



**HAL**  
open science

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

Jean Delmotte, Camille Pelletier, Benjamin Morga, Richard Galinier, Bruno Petton, Jean-Baptiste Lamy, Oliver Kaltz, Jean-Christophe Avarre, Caroline Montagnani, Jean-Michel Escoubas

### ► To cite this version:

Jean Delmotte, Camille Pelletier, Benjamin Morga, Richard Galinier, Bruno Petton, et al.. Phylogeography of infectious disease: genetic diversity and connectivity of the Ostreid herpesvirus 1 population in France. 2021. hal-03217723

**HAL Id: hal-03217723**

**<https://hal.science/hal-03217723v1>**

Preprint submitted on 5 May 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Jean Delmotte<sup>1</sup>, Camille Pelletier<sup>2</sup>, Benjamin Morga<sup>2</sup>, Richard Galinier<sup>3</sup>, Bruno Petton<sup>4</sup>, JeanBaptiste Lamy<sup>2</sup>, Oliver Kaltz<sup>5</sup>, Jean-Christophe Avarre<sup>5</sup>, Caroline Montagnani<sup>1\*</sup>, JeanMichel Escoubas<sup>1\*</sup>

<sup>1</sup>IHPE, Univ. Montpellier, CNRS, Ifremer, UPVD, F-34095 Montpellier, France

<sup>2</sup>Ifremer, RBE-SG2M-LGPMM, Station La Tremblade, F-17390 La Tremblade, France

<sup>3</sup>IHPE, Univ. Montpellier, CNRS, Ifremer, UPVD, F-66000 Perpignan, France

<sup>4</sup>Ifremer, LEMAR UMR 6539 (Université de Bretagne Occidentale, CNRS, IRD, Ifremer), F-29840 Argenton-en-Landunvez, France

<sup>5</sup>ISEM, IRD, CNRS, EPHE, Univ. Montpellier, F-34095 Montpellier, France

Running Head: Phylogeography of *Ostreid herpesvirus 1*

\*Corresponding authors: [Caroline.Montagnani@ifremer.fr](mailto:Caroline.Montagnani@ifremer.fr);

[Jean.Michel.Escoubas@ifremer.fr](mailto:Jean.Michel.Escoubas@ifremer.fr)

### Abstract

The genetic diversity of viral populations is a key to understanding their phylogeographic and dissemination history of viruses, but studying the diversity of whole genomes from natural populations remains a challenge. Molecular ecology approaches are

commonly used for RNA viruses harboring small genomes, but have only rarely been applied to DNA viruses with large 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 genomes. Combining wholegenome comparisons with phylogenetic analysis and quantification of major and minor variants, we assessed the connectivity of OsHV-1 viral populations between the three oysterfarming 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 molecular ecology approaches can be applied to large-genome viruses to determine the extent of their genetic diversity and better understand the spread of viral populations in natural environments.

### **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 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 oysterfarming sites.

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

## **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 (1), 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 (2, 3). 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 (4-6). For example, human cytomegalovirus shows considerable inter-host and intra-host genetic divergence across tissue compartments and times of infection (7, 8). In addition, the evolution of a disease is

sometimes not fully explained by intrinsic host factors, but by the genotypic diversity of herpesviruses (9).

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 (10-12). One illustration is the Pacific oyster mortality syndrome (POMS), which threatens global *Crassostrea gigas* oyster production, a main sector in aquaculture worldwide (reviewed in (13)). Since 2008, this syndrome has caused mass mortality in cultivated oysters around the world, from Europe to America and Asia (14-22). 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) (23-27). Two major genetic factors seem to affect the severity of POMS: the genetic background of the oysters (28-30) and OsHV-1 genetic diversity (31, 32). Since the characterization of the emergent genotype OsHV-1  $\mu$ Var, associated with the 2008 high mortality events (27), several variants of the  $\mu$ Var genotype have been described (reviewed in (16)). 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 (33). 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  $\mu$ Var and its comparison with the reference genome published in 2005 (34) 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 (35). 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. To date, the entire genomes of only four OsHV-1 infecting *C. gigas* oysters are available, namely the reference genome (AY509253), two OsHV-1  $\mu$ Var ( $\mu$ Var A KY242785.1 and  $\mu$ Var B KY271630.1) and more recently OsHV-1-PT (MG561751.2) (14, 34, 35). These four assemblages show significant genomic diversity; however, they tell us very little about the geographic distribution and dissemination of the virus.

As shown with the H1N1 influenza virus (36), the dengue virus (37), the Zika virus (38), the AIDS virus (39) and more recently SARS-CoV-2 (40), 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  $\mu$ Var in natural viral populations both within and between host individuals, to better understand the dissemination and evolution of OsHV-1  $\mu$ Var populations. Using a deepsequencing approach conducted at the individual level and novel bioinformatics analyses, we identified 21 new complete genome sequences of OsHV-1  $\mu$ Var 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. This phylogeographic 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, for revealing how the genetic relationships shape 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. The cumulative mortalities for the three batches of oysters were 85%, 81% and 70% for Thau (Th), Marennes-Oléron (Mo) and Brest (Br), respectively (Fig. 1A). No mortality was observed in the control group.

