Genetic differentiation and host specialization among OsHV-1 infecting two oyster species in France

Camille Pelletier[¶1], Germain Chevignon[1], Nicole Faury [1], Isabelle Arzul[1], Céline Garcia[1], Bruno Chollet[1], Tristan Renault[2], Benjamin Morga[&*1], Maude Jacquot[&*1].

Affiliations

[1] Ifremer, RBE-ASIM, Unité Adaptations Santé des Invertébrés Marins, Avenue de Mus de Loup, 17390 La Tremblade, France.

[2] Ifremer, Département Ressources Biologiques et Environnement, Nantes, France

*** Corresponding authors and co-last authors:**

Maude Jacquot

E-mail address: maude.jacquot@ifremer.fr,

Tel +33 5 46 76 26 82

Benjamin Morga

E-mail address: Benjamin.morga@ifremer.fr,

Tel +33 5 46 76 26 49

Fax +33 5 46 76 26 11

[¶] These authors contributed equally to this work

[&] These authors contributed equally to this work

Author contributions:

Conceptualization: IA, BM and CP. Resources: IA, CG, TR, BM. Performed histological analyses: BC, CG. Performed molecular biology analyses: NF. Investigation: CP. Methodology, supervision and revision: GC, MJ, BM. Writing – original draft preparation: CP. Writing – review and editing: CP, GC, MJ, BM. Revised and approved the paper: All.

Abstract

Cross-species transmission is one of the most significant causes of disease emergence in humans and other species. The Ostreid Herpesvirus type 1 has mostly been detected in the Pacific oyster *Magallana gigas* in the context of mortality events. However, it has been detected in other mollusc species including the flat oyster *Ostrea edulis* which raise questions about OsHV-1 host specialization. This study looked at the genetic differentiation of OsHV-1 in Pacific and European flat oysters, and the mechanisms that might underlie this genetic differentiation. To this end, high-throughput deep sequencing was used to analyze forty OsHV-1 genomes to characterize the genetic diversity and the degree of genetic isolation of OsHV-1 viruses from both *O. edulis* and *M. gigas* in France as well as to assess the frequency of the virus cross-species transmission. To achieve this, comparative and population genomic, phylogenetic, and phylodynamic approaches were employed. Our findings highlight the significant influence of host species on the genetic diversity and differentiation among OsHV-1. The emergence of OsHV-1 in France seems to be traced back to the introduction of *M. gigas*, and the virus has since evolved into two distinct lineages following a cross-species event. Moreover, our analysis revealed selection signals in regions responsible for crucial biological functions and domains, indicating the potential for OsHV-1 to adapt and specialize in response to its respective host species. To get further insights into evolutionary processes and host-virus interactions, future research should focus on investigating the coevolution between OsHV-1 and a wider range of host species using phylogenetic approaches.

Author Summary

In 1972, a herpes-like virus was discovered for the first time infecting *Crassostrea virginica* in the USA, during the course of laboratory experiments on adult oysters. Since the beginning of 1990s, a herpesvirus identified as Ostreid herpes virus type 1 (OsHV-1) has caused large-scale deaths in the Pacific oyster

(*M. gigas*) and has been detected in other species including scallops, clams, and flat oysters. A recent study examined different lineages of OsHV-1 infecting various host species, raising questions about how often the virus is transmitted from one species to another and whether it is specialized to a particular host, affecting co-culture systems. To investigate this, advanced tools in genetics and computational analysis were used to determine the frequency and impact of cross-species transmission events between the Pacific oyster and the European flat oyster. Our findings indicate that these events are uncommon but have given rise to a distinct lineage of OsHV-1 that specifically infects the flat oyster, suggesting specialization to its host. This study offers valuable insights into the evolutionary history of this viral pathogen and its interactions with different host species.

Introduction

The study of multi-host-pathogen systems is crucial to improve our understanding of the spread and emergence of infectious diseases, and to develop effective strategies to address animal health problems. Many pathogens are able of multiplying themselves in multiple host species, and multiple pathogens genotypes may exhibit different phenotypic traits [1,2]. Understanding the partitioning of pathogens genetic diversity among host species is particularly important for disease transmission when pathogens have the ability to infect both wildlife and livestock [2,3]. Disentangling epidemiological cycles in such systems is often challenging, yet investigating how the genetic diversity of pathogens is partitioned among hosts species is key to properly assess and manage disease risk [1,4]. Pathogen genotype frequencies within host populations can vary depending on, for example, their transmission efficiency within and between host populations [5,6], while barriers to transmission may result in population-specific patterns of infection for different genotypes [7–9].

Herpesviruses are an order of double-stranded DNA viruses commonly found across multiple species of invertebrates (*e.g.* mollusks) and vertebrates (*e.g.* fish, birds, pigs, horses, humans). However, they are characterized by their narrow host range and are known to coevolve with their hosts, leading to a long-standing relationship between virus and host species [10–12]. Nevertheless, there have been several reports of cross-species transmission (CST) events in both vertebrates (*e.g.* transmission of cercopithecine herpesvirus 1 from primates to humans [13]) and invertebrates such as mollusks (*e.g.* transmission of ostreid herpesvirus 1 from Pacific oysters to scallops or flat oysters [10,11]). Herpesviruses affecting mollusks are restricted to the *Malacoherpesviridae* family which currently includes two species, the abalone herpesvirus (AbHV) infecting gastropods and the ostreid herpesvirus 1 (OsHV-1) infecting bivalves. Several OsHV-1 variants have been reported infecting a wide range of host species [14–17]. Indeed, these variants can infect and cause mortalities in many bivalve species, including the scallops *Pecten maximus* [15] and *Chlamys farreri* [18], the clams *Anadara broughtonii* [19], *Ruditapes philippinarum* [20,21] and *Ruditapes decussatus* [20,21], as well as the oysters *Magallana angulata* [16], *Crassostrea virginica* [22], *Ostrea edulis* [16,23,24] and *Magallana gigas* [25–27].

In France, two oyster species have been shown to be infected by OsHV-1: the native European flat oyster *O. edulis* and the introduced Pacific oyster *M. gigas*. Pacific oysters were introduced in France at the end of the 1960s in response to the dramatic decline of the Portuguese oyster (*M. angulata*) populations [28]. *M. gi*gas is currently the most widely cultivated marine mollusk species, in France, in Europe and worldwide [29]. While herpes like viruses were first observed in *C. virginica* adults in the United States of America in 1970 , it has been detected in France in *M. gigas*, in a hatchery, approximately 20 years after the introduction of *M. gigas* [27]. OsHV-1 induces a major threat to juvenile Pacific oysters and has significant economic impacts in many production areas all around the world [18,27,30–33]. On the other hand, OsHV-1 does not impacted the flat oyster production currently, as it only leads to rare symptoms in few cases accompanied by mortalities.

Traditionally, OsHV-1 genotypes have been characterized by targeted amplicon sequencing [15,15,24,33–35]. This approach allowed the characterization of OsHV-1 Var detected in clams *R. philippinarum* in 1997 [16] and scallops *P. maximus* in 2000 [15]. This genotype has never been detected in wild clams or scallops as very limited detection of OsHV-1 have been carried out on these species in natural environment. Targeted approaches also allowed the characterization of OsHV-1 µVar, the genotype mostly detected in Europe currently. This genotype was first detected in spring 2008 and has been involved in mass mortalities of the Pacific oyster juveniles [35].

According to these observations, and because herpesviruses generally have a limited ability to infect a wide range of hosts, the assumption of OsHV-1 host specialization cannot be ruled out [16] despite the fact that several viral genotypes have been detected sporadically in various bivalve host species worldwide [15,16,24,36]. Indeed, the genetic information obtained so far suggests that different host species may be infected by distinct genotypes of OsHV-1 [19,37], indicating a complex epidemiological cycle that involves a wide range of host species that still needs to be clarified.

