Genetic diversity and connectivity of the Ostreid herpesvirus 1 populations in France: A first attempt to phylogeographic inference for a marine mollusc disease

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

[1]IHPE, CNRS, Ifremer, UPVD, University of Montpellier, Montpellier F-34095, France, [2]Ifremer, RBE-ASIM, Station La Tremblade, La Tremblade F-17390, France, [3]IHPE, CNRS, Ifremer, UPVD, University of Montpellier, Perpignan F-66000, France, [4]Ifremer, CNRS, IRD, Ifremer, LEMAR UMR 6539 Université de Bretagne Occidentale, Argenton-en-Landunvez F-29840, France and [5]ISEM, IRD, CNRS, University of Montpellier, Montpellier F-34095, France

[†]Authors contributed equally to this work.
[‡]https://orcid.org/0000-0001-6899-7052
[§]https://orcid.org/0000-0001-5945-4689
[**]https://orcid.org/0000-0002-6138-1148
*Corresponding authors: E-mail: Maude.Jacquot@ifremer.fr; Jean.Michel.Escoubas@ifremer.fr

Abstract

The genetic diversity of viral populations is a key driver of the spatial and temporal diffusion of viruses; yet, studying the diversity of whole genomes from natural populations still 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 (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 an innovative combination of bioinformatics tools, we *de novo* assembled twenty-one OsHV-1 new genomes. Combining quantification of major and minor genetic variations, 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. We demonstrate that phylodynamic approaches can be applied to aquatic DNA viruses to determine how epidemiological, immunological, and evolutionary processes act and potentially interact to shape their diversity patterns.

Key words: herpesvirus; OsHV-1; Pacific oyster; phylogeography; non-redundant genomes; genetic diversity; viral spread; minor variations.

1. 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 variations, known as viral quasispecies. This genetic organization currently encumbers our understanding of viral diseases and their evolution, and it impedes the straightforward characterization of virus population structure (Vignuzzi et al. 2006; Lauring and Andino 2010). Studies have shown that the level of genetic diversity within these viral populations likely influences viral pathogenicity, dissemination, and

host immune evasion. 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 to certain large DNA viruses, such as herpesviruses, whose genomic variability rivals that of many RNA viruses (Renzette et al. 2013; Hammoumi et al. 2016; Renner and Szpara 2018). For example, human cytomegalovirus shows considerable inter-host and intra-host genetic divergence across tissue compartments and times of infection (Renzette et al. 2014, 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 (Stentiford et al. 2012; Pernet et al. 2016; Burge, Shore-Maggio, and Rivlin 2017). 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 (Friedman et al. 2005; Moss et al. 2007; Vasquez-Yeomans, Garcia-Ortega, and Caceres-Martinez 2010; Peeler et al. 2012; Jenkins et al. 2013; Barbosa Solomieu, Renault, and Travers 2015; Gittenberger, Voorbergen-Laarman, and Engelsma 2016; Abbadi et al. 2018; Barbieri et al. 2019). In 2010, the causative agent of these massive mortalities was identified: it is an emerging genotype of a herpeslike virus named Ostreid herpesvirus 1 (OsHV-1) (Segarra et al. 2010; Pernet et al. 2012; Renault et al. 2012; Paul-Pont, Dhand, and Whittington 2013; Petton et al. 2013). Two major genetic factors seem to affect the severity of POMS: the genetic background of the oysters (Dégremont, Nourry, and Maurouard 2015; de Lorgeril et al. 2018, 2020) and OsHV-1 genetic diversity (Martenot et al. 2011; Delmotte et al. 2020; Friedman et al. 2020). 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, Renault, and Travers 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 polymerase chain reaction (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, Prearo, and Houssin 2017). To date, no study has specifically looked at the links between the geographic distribution of OsHV-1 and its genetic diversity at the whole-genome level. 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) (Davison et al. 2005; Burioli, Prearo, and Houssin 2017; Abbadi et al. 2018). 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 Acquired Immune Deficiency Syndrome virus (Faria et al. 2014), and more recently Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-Cov-2) (Martin, VanInsberghe, and Koelle 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 in viral populations of OsHV-1 μVar encountered in oyster-farming areas both within and

between host individuals, to better understand the dissemination 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 twenty-one 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 trait approach allows us to propose a scenario explaining the phylogenetic relationships between viral populations encountered in the three farming areas. This novel combination of approaches, together with epidemiological information, holds promise for understanding OsHV-1 evolution and epidemic dynamics and for identifying transmission patterns between oyster-producing regions. These data can be highly useful for developing novel disease management strategies.

2. Results
2.1 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 per cent 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 $\pm$16.7 SD; Fig. 1B; Supplementary 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, Prearo, and Houssin 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 per cent) for MO ind3 to 11.94 M reads (13.21 per cent) for Th ind10 (Fig. 1B; Supplementary Table S1). Therefore, OsHV-1 genome coverage varied from 480× to 8,745× for MO ind3 and Th ind10, respectively.

A rarefaction analysis was run using OsHV-1 μVar A as a reference genome 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× (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 deduced from rarefaction analysis between the three environments showed that (1) the number of SNPs in viruses from MO (170 ± 27.7) was significantly higher than that of viruses from Br (136.4 ± 5.9; $P \le 0.001$) and Th (143.8 ± 5.8; $P \le 0.05$) and (2) 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; $P \le 0.0001$) (Fig. 1D; Supplementary 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.
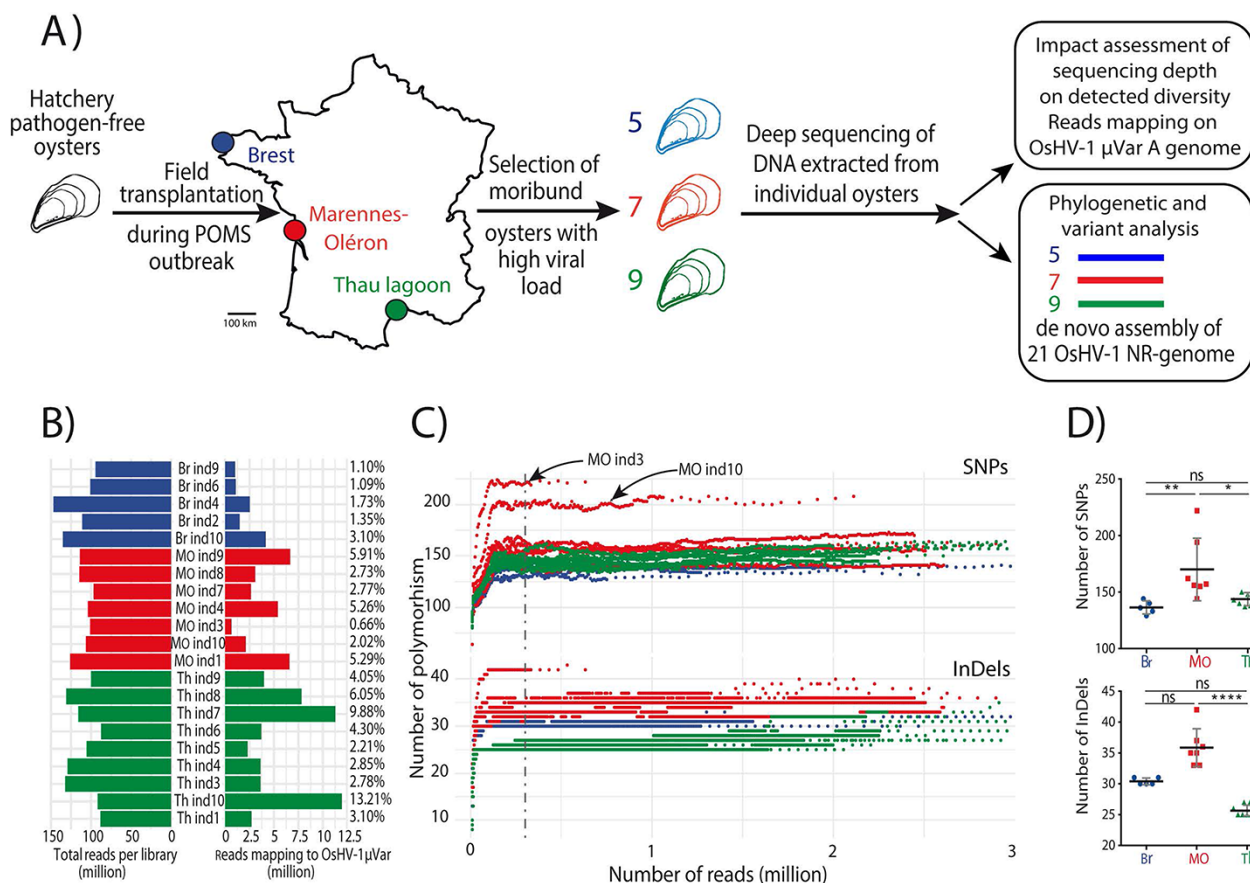
Figure 1. Variability in sequencing depth does not impair the analysis of OsHV-1 diversity. (A) Schematic representation of the sampling of OsHV-1 genetic diversity in the three main French oyster-farming areas during the POMS event in 2018 (see Section 4 for more details). Each color represents a farming area, with Br in blue (five individual samples), MO in red (seven individual samples), and Th Lagoon (Th) in green (nine 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 showing the accumulation of polymorphisms (SNPs and InDels top and bottom panels, respectively) during iterative variant calling analysis of the twenty-one libraries on the OsHV-1 μVar A genome. 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, $*P \leq 0.05$, $**P \leq 0.01$ $****P \leq 0.0001$).