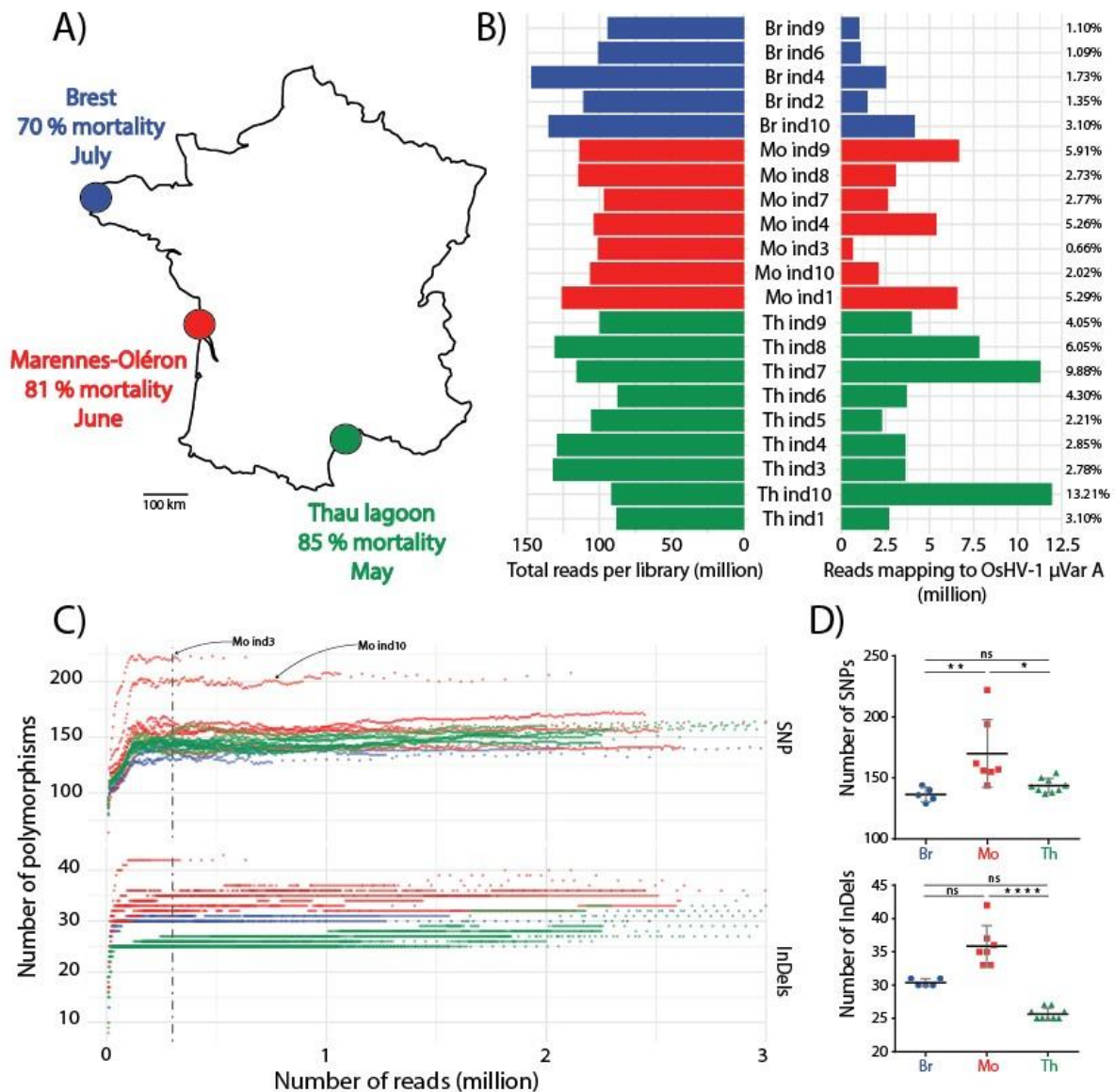
For each individual oyster, deep sequencing produced from 86.2 to 145.9 M reads per sample (average 109.7 M reads  $\pm$ 16.7 SD, Fig.1B. and Table S1). To evaluate OsHV-1 sequencing depth in each sample, sequencing reads were mapped on the OsHV-1  $\mu$ Var A genome (KY242785.1) (35). The number of reads mapping to the OsHV-1  $\mu$ Var 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 ( $170 \pm 27.7$ ) was significantly higher than that of viruses from Br ( $136.4 \pm 5.9$ ;  $P \leq 0.001$ ) and Th ( $143.8 \pm 5.8$ ;  $P \leq 0.05$ ) and ii) the number of InDels in viruses from Mo ( $35.9 \pm 3.1$  SD) was also significantly higher than that of viruses from Th ( $25.7 \pm 0.9$ ;  $P \leq 0.0001$ ) (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) Sampling locations of the Ostreid herpesvirus-1 (OsHV-1) infected oysters during the Pacific oyster mortality syndrome (POMS) event in 2018. 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). Below each farming area, the percentage of mortality within the site is given. (B) Quantification of the number of reads in



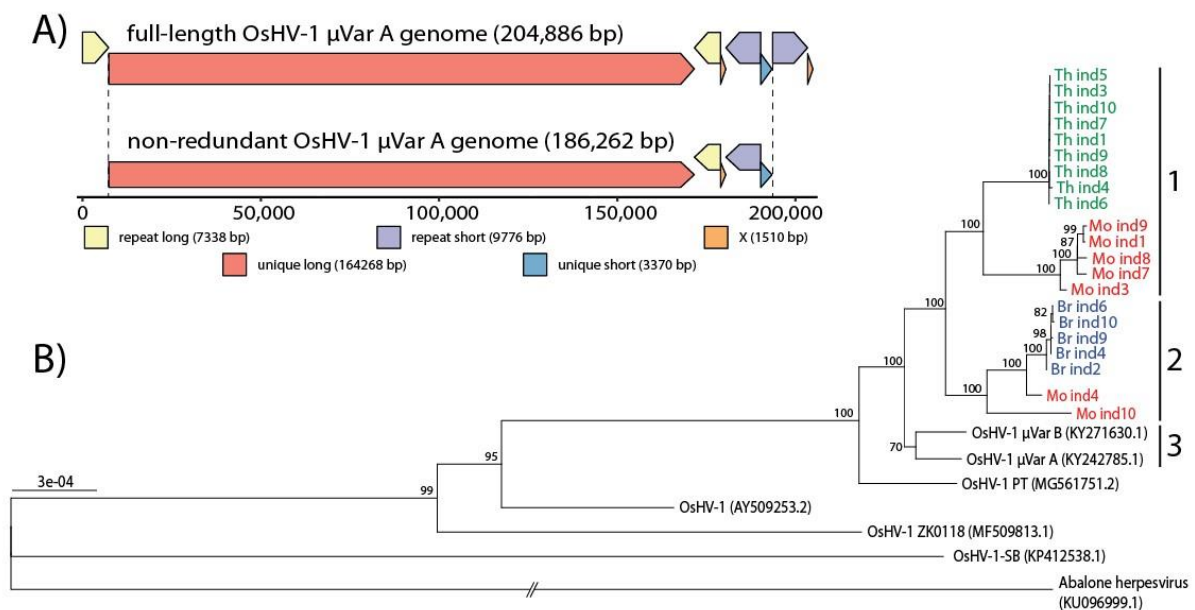
the library and the number of reads mapping to the OsHV-1  $\mu$ Var A genome. Numbers on the right side of the figure indicate the percentages of reads mapping to the OsHV-1  $\mu$ Var A genome. (C) Rarefaction curves using the OsHV-1  $\mu$ Var A genome as the reference sequence. The top panel shows 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 $\mu$ Var genotype.**

We assembled non-redundant 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  $\mu$ Var A genome contains two unique regions and three repeats: the unique long region (UL, 164,268 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 non-redundant OsHV1  $\mu$ Var 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 7 other published genomes belonging to the *Malacoherpesviridae* family: six *Ostreid herpesvirus 1*

(genus *Ostreavirus*) and one *Halitid herpesvirus 1* (abalone herpesvirus (AbHV-1), genus *Aurivirus*).