One significant drawback of targeted approaches is that they cannot comprehensively capture the entire genome, leading to an incomplete interpretation of the genomic information [38,39]. Advances in sequencing technologies allowed for the acquisition of the first OsHV-1 genome obtained from *M. gigas* larvae collected in 1995. This genome has since then been used as a reference genome [12]. To date, 52

OsHV-1 additional genomes have been sequenced from four host species: *C. farreri*, *A. broughtonii*, *O. edulis* and *M. gigas* [18,19,25,37,40,40,41]. The genome of the virus is linear and its size ranges from 199 kb to 211 kb. It contains from 123 to 128 predicted ORFs, of which only 57 have putative functions [12,18,19,25,40,41]. OsHV-1 genomic architecture is similar to some other herpesvirus genomes (i.e. Herpesvirus group E such as HSV-1) with two invertible unique regions (U_L for Unique Long and U_S for Unique Short) surrounded by inverted repeats (TR_L/IR_L and TR_S/IR_S). An additional unique sequence (called X) can be found between IR_L and IR_S, resulting in the overall genomic organization that can be resumed as TR_L-U_L-IR_L-X-IR_S-U_S-TR_S [12]. Preliminary analyses of a few number of OsHV-1 genomes, isolated from different bivalve species, revealed a relatively high genotypic diversity [18,19,37,41]. More specifically, comparative genomic analyses of viruses infecting *O. edulis* and *M. gigas* showed the presence of two large deletion areas in the U_L region as well as host-specific polymorphisms along the genome [37]. These results suggested that OsHV-1 evolution is strongly influenced by the host species [37]. However, the low sampling size of previous studies did not allow the application of population genetic and phylodynamic analyses and still little is known about the level of genetic isolation of OsHV-1 genotypes infecting *O. edulis* and *M. gigas* species along with about the frequency of cross-species transmission events [37].

In the present study, high-throughput deep sequencing was used to characterize the genetic diversity of OsHV-1 found in both *O. edulis* and *M. gigas* in France and Sweden. To this end, 40 new OsHV-1 were sequenced and assembled. Using these data and previously gathered information about genotypes circulating in different host species, our aims were to (*i*) describe OsHV-1 genomic diversity patterns in both species, (*ii*) estimate the genetic isolation between viruses infecting both species and (*iii*) determine if OsHV-1 cross-species transmission occurred and at which frequency. To do so, a combination of comparative and population genomic, phylogenetic and phylodynamic approaches were applied and discussed results in the context of current knowledge of OsHV-1 molecular epidemiology.

Materials and Methods

OsHV-1 sampling

To study the diversity of OsHV-1 in different host species, archived samples from Ifremer National NRL and EURL collection were used. OsHV-1 viruses were sampled from wild Pacific oysters *M. gigas* (seventeen samples) collected in the field during OsHV-1 infection outbreaks between 2005 and 2010 from individuals of different developmental stages (larvae, spats and adults) at five sites along the French coast (Antioche inlet, Palavas lagoon, Thau lagoon, Brittany inlet, Pirou Agon-Coutainville). Wild flat oysters *O. edulis* (twenty-three samples) collected in the field from April to August between 2007 and 2009 in the Quiberon bay as describe by [42] and one sample has been collected in Sweden (Table S1, Fig 1A). This last sample was used to compare the viral genotype from France and Sweden within *O. edulis* individuals to widen the geographical area studied. Criterion used to select *M. gigas* samples was *i)* the geographic proximity to *O. edulis* samples and *ii)* the year of samples collection, which covers the *O. edulis* samples collection period. Before DNA extraction, animals have been stored at -80°C or in ethanol from 6 to 14 years.

In addition, 22 *O. edulis* pools of larvae were sampled from adults' gills and were preserved in Davidson's fixative for histological analyses.

Samples were named according to the nomenclature recommended by [43] which is made of the virus name, the sampling location, the date of collection, the host species, and the sample number (*e.g.* OsHV-1/Rivedoux.FRA/2008/Cg-S-053 for the sample number 053 collected in 2008, in Rivedoux, from *M. gigas* spat). However, hereafter, shortened names have been used to facilitate reading (*e.g.* CG-08-Ri-S-053 for the same sample, see Table S1 for correspondence).

Histological assessment of tissue lesions during OsHV-1 infection

OsHV-1 is responsible for tissue lesions that have already been described in Pacific oysters [44]. Lesions in *O. edulis* tissues related to OsHV-1 infection have been rarely described by histology [45–47]. In order to detect tissue lesions and assess the presence of OsHV-1 within flat oyster tissues, a histological

analysis has been performed. Six pools of *Ostrea edulis* larvae collected in the Quiberon bay and previously stored in Davidson's fixative (see above) were transferred to 70% ethanol after 24-48h of fixation before processing standard histological protocol of paraffin inclusion [48]. Embedded paraffin blocks were then sectioned (2-3µm) and stained with hematoxylin and eosin. Histological sections were examined for pathological conditions under light microscopy (Olympus BX63) and photography were taken with an associate camera (Olympus DP74).

DNA extraction, viral copy numbers quantification and sequencing

Depending on host species, two different methods were used for samples DNA extraction (Fig 1C). For *M. gigas*, DNA was extracted from mantle of moribund individual spats, from pool of larvae or from pool of shred oysters using the Qiamp DNA Mini kit (Qiagen) according to the manufacturer's protocol. For *O. edulis*,DNA was extracted from gills of individual adults or from pool of larvae using the MagAttract® HMW DNA kit (Qiagen) according to the manufacturer's protocol.

DNA purity and concentration were assessed using a Nano-Drop ND-1000 spectrometer (Thermo Scientific) and Qubit® dsDNA BR assay kits (Molecular Probes Life Technologies), respectively. Then quantification of viral copy numbers was carried out in duplicate by quantitative PCR using a Mx3005 P Thermocycler (Agilent) as described by [49]. Each reaction contained 1µl of Milli-Q Water, 2µl of each primer at 5,5 µM, OsHVDP For (forward) 5′-ATTGATGATGTGGATAATCTGTG-3′ and OsHVDP Rev (reverse) 5′-GGTAAATACCATTGGTCTTGTTCC-3′, 10µl of Brilliant III Ultra-Fast SYBR® Green PCR Master Mix (Agilent), and finally 25 ng of DNA sample from infected oysters or Milli-Q water (non-template control) in a total volume of 20µl. The qPCR program consisted in 3 min at 95 °C followed by 40 cycles of amplification at 95 °C for 5s and 60 °C for 20s. Results were expressed as the log-transformed copy number of viral DNA per nanogram of extracted DNA ($cp.ng^{-1}$).

Illumina DNA-Seq library preparation and sequencing were performed at four sequencing platforms (see Table S1 for details about libraries preparation and sequencing technologies). Libraries were prepared using (*i)* True seq DNA PCR free kit, (*ii)* Shotgun PCR-free library preparation kit (Lucigen) or (*iii*) Illumina genomic PCR-free kit. The sequencing step was performed using (*i*) HiSeq™ 2500

device (paired-ends, 100 pb) by GenoToul (Toulouse, France), (*ii*) HiSeq™ 4000 device (paired-ends, 150 bp) by the Lille Integrated Genomics Advanced Network for personalized medicine (LIGAN-PM) sequencing platform (UMR 8199, CNRS, Lille, France) or (*iii)* NovaSeq™ 6000 device (paired-ends, 150 bp) by the Genome Quebec Company (Genome Quebec Innovation Center, McGill University, Montreal, Canada) or by Fasteris Life Science Genesupport (Plan-les-Ouates, Switzerland).