2.2 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 (UL) region (164,268 bp), flanked by two inverted repeats (repeat long (RL), 7,338 bp), and the unique short (US) region (3,370 bp), flanked by two inverted repeats (repeat short (RS), 9,777 bp), and two X regions (1,510 bp) (Fig. 2A, upper panel). Considering the high sequence identity within each set of repeats (between 99.9 and 100 per cent), removing one of the repeats makes it possible to reliably map the reads from these repeated regions without affecting the overall genome composition. Removing a copy of each repeated region led to an NR OsHV-1 μVar A genome (NR-genome) of 186,262 bp, which represents 91 per cent of the full-length genome and contains all the coding sequences (Fig. 2A, lower panel).

Using this approach, twenty-one NR-genomes were assembled *de novo* from the sequencing data generated during this study, as well as six other OsHV-1 (belonging to the *Malacoherpesviridae* family and *Ostreavirus* genus) published genomes. A phylogenetic analysis of these twenty-seven NR-genomes revealed that the twenty-one new OsHV-1 isolates were phylogenetically closer to OsHV-1 μVar A and B isolates (Supplementary Fig. S1) than to the OsHV-1 reference genome (AY509253.2), suggesting that they belong to the μVar genotype.

Among newly sequenced isolates, those originating from Th and Br were grouped into two distinct clusters reflecting their geographic origin (Fig. 2B, Clusters 1 and 2). Conversely, MO OsHV-1 isolates were distributed over these two clusters: five in Cluster 1 and two in Cluster 2. The phylogenetic distance among Th or Br isolates (average $2.47 \times 10^{-5} \pm 4.68 \times 10^{-5}$ SD and $1.43 \times 10^{-4} \pm 7.91 \times 10^{-5}$ SD, respectively) was lower than that observed among the MO isolates (average $5.00 \times 10^{-3} \pm 4.21 \times 10^{-3}$), indicating greater viral diversity in MO than in the other two farming areas.

2.3 Comparative genomics of the twenty-one NR genomes confirms that OsHV-1 diversity is higher in MO than in the other two farming areas

To compare the genetic diversity at the whole-genome scale, we used the twenty-one NR-genomes to construct an average NR genome (Avg-genome) (see Section 4.5 for more details).
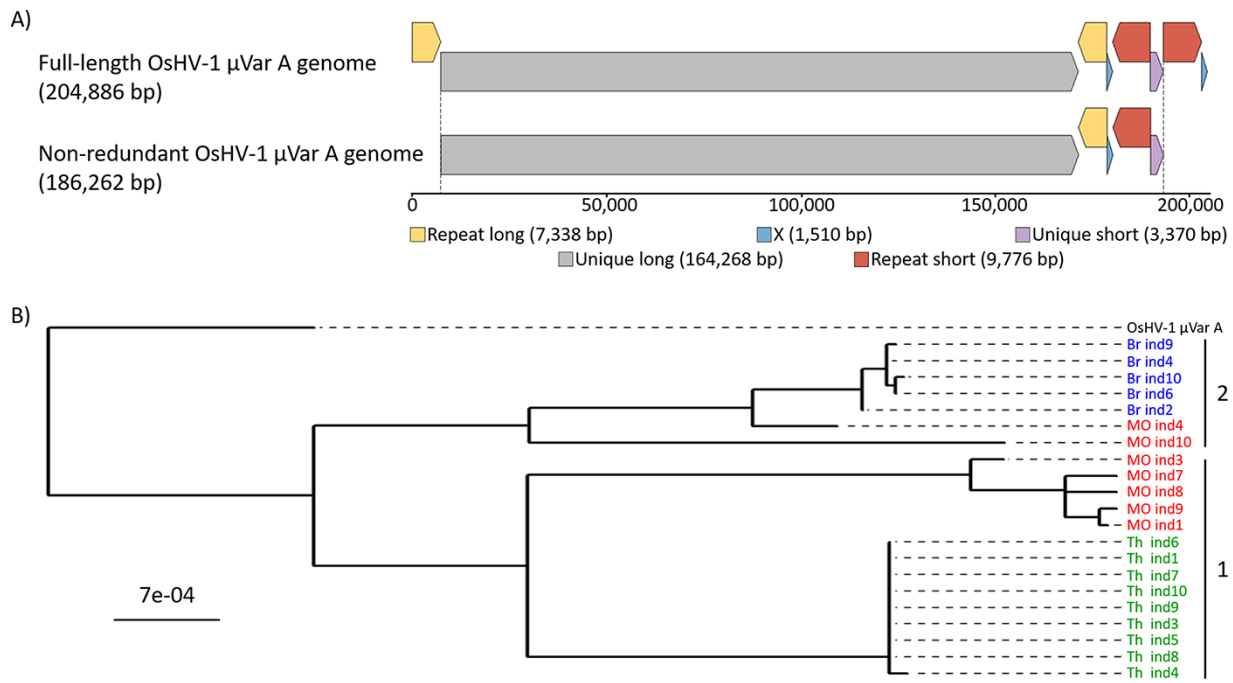
Figure 2. Phylogenetic relationship between the twenty-one *de novo* assembled and other OsHV-1 NR genomes. (A) Full-length (upper panel) and NR (lower panel) genomes of OsHV-1 μVar A (KY242785). The OsHV-1 μVar A genome consists of two unique regions, referred to as UL (in grey) and US (in purple), and three repeat regions, referred to as RL (in yellow), RS (in orange), and X region (X, in blue). The first two repeat regions each have copies at a terminal locus (TRL and TRS) and an internal locus (IRL and 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 the 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) Shown is a maximum-likelihood phylogenetic tree of the twenty-two OsHV-1 NR-genomes rooted in OsHV-1 μVar A. The twenty-one isolates from this study are colored according to their geographic origin, Br in blue, MO in red, and Th Lagoon in green. Within those, two clusters were defined based on genetic distances (numbered on the right).

The comparison of the twenty-one NR-genomes to the Avg-genome identified a total of 399 genetic variations (SNPs and InDels, hereafter called variations), and Spearman's correlation coefficient (-0.382; $P = 0.087$) indicates that there is no relationship between the viral load and the number of variations per sample (Supplementary Fig. S2).

The 399 variations spread over 310 positions, and twenty-two of them contained from two to seven allelic variations (Supplementary Table S3). Prediction of the possible functional effects of the 399 variations revealed that 216 (54.1 per cent) were distributed over 94 ORFs (summarized in Supplementary Fig. S3 and detailed in Supplementary Table S3). The twenty-one variations predicted to have a high impact (frameshift or start lost) were distributed over eleven ORFs. Unfortunately, these results are not very informative, given that six ORFs were annotated 'unknown' due to the absence of homologies with any known proteins and the five others were annotated according to domain homologies suggesting a protein localization or a putative function.