A phylogenetic analysis of these 28 NR-genomes revealed three main polyphyletic clusters (Fig. 2B). AbHV-1, the sole herpesvirus isolated from a gastropod, was identified as an outgroup, whereas the two herpesviruses isolated from bivalves other than oysters (OsHV-1-SB isolated from the clam *Scapharca broughtonii* and OsHV-1 ZK0118 isolated from the scallop *Azumapecten farreri*) formed two separate branches. Among the viruses isolated from *C. gigas*, the 21 new OsHV-1 isolates were phylogenetically closer to OsHV-1  $\mu$ Var A and B isolates (Fig. 2B cluster 3) than to the OsHV-1 reference genome (AY509253.2), suggesting that they belong to the  $\mu$ Var genotype. Th and Br isolates grouped into two distinct clusters reflecting their geographic origin (Fig. 2B, cluster 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  $5.33 \times 10^{-6} \pm 1.99 \times 10^{-6}$  SD and  $2.16 \times 10^{-5} \pm 1.43 \times 10^{-5}$  SD, respectively) was lower than that observed within the Mo isolates (average  $6.16 \times 10^{-4} \pm 5.29 \times 10^{-4}$ ), indicating greater viral diversity in Mo than in the other two farming areas.



**Fig. 2: Phylogenetic relationship between the 21 *de novo* assembled non-redundant genomes and other *Malacoherpesviridae* viruses.** (A) Full-length (upper panel) and nonredundant (NR) (lower panel) genomes of OsHV-1  $\mu$ Var A (KY242785). The OsHV-1  $\mu$ Var 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 Lagoon (Th) in green. Three clusters were defined based on the clustering of genetic distances (numbered on the right from 1 to 3).

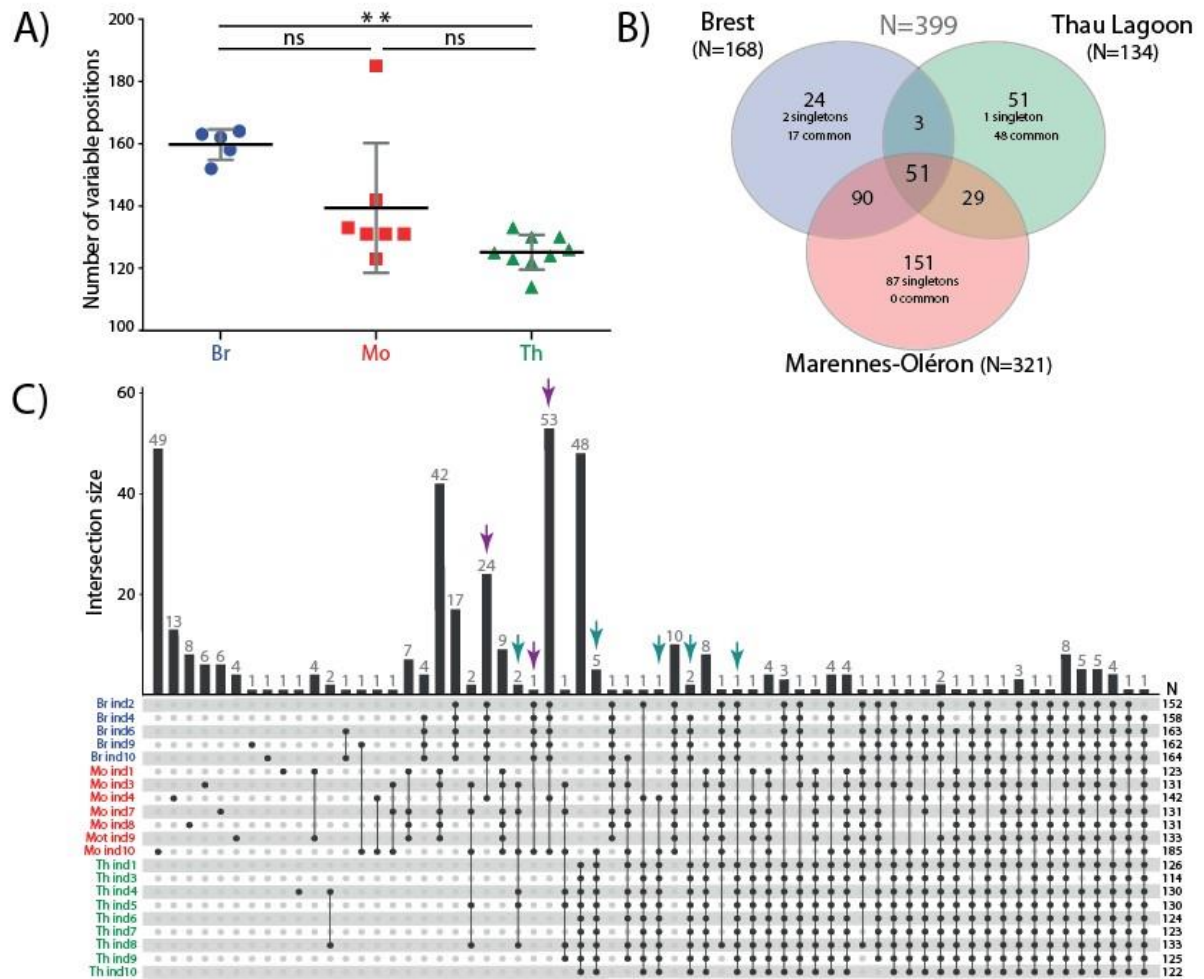
### **3) Comparative genomics of the 21 non-redundant genomes confirms that OsHV-1 diversity is higher in Marennes-Oléron than in the other two farming areas.**

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 ( $159.8 \pm 4.9$ ) than from Th ( $125 \pm 5.6$ ;  $P$

< 0.01), whereas the average number of variants in Mo ( $139.4 \pm 20.8$ ) 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: nonsignificant, \*\*  $P < 0.01$ ). (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.