De novo assembly of OsHV-1 genomes

De novo OsHV-1 genome assemblies were obtained by processing sequencing reads as previously described (37,38 and see 48 for details on the bioinformatic pipeline). Briefly, for each library, raw reads were quality filtered (PHRED score > 31) and adapters trimmed using Fastp v0.20.1 (with parameter -q 31; [52]. In order to estimate the number of OsHV-1 reads and coverage in each sample, filtered reads were aligned to the reference genome of OsHV-1 (accession number: NC_005881.2; [12] using Bowtie2 v2.4.1 (parameter --local and --no-unal; 50). Meanwhile, reads were aligned to the reference genome of *M. gigas* (accession number: GCF_902806645.1, [54]) or *O. edulis* (accession number: GCA_024362755.1**,** [55]) according to the species from which the sample was collected, using Bowtie 2 v2.4.1 (parameter --local). Non-oyster reads were selected with SAMtools v1.9 (option view -f 4,[56]) and Seqtk v1.2 subseq (default parameters; [57]). Non-oyster reads were then *de novo* assembled with metaSPAdes v3.14.0 (parameter --meta; [58]). OsHV-1 contigs were selected using BLASTn v2.12.0 (options -max_target_seqs 1, -evalue "1e-10"; [59]) against the OsHV-1 reference genome. OsHV-1 contigs were then manually re-ordered in Geneious Prime v2022.2.2 by visualizing reads alignment on the *de novo* assembled contigs to identify overlaps between trimmed reads at each contig extremities. To avoid inverted repeat regions reconstruction issues for all subsequent analyses, every OsHV-1 genome was reconstructed as non-redundant (NR) genome containing only one copy of each duplicated regions (IR_L and IR_S) as described by [60], [50] and [51]. NR-genomes were validated and corrected by similarity matrix visualization of each genome to itself using the dotplot tool (based on the EMBOSS v6.5.7 tool dottup) implemented in Geneious Prime v2022.2.2. To verify and clean the final assemblies as well as to assess coverage uniformity, quality filtered reads were aligned back to the new constructed OsHV-1 NR-genomes using Bowtie 2 v2.4.1 (parameter --end-to-end). Raw data as well as

reconstructed genomes have been deposited on the NCBI database under Bioproject PRJNA999485 accession numbers SAMN36736021 to SAMN36736060 for future reference and accessibility (Table S1).

Genomic architecture and ORF predictions

The genomic annotation in this study has been carried out at both genome and loci scales, including the identification of genomic regions (U_L, IR_L, X, IR_S, U_S), ORF prediction and functional annotation, signal peptides and transmembrane domain prediction.

Genomic annotations

At the genome scale, the structural genomic regions annotation, was performed using a combination of tools. Firstly, BLASTn v2.12.0 [59] implemented in Geneious Prime v2022.2.2 was used to compare each OsHV-1 NR-genome assembled in this study with the OsHV-1 reference genome (NC_005881.2) and transfer raw annotation to assembled NR-genome. Secondly, the structural annotations were corrected using coverage plots generated after remapping reads to control and clean the assembly. By observing where the coverage levels experienced changes or abrupt shifts, it allows the identification and correction of the exact boundaries between repeated and unique regions.

ORF, signal peptides and transmembrane domains predictions

To ensure consistency in the predictions across different genomes and to facilitate comparability of polymorphisms and gene positions, an ancestral sequence of OsHV-1 for each host species was reconstructed. To this end, first multiple alignment of genomes sampled from *M. gigas* (twenty genomes) or from *O. edulis* (twenty-three genomes) was performed using MAFFT v1.4.0 [61]. Analyses were performed using the model of nucleotide substitution HKY+G+I based on Akaike Information Criterion (AIC) obtained using MEGA v11.0.11 [62]. Next, ancestral sequences of studied OsHV-1 genomes for both host species were reconstructed using Beast v1.10.4 [63]. Different combinations of coalescent models (*i.e.* constant size, GMRF Bayesian skyride, expansion growth, and exponential growth) and molecular clocks (*i.e.* strict clock and uncorrelated relaxed clock) were tested to find out the most appropriate model. The maximum likelihood estimator (MLE) calculated by path sampling

(PS) and stepping-stone sampling (SSS) were compared to select the most appropriate model. According to results obtained, ancestral sequences were reconstructed using a strict clock and a constant size coalescent model for the *M. gigas* group, and an uncorrelated relaxed clock and an expansion growth model for the *O. edulis* group and MCMC chains of 10,000,000 steps were ran and sub-sampled every 1,000,000 generations.

Subsequently, local predictions of genes, signal peptides and transmembrane domain on the two reconstructed ancestral sequences were performed. ORFs were predicted with Prodigal v2.6.3 [64] with default parameters, a tool already used for double-stranded DNA viruses infecting eukaryotes [65]. Functional annotations of predicted proteins were performed with EggNOG-mapper v2.1.8 [66]. Signal peptides were predicted using SignalP v6.0 [67] and transmembrane domains using DeepTMHMM v1.0.24 [68].

Within individual's genomic diversity

Analysis of polymorphisms rarefaction

To verify that sequencing errors was not a major driver of intra-individual diversity and investigate the degree to which sequencing effort affected the number of intra-individual variations detected, a rarefaction analysis on the number of detected variations was performed for each sequencing library [50,69]. To this end, an iterative variant calling analysis was performed on subsampled reads that were mapped on the previously assembled genomes. In order to accelerate analyses, a nonlinear subsampling incrementation was performed based on 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 50%, 80% and 100% of the total number of reads of libraries. After each subsampling/variant calling step, variants were classified as SNPs (single nucleotide polymorphism) and InDels (insertion or deletions) and counted in order to increment rarefaction curves.

Variant calling

Then, to determine the genomic diversity within samples, variations were called against each of the 40 genomes previously assembled, using Freebayes v1.3.5, [70] 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 and --gvcf. All the variations were then normalized, multiallelic variants and biallelic blocks were decomposed and then sorted and deduplicates (using programs normalize, decompose, decompose_blocksub, sort and uniq) with vt v2015.11.10 [71] using default parameters. Polymorphisms detected by Freebayes were mapped on the corresponding consensus, then all the forty-one consensuses with their corresponding annotated polymorphisms were aligned using MAFFT v1.4.0 [61] implemented in Geneious Prime to re-coordinate all polymorphisms. Re-coordinated polymorphisms were then used to create unique identifier as described by [50]. Briefly, unique identifiers are composed of *(i*) the position of the polymorphism along the multiple alignment, (*ii*) the reference allele (REF) and *(iii*) the alternative allele (ALT) (*i.e.* position_REF >ALT).

Statistical description of the intra-individual diversity

A MultiDimensional Scaling (MDS) plots was used to examine the impact of host species on the genetic diversity of OsHV-1. The first MDS analysis summarized the allelic frequencies of each SNP within the sequencing libraries, allowing to explore the distances between pairs of SNPs in terms of their allelic frequencies within each library. This information was linked to those of host species for the visualization step. The second MDS was performed using a spearman correlation matrix based on individuals' genetic similarities. The number of clusters of individuals were searched using the kmeans method and were represented on the MDS plot. Then, a genotype association analysis was performed using the co-occurrence matrix of genotypes within individuals, as well as the adjacency and incidence of vertices on each other. Connections between edges and their weights were used to represent the strength of interaction between individuals in the MDS graph.

Both MDS and plotting were performed in R v4.2.2 using stats v4.2.2 [72], igraph v1.3.5 [73] and ggplot2 v3.4.0 [74] packages.

To examine the density of SNPs within each genomic region (*i.e.* U_L, IR_L, X, IR_S, U_S), the ratio of the number of SNPs detected within a region to the length of the region ($\frac{Number\ of\ SNPs}{Genomic\ region\ length}$) has been calculated.

Whole genome alignment and recombination detection

Identification and description of OsHV-1 structural variations

To compare the genomic architecture between viral genotypes and determine large-scale evolutionary events, the forty reconstructed NR-genomes were aligned with three published OsHV-1 genomes (OsHV-1 µVarA, *M. gigas*, 2010, France: KY242785, OsHV-1 µVarB, *M. gigas*, 2011, Ireland: KY271630, and OsHV-1 Ref., *M. gigas*, 1995, France: NC_005881.2) using Mauve v1.1.3 (progressive Mauve algorithm, with default parameters, [75]). These three genomes were available in public databases and were also assembled from Pacific oysters. Furthermore, the genotypes are very similar to those detected in our libraries. However, the three published genomes assembled from clams or scallops were not include as they are too distant from the genotypes studied in this research, and not enough genomes are available for these species. Our multiple alignment of the forty-three OsHV-1 genomes allowed us to identify typical genomic structures based on rearranged blocks and to classify genomes according to their structure. Specifically, five distinct genomic structures were identified in this study.

To plot the five types of genomic structures observed in our samples, pairwise alignments of the five representative genotypes against the OsHV-1 reference genome (NC_005881.2) were conducted using minimap2 v2.24 (parameters -ax asm5 --eqx; [76]). Alignments were then sorted and indexed using SAMtools v1.14 [56]. Structural rearrangements were characterized using Syri v1.5.4 [77] and plot with the python package plotsr v0.5.3 [78].