The average number of variations per genome was significantly lower in samples from Th (63.7 ± 2.7) than from MO (92.1 ± 25.8; $P < 0.05$) and Br (120.8 ± 2.8; $P < 0.001$), whereas the average number of variations in MO was not significantly different from Br (Fig. 3A, Supplementary Table S3). However, viruses from MO showed the highest diversity, with 290 variations accounting for 72.7 per cent of the total diversity. Among the variations, 187 (64.5 per cent) were specific to MO, and 114 were observed in one sample only (singletons) (Fig. 3B, Supplementary Table S3). None of these 187 variations were shared by all the samples from this area. The majority of MO singleton variations were observed in two samples: ind10 (fifty-six) and ind4 (twenty-three) (Fig. 3C). For

Th and Br farming areas, 82 and 132 variations were identified, among which 70 and 38 were area-specific, respectively (Fig. 3B, Supplementary Table S3). In contrast within MO, 57.9 per cent (22/38) and 75.5 per cent (53/70) of the variations were common to all the samples of Br and Th, respectively (Fig. 3B, Supplementary Table S3). Among the ninety-two variations shared by Br and MO, eighty-seven (94.6 per cent) were shared by two MO individuals (ind10 and ind4) and all Br individuals (Fig. 3C, purple arrow). These two MO individuals also shared five variations with all Th individuals (Fig. 3C, red arrows).

Interestingly, these major variations were only characterized in a small number of individuals within the samples analyzed; indeed, 129 (32 per cent) of them are found in only one individual, and only 2 (0.5 per cent) are shared by more than 50 per cent of individuals.

2.4 Within-individuals diversity of OsHV-1: the minor variations are more frequent in MO than in the other two farming areas

To fully characterize OsHV-1 genetic diversity, we performed variant calling analyses, using the Avg-genome as the reference, to quantify both major (frequency >50 per cent) and minor (frequency <50 per cent) variations within each of the twenty-one individual samples. Among the 399 variations identified as major variations, 132 had variable frequencies (above or below 50 per cent) across samples; moreover, 208 additional minor variations were identified (Fig. 4A, Supplementary Table S4). These 208 minor variations are spread over 164 positions of the Avg-genome, and 17 of these positions are showing 2–12 allelic variations (Supplementary Table S4). SnpEff predicts that 72 (18.0 per cent) of the 208
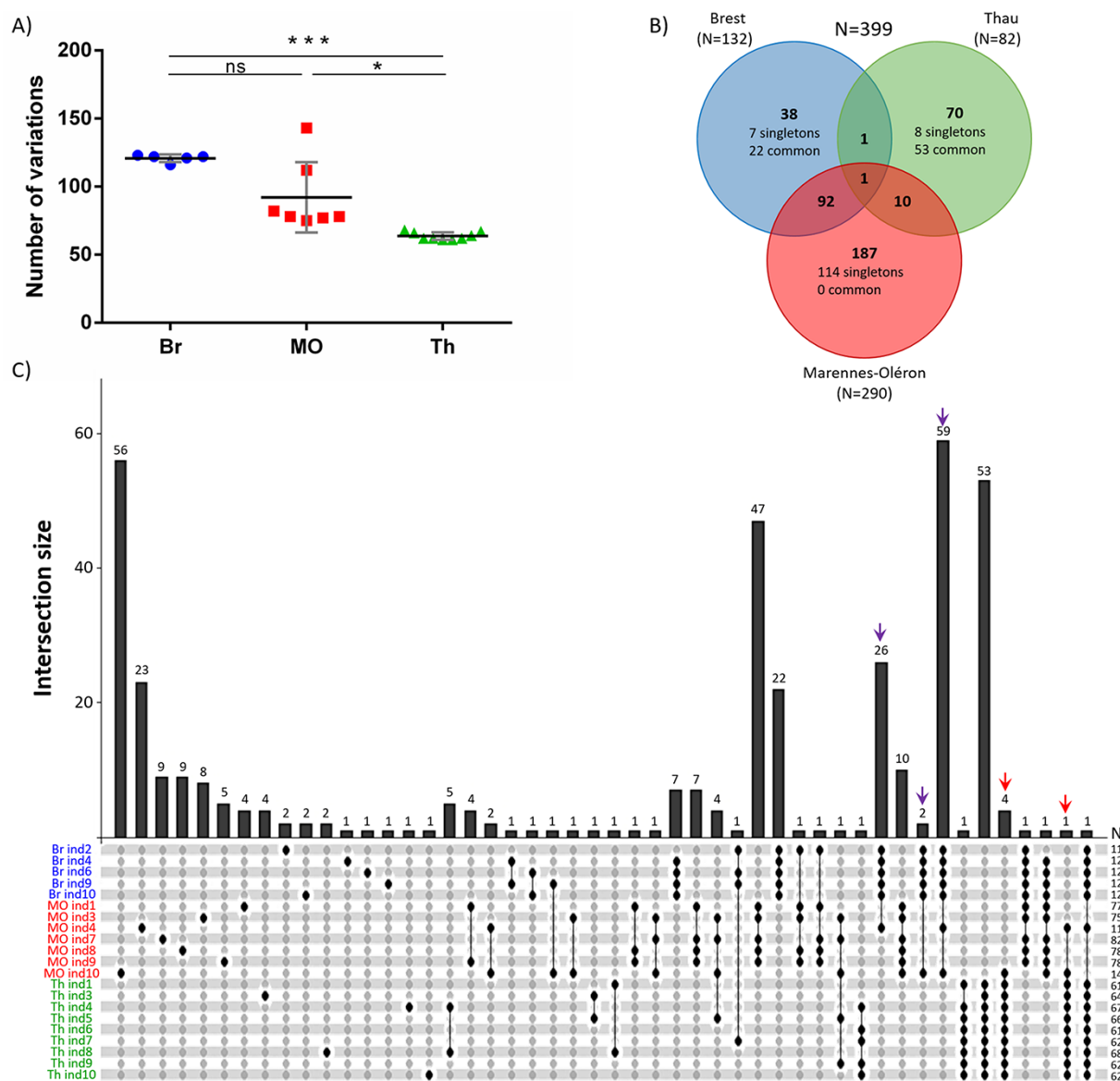
Figure 3. Comparative genomics of the twenty-one NR genomes of OsHV-1 isolated from three oyster-farming areas. (A) Distribution of the 399 variations between the twenty-one individuals and the three farming areas. Significance was calculated according to Kruskal–Wallis and Dunn's multiple comparison tests (ns: non-significant, $*P < 0.05$, $***P < 0.001$). (B) Venn diagram summarizing the distribution of variations between the three farming areas. Color depends on the farming area with Br in blue, MO in red, and Th Lagoon in green. The numbers of variations per farming area are indicated in parentheses. Singletons are variations that occur in only one sample, and the term 'common' indicates variations present in all the sampled individuals at a given location. (C) UpSet plot of variation distribution among the twenty-one NR-genomes. The bottom panel shows the specific combinations (or intersections), and the vertical bars indicate the number of variations within these combinations. Colored arrows on the top indicate variations 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 variations per sample.

variations were distributed over 34 predicted ORFs (summarized in Supplementary Fig. S4 and detailed in Supplementary Table S4). The seven variations predicted to have a high impact (frameshift or start gained) were distributed in six ORFs. But, again, these results were not very informative, given that four of these ORFs were annotated 'unknown' and the two others were annotated according to domain homologies suggesting a membrane location of the predicted protein.

Frequencies of the 208 strictly minor variations ranged from 5.0 to 48.7 per cent (mean $= 9.57 \pm 6.52$ per cent SD). The number of minor variations was lower in oysters from Br (88) than in oysters from Th (98) or MO (114) (Fig. 4B, Supplementary Table S4).

Fifty-five of these 208 minor variations (24.5 per cent) were found in all three farming areas. Moreover, all site-specific variations were found in only one individual oyster (singletons).