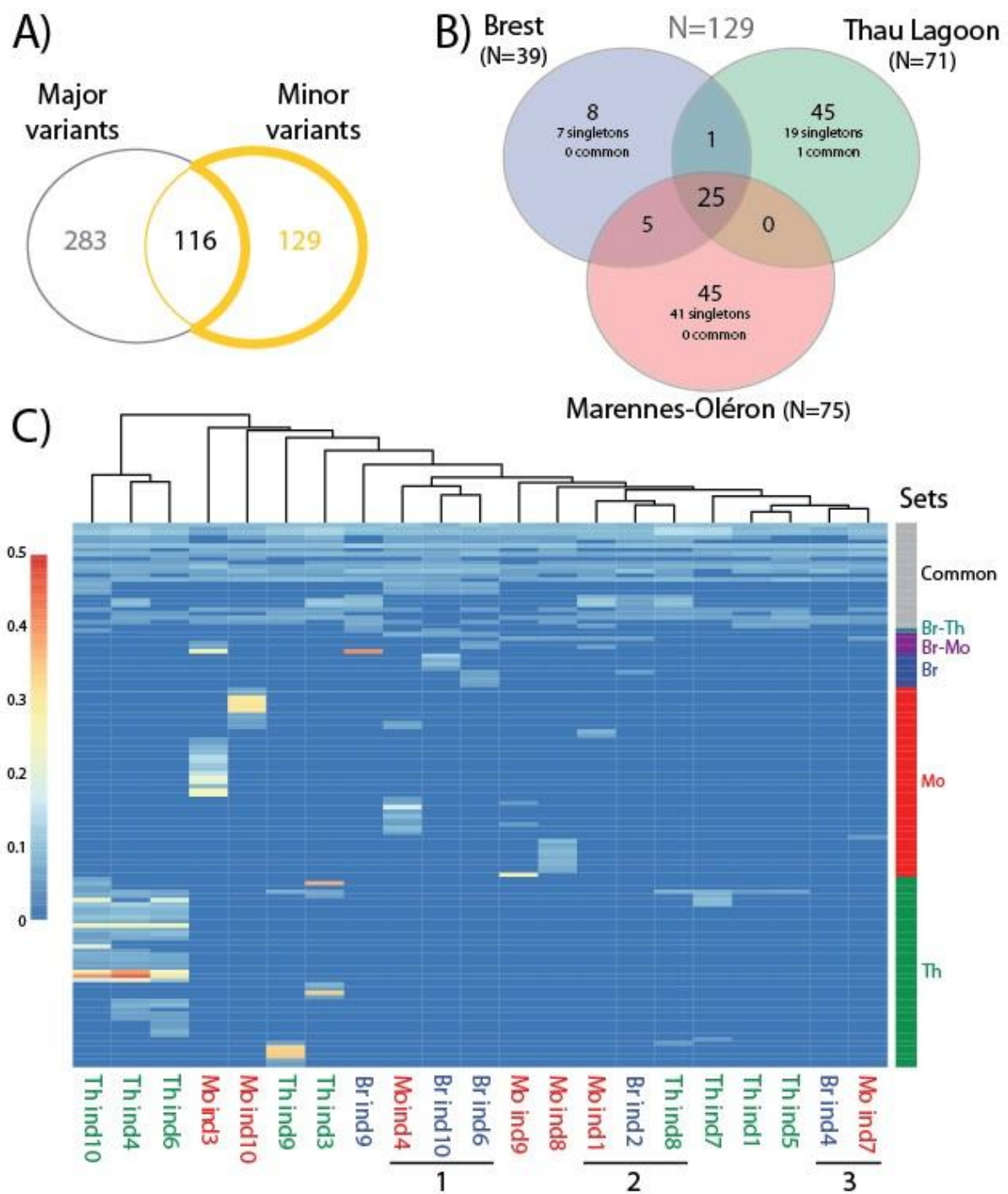
#### **4) Diversity of minor OsHV-1 variants is higher in Marennes-Oléron than in the other two farming areas.**

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%  $\pm$  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 sitespecific 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 OsHV1 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, Marenes-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.

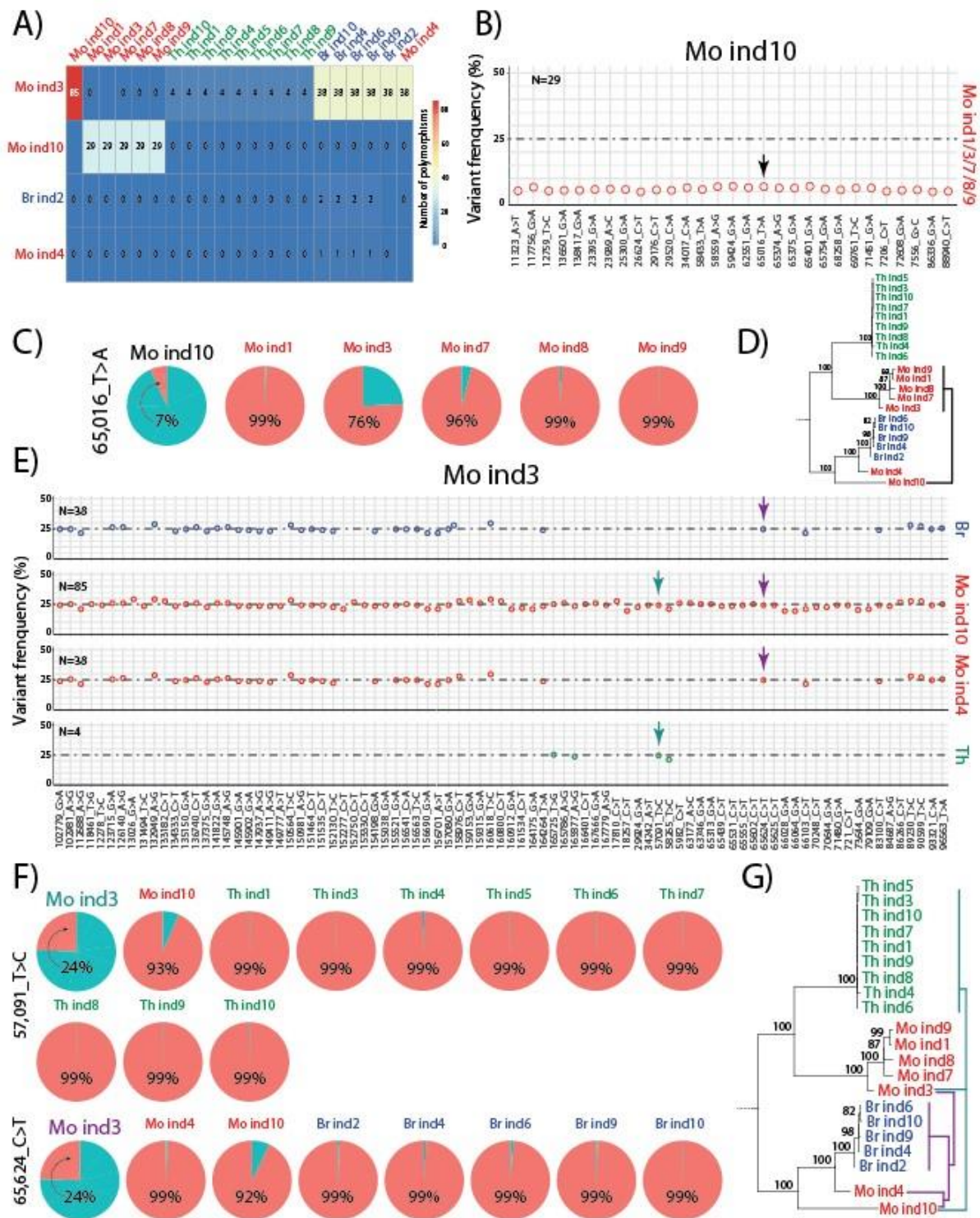
### **5) Some minor OsHV-1 variants from the Marennes-Oléron farming area are major variants in the other two farming areas.**