Manual restructuring of OsHV-1 genomes and multiple alignment

As some genomes contained large rearrangements, thirty assembled genomes were manually modified, from the Mauve multiple alignment, in Geneious Prime to ensure that the genomic architecture was identical for every genome, making them comparable to each other. Genomes were then aligned using MAFFT v1.4.0 [61] that allowed the computation of nucleotide identity percentages between pair of genomes.

Recombination events detection

Gubbins v3.3 [79] with default settings was used to identify regions that might be impacted by recombination events among the assembled OsHV-1 NR-genomes. Additionally, to confirm results obtained with Gubbins, linkage disequilibrium values (D') were computed to test for random association of allelic distribution and then explore the possibility of recombination in our dataset. D' values were computed in DNAsp6 [80] based on informative sites and significance was assessed using a Fisher's exact test. D^{2}' values were then plotted as a heatmap in R using the heatmap function from the ggplot package v3.4.0 [74]. All genomic regions detected as potentially recombinant were removed from the downstream analyses (Table S5).

Consensus based diversity computation

Population structure and evolution analysis through phylogenetics

First, a phylogenetic analysis has been performed to explore the population structure and evolutionary relationships among OsHV-1 genotypes. A phylogenetic search that applied a maximum-likelihood approach was performed on the alignment of the forty-three genomes using PhyML v3.3.2 [81] with 1000 bootstraps replications. The best fitting model of evolution for the samples was found using MEGA v11.0.11 [62] based on AIC. Therefore, a HKY+G+I model was used. Results were visualized using FigTree v1.4.4 (available at http://tree.bio.ed.ac.uk/software/figtree/).

Investigating genetic differentiation between groups

Then, description of consensus-based diversity was done using population genetic statistics such as the deviation of allelic frequencies from the neutrality equilibrium (Tajima's D) and the differentiation level of viruses infecting the two host populations (G_{ST}).

To obtain insights into deviation from demographic equilibrium and/or selective neutrality, Tajima's D values were computed in DNAsp6 v6.12.03 [80] for 1kb-windows along the alignment in order to standardize the amount of available information among analyses. Tajima's D values were calculated for the overall population (forty-three genomes) as well as for samples from each host species separately.

Under the neutral evolutionary model, Tajima's D is expected to be null. At a locus scale, positive Tajima's D values can be induced by balancing selection, due to the maintenance of highly divergent variants and, negative Tajima's D values can be the result of purifying selection or of a recent selective sweep, which both result in an excess of weakly divergent alleles. At the whole genome scale, positive Tajima's D values reflect a decrease of the group effective size thus favoring intermediate frequency alleles, while negative Tajima's D values suggest an increase of the population effective size implying expansion of rare alleles.

Then differentiation measures were computed between groups as they reflect the genetic variability among groups of individuals relative to the variability of a whole sample. The pairwise Nei G_{ST} method was used and, per locus G_{ST} values were computed for 1-kb windows with the option diff_stats in R using the package mmod v1.3.3 [82]. G_{ST} values were also computed for 100-bp windows within 1-kb windows that were detected as differentiated to define precisely regions involved in differentiation and facilitated the link with ORF potential function for biological hypothesis assessment.

An AMOVA was conducted to investigate the possible factors contributing to the structuring of OsHV-1 diversity at the consensus level. The global alignment of genomes was used for the AMOVA, following the same procedure as described in the previous paragraph (see "Within individual's genomic diversity" for more details).

Selective sweeps detection

To explore the potential effects of natural selection, signals of selective sweeps were searched in our datasets. A selective sweep occurs when a new beneficial mutation increases in frequency and becomes fixed, resulting in a reduction or elimination of genetic variation near the mutation. RAiSD v2.9 [83] was used to detect selective sweep signatures along OsHV-1 genomes.

Ancestral traits reconstruction

In order to analyze the temporal signal and the 'clocklikeness' of the ML tree and first explore the evolution rate of OsHV-1, TempEst v1.5.3 [84] has been used to plot the divergence of each tip from the root over time. The root-to-tip divergence (RTTD) plot has been performed separately for *O. edulis*

group (twenty-three genomes) and for the two subclusters of the OsHV-1 *M. gigas* group (respectively nine and eleven genomes) using a midpoint rooted tree and a best fitting root approaches for comparisons.

Analysis were performed using a HKY+G+I model of nucleotide substitution based on AIC obtained using MEGA v11.0.11 [62]. As before, different combinations of coalescent models (*i.e.* constant size, GMRF Bayesian skyride, and exponential growth) and molecular clocks (*i.e.* strict clock and uncorrelated relaxed clock) were tested to find the most appropriate model for our dataset. The maximum likelihood estimator (MLE) calculated by path sampling (PS) and stepping-stone sampling (SSS) for each model were compared. According to those, an uncorrelated relaxed molecular clock and a constant size coalescent model has been used accordingly to MLE and a MCMC chain of 10,000,000 steps was run and sub-sampled every 1,000,000 generations using Beast v1.10.4 [63]. The effective sample size, representing the mixing of the population (Ne > 200), and the stationary distribution of the trace were checked in Tracer v1.7.2 [85] for all parameters, using an initial burn-in of 10%. Finally, maximum clade credibility (MCC) trees were generated from tree output files from BEAST in TreeAnnotator v1.10.4 [63], and annotated in FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Modeling of demographic evolution

To assess the historical population dynamics of OsHV-1 in both host species, a demographic analysis was conducted by reconstructing past changes in population size over time. To do so, the log files of the trees that were previously generated to reconstruct ancestral sequences of OsHV-1 in both host species were used. The demographic history was then reconstructed using a GMRF Skyride model in Tracer v1.7.2 [85].

Results

Detection of OsHV-1 in *O. edulis* and *M. gigas* collection samples

The number of viral copies varied from 1.47×10^3 to 3.07×10^8 copies per ng of DNA, with an average quantity of 1.34×10^7 copies per ng of DNA for the samples collected from *O. edulis* (Fig S1, Table S1).

For samples collected from *M. gigas*, the viral copy numbers ranged from 2.21×10^4 to 4.21×10^7 copies per ng of DNA, with an average of 2.81×10^6 copies per ng of DNA (Fig S1, Table S1).

As the minimum threshold for efficient Illumina sequencing of virus particles is 10^3 copies, all samples were found to be suitable for sequencing [37,50,51].

In addition to qPCR quantification, the lesions related to the viral infection in flat oyster tissues has been highlighted because few studies of lesions exist, particularly in flat oyster larvae. Histological observations revealed the presence of apparently healthy flat oyster larvae (Fig S2A), larvae with hemocytes infiltration and necrosis lesions (Fig S2B to S2D) and also dead larvae (Fig S2E). On moribund larvae, hemocyte infiltrations were light and generally located in the mantle (Fig S2B); concerning necrosis, it was the main lesion observed and essentially located in the mantle and velum (Fig 2C). It was associated with the presence of cells with abnormal nuclei; these abnormalities were characterized by the presence of perinuclear chromatin and an amorphous matrix in the central area of nucleus or abnormal enlarged nuclei (Fig S2C). In some disease larvae, necrosis affected all tissues (Fig S2D). Dead larvae appeared in decomposition state with a lysis of all tissues and a loss of tissue architecture (Fig S2E).

Sequencing statistics

Sequencing depth and the viral genome coverage were correlated to the total number of reads for the different libraries and to the number of copies of OsHV-1 copies (Fig S1). The Illumina DNA sequencing step produced from 18.9 to 290.9 million reads for a given sample (average 112.2 M reads ± 60.3 SD, Fig S1 and Table S1). Host reads represented from 5.2% to 91.2% of the total reads. The percentage of reads mapped to OsHV-1 varied from 0.013% to 38.51%, and the genome coverage varied accordingly from 12 to 26,790 reads depth. Using this sequencing dataset, 40 OsHV-1 NR-genomes were assembled *de novo*. The length of the assembled NR-genomes varied from 185.9 kbp to 190.2 kbp.

To verify that sequencing errors were not a major driver of intra-individual diversity and investigate the degree to which sequencing effort affected the number of intra-individual variations detected, a rarefaction analysis on the number of detected variations was performed. Rarefaction curves reached a

plateau at 150,000 reads corresponding to an average of 110 reads depth (Fig S3). Therefore, this value was used as the minimum number of reads necessary to accurately detect SNPs in our samples. 12 libraries did not reach this threshold and were excluded from all analyses focusing on intra-sample diversity. For others libraries, when a high number of sequences were used, no additional variations were detected, which would have been the case if there was an accumulation of sequencing errors.