A hierarchical clustering analysis applied to the minor variations showed that individuals did not cluster according to their geographic origin (Fig. 4C). In some cases, we even observed two individuals from different farming areas clustering together (MO ind10 and Br ind2; Fig. 4C), suggesting that some OsHV-1 found in these individuals had a common origin. Figure 4C also revealed that the fifty-one variations shared by the three farming areas (common) were found in a majority of sampled individuals (mean $= 13.9 \pm 5.7$ SD), suggesting that they had not arisen within
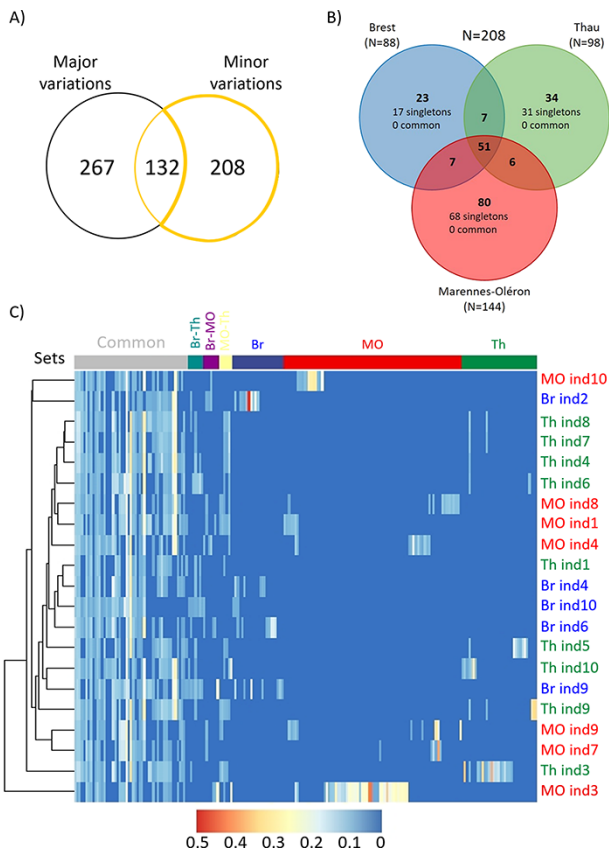
Figure 4. Variant calling analysis and minor variation distribution in the three farming areas. (A) Quantification of both major and minor variations using a variant calling analysis using the Avg-genome as the reference. (B) Venn diagram summarizing the distribution of the 208 strictly minor variations (frequency <50 per cent) between the three farming areas (Br in blue, MO in red, and Th Lagoon in green). The number of minor variations within each farming area is indicated in parentheses. Singletons correspond to variations that occur in only one oyster sample, whereas the term 'common' corresponds to variations present in all the oyster samples within the dataset. (C) Heatmap of the allelic frequency of strictly minor variations across the twenty-one samples with a gradient from 0 (blue) to 0.5 (red). Hierarchical clustering using the Euclidean distance of the allele frequency is displayed on the top. Each line of the heatmap corresponds to a strictly minor variation. Variation sets on the top are derived from the Venn diagram. Samples are indicated on the right and colored according to their geographic origin.

individual oysters, but had been transmitted horizontally to other oysters.

2.5 Distribution of within and between individual genetic variations among samples reveals oyster-farming areas' connectivity

The 132 variations that had variable frequencies across samples (Fig. 4A) were present as minor variations 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 125 of these minor variations. Sample MO ind10 contained thirty minor variations (with a frequency ranging from 5 to 7.2 per cent; Fig. 5B; Supplementary Table S5) that were major variations in samples MO ind1/3/7/8/9. For instance, the variation at position 65,003 (T > A) of the Avg-genome had a frequency of 7 per cent in MO ind10, whereas it varied from 76 to 99 per cent in MO ind1/3/7/8/9 (Fig. 5C). Similar fluctuations in variation frequency were observed for the other twenty-nine variations

(Supplementary Table S5). These thirty variations highlighted a clear genetic connectivity 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 ninety-five minor variations that corresponded to major variations in MO ind10, forty-six that corresponded to major variations in MO ind4 and all Br samples, and four that corresponded to major variations in all Th samples (Fig. 5A and E). For instance, the variation at position 57,080 (T > C) in the Avg-genome had a frequency of 24 per cent in MO ind3, whereas its frequency was 93 per cent in MO ind10 and over 99 per cent in all Th samples (Fig. 5F; Supplementary Table S5). Another example is the variation at position 65,610 (C > T) in the Avg-genome: its frequency grew from 24 per cent in MO ind3 to 92 per cent in MO ind10, and 99 per cent in MO ind4 and all Br samples (Fig. 5F, Supplementary Table S5). Again, minor variations highlighted genetic connectivity between samples from different farming areas and from phylogenetic Clusters 1 and 2 (Fig. 5G; Supplementary Table S5).

2.6 Ancestral state reconstruction of discrete trait
The analysis of minor variations 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 indicate that all the OsHV-1 variations characterized in this study originated from the MO farming area. This is congruent with ancestral state reconstruction of the discrete locations (Fig. 6A, Supplementary Fig. S5) that showed evidence for MO being the ancestral location of the twenty-one OsHV-1 isolates sequenced. Moreover, the Bayesian stochastic search variable selection approach suggested significant transition rates of OsHV-1 between the three sampling sites regardless of the alignment used to perform the analysis, i.e. full NR-genomes or NR-genomes with regions potentially impacted by recombination events removed (Supplementary Table S6). Indeed, two rates were inferred with decisive supports: from MO to Th (Bayes factor (BF) = 3,683) and from MO to Br (BF = 11,051) (Fig. 6B).