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).

Altogether, 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.



**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.

## 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 (41). 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  $\approx 3000X$ . 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 (42), 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 (43, 44).

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  $\mu$ Var A. To solve this problem, we used the approach proposed by Morse *et al.* (45): 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 intraindividual (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 (4). They also suggest that all OsHV-1 characterized during this study share a common ancestor with OsHV-1  $\mu$ Var characterized by Burioli *et al.* in Normandy (France), reinforcing the idea that genotypes closely related to  $\mu$ Var have replaced the reference genotype and are predominant in oyster-farming areas along the French coasts (35, 46). Our findings also showed that OsHV-1  $\mu$ Var populations are much more diverse than previously thought. Phylogenetic analyses corroborated results from Delmotte *et al.* showing that OsHV-1  $\mu$ Var from the Bay of Brest and Thau Lagoon constitute two distinct viral populations (31). Clearly, it is now necessary to change the common practice of relying on the generation of consensus genomes to study the epidemiology of DNA viruses, because they do not accurately depict the actual genetic diversity (5). For this reason, we developed this novel approach, applicable to any other DNA virus model.

Our approach demonstrated that OsHV-1 diversity in Marennes-Oléron Bay was higher than in the Bay of Brest and Thau Lagoon. This variability may be related to the fact that the Marennes-Oléron Bay has become one of the largest French oyster-farming areas, with chronic overstocking (47). 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 infection, and in turn increase its genetic diversity (48).

Our approach also made it possible to identify major and minor variants, and variants with frequencies varying dramatically among sampled oyster individuals. Thus, ultra-deep sequencing can support the reconstruction of the true structure of a viral population, and can even highlight potential genetic links between individuals from the same farming area or between individuals from the three farming areas (Fig. 5 D and G, respectively). These genetic links, combined with whole-genome phylogeny, can be used to infer the dynamics of dispersal of OsHV-1 populations across the three farming areas (Fig. 6). Samples from Marennes-Oléron

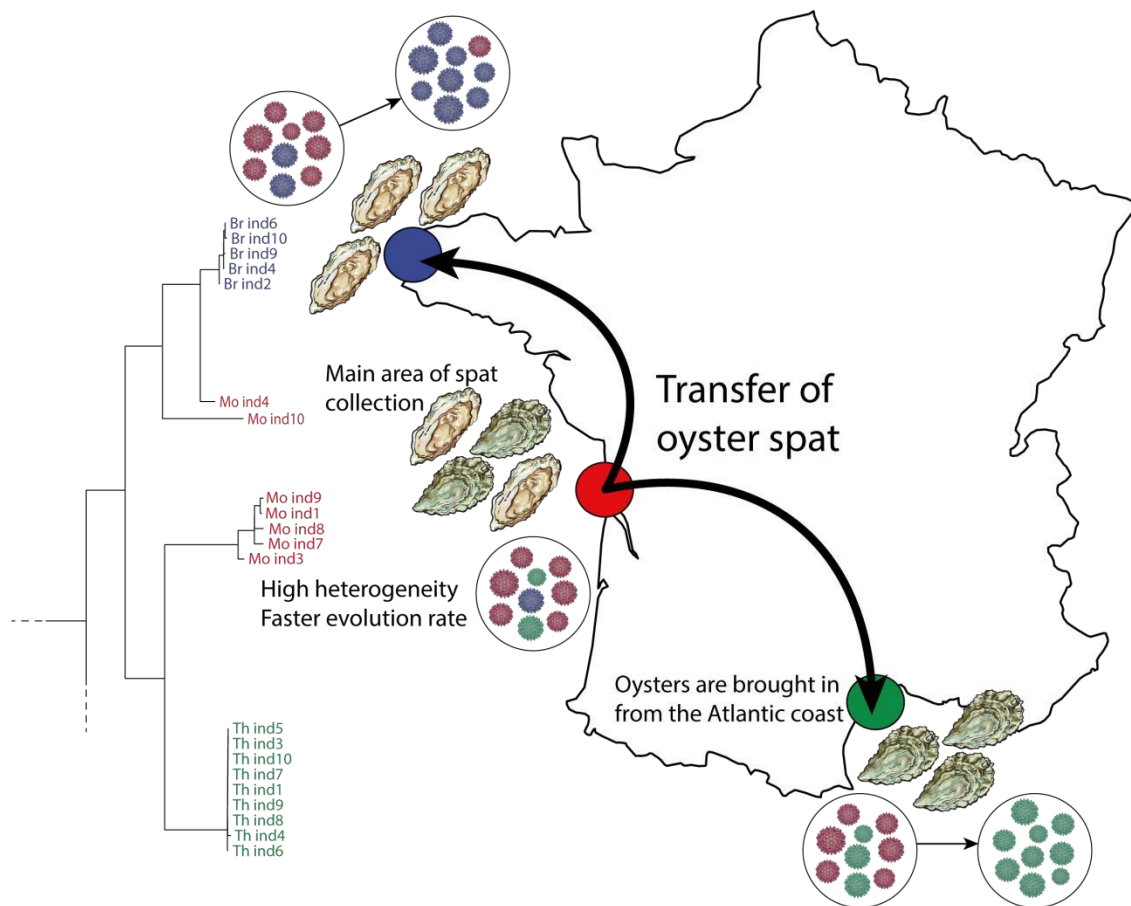


have the most heterogeneous viral populations with the fastest evolution rate. Moreover, major variants characterized in the other two farming areas were also present in Marennes-Oléron, as minor variants. This suggests that Marennes-Oléron farming area is the main source of OsHV-1 viral diversity, from 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 (47). 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.*, (49)). Viral populations originating from the Marennes-Oléron Bay are then subsequently shaped by local evolutionary and ecological processes in the different farming areas. The next challenge is to characterize these evolutionary and ecological processes that affect OsHV-1 genetic diversity and, in turn, their impact on OsHV-1 virulence in the different farming areas. Given that the fate of new genetic variants is largely determined by host selection, one hypothesis is that oyster populations located in the Bay of Brest and Thau Lagoon 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 (50-52). 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 (41). Furthermore, the relationship between environmental factors and mortality events in Pacific oysters has been well documented (reviewed in (53)). 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  $\mu$ Var populations may shed new light on the processes of variant selection.

