Journal of Virology *86*, 5942-5947.

Renner, D.W., and Szpara, M.L. (2018). Impacts of Genome-Wide Analyses on Our Understanding of Human Herpesvirus Diversity and Evolution. J Virol *92*.

Renzette, N., Gibson, L., Bhattacharjee, B., Fisher, D., Schleiss, M.R., Jensen, J.D., and Kowalik, T.F. (2013). Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. PLoS Genet *9*, e1003735.

Renzette, N., Gibson, L., Jensen, J.D., and Kowalik, T.F. (2014). Human cytomegalovirus intrahost evolution-a new avenue for understanding and controlling herpesvirus infections. Curr Opin Virol *8*, 109-115.

Renzette, N., Pokalyuk, C., Gibson, L., Bhattacharjee, B., Schleiss, M.R., Hamprecht, K., Yamamoto, A.Y., Mussi-Pinhata, M.M., Britt, W.J., Jensen, J.D., *et al.* (2015). Limits and patterns of cytomegalovirus genomic diversity in humans. Proc Natl Acad Sci U S A *112*, E4120-4128.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet *16*, 276-277.

Rohfritsch, A., Bierne, N., Boudry, P., Heurtebise, S., Cornette, F., and Lapegue, S. (2013). Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, Crassostrea gigas. Evolutionary applications *6*, 1064-1078.

Sanjuán, R., and Domingo-Calap, P. (2019). Genetic Diversity and Evolution of Viral Populations.

Segarra, A., Pepin, J.F., Arzul, I., Morga, B., Faury, N., and Renault, T. (2010). Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, Crassostrea gigas, in France in 2008. Virus research *153*, 92-99.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962.

Stentiford, G.D., Neil, D.M., Peeler, E.J., Shields, J.D., Small, H.J., Flegel, T.W., Vlak, J.M., Jones, B., Morado, F., Moss, S., *et al.* (2012). Disease will limit future food supply from the global crustacean fishery and aquaculture sectors. J Invertebr Pathol *110*, 141-157.

Suchard, M.A., Lemey, P., Baele, G., Ayres, D.L., Drummond, A.J., and Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol *4*, vey016.

Tan, A., Abecasis, G.R., and Kang, H.M. (2015). Unified representation of genetic variants. Bioinformatics *31*, 2202-2204.

Theze, J., Li, T., du Plessis, L., Bouquet, J., Kraemer, M.U.G., Somasekar, S., Yu, G., de Cesare, M., Balmaseda, A., Kuan, G., *et al.* (2018). Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. Cell Host Microbe *23*, 855-864 e857.

Vasquez-Yeomans, R., Garcia-Ortega, M., and Caceres-Martinez, J. (2010). Gill erosion and herpesvirus in Crassostrea gigas cultured in Baja California, Mexico. Dis Aquat Organ *89*, 137-144.

Vendrami, D.L.J., Houston, R.D., Gharbi, K., Telesca, L., Gutierrez, A.P., Gurney-Smith, H., Hasegawa, N., Boudry, P., and Hoffman, J.I. (2019). Detailed insights into pan-European population structure and inbreeding in wild and hatchery Pacific oysters (Crassostrea gigas) revealed by genome-wide SNP data. Evolutionary applications *12*, 519-534.

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., and Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature *439*, 344-348.

Webb, S.C., Fidler, A., and Renault, T. (2007). Primers for PCR-based detection of ostreid herpes virus-1 (OsHV-1): Application in a survey of New Zealand molluscs. Aquaculture *272*, 126-139.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag, New York).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., *et al.* (2019). Welcome to the Tidyverse. Journal of Open Source Software *4*, 1686.

Wickham, H., François, R., Henry, L., and Müller, K. (2018). dplyr: A Grammar of Data Manipulation (RStudio).

Yu, G., Smith, D.K., Zhu, H., Guan, Y.Y., Lam, T.T.Y., and McInerny, G. (2016). GGTREE : an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution *8*, 28-36.

Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., *et al.* (2012). The oyster genome reveals stress adaptation and complexity of shell formation. Nature *490*, 49-54.