Intra-sample OsHV-1 diversity

A total of 986 "intra-sample" polymorphisms, having an allelic frequency greater than 10%, was detected across 776 polymorphic sites along genomes (Table S2, Fig 2A). In the genome group sampled from *O. edulis*, 536 polymorphisms were detected at 430 variable positions, out of which 67 positions carry more than two alleles. On the other hand, the genome group sampled from *M. gigas* exhibited 434 polymorphisms found at 327 variable positions, with 61 positions carrying more than two alleles. Additionally, 92 polymorphisms located on 51 variable positions were shared between genome groups of both species.

Out of the 536 polymorphisms detected in *O. edulis* libraries, 336 were located within 58 predicted ORFs (Table S3). More precisely, 161 polymorphisms were found in ORFs coding for proteins with unknown functions, while the remaining polymorphisms were identified in ORFs coding for proteins with known domains or functions (Table S3). In detail, 105 polymorphisms were detected in ORF coding for putative transmembrane proteins (in ORFs 22, 41, 54, 57, 59, 65, 72, 77, and 111), 26 polymorphisms in ORF coding for putative DNA primase (ORFs 66 and 7), 15 SNPs in ORF coding for putative DNA packaging protein (ORF 109), 6 polymorphisms in ORF coding for DNA polymerase (in ORF 100), or 6 polymorphisms in ORF coding for Zinc-finger RING type protein, (in ORFs 9, 38, 97, and 117).

In viral genotypes sampled from *M. gigas*, 203 out of the 434 detected polymorphisms were located within 44 predicted ORFs. 57 polymorphisms were found in 21 ORFs coding for proteins with unknown functions, while the remaining 148 polymorphisms were identified in ORFs coding for proteins with known domains or functions. More precisely, 81 polymorphisms were detected in ORFs coding for

transmembrane proteins (in ORFIN4, ORF 22, 41, 65, 68, and 77), 48 polymorphisms were detected in ORFs coding for membrane proteins (in ORF 16, 36, and 63) and 6 polymorphisms in ORFs coding for Zinc-finger, RING type. The remaining polymorphisms were detected in ORFs coding for DNA primase, DNA packaging, or DNA polymerase (Table S3).

In proportion to the size of the regions, the U_L region exhibits fewer polymorphisms (0.14% for *M. gigas* samples and 0.19% for *O. edulis* samples) compared to the other regions (Table S4). In the *M. gigas* samples, the genomic regions with the highest SNP density are the IR_L and IR_S regions, containing 0.81% and 0.57% of the SNPs, respectively, while 0.53% of the X and U_S regions contain SNPs. In the *O. edulis* samples, the U_S, IR_L, and X regions exhibit the highest SNP density, containing 1.03%, 0.99%, and 0.72% of the SNPs, respectively, while only 0.36% of the IR_S region contains SNPs.

Interestingly, the polymorphisms shared by samples from both oyster species are located in the repeated regions IR_L and IR_S, particularly in regions that exhibit a GC content greater than 50% (Fig S2B). No polymorphisms are detected in the X region, except in the homopolymers at the junctions with the repeated regions, which demonstrate an accumulation of SNPs.

The multidimensional scaling (MDS) analysis of allelic frequencies of polymorphic sites showed alleles frequencies in samples from *O. edulis* are different from those in samples from *M. gigas*, and indicate a clear structuring of intra-sample variations according to host species (Fig 3A). However, 92 allelic frequencies were more similar within both species compared to others, indicating that some positions are variable independently of the species studied (Fig 3A).

The genotype association graph, coupled with an MDS analysis, summarized the two-dimensional comparison of the co-occurrence frequency of each polymorphism inside and between both genotype groups (Fig 3B). The modularity values of a network indicate the strength of connections within and between communities, and thus, provide information about the quality of network division. The analysis of alleles co-occurrence revealed the presence of three distinct groups based on modularity measurements. Out of these three groups, two clusters corresponding to *M. gigas* samples (represented by the dark and light green clusters on Fig 3B) and one cluster corresponding to *O. edulis* (represented

by the purple cluster on Fig 3B). The dark green cluster of *M. gigas* samples comprises samples collected from various developmental stages between 2005 and 2010 in Charente-Maritime, the Breton inlet, and Normandie. On the other hand, the light green cluster of *M. gigas* samples consists of samples collected between 2006 and 2008 in Palavas and Thau lagoons, and the Breton inlet, encompassing all developmental stages.

Structural variations

Using a whole genome alignment approach, a structural analysis of all OsHV-1 genomes assembled from *M. gigas* and *O. edulis* samples has been performed (Fig S4). The whole genome comparison revealed six types of genomic architecture (Fig 4). The first one includes a group of four genomes obtained from *M. gigas* and similar to the reference genome described by [12], characterized by NR-genomes of approximately 190 kbp. The five other structural variants were identified and described in comparison with the OsHV-1 reference genome [12]. First, seven NR-genomes of 190 kbp collected from *M. gigas* differ from OsHV-1 reference genome by a reversion of 8,530 bp in the U_L region between ORF 36 which is disrupted (potentially not functional) and ORF 43. Second, a group of nine NR-genomes of 204 kbp sampled from *M. gigas* were analogous to the µVar genotype described by [25] which contains four deletions in the U_L region (ranging from 606 bp to 3,550 bp), one in the IR_L (725 bp) region, and one insertion of 2,671 bp in the U_L region between ORF 43 and 44, for a total of approximately 10 kbp insertions or deletions. Third, one NR-genome of 192 kbp, sampled from *O. edulis*, contains a reverse translocation of the first part (between ORF 6 and ORF 4 which is disrupted) of the TR_L region (7 kbp) and an insertion of 2,204 bp upstream to the U_L region. In addition, this genome contains multiple large deletions (6 deletions for a total of 5,303 bp) and insertions (5 insertions for a total of 4,017 bp) for an average of 13 kbp insertions or deletions. Fourth, a group of twenty NR-genomes of 185 kbp and sampled from *O. edulis*, contains a reverse translocation of the first part of the TR_L region (4 kbp). In addition, these genomes contain multiple large deletions (11 deletions for a total of 6,423 bp) and insertions (4 insertions for a total of 1,133 bp) for an approximate 11 kbp InDels in total. Finally, a group of two NR-genomes contains the same reverse translocation, deletions, and insertions detected as those detected in the fourth group of twenty NR-genomes assembled from *O.*

edulis, but in addition, an inversion of 1,470 bp was detected in the UL region (ORF 50 disrupted) for a total of approximately 12 kbp variations.

To ensure that homologous regions are all in phase, and that multiple alignment and phylogenetic analyses can be carried out, genomic regions of the forty-three OsHV-1 NR-genomes have been manually rearranged. The recombination analysis identified the occurrence of 17 potential recombination events along OsHV-1 genomes, that may have impacted two genomic regions, as shown in Fig S5. The first potential recombination event was located in the U_L region between 61,474pb and 68,405bp, specifically between a microsatellite $(TA)_n$ present in the viral genome collected from *O. edulis* s and a stem loop predicted to be the probable origin of replication of OsHV-1. The other four recombination events detected were located in the repeated regions. These potential recombination events have been detected respectively between 173,860 and 175,357 bp at the junction between U_L and IR_L regions, between 178,966 and 180,440 bp and from 184,844 to 185,190 bp both in the IR_L region, and finally from 194,785 to 195,422 bp located in the IR_S region.

D'^2 values were mostly close to 1, indicating an absence of linkage disequilibrium (Fig 5). However, one position close to the start of the genome within the ORF 6 (position 3,108), as well as two other positions at the end of the U_L respectively within ORF 88 and ORF 104 (positions 137,500 and 164,760), one position in the IR_L within ORF 114 (position 182,122), and one position in the IR_S (position 195,034) regions had very low linkage disequilibrium values, close to 0. None of the detected low D'^2 values was in a homopolymer region but two low D'^2 positions (positions 182,122 and 195,034) correspond to recombination breakpoints detected by Gubbins (Table S5).