3. Discussion
The accurate description of viral genetic diversity (major and minor variations) 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 infected moribund oysters. To optimize the characterization of OsHV-1 variations, 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 ≈3,000×. However, because OsHV-1 coverage varied among samples (ranging from 480× to 8,745×), 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
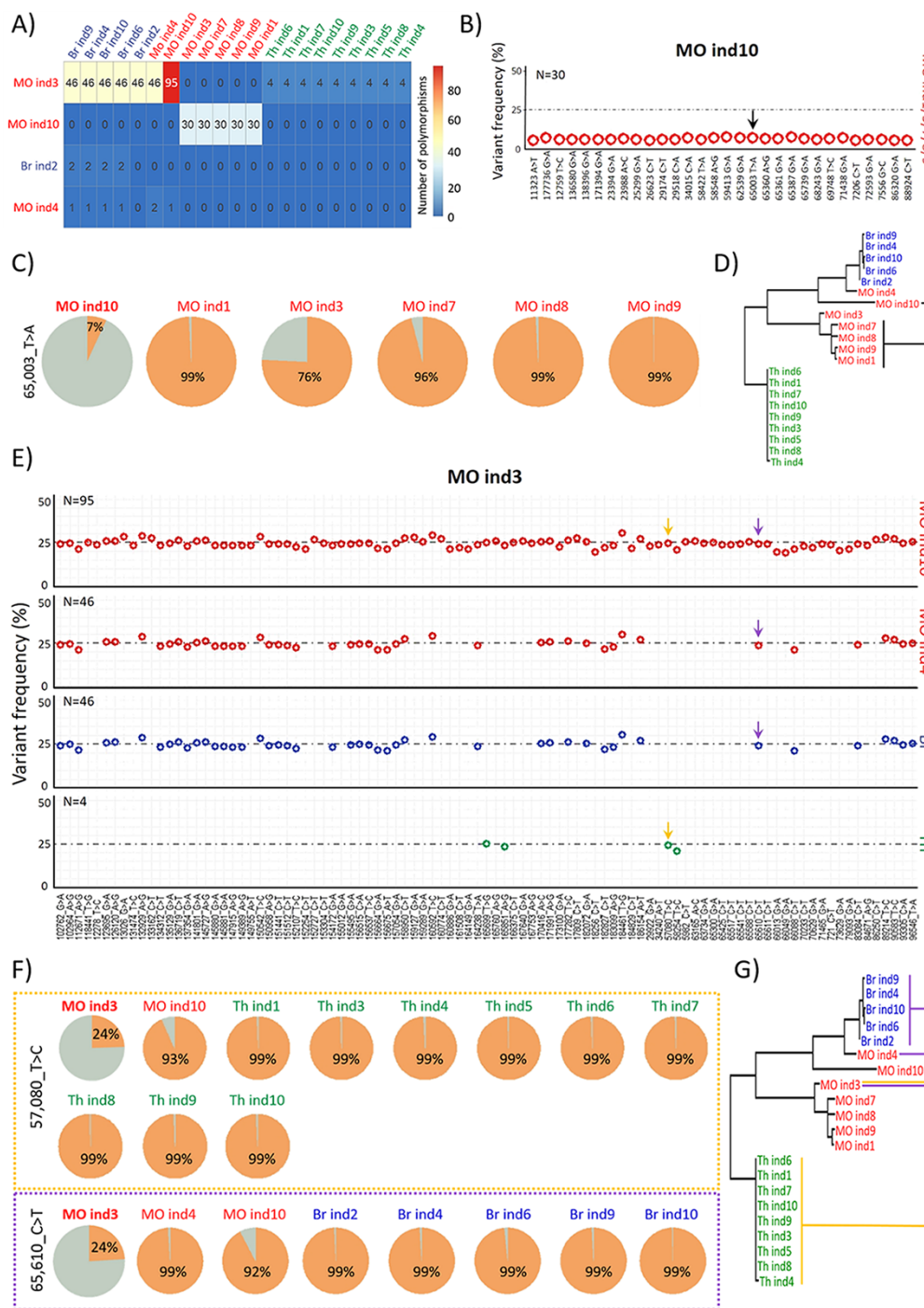
Figure 5. Some minor OsHV-1 variations found in the MO farming area are major variations in the other two farming areas. (A) Matrix of the 132 variations, which were major or minor, according to the sample. Rows display the four samples in which the 132 variations were minor, and columns show all twenty-one samples. Intersections give the number of shared variations with a minor (row) and major (column) frequency. For self-comparisons, no number is given. (B) Dot plot of the 30 minor variations of MO ind10, which were major variations in MO ind1/3/7/8/9. Positions on the Avg-genome and the type of SNPs are indicated on the x-axis. Variation frequency (in per cent) is indicated in ordinate. The black arrow indicates the variation that is detailed in (C). (C) Pie charts of the frequency of the variation found at position 65,003 of the NR-genome in MO ind1/3/7/8/9/10. (D) Phylogenetic tree of the twenty-one *de novo* assembled NR-genomes rooted in the midpoint. The black line shows the genetic connectivity established between MO ind10 and MO ind1/3/7/8/9 from variation frequencies at the thirty positions indicated in (B). (E) Dot plot showing the position and type of SNPs (x-axis) and variation frequency (y-axis) of the ninety-five minor variations of MO ind3 that were present as major variations in MO ind10 (ninety-five variations), in MO ind4 (forty-six variations), as well as in all individuals from Br (forty-six variations) and Th (four variations). Yellow and purple arrows indicate the two variations (57,080_T > C and 65,610_C > T, respectively) that are detailed in (F). (F) The minor variation found in MO ind3 at position 57,080 of the Avg-genome corresponds to major variations in MO ind10 and all Th samples (upper panel, framed in yellow). The minor variation found in MO ind3 at position 65,610 of the Avg-genome corresponds to major variations in MO ind4/10 and all Br samples (lower panel, framed in purple). (G) Phylogenetic tree of the twenty-one *de novo* assembled NR-genomes rooted in the midpoint. The yellow lines show the genetic connectivity established between MO ind3 and MO ind10 plus all Th samples based on the study of variation frequencies at the four positions indicated in (E). The purple lines show the genetic connectivity established between MO ind3 and MO ind4, MO ind10 plus all Br samples based on the study of variation frequencies at the ninety-five (MO ind4) and forty-six (MO ind10 and all Br samples) positions indicated in (E).
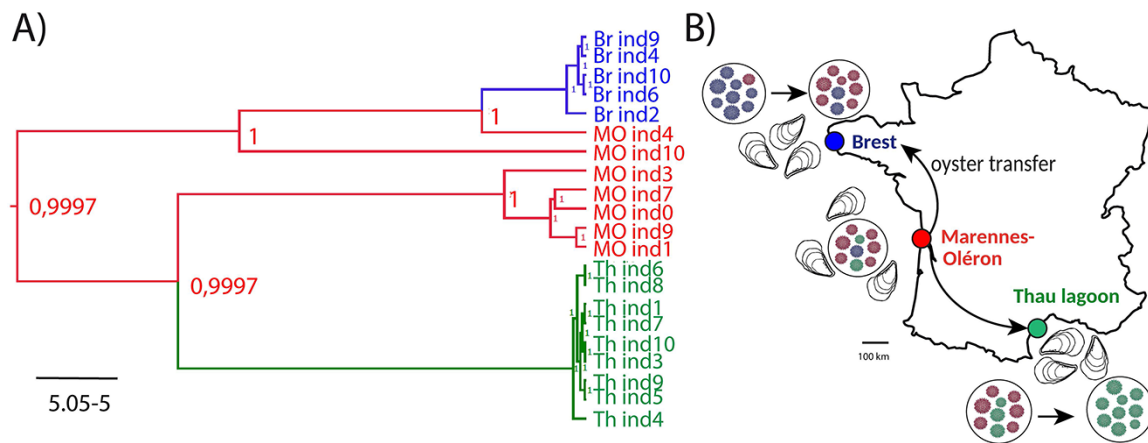
Figure 6. Graphical representation of OsHV-1 population connectivity in France. (A) MCC phylogeny of NR-genomes of OsVH-1 isolates and ancestral state reconstruction of the discrete locations: Br, MO and Th lagoon. The color of the branches represents the most probable location state of their descendent 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 the average number of substitutions per time unit. (B) Bayesian stochastic search variable selection was used to infer dispersal routes with non-zero rates using BFs. Routes that were inferred to have decisive support (BF > 100) are plotted on map (black arrows).

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 (Domingo, Sheldon, and Perales 2012; Al Khatib et al. 2020).

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 per cent of the genome of OsHV-1 μVar A. To solve this problem, we used an approach commonly used for herpesvirus genome analyses (Cunningham et al. 2010; Morse et al. 2017): contigs generated by *de novo* assembly were ordered using a reference-based approach to ultimately assemble NR genomes in which only one copy of each repeated element was kept. These NR 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 variations 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 herpesvirus, the Cyprinid herpesvirus 3 (Hammoumi et al. 2016). They also suggest that all OsHV-1 characterized in our 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, Prearo, and Houssin 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 Br and Th 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 for intra-host diversity

characterization in order to accurately study the epidemiology of DNA viruses (Renner and Szpara 2018).

Here, we identified major and minor variations within viral populations that substantially vary in frequency among sampled oyster individuals. Thus, ultra-deep sequencing can support the reconstruction of the true structure of a viral population and inform on relationships between individuals from the same farming area or between individuals from the three farming areas (Fig. 5D and G, respectively). Ancestral state reconstruction, used to infer the dynamics of dispersal of OsHV-1 populations across the three farming areas (Fig. 6A), along with the distribution of major and minor variations, suggests that the MO farming area is the main source of OsHV-1 viral diversity, from where variations are dispersed to the other farming areas. This hypothesis is in agreement with what is known about oyster transfers. Indeed, MO 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. 6B). 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 the transportation of infected oysters, similar to the spread of other diseases via livestock transfer (e.g. Keeling et al. 2001). Effectively several studies showed that wild-caught spats or oysters surviving an OsHV-1 infection contain the virus and are therefore able to transmit it when cohabiting with uninfected oysters (Schikorski et al. 2011; Dégremont and Benabdelmouna 2014; Agnew et al. 2020).

OsHV-1 diversity found in the MO Bay was higher than in the Bay of Br and Th Lagoon. This may be related to the fact that the MO Bay has become one of the largest French oyster-farming areas, with intensive shellfish cultivation and chronic overstocking (Buestel et al. 2009). While we acknowledge the potential bias resulting from small sampling size, this is coherent with large viral population size and increased opportunities for recombination. Mutation and recombination rates being good prognostic markers for the possible emergence of phenotypically different viruses, studying them at the different farming areas would help to better prevent future new OsHV-1 variation outbreaks. Moreover, a network analysis of oyster movements revealed that oysters are

moved up to nine 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).