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 (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 and the Bay of Brest (Br) (Logonna-Daoulas, lat.: 48.335263 long.: -4.317922) 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 (55, 56). 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 consisted of oysters that had not been transplanted in natural environment. Monitoring was stopped after three consecutive days without 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 moribund oysters using the MagAttract<sup>®</sup> 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<sup>®</sup> 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  $\mu$ L, consisting of 0.5  $\mu$ L of DNA (40 ng/ $\mu$ L) and 1  $\mu$ L of LightCycler 480 SYBR Green I Master mix (Roche) containing 0.5  $\mu$ M 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): Fw-5'-ATTGATGATGTGGATAATCTGTG-3' and Rev-5'-GGTAAATACCATTGGTCTTGTTC-3' (57). 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<sup>5</sup> 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 evaluated using FastQC v0.11.8 (58), and sequence adapters were removed using trimmomatic 0.39 (59). 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 (60) and Seqkit v0.12.0 (61) 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 (62)) using Bowtie2 4.8.2 (63), 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 --onlyassembler options (64). Assembled scaffolds were aligned to the OsHV-1  $\mu$ Var A genome (GenBank no. KY242785.1) using BLAST 2.9.0 (65) with an e-value of 0.00001. Scaffolds were then extended using SSPACE v3.0 (66). 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, (34)) 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 (61). 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) (45). 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**

To assess the relationships between OsHV-1 genomes, the 21 NR-genome assemblies were aligned using MAFFT v7.455 with default parameters (67). A consensus non-redundant genome (C-NR-genome) was subsequently built using the “cons” argument from the EMBOSS

suite v6.6.0.0 with the default settings (68). Then the 21 NR-genomes were compared to the C-NR-genome using MUMer4 4.0.0beta2 (69) with the option “nucmer -c 100 -l 15 -f” to identify all the variable positions between the 21 NR-genomes. A maximumlikelihood phylogenetic inference was conducted on the 21 NR-genomes and 7 NR-genomes derived from previously published OsHV-1 genomes (AY509253, KU096999, KP412538, KY242785, KY271630, MG561751, MF509813) using IqTree 1.6.7 (70) with 1000 bootstrap replications. The NR genome of the abalone herpesvirus (AbHV-1, KU096999) was used as an outgroup. The ModelFinder implementation (71) resulted in the following best-fit model according to Bayesian information criterion: Transversion model + empirical base frequencies + gamma model with 4 rate categories. Results were locally visualized using the R package ggTree v2.0.4 (72).

### **Variant calling analysis**

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 (73), with the following settings: --use-mapping-quality, --min-repeatentropy 1, --haplotype-length 0, --min-alternate-count 5, --pooled-continuous, --hwe-priorsoff, --allele-balance-priors-off. The resulting variant calling outputs were normalized using BCFtools v- (74), decomposed with vt v1.0.0 (75) 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 (76) operation and visualized with the Venn Diagram package (77), UpSetR v1.4.0 (78) or pheatmap v1.0.12.

### **Downstream analyses**

All downstream analyses (tables, graphs, plot creation and edition) were performed locally in R 3.6 (76) on R studio IDE (79) with an extensive use of Dplyr v1.0.0 (80), and ggplot2 (81) from the tidyverse v1.3.0 package (82). Genome visualization was carried out using gggenes 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**

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 project. This study is set within the framework of the "Laboratoires d'Excellences (LABEX)" TULIP (ANR-10-LABX-41).

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 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

1. Nkili-Meyong AA, Bigarre L, Labouba I, Vallaëys T, Avarre JC, Berthet N. 2016. Contribution of Next-Generation Sequencing to Aquatic and Fish Virology. *Intervirology* 59:285-300.
2. Lauring AS, Andino R. 2010. Quasispecies Theory and the Behavior of RNA Viruses. *Plos Pathogens* 6.



3. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439:344-348.
4. Hammoumi S, Vallaeyts T, Santika A, Leleux P, Borzym E, Klopp C, Avarre JC. 2016. Targeted genomic enrichment and sequencing of CyHV-3 from carp tissues confirms low nucleotide diversity and mixed genotype infections. *PeerJ* 4:e2516.
5. Renner DW, Szpara ML. 2018. Impacts of Genome-Wide Analyses on Our Understanding of Human Herpesvirus Diversity and Evolution. *J Virol* 92.
6. Renzette N, Gibson L, Bhattacharjee B, Fisher D, Schleiss MR, Jensen JD, Kowalik TF. 2013. Rapid intrahost evolution of human cytomegalovirus is shaped by demography and positive selection. *PLoS Genet* 9:e1003735.
7. Renzette N, Gibson L, Jensen JD, Kowalik TF. 2014. Human cytomegalovirus intrahost evolution-a new avenue for understanding and controlling herpesvirus infections. *Curr Opin Virol* 8:109-15.
8. Renzette N, Pokalyuk C, Gibson L, Bhattacharjee B, Schleiss MR, Hamprecht K, Yamamoto AY, Mussi-Pinhata MM, Britt WJ, Jensen JD, Kowalik TF. 2015. Limits and patterns of cytomegalovirus genomic diversity in humans. *Proc Natl Acad Sci U S A* 112:E4120-8.
9. Akhtar LN, Bowen CD, Renner DW, Pandey U, Della Fera AN, Kimberlin DW, Prichard MN, Whitley RJ, Weitzman MD, Szpara ML. 2019. Genotypic and Phenotypic Diversity of Herpes Simplex Virus 2 within the Infected Neonatal Population. *mSphere* 4.
10. Pernet F, Lupo C, Bacher C, Whittington RJ. 2016. Infectious diseases in oyster aquaculture require a new integrated approach. *Philos Trans R Soc Lond B Biol Sci* 371.