A total of 11,184 bp distributed within 13 regions potentially involved in recombination were removed from the alignment for the rest of the analysis. The suppressed regions are referenced in the table S5. More precisely, 541 SNPs have been removed from consensus-scale analyses.

A maximum likelihood phylogeny was then generated using a HKY+G+I model using PhyML [81]. This tree showed that OsHV-1 genomes sampled from *M. gigas* were genetically distant from those

sampled from *O. edulis* (Fig S5). In the present study, the viral genotypes were clearly structured according to host species, with two main clusters identified (Fig S5). Within each cluster, two subclusters were identified. In the *M. gigas* cluster, subcluster 1.1 corresponded to samples related to the OsHV-1 reference genome sampled between 1994 and 2008, while subcluster 1.2 corresponded to samples corresponding to µVar genotype sampled between 2008 and 2010. In the *O. edulis* cluster, subcluster 2.2 corresponded to samples collected between 2007 and 2009 in the Quiberon bay, while subcluster 2.1 corresponded to the sample collected in Sweden in 2012. The clustering was found to be primarily based on host species and then on the sampling date (Fig S5). The multiple alignment of the forty-three genomes enabled us to determine the similarity percentages between every pair of genomes (Fig S6, Table S6). To provide an overview of results, the average percentage of identities for each cluster and between clusters has been computed (Table 1). Identity percentages within clusters were close to 100%, indicating a high similarity between genomes within a cluster. In contrast, the average identity percentage varied between 96% (cluster 2.2 compared to 2.1) and 91% (cluster 1.1 compared to 2.1) when comparing pairs of genomes belonging to different clusters.

Table 1. Average percentage identities for each cluster and between clusters based on pairwise distances computing using MAFFT.

	Cluster 1.1	Cluster 1.2	Cluster 2.1	Cluster 2.2
Cluster 1.1	99,98%	94,66%	91,07%	92,51%
Cluster 1.2	94,66%	99,90%	92,96%	94,58%
Cluster 2.1	91,07%	92,96%	100%	96,03%
Cluster 2.2	92,51%	94,58%	96,03%	99,52%

Consensus-based diversity analysis

An AMOVA has been performed to test the structuring factors independently of the observed genetic differences (Table 2). The factors have been tested separately because of missing metadata. The results indicating that 94.5% and 95.3% of the variance are explained respectively by host species and sampling origin while sampling year and host developmental stage explain only 40% and 66.4 % of the variance observed. In addition, 36% of the molecular variance is explained by the sequencing technology, there is therefore not much impact of sequencing technology on the consensus scale.

Table 2. Analysis of Molecular Variance (AMOVA) to determine the impact of host species, origin, year of sampling and host developmental stage on the OsHV-1 diversity at the genomic scale.

	DF	*SS*	*MS*	*Est. var.*	*%*	*PhiPT*
Host species	1	2.87e-04	2.87e-04	1.34e-05	94.5%	0.945
Sampling year	8	0.000152	1.89e-05	3.29e-06	40%	0.4
Origin	11	0.000304	2.76e-05	9.88e-06	95.3%	0.953
Host developmental stage	3	0.000196	6.54e-05	6.23e-06	66.4%	0.664
Sequencer	4	9.58e-05	2.39e-05	3.31e-06	36%	0.36

DF: degrees of freedom, SS: sum of square deviation, MS: mean square deviation, Est. var.: estimated variance, %: percentage of differentiation explained by the factor, PhiPT: population differentiation statistics.

The diversity analysis based on every NR consensus genomes was completed with the computation of commonly used measures in population genetics, including the fixation index G_{ST} and Tajima's D (Fig 6). The genetic differentiation between species was estimated based on SNPs, and the global pairwise G_{ST} was computed for the alignment of both OsHV-1 genomes from *O. edulis* and *M. gigas* (Fig 6A). The differentiation value between the two host species computed with a pairwise Nei G_{ST} was 0.745, indicating the presence of different fixed alleles in both groups. To determine regions that were highly differentiated between groups, G_{ST} was computed for 1-kb windows along the alignment of OsHV-1 genomes. G_{ST} values varied from 0.005 to 1 along genomes and were, on average, equal to 0.545. A threshold of G_{ST} = 0.6 was used to distinguish between highly differentiated and less-differentiated regions. Results show 34 out of 200 windows of 1-kb having a G_{ST} higher than 0.6 indicating high level of genetic differentiation of these regions (Fig 6A). Among these 34 windows, 29 were located within the U_L region and 5 were located in the IR_L region (Table S7). X, IR_S and U_S regions seemed to be less differentiated according to host species than the rest of the genome. Additionally, 100-bp windows were computed within these 1-kb windows to define more precisely positions of SNPs that were differentiated. 87 100-bp windows out of 340 were detected as highly differentiated (G_{ST} > 0.6). These windows were located within 2 intergenic regions and 34 ORFs coding for proteins with mainly unknown functions (18 ORFs, see table S5), secreted proteins (ORFs 68, 88), proteins containing

domains such as Zing-finger, Ring type proteins (ORFs 53, 97, 106), proteins with transmembrane domain (ORFs 22, 54, 59, 65.1, 68, 77, 84, 88), inhibitor of apoptosis (ORFs 87, 106), ribonucleotide reductase (ORF 20), DNA replication origin binding (ORF 115), large subunit DNA-packaging terminase (ORF 109).

Tajima's D values were calculated for 1-kb windows along the alignment to identify selection patterns along OsHV-1 genomes. At the overall group level, most of the Tajima's D values were positive (Fig 6B), ranging from -1.87 to 4.44 with an average of 0.673. Later, Tajima's D was computed for each viral group individually. Within *O. edulis* OsHV-1 Tajima's D values were mostly negative, ranging from -2.02 to 0.54, with an average of -0.851. Within *M. gigas*, Tajima's D values were predominantly positive, varying from -1.97 to 2.92, with an average of 0.627, which was comparable to the values observed for the overall population.

Tajima's D values for the *O. edulis* viral group suggest a selective sweep might have recently occured along OsHV-1 genome. This was confirmed by RaISD results. In the *O. edulis* viral group, 168 regions with μ-statistic values ranging from 1 to 8 were found to have been impacted by selective sweeps. For *M. gigas* viral group, 587 regions with μ-statistic values ranging from 3 to 8 were affected (Fig 7A). Regions impacted by selective sweeps in both groups were mainly located within ORFs coding for proteins with unknown functions. Furthermore, 8.3% to 9.9% of the selective sweep regions were located within intergenic regions (Table S8). The main domain impacted by selective sweep appear to be transmembrane protein (18.5% for *O. edulis* and 17.5% for *M. gigas*), Zinc-finger, RING-type (5.9% and 5.5% of the detected selective sweep regions), and viral maintenance functions such as DNA primase, polymerase, or packaging (Fig 7B).

Ancestral states reconstruction

R^2 values of regression of root-to-tip genetic distance against sampling times are provided as a measure of how OsHV-1 evolution follows a clock-like pattern. According to R^2 values in best fitting root models, OsHV-1 harbors a sufficient temporal signal to perform molecular clock-based approaches to estimate evolutionary rates over time.

The comparison of MLE values, with higher values reflecting the most appropriate models, has led to the selection of an uncorrelated relaxed molecular clock and a constant size coalescent model (Table 3).

Table 3: maximum likelihood estimation values used to compare the different demographic and clock models.

	Constant size		*GMRF*		*Exponential growth*	
	Strict clock	*Uncorrelated relaxed clock*	*Strict clock*	*Uncorrelated relaxed clock*	*Strict clock*	*Uncorrelated relaxed clock*
PS	-1293,121731	**-1279,27**	-1312,775139	-1312,436377	-1294,032584	-1293,632567
SSS	-1293,270332	**-1279,441704**	-1312,798034	-1312,384202	-1294,529875	-1294,145544

Values in bold correspond to the maximum MLE used to select the model for ancestral state. reconstruction with Beast.

Moreover, the evolutionary rate has been computed for the global phylogeny and for each host subcluster. For the global phylogeny, the evolutionary rate has been estimated at 1.50E-02 nucleotide substitutions per site per year. Within the cluster from *M. gigas*, the evolutionary rate was 2.37E-02 nucleotide substitutions per site per year. In comparison, within the genotypes of OsHV-1 infecting *O. edulis*, the evolutionary rate was slightly lower with a value of 1.20E-02 nucleotide substitutions per site per year.