One hypothesis to explain the lower diversity in the Bay of Br and Th Lagoon samples could be that only few oysters were infected during field transplantation, and then the infection would be spread in the experimental tanks. However, this hypothesis is unlikely: first because dead oysters were observed in the batches as soon as they were recovered from the field, which clearly shows that they were infected, and secondly because genetic variations present only in one individual (singletons) were characterized in samples from Br and Th. Given that the fate of new genetic variations is largely determined by host selection, another hypothesis to explain the lower diversity in the Bay of Br and Th Lagoon is that oyster populations exert a bottleneck on viral diversity imported from MO. Indeed, it can be speculated that some OsHV-1 variations imported from MO have not found oysters with a genetic background compatible with their replication and survival. 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 (Rohfritsch et al. 2013; Vendrami et al. 2019; Lapegue et al. 2020). 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, Nguyen, and Merien 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 variation selection. Therefore, a larger 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 (Pybus et al. 2012; Dellicour, Rose, and Pybus 2016).

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

4. Materials and methods
4.1 OsHV-1 field infection and oyster sampling
All *C. gigas* oysters used in the present study were produced in August 2017 in the Ifremer experimental facilities located in Argenton (Brittany, France), as SPF juveniles (Petton et al. 2015). This SPF oyster cohort was generated from 132 genitors (108 females and 24 males) derived from wild-caught spat collected between 2012 and 2015 (33 genitors for each year). To sample

OsHV-1 diversity, SPF oysters originating from the same batch were transplanted into three oyster-farming areas (~1,000 individuals per farm; average weight 1.5 g/oyster) during a disease outbreak when seawater temperature was above 16°C: (1) the Th Lagoon (Mèze, lat.: 43.379 and long.: 3.571) in May 2018, (2) the MO Bay (La Floride, lat.: 45.803 and long.: −1.153) in June 2018, and (3) the Bay of 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 (Petton et al. 2015; Dupont et al. 2020). For this reason, oysters were transferred back to laboratory facilities 5 days after transplantation. Dead oysters were removed from each batch, and the remaining oysters were then maintained in seawater tanks at 20°C (density of 20–25 oysters/l), 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 into 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.

4.2 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® double-stranded DNA 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 (Eurogentec 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, Fidler, and Renault 2007). The following program was applied: enzyme activation at 95°C for 10 min followed by forty 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 (five from Br, seven from MO, and nine 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).

4.3 Raw read quality control 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, Lohse, and Usadel 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://broad-institute.github.io/picard/ accessed 21 April 2022). Then, viral reads were extracted using KrakenUniq 0.5.8 (Breitwieser, Baker, and Salzberg 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.

4.4 Rarefaction analysis

In order to verify that OsHV-1 sequencing depth was sufficient to accurately characterize whole viral genomic diversity, we performed a rarefaction analysis for each of the twenty-one libraries. To this end, we performed an iterative variant calling analysis by subsampling and mapping the reads on the OsHV-1 *μ*Var A genome (GenBank: KY242785). In order to accelerate analyses, a nonlinear subsampling incrementation was performed. We started by subsampling 1,000 reads until reaching 10 per cent of the total number of reads of the library analyzed. Then, we subsampled 4,000, 11,000, 21,000, and 41,000 reads until reaching 30, 50, 80, and 100 per cent of the total number of reads of the library, respectively. After each step of subsampling/variant calling, variants were compared to the previously identified ones, and the new ones were classified in SNPs and InDels in order to increment rarefaction curves.

4.5 *De novo* assembly of NR 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 *μ*Var 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 9,777 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 and 4 kbp and made it possible to select the US region. The second group was composed of sequences with a size between 4 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 and made it possible to select the UL region. We used these three groups of scaffolds to construct 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 the assembly of sequences in the wrong orientation. Using this strategy, twenty-one OsHV-1 NR-genomes were assembled from the twenty-one individual oyster samples.

4.6 Whole-genome comparisons and phylogenetic analyses

The twenty-one NR-genomes were then aligned with the six NR-genomes derived from previously published OsHV-1 genomes (AY509253, KU096999, KP412538, KY242785, KY271630, MG561751 and MF509813) using MAFFT v7.475 with default parameters (Katoh et al. 2002). A maximum-likelihood phylogenetic inference was conducted on the twenty-seven NR-genomes using PhyML v3.1 (Guindon et al. 2009) with 100 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's information criterion. Similarly, in order to get a finer picture of phylogenetic relationships among samples, the twenty-one NR-genomes were aligned with OsHV-1 *μ*Var A (KY242785) genome only, and a maximum-likelihood tree was obtained using an HKY+G+I substitution model. The results were locally visualized using the R package ggTree v2.0.4 (Yu et al. 2016) and FigTree v1.4.4 (available at http://tree.bio.ed.ac.uk/software/figtree/ accessed 21 April 2022).

4.7 Genomic variability within and between farming areas

An Avg-genome of the twenty-one newly generated NR-genomes was built using the 'cons' argument from the EMBOSS suite v6.6.0.0 with the default settings (Rice, Longden, and Bleasby 2000). The NR-genomes were compared to the Avg-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 twenty-one NR-genomes.

To assess the genomic variability between the farming areas, both minor and major variations of each sequencing library (twenty-one) were called on the Avg-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, Abecasis, and Kang 2015), and split with vcflib v1.0.0. Considering the very rare occurrence of multi-nucleotide polymorphisms, these latter were counted as InDels. Variable positions with a frequency of >50 per cent (major variations) 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 minor variations (with a frequency of <50 per cent). A unique identifier composed of the position and sequence information (e.g. 128,737_A > G) of each minor variation was created. Set analyses of nucleotide variations were performed with a matrix approach 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. 2022) 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, 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.

4.8 Genomic annotation and variation effect predictions

Genomic annotation was performed using five gene prediction softwares: Prodigal (Hyatt et al. 2010), FragGeneScan (Rho et al., 2010; version 1.31), GeneMark with heuristic models

(Zhu, Lomsadze, and Borodovsky 2010; version 3.25), GeneMarkS for virus (Besemer, Lomsadze, and Borodovsky 2001; version 4.28), and finally ORF finder implemented in Geneious Prime (version 2022.0.1). All identified ORFs were manually sorted in Geneious Prime, and only those predicted by at least three gene predictors were retained for genomic annotation. Variant calling files (VCF) of major and minor variations were used to generate two global VCF containing the coordinates of the variations according to the Avg-genome. SnpEff (Cingolani et al. 2012; version 4.1l) was then used to predict the effect of major and minor variations using a home-made database constructed with the Avg-genome and its annotations.

4.9 Ancestral state reconstruction of discrete trait

To infer the geographic origin of our sampled isolates, sampling sites (Th, MO and Br) were modeled 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. 2014) 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 GTR+I+G model of nucleotide substitution. A strict molecular clock and a constant coalescent model were used, and an Markov chain Monte Carlo 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 per cent of the samples as burn-in. The BEAGLE library was used to increase computational speed (Ayres et al. 2012). BFs 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 BF > 100 (Jeffreys 1961).

To investigate the possible impact of recombination on the ancestral state reconstruction, the analysis was repeated on an alignment where genomic regions potentially impacted by recombination events were removed. To do so, we 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-genomes 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 three recombination events and removed from the alignment (213 SNPs remaining after pairwise deletion at gapped sites, 46 after complete deletion) before running jModelTest, BEAST, and SpreaD3 again.

Maximum clade credibility (MCC) trees were 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/ accessed 21 April 2022).

Data availability

The datasets generated from this study can be found in the Sequence Read Archive database BioProject accession number PRJNA681599 with submission ID SUB8642385. The twenty-one OsHV-1 genomes generated in this study are available in GenBank under the following accession numbers: Brest samples (from OM811577 to OM811581), Marennes-Oléron samples (from OM811582 to OM811588) and Thau samples (from OM811589 to OM811597). All the scripts generated in this study are freely available at https://github.com/propan2one/OshV-1-molepidemio accessed 21 April 2022. Complementary information is available from the corresponding authors upon reasonable request.