11. Stentiford GD, Neil DM, Peeler EJ, Shields JD, Small HJ, Flegel TW, Vlak JM, Jones B, Morado F, Moss S, Lotz J, Bartholomay L, Behringer DC, Hauton C, Lightner DV. 2012. Disease will limit future food supply from the global crustacean fishery and aquaculture sectors. *J Invertebr Pathol* 110:141-57.
12. Burge CA, Shore-Maggio A, Rivlin ND. 2017. Ecology of Emerging Infectious Diseases of Invertebrates. *In* Hajek AE (ed), *Ecology of Invertebrate Diseases* doi:doi.org/10.1002/9781119256106.ch16. John Wiley & Sons Ltd.
13. Petton B, Destoumieux-Garzon D, Pernet F, Toulza E, de Lorgeril J, Degremont L, Mitta G. 2021. The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions. *Front Immunol* 12:630343.
14. Abbadi M, Zamperin G, Gastaldelli M, Pascoli F, Rosani U, Milani A, Schivo A, Rossetti E, Turolla E, Gennari L, Toffan A, Arcangeli G, Venier P. 2018. Identification of a newly described OsHV-1 microvar from the North Adriatic Sea (Italy). *J Gen Virol* 99:693-703.
15. Barbieri ES, Medina CD, Vazquez N, Fiorito C, Martelli A, Wigdorovitz A, Schwindt E, Morga B, Renault T, Parreno V, Baron PJ. 2019. First detection of Ostreid herpesvirus 1 in wild *Crassostrea gigas* in Argentina. *J Invertebr Pathol* 166:107222.
16. Barbosa Solomieu V, Renault T, Travers MA. 2015. Mass mortality in bivalves and the intricate case of the Pacific oyster, *Crassostrea gigas*. *J Invertebr Pathol* 131:2-10.
17. Friedman CS, Estes RM, Stokes NA, Burge CA, Hargove JS, Barber BJ, Elston RA, Burrenson EM, Reece KS. 2005. Herpes virus in juvenile Pacific oysters *Crassostrea gigas* from Tomales Bay, California, coincides with summer mortality episodes. *Dis Aquat Organ* 63:33-41.

18. Gittenberger A, Voorbergen-Laarman MA, Engelsma MY. 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.
19. Jenkins C, Hick P, Gabor M, Spiers Z, Fell SA, Gu X, Read A, Go J, Dove M, O'Connor W, Kirkland PD, Frances J. 2013. Identification and characterisation of an ostreid herpesvirus-1 microvariant (OsHV-1 micro-var) in *Crassostrea gigas* (Pacific oysters) in Australia. *Dis Aquat Organ* 105:109-26.
20. Moss JA, Burreson EM, Cordes JF, Dungan CF, Brown GD, Wang A, Wu X, Reece KS. 2007. Pathogens in *Crassostrea ariakensis* and other Asian oyster species: implications for non-native oyster introduction to Chesapeake Bay. *Dis Aquat Organ* 77:207-23.
21. Peeler EJ, Reese RA, Cheslett DL, Geoghegan F, Power A, Thrush MA. 2012. Investigation of mortality in Pacific oysters associated with Ostreid herpesvirus-1 mu Var in the Republic of Ireland in 2009. *Preventive Veterinary Medicine* 105:136-143.
22. Vasquez-Yeomans R, Garcia-Ortega M, Caceres-Martinez J. 2010. Gill erosion and herpesvirus in *Crassostrea gigas* cultured in Baja California, Mexico. *Dis Aquat Organ* 89:137-44.
23. Paul-Pont I, Dhand NK, Whittington RJ. 2013. Spatial distribution of mortality in Pacific oysters *Crassostrea gigas*: reflection on mechanisms of OsHV-1 transmission. *Diseases of Aquatic Organisms* 105:127-138.
24. Pernet F, Barret J, Le Gall P, Corporeau C, Dégremont L, Lagarde F, Pépin J-F, 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.

25. Petton B, Pernet F, Robert R, Boudry P. 2013. Temperature influence on pathogen transmission and subsequent mortalities in juvenile Pacific oysters *Crassostrea gigas*. *Aquaculture Environment Interactions* 3:257-273.
26. Renault T, Moreau P, Faury N, Pepin JF, Segarra A, 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.
27. Segarra A, Pepin JF, Arzul I, Morga B, Faury N, Renault T. 2010. Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. *Virus Res* 153:92-9.
28. de Lorgeril J, Lucasson A, Petton B, Toulza E, Montagnani C, Clerissi C, VidalDupiol J, Chaparro C, Galinier R, Escoubas J-M, Haffner P, Dégremont L, Charrière GM, Lafont M, Delort A, Vergnes A, Chiarello M, Faury N, Rubio TP, Leroy MA, Pérignon A, Régler D, Morga B, Alunno-Bruscia M, Boudry P, Le Roux F, Destoumieux-Garzón D, Gueguen Y, Mitta G. 2018. Immune-suppression by OsHV-1 viral infection causes fatal bacteraemia in Pacific oysters. *Nature Communications* 9:4215.
29. de Lorgeril J, Petton B, Lucasson A, Perez V, Stenger PL, Degremont L, Montagnani C, Escoubas JM, Haffner P, Allienne JF, Leroy M, Lagarde F, Vidal-Dupiol J, Gueguen Y, Mitta G. 2020. Differential basal expression of immune genes confers *Crassostrea gigas* resistance to Pacific oyster mortality syndrome. *BMC Genomics* 21:63.
30. Dégremont L, Nourry M, 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.

31. Delmotte J, Chaparro C, Galinier R, de Lorgeril J, Petton B, Stenger PL, Vidal-Dupiol J, Destoumieux-Garzon D, Gueguen Y, Montagnani C, Escoubas JM, Mitta G. 2020. Contribution of Viral Genomic Diversity to Oyster Susceptibility in the Pacific Oyster Mortality Syndrome. *Frontiers in Microbiology* 11:1579.
32. Martenot C, Oden E, Travaille E, Malas JP, 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.
33. Burioli EA, Prearo M, Riina MV, Bona MC, Fioravanti ML, Arcangeli G, 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.
34. Davison AJ, Trus BL, Cheng N, Steven AC, Watson MS, Cunningham C, Le Deuff RM, Renault T. 2005. A novel class of herpesvirus with bivalve hosts. *J Gen Virol* 86:41-53.
35. Burioli EAV, Prearo M, Houssin M. 2017. Complete genome sequence of Ostreid herpesvirus type 1 microVar isolated during mortality events in the Pacific oyster *Crassostrea gigas* in France and Ireland. *Virology* 509:239-251.
36. Rambaut A, Pybus OG, Nelson MI, Viboud C, Taubenberger JK, Holmes EC. 2008. The genomic and epidemiological dynamics of human influenza A virus. *Nature* 453:615-9.
37. Allicock OM, Lemey P, Tatem AJ, Pybus OG, Bennett SN, Mueller BA, Suchard MA, Foster JE, Rambaut A, Carrington CV. 2012. Phylogeography and population dynamics of dengue viruses in the Americas. *Mol Biol Evol* 29:1533-43.
38. Theze J, Li T, du Plessis L, Bouquet J, Kraemer MUG, Somasekar S, Yu G, de Cesare M, Balmaseda A, Kuan G, Harris E, Wu CH, Ansari MA, Bowden R, Faria NR, Yagi S, Messenger S, Brooks T, Stone M, Bloch EM, Busch M, Munoz-Medina JE,