The most recent common ancestor (MRCA) of the OsHV-1 genotypes was estimated to have existed in the 1990's (95% HPD interval [1989,1995]) and is associated with *M. gigas*. Then, the actual genotype µVar detected in France seems to have emerged around 2007 (95% HPD interval [2006,2008]). In parallel, one event of CST from *M. gigas* to *O. edulis* has been inferred to have occurred between 1998 (95% HPD interval [1993,2004]) and 2004 (95% HPD interval [2001,2008]) follow-up by a differentiation and a specialization of the lineages infecting *O. edulis*.

Demographic evolution over time within both viral groups

A model of the effective size of the viral group over time was constructed using output trees from ancestral state reconstruction. The viral populations sampled from *O. edulis* showed a constant decrease in "population" size from 1965 to 2002 (Fig 8C). Between approximately 2002 and 2007, the decrease in "population" size was drastic and was quickly followed by a rapid expansion step. At present, the

viral group is still experiencing a slight, constant decrease in "population" size. In contrast, the *M. gigas* viral group exhibited a fluctuating step of "population" size between 1982 and 1990, followed by a constant decrease in "population" size until 2008, which corresponds to the detection of the µVar genotype (Fig 8B). After 2008, this group began to expand. Moreover, these results suggest that the viral populations sampled from both oyster species was present before the first detection of OsHV-1 made in 1991 in France [26].

Discussion

Several approaches have been used to either detect or characterize OsHV-1 genotypes infecting *M. gigas*, including histology and microscopy [16,36,45,46], PCR [16,36] as well as restriction length polymorphisms (RFLP) [16]. However, this is a recent genomic study that revealed a potential structuring of OsHV-1 lineages according to host species [37]. This raised questions about OsHV-1 host specificity, transmission cycles and cross-species transmission (CST). Here, the diversity of OsHV-1 genomes between *M. gigas* and *O. edulis* collected from 2005 to 2012 in a dataset were compared, at both the intra-sample and global genomic scale, and gained insights into evolutionary history of OsHV-1 and CST frequency between both host species.

OsHV-1 population structure and differentiation driven by host species

Previous studies have provided preliminary evidences that the genetic diversity of OsHV-1 is influenced by various factors, including geographical location [37,50], time [25,37,86], and host species [37].

Our analysis of structural variations identified six distinct variant genotype structures in our dataset, differentiated by large Indels, genomic translocations, and reversions mainly in the U_L and IR_L regions, which resulted in total size variations of approximately 6 kbp between genotypes. Our findings are consistent with those reported by Morga-Jacquot et *al*. (2021), who previously reported on the evolution of OsHV-1 and highlighted the prevalence of genomic deletions as a driving force [37]. Consistent with their previous observations, our dataset also exhibited deletions in certain regions, including ORFs 11,

35 to 37, 48, and 62 to 63, and probably underestimated because of genomes non-redundancy. The exact mechanisms underlying the generation of these structural variants are not yet understood. However, it is plausible that recombination or rearrangement events, which are common in herpesviruses, play a crucial role in their appearance. Similarly, studies based on both Sanger sequencing [12] and recently Nanopore long reads sequencing [51] have demonstrated the presence of four major genomic isomers of the OsHV-1 genome, considered as structural variations, that were present in stoichiometric proportions. In herpesviruses, and especially in the herpes simplex virus 1, the presence of isomeric genomes is closely related to the replication and recombination mechanisms [87–90] and occurs early during the herpesvirus life cycle. During DNA replication, numerous random double-strand breaks efficiently initiate recombination events and give rise to new genomic rearrangements [91]. Importantly, these structural variants appear to be significant and not detrimental to the maintenance and replication stages of the viral life cycle. On the contrary, these variations may even be an advantage, as they potentially allow the emergence of new viral lineages and the adaptation of genotypes to new hosts [92].

Interestingly, OsHV-1 lineages sampled within the same species were genetically very close to each other (respectively 94% and 96% identities for the genomes collected from *M. gigas* and *O. edulis*), while lineages sampled from *O. edulis* were significantly different from those detected in *M. gigas* (between 91% and 94% identities). The genetic differentiation of OsHV-1 observed here is consistent with previous studies that have examined the genetic distances between viral genomes infecting closely related host species [93,94]. Similarly, Kolb *et al*. described the phenomenon of host shift in herpes simplex virus infecting old world monkeys, followed by co-evolution with its specific host, and demonstrated a strong correlation between viral genotype distance and phylogenetic host distance [93]. However, within the samples collected from *M. gigas*, there is a strong clustering of genotypes according to sampling date, corresponding to the reference (pre-2008) and µVar (post-2008) genotypes. Additionally, the µVar genotypes were not detected in *O. edulis* samples investigated in this study and collected in the absence of OsHV-1 mortality episodes.

Host shifts have been identified as a key driver of genomic rearrangements and recombination events that facilitate host adaptation and specialization. An experimental evolution study has shown that

structural rearrangements of the 3'-UTR of the Dengue virus genome were observed after 10 generations, leading to differential viral groups within mosquitoes and human [95]. These rearrangements could potentially confer a high capacity for rapid adaptation and evolution in response to host shifts. Although our analysis of OsHV-1 identified potential recombination events, only a limited number of them were detected. Specifically, five recombination events were identified between the stem-loop regions and a microsatellite $(TA)_n$ located in the U_L region, as well as in the homopolymer regions at the junctions of the X with repeated regions. Notably, it has been demonstrated that a bias in the detection of recombination breakpoints could occur in repetitive regions, intergenic regions, and areas with high G+C content [96]. This finding provides a potential explanation for the occurrence of only five recombination events observed in our study. Moreover, the frequency of recombination events in herpesviruses is not well-known [97]. However, studies have demonstrated numerous recombination events between Herpes simplex virus 1 and 2 [98,99] as well as between strains of Human cytomegalovirus (HCMV). These recombination events were detected throughout the genome and mainly in unique long regions [100,101] as observed in our dataset. The genome of OsHV-1 may have undergone evolutionary processes that have led to these rare occurrences of recombination and potentially contributed to the adaptation of each lineage to its respective host [102]. In order to fully understand the implications of the observed rearrangements and recombination events within the genome of OsHV-1, additional investigations are required. Specifically, it is crucial to examine additional OsHV-1 samples sourced from diverse geographic locations and host species. This would help to determine whether similar patterns of genomic rearrangement and selection exist across different populations. Additionally, conducting experimental infections studies aiming at elucidating the functional consequences of these genetic modifications would provide valuable insights into their biological significance.

Among OsHV-1 clusters, contrasted differentiation and selection patterns are observed along the genome. In particular, regions encoding for essential biological functions or domains involved in viral particle binding to host cells, DNA synthesis, replication and packaging, as well as trans- and membrane proteins may be critical for adapting to new host species. Indeed, viral entry proteins play a crucial role

in enhancing viral fitness, and a single amino acid mutation within these proteins can alter the virus's host range [103,104]. Furthermore, it has been shown that domains, such as envelope domains (including membrane glycoproteins and transmembrane receptors), auxiliary domains (such as Zinc-finger, RING type, dinucleoside kinase) and modulation and control domains (such as Interleukin, Interferon-regulatory factor or Zinc-finger) have been gained, duplicated, or lost during the evolution of Herpesviridae [105]. Many of these acquired domains allowed viruses to bind specifically to host cells and to evade or modulate the host immune system. In our study, most of the selection signals have been detected within these domains, suggesting a potential adaptation of essential proteins leading to a specialization of viral genotypes to their host species. Furthermore, the host species was found to play a significant role in driving viral diversity through selection at the consensus scale [95]. The selective pressure imposed by different host species has resulted, for example, in a diversification of Dengue viral lineages in mosquitoes and a reduction of viral population diversity in mammals [95].