Supplementary data

Supplementary data is available at *Virus Evolution* online.

Author's contribution

J.-C.A., O.K., B.M., J.D., and J.-M.E. were involved in the study conception and design. B.M., B.P., C.P., J.D., and J.-M.E. were involved in the collection of samples and in the experimental work. J.D., C.P., M.J., and J.-M.E. were involved in bioinformatics and statistical analyses. J.D., R.G., C.M., M.J., and J.-M.E. drafted the manuscript, and all authors revised and approved the final manuscript.

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 genomes generally have lower nucleotide diversity than RNA viruses. In this study, by combining the use of both major and minor genetic variations with phylogeographic analyses of the oyster herpesvirus OsHV-1, we highlight genealogical relationships that are not 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.

References

Abbadi, M. et al. (2018) 'Identification of a Newly Described OsHV-1 Microvar from the North Adriatic Sea (Italy)', *The Journal of General Virology*, 99: 693–703.

Agnew, M. V. et al. (2020) 'Differential Mortality and High Viral Load in Naive Pacific Oyster Families Exposed to OsHV-1 Suggests Tolerance Rather Than Resistance to Infection', *Pathogens*, 9: 1057–75.

Akhtar, L. N. et al. (2019) 'Genotypic and Phenotypic Diversity of Herpes Simplex Virus 2 within the Infected Neonatal Population', *mSphere*, 4: e00590–18.

Al Khatib, H. A. et al. (2020) 'Within-Host Diversity of SARS-CoV-2 in COVID-19 Patients with Variable Disease Severities', *Frontiers in Cellular and Infection Microbiology*, 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: 1148–68.

Allicock, O. M. et al. (2012) 'Phylogeography and Population Dynamics of Dengue Viruses in the Americas', *Molecular Biology and Evolution*, 29: 1533–43.

Altschul, S. F. et al. (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology*, 215: 403–10.

Andrews, S. (2010) *FastQC: A Quality Control Tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ accessed 21 Apr 2022.

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

Ayres, D. L. et al. (2012) 'BEAGLE: An Application Programming Interface and High-performance Computing Library for Statistical Phylogenetics', *Systematic Biology*, 61: 170–3.

Barbieri, E. S. et al. (2019) 'First Detection of Ostreid Herpesvirus 1 in Wild *Crassostrea gigas* in Argentina', *Journal of Invertebrate Pathology*, 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*', *Journal of Invertebrate Pathology*, 131: 2–10.

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001) 'GeneMarkS: A Self-training Method for Prediction of Gene Starts in Microbial Genomes. Implications for Finding Sequence Motifs in Regulatory Regions', *Nucleic Acids Research*, 29: 2607–18.

Bielejec, F. et al. (2016) 'SpreaD3: Interactive Visualization of Spatiotemporal History and Trait Evolutionary Processes', *Molecular Biology and Evolution*, 33: 2167–9.

Boetzer, M. et al. (2011) 'Scaffolding Pre-assembled Contigs Using SSPACE', *Bioinformatics*, 27: 578–9.

Bolger, A. M., Lohse, M., and Usadel, B. (2014) 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data', *Bioinformatics*, 30: 2114–20.

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: 198.

Buestel, D. et al. (2009) 'History, Status, and Future of Oyster Culture in France', *Journal of Shellfish Research*, 28: 813–20.

Burge, C. A., Shore-Maggio, A., and Rivlin, N. D. (2017) 'Ecology of Emerging Infectious Diseases of Invertebrates'. In: Hajek, A. E. (ed.) *Ecology of Invertebrate Diseases*, pp. 587–612. John Wiley & Sons Ltd: Hoboken, New Jersey.

Burioli, E. A. et al. (2016) 'Ostreid Herpesvirus Type 1 Genomic Diversity in Wild Populations of Pacific Oyster *Crassostrea gigas* from Italian Coasts', *Journal of Invertebrate Pathology*, 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: 239–51.

Burioli, E. A. V. et al. (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*, 41: 1759–69.

Chase, J. M. et al. (2018) 'Embracing Scale-dependence to Achieve a Deeper Understanding of Biodiversity and Its Change across Communities', *Ecology Letters*, 21: 1737–51.

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.

Cingolani, P. et al. (2012) 'A Program for Annotating and Predicting the Effects of Single Nucleotide Polymorphisms, SnpEff: SNPs in the Genome of Drosophila Melanogaster Strain W(1118); Iso-2; Iso-3', *Fly*, 6: 80–92.

Cunningham, C. et al. (2010) 'Sequences of Complete Human Cytomegalovirus Genomes from Infected Cell Cultures and Clinical Specimens', *The Journal of General Virology*, 91: 605–15.

Darriba, D. et al. (2012) 'jModelTest 2: More Models, New Heuristics and Parallel Computing', *Nature Methods*, 9: 772.

Davison, A. J. et al. (2005) 'A Novel Class of Herpesvirus with Bivalve Hosts', *The Journal of General Virology*, 86: 41–53.

de Lorgeril, J. et al. (2018) 'Immune-suppression by OsHV-1 Viral Infection Causes Fatal Bacteraemia in Pacific Oysters', *Nature Communications*, 9: 4215.

——— et al. (2020) 'Differential Basal Expression of Immune Genes Confers *Crassostrea gigas* Resistance to Pacific Oyster Mortality Syndrome', *BMC Genomics*, 21: 63.

Dégremont, L., and Benabdelmouna, A. (2014) 'Mortality Associated with OsHV-1 in Spat *Crassostrea gigas*: Role of Wild-Caught Spat in the Horizontal Transmission of the Disease', *Aquacultural International*, 22: 1767–81.

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: 111–21.

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. 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', *Microbiology Molecular Biology Reviews*, 76: 159–216.

Dupont, S. et al. (2020) 'Oyster Hemolymph Is a Complex and Dynamic Ecosystem Hosting Bacteria, Protists and Viruses', *Animal Microbiome*, 2: 12.

Faria, N. R. et al. (2014) 'HIV Epidemiology. The Early Spread and Epidemic Ignition of HIV-1 in Human Populations', *Science*, 346: 56–61.

Friedman, C. S. et al. (2005) 'Herpes Virus in Juvenile Pacific Oysters *Crassostrea gigas* from Tomales Bay, California, Coincides with Summer Mortality Episodes', *Diseases of Aquatic Organisms*, 63: 33–41.

——— et al. (2020) 'Unraveling Concordant and Varying Responses of Oyster Species to Ostreid Herpesvirus 1 Variants', *The Science of the Total Environment*, 739: 139752.

Garrison, E., and Marth, G. (2012) *Haplotype-based Variant Detection from Short-read Sequencing*. arXiv:1207.3907 [q-bio.GN].

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: 105–9.

Guindon, S. et al. (2009) 'Estimating Maximum Likelihood Phylogenies with PhyML', *Methods in Molecular Biology (Clifton, N.J.)*, 537: 113–37.

Hammoumi, S. et al. (2016) 'Targeted Genomic Enrichment and Sequencing of CyHV-3 from Carp Tissues Confirms Low Nucleotide Diversity and Mixed Genotype Infections', *PeerJ*, 4: e2516.

Hyatt, D. et al. (2010) 'Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification', *BMC Bioinformatics*, 11: 119.

Jeffreys, H. (1961) *Theory of Probability*, 3rd edn. Clarendon Press: Oxford.

Jenkins, C. et al. (2013) 'Identification and Characterisation of an Ostreid Herpesvirus-1 Microvariant (Oshv-1 Micro-var) in *Crassostrea gigas* (Pacific Oysters) in Australia', *Diseases of Aquatic Organisms*, 105: 109–26.

Katoh, K. et al. (2002) 'MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform', *Nucleic Acids Research*, 30: 3059–66.

Keeling, M. J. et al. (2001) 'Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape', *Science*, 294: 813–7.

Langmead, B., and Salzberg, S. L. (2012) 'Fast Gapped-Read Alignment with Bowtie 2', *Nature Methods*, 9: 357–9.