- Gonzalez-Bonilla CR, Wolinsky S, Lopez S, Arias CF, Bonsall D, Chiu CY, Pybus OG. 2018. Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico. *Cell Host Microbe* 23:855-864 e7.
39. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pepin J, Posada D, Peeters M, Pybus OG, Lemey P. 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 346:56-61.
40. Martin MA, VanInsberghe D, Koelle K. 2021. Insights from SARS-CoV-2 sequences. *Science* 371:466-467.
41. Sanjuán R, Domingo-Calap P. 2019. Genetic Diversity and Evolution of Viral Populations. doi:10.1016/b978-0-12-809633-8.20958-8.
42. Chase JM, McGill BJ, McGlinn DJ, May F, Blowes SA, Xiao X, Knight TM, Purschke O, Gotelli NJ. 2018. Embracing scale-dependence to achieve a deeper understanding of biodiversity and its change across communities. *Ecol Lett* 21:17371751.
43. Al Khatib HA, Benslimane FM, Elbashir IE, Coyle PV, Al Maslamani MA, Al-Khal A, Al Thani AA, Yassine HM. 2020. Within-Host Diversity of SARS-CoV-2 in COVID-19 Patients With Variable Disease Severities. *Front Cell Infect Microbiol* 10:575613.
44. Domingo E, Sheldon J, Perales C. 2012. Viral Quasispecies Evolution. *Microbiology and Molecular Biology Reviews* 76:159-216.
45. Morse AM, Calabro KR, Fear JM, Bloom DC, McIntyre LM. 2017. Reliable Detection of Herpes Simplex Virus Sequence Variation by High-Throughput Resequencing. *Viruses* 9.
46. Burioli EAV, Varello K, Lavazza A, Bozzetta E, Prearo M, 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. *J Fish Dis*  
doi:10.1111/jfd.12883.
47. Buestel D, Ropert M, Prou J, Gouilletquer P. 2009. HISTORY, STATUS, AND FUTURE OF OYSTER CULTURE IN FRANCE. *Journal of Shellfish Research* 28:813-820.
  48. Lupo C, Ezanno P, Arzul I, Garcia C, Jadot C, Joly J-P, Renault T, Bareille N. How network analysis of oyster movements can improve surveillance and control programs of infectious diseases?, p. *In* (ed), *Frontiers*,
  49. Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, Cornell SJ, Kappey J, Wilesmith J, Grenfell BT. 2001. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* 294:813-7.
  50. Lapegue S, Heurtebise S, Cornette F, Guichoux E, Gagnaire PA. 2020. Genetic Characterization of Cupped Oyster Resources in Europe Using Informative Single Nucleotide Polymorphism (SNP) Panels. *Genes (Basel)* 11.
  51. Rohfritsch A, Bierne N, Boudry P, Heurtebise S, Cornette F, Lapegue S. 2013. Population genomics shed light on the demographic and adaptive histories of European invasion in the Pacific oyster, *Crassostrea gigas*. *Evol Appl* 6:1064-78.
  52. Vendrami DLJ, Houston RD, Gharbi K, Telesca L, Gutierrez AP, Gurney-Smith H, Hasegawa N, Boudry P, Hoffman JI. 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.
  53. Alfaro AC, Nguyen TV, 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.



54. Avarre JC. 2017. Editorial: Molecular Tracing of Aquatic Viruses: Where Epidemiology Needs to Meet Genomics. *Front Microbiol* 8:1498.
55. Petton B, Bruto M, James A, Labreuche Y, Alunno-Bruscia M, 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.
56. Dupont S, Lokmer A, Corre E, Auguet JC, Petton B, Toulza E, Montagnani C, Tanguy G, Pecqueur D, Salmeron C, Guillou L, Desnues C, La Scola B, Bou Khalil J, de Lorgeril J, Mitta G, Gueguen Y, Escoubas JM. 2020. Oyster hemolymph is a complex and dynamic ecosystem hosting bacteria, protists and viruses. *Animal Microbiome* 2.
57. Webb SC, Fidler A, 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.
58. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
59. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-20.
60. Breitwieser FP, Baker DN, Salzberg SL. 2018. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biology* 19.
61. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* 11:e0163962.
62. Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PW, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, Hedgecock D, Xu Z, Liu Y, Domazet-Lošo T, Du Y, Sun X, Zhang S, Liu B, Cheng P, Jiang X, Li J, Fan D, Wang W, Fu W, Wang T, Wang B, Zhang J, Peng Z, Li Y, Li

- N, Wang J, et al. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* 490:49-54.
63. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357-9.
64. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 5:824-834.
65. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403-10.
66. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27:578-9.
67. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30:3059-66.
68. Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16:276-7.
69. Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol* 14:e1005944.
70. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268-74.
71. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587-589.

72. Yu G, Smith DK, Zhu H, Guan YY, Lam TTY, 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.
73. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv.org.
74. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 32:1749-51.
75. Tan A, Abecasis GR, Kang HM. 2015. Unified representation of genetic variants. *Bioinformatics* 31:2202-4.
76. R Core Team. 2018. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
77. Chen H, Boutros PC. 2011. VennDiagram: a package for the generation of highlycustomizable Venn and Euler diagrams in R. *BMC Bioinformatics* 12:35.
78. Lex A, Gehlenborg N, Strobel H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of Intersecting Sets. *IEEE Trans Vis Comput Graph* 20:1983-92.
79. R-Studio-Team. 2015. RStudio: Integrated Development Environment for R., RStudio Inc., Boston, MA.
80. Wickham H, François R, Henry L, Müller K. 2018. dplyr: A Grammar of Data Manipulation, RStudio, <https://dplyr.tidyverse.org/reference/dplyr-package.html>.
81. Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York,
82. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Golemund G, 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