Previous studies have not specifically investigated the intra-sample genetic diversity between infected host species but were more interested in comparing viral diversity in different organs or different individuals (e.g. vaccinated or unvaccinated) [106,107]. However, it is interesting to study within-host viral diversity and compare it to better understand minor diversity, differentiation and selection pressure. To investigate OsHV-1 diversity and differentiation according to host species, global scale genomic analyses were completed by within samples and within clusters. At the intra-sample scale, a clear contrast in the polymorphisms detected between host species was observed. Only 92 SNPs out of 1,051 were shared between both groups, and they were located within repeated regions and homopolymers at the junction between X and the repeated regions of the OsHV-1 genome. The accumulation of polymorphisms within the repeated regions at low frequencies has already been observed [37]. There are several possible explanations for these results. Firstly, these regions are duplicated in the virus genome. Repeated regions are known, in herpesviruses, to allow rearrangements of single regions by pairing between these regions [91]. In addition, these regions contain multiple small repeated sequences that can vary in size and number from one genome to another. Some of the polymorphisms detected in these regions can be explained by the loss or gain of some of these repeats. Secondly, the use of a non-

redundant genome to study the virus nucleotide diversity may induce a bias. Indeed, both the (TR_L/IR_L and IR_S/TR_S) repeat regions may contain polymorphism, which is detected here in a single region. Finally, GC-rich regions or regions containing various repeats, hairpin structures or long homopolymer stretches are notoriously difficult to sequence [108]. Repeated regions combine several of these problematic structures such as GC-rich, multiple repeats, and long homopolymer stretches. Polymorphisms detected in these regions are therefore possibly difficult to interpret and their accumulation may be linked to sequencing errors due to the difficult template. Indeed, these regions are difficult to access by the polymerases used during sequencing steps due to DNA methylations or folds. Sequencing technologies that do not use polymerases, such as Nanopore, could potentially solve these sequencing problems.

The genotype association graph used to compare the co-occurrence frequency of each polymorphism within and between both genotype groups revealed three distinct groups: two corresponding to *M. gigas* and one corresponding to *O. edulis* population. These results suggested that genotypes present in *M. gigas* displayed a higher level of genetic diversity compared to those originating from *O. edulis*, that could be explained by host populations size and species culture. *M. gigas* is currently the most widely cultivated oyster in the world, with enormous population sizes, both locally within oyster beds or in the various production basins, and worldwide. Moreover, this species is often transferred from one basin to another, particularly in France due to breeding practices [109]. In the case of the flat oyster, populations are small in France, mainly located in Brittany for a long time, and are transferred locally between North and South Brittany during their growth. It is particularly true for the flat oyster samples used in this study, because they were collected from natural populations which are not transferred. Moreover, the populations are distributed in "patches", which does not allow direct connections between the different populations and limit viral transmission [110].

The fine-scale diversity observed within OsHV-1 provides further support to findings obtained from consensus-based analyses and confirms the hypothesis that the virus is differentiated based on host species and has undergone coevolution with its respective new host following adaptation. An important avenue for future research would be to investigate the coevolution between OsHV-1 and multiple host

species using phylogenetic approaches. This would involve sampling different host species from various locations to obtain a comprehensive understanding of the coevolutionary dynamics between the virus and its hosts and give further insights into the evolutionary processes and host-virus interactions shaping the viral diversity pattern.

As mentioned earlier, our sample selection aimed to maximize heterogeneity to explore the potential effects of various factors on viral diversity. However, it is crucial to acknowledge that the variability in sampling host stage, dates, and origin may have introduced biases in our results [111,112]. For example, this variability could have led to the overrepresentation or underrepresentation of specific viral genotypes or subgroups associated with particular factors such as host developmental stage variability or sampling time after infection, thereby compromising the accuracy of our representation of the viral diversity or have significant impact on ancestral states reconstruction [113]. Furthermore, the heterogeneity of our dataset introduces challenges in disentangling confounding effects among factors, making it difficult to differentiate between clusters parameters and establish clear causal relationships between variables and their impact on the viral diversity. This complexity can hinder our ability to draw definitive conclusions about the specific factors influencing viral dynamics. Moreover, comparisons between samples may be limited due to inherent differences, such as temporal dynamics in viral spread or diversity across species, host stages, or sampling years. These differences introduce additional complexities and may hinder direct comparisons between samples, potentially impacting the generalizability of our findings.

CST frequency and OsHV-1 evolutionary history

Herpes like viruses particles in bivalves were first observed by Farley *et al*. in 1972 infecting adults of the *C. virginica* oyster species in USA during laboratory experiments [22]. Shortly after, infections of another oyster species, *O. edulis*, were detected in the United Kingdom [12]. Since these initial detections, OsHV-1 infections have been reported in various marine bivalve species, including *M. gigas* [25–27], *P. maximus* [15], *R. philippinarum* [20,21] and *R. decussatus* [20,21].

OsHV-1 is a measurably evolving virus [37]. At the global genomic scale, the mean evolution rate inferred in this study was 7.19E-05 nucleotide substitutions per site per year for the global phylogeny which is comparable to the one inferred previously [37]. Within the *M. gigas* cluster, mutation rates vary from 5.99E-05 to 3.29E-05 nucleotide substitutions per site per year while within the *O. edulis* cluster, mutation rate is slightly lower, estimated at 1.16E-05 nucleotide substitutions per site per year. These results were expected, since the effective population size of *M. gigas* is larger than that of *O. edulis*, which are now in decline. Nevertheless, the evolution rate of OsHV-1 is much higher than the evolution rate estimated for members of the Herpesvirales order under co-speciation respectively, 3.0E-09 nucleotide substitutions per site per year for mammalian and avian herpesviruses [114], 4.4E-09 nucleotide substitutions per site per year for herpesviruses collected only from mammals [115] and 4.37E-09 nucleotide substitutions per site per year for cyprinivirus [116]. These differences in evolution rates can be explained by the oyster farming conditions and by wild population sizes. In fact, oysters live in close proximity to one another, in large population densities, which may facilitate virus transmission between oysters. It is well-known that the easier viral transmission is, the more viruses evolve and generate diversity because of increased replication rates [117]. However, in this case, *M. gigas* virus groups are much larger than those of *O. edulis*, so the rate of evolution should be much higher for viruses infecting *M. gigas*, unless a second host with large populations is infected by the *O. edulis* genotypes and maintains a high mutation rate. A second hypothesis to explain the higher mutation rates in OsHV-1 flat oyster cluster than in Pacific oyster cluster is closely related to the reproduction cycle of the flat oyster, which stores larvae in its shell before releasing them to fix on a substrate [118]. During this period, when the larvae are contained within the mother's shell, larval densities are extremely high, which facilitates viral transmission and influences the mutation rate of the viral genome.

The analysis of ancestral traits indicated that OsHV-1 has evolved and adapted to two distinct lineages, one infecting *M. gigas* and the other infecting *O. edulis*. These two lineages clustered together based on host species and year of sampling. The MRCA of these two lineages was traced back to *M. gigas*, suggesting that this host species may have played a crucial role in the evolution and emergence of the reference genotype. Intriguingly, the reconstruction of ancestral traits using genomes with and without

recombinant regions produced two different tree topologies, indicating the possibility of coinfection of a single oyster by the two genotypes and the occurrence of a recombination event between these genotypes [119], which could have contributed to the evolution and emergence of the virus infecting *O. edulis*. Furthermore, although the first event of CST led to the emergence of two highly distinct genotypes, no subsequent CST or inter-species recombination events have been observed while the two host species *O. edulis* and *M. gigas* are living in nearby environment. As the only three other OsHV-1 genomes assembled from infected *C. farreri* (MF509813 and GQ153938.1) and *A. broughtonii* (KP412538.1) were not collected in France and are genetically distant from the genotypes sampled from *M. gigas* and *O. edulis* [37,51], they were not included in the study. Moreover, no other samples collected from different host species were available to complete the dataset. These results provide an initial insight into the evolutionary history of OsHV-1, but need to be supplemented by the contribution of other viral genomes sampled from different species to understand the involvement of different host species in viral diversity and infer the full range of CST events.

The frequency of CST occurrences is determined by the presence of facilitators or barriers encountered by the virus [104,120]. CST events are particularly frequent in RNA viruses because of their rapid evolution [102]. One of the most well-known viruses that can cross-species barriers is the Influenza virus, which is capable of rapidly adapting to different host species. The success of these viruses in infecting new hosts has been linked to their ability to adapt to a wide range of host species, notably by changing their virus surface proteins [121] and by using different strategies to evade the host's immune system, such as the production of non-structural proteins that interfere with the