Lapegue, S. et al. (2020) 'Genetic Characterization of Cupped Oyster Resources in Europe Using Informative Single Nucleotide Polymorphism (SNP) Panels', *Genes*, 11: 451–68.

Lauring, A. S., and Andino, R. (2010) 'Quasispecies Theory and the Behavior of RNA Viruses', *PLoS Pathogens*, 6: e1001005.

Lemey, P. et al. (2014) 'Unifying Viral Genetics and Human Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2', *PLoS Pathogens*, 10: e1003932.

——— et al. (2009) 'Bayesian Phylogeography Finds Its Roots', *PLoS Computational Biology*, 5: e1000520.

Lex, A. et al. (2014) 'UpSet: Visualization of Intersecting Sets', *IEEE Transactions on Visualization and Computer Graphics*, 20: 1983–92.

Lupo, C. et al. (2016) How Network Analysis of Oyster Movements Can Improve Surveillance and Control Programs of Infectious Diseases? AquaEpi (Aquatic Animal Epidemiology), Oslo, Norway.

Marcais, G. et al. (2018) 'MUMmer4: A Fast and Versatile Genome Alignment System', *PLoS Computational Biology*, 14: e1005944.

Martenot, C. et al. (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. et al. (2015) 'RDP4: Detection and Analysis of Recombination Patterns in Virus Genomes', *Virus Evolution*, 1: vev003.

Martin, M. A., VanInsberghe, D., and Koelle, K. (2021) 'Insights from SARS-CoV-2 Sequences', *Science*, 371: 466–7.

Morga, B. et al. (2021) 'Genomic Diversity of the Ostreid Herpesvirus Type 1 across Time and Location and among Host Species', *Frontiers in Microbiology*, 12: 711377.

Morse, A. M. et al. (2017) 'Reliable Detection of Herpes Simplex Virus Sequence Variation by High-Throughput Resequencing', *Viruses*, 9: 226.

Moss, J. A. et al. (2007) 'Pathogens in *Crassostrea ariakensis* and Other Asian Oyster Species: Implications for Non-native Oyster Introduction to Chesapeake Bay', *Diseases of Aquatic Organisms*, 77: 207–23.

Narasimhan, V. et al. (2016) 'BCFtools/RoH: A Hidden Markov Model Approach for Detecting Autozygosity from Next-generation Sequencing Data', *Bioinformatics*, 32: 1749–51.

Nkili-Meyong, A. A. et al. (2016) 'Contribution of Next-Generation Sequencing to Aquatic and Fish Virology', *Intervirology*, 59: 285–300.

Nurk, S. et al. (2017) 'metaSPAdes: A New Versatile Metagenomic Assembler', *Genome Research*, 5: 824–34.

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: 127–38.

Peeler, E. J. et al. (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: 136–43.

Pernet, F. et al. (2012) 'Mass Mortalities of Pacific Oysters *Crassostrea gigas* Reflect Infectious Diseases and Vary with Farming Practices in the Mediterranean Thau Lagoon, France', *Aquaculture Environmental Interactions*, 2: 215–37.

——— et al. (2016) 'Infectious Diseases in Oyster Aquaculture Require a New Integrated Approach', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371: 1689.

Petton, B. et al. (2015) '*Crassostrea gigas* Mortality in France: The Usual Suspect, a Herpes Virus, May Not Be the Killer in This Polymicrobial Opportunistic Disease', *Frontiers in Microbiology*, 6: 686.

——— et al. (2021) 'The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions', *Frontiers in Immunology*, 12: 630343.

——— et al. (2013) 'Temperature Influence on Pathogen Transmission and Subsequent Mortalities in Juvenile Pacific Oysters *Crassostrea gigas*', *Aquaculture Environment Interactions*, 3: 257–73. Inter-reseach science publisher.

Pybus, O. G. et al. (2012) 'Unifying the Spatial Epidemiology and Molecular Evolution of Emerging Epidemics', *Proceedings of the National Academy of Sciences of the United States of America*, 109: 15066–71.

R Core Team (2018) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing: Vienna, Austria.

Rambaut, A. et al. (2008) 'The Genomic and Epidemiological Dynamics of Human Influenza A Virus', *Nature*, 453: 615–9.

Renault, T. et al. (2012) 'Analysis of Clinical Ostreid Herpesvirus 1 (*Malacoherpesviridae*) Specimens by Sequencing Amplified Fragments from Three Virus Genome Areas', *Journal of Virology*, 86: 5942–7.

Renner, D. W., and Szpara, M. L. (2018) 'Impacts of Genome-Wide Analyses on Our Understanding of Human Herpesvirus Diversity and Evolution', *Journal of Virology*, 92: e00908–17.

Renzette, N. et al. (2013) 'Rapid Intrahost Evolution of Human Cytomegalovirus Is Shaped by Demography and Positive Selection', *PLoS Genetics*, 9: e1003735.

——— et al. (2014) 'Human Cytomegalovirus Intrahost Evolution-A New Avenue for Understanding and Controlling Herpesvirus Infections', *Current Opinion in Virology*, 8: 109–15.

——— et al. (2015) 'Limits and Patterns of Cytomegalovirus Genomic Diversity in Humans', *Proceedings of the National Academy of Sciences of the United States of America*, 112: E4120–8.

Rho, M., Tang, H., and Ye, Y. (2010) 'FragGeneScan: predicting genes in short and error-prone reads', *Nucleic Acids Research*, 38: e191.

Rice, P., Longden, I., and Bleasby, A. (2000) 'EMBOSS: The European Molecular Biology Open Software Suite', *Trends in Genetics: TIG*, 16: 276–7.

Rohfritsch, A. et al. (2013) 'Population Genomics Shed Light on the Demographic and Adaptive Histories of European Invasion in the Pacific Oyster, *Crassostrea gigas*', *Evolutionary Applications*, 6: 1064–78.

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

Sanjuán, R., and Domingo-Calap, P. (2019) 'Genetic Diversity and Evolution of Viral Populations', *Encyclopedia of Virology*, 2021: 53–61.

Schikorski, D. et al. (2011) 'Experimental Ostreid Herpesvirus 1 Infection of the Pacific Oyster *Crassostrea gigas*: Kinetics of Virus DNA Detection by q-PCR in Seawater and in Oyster Samples', *Virus Research*, 155: 28–34.

Segarra, A. et al. (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: 92–9.

Shen, W. et al. (2016) 'SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation', *PLoS One*, 11: e0163962.

Stentiford, G. D. et al. (2012) 'Disease Will Limit Future Food Supply from the Global Crustacean Fishery and Aquaculture Sectors', *Journal of Invertebrate Pathology*, 110: 141–57.

Suchard, M. A. et al. (2018) 'Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10', *Virus Evolution*, 4: vey016.

Tan, A., Abecasis, G. R., and Kang, H. M. (2015) 'Unified Representation of Genetic Variants', *Bioinformatics*, 31: 2202–4.

Theze, J. et al. (2018) 'Genomic Epidemiology Reconstructs the Introduction and Spread of Zika Virus in Central America and Mexico', *Cell Host & Microbe*, 23: 855–64 e857.

Vasquez-Yeomans, R., Garcia-Ortega, M., and Caceres-Martinez, J. (2010) 'Gill Erosion and Herpesvirus in *Crassostrea gigas* Cultured in Baja California, Mexico', *Diseases of Aquatic Organisms*, 89: 137–44.

Vendrami, D. L. J. et al. (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: 519–34.

Vignuzzi, M. et al. (2006) 'Quasispecies Diversity Determines Pathogenesis through Cooperative Interactions in a Viral Population', *Nature*, 439: 344–8.

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: 126–39.

Wickham, H. (2016) *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag: New York.

——— et al. (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4: 1686.

——— et al. (2022) *Dplyr: A Grammar of Data Manipulation.* RStudio.

Yu, G. et al. (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. et al. (2012) 'The Oyster Genome Reveals Stress Adaptation and Complexity of Shell Formation', *Nature*, 490: 49–54.

Zhu, W. H., Lomsadze, A., and Borodovsky, M. (2010) 'Ab Initio Gene Identification in Metagenomic Sequences', *Nucleic Acids Research*, 38: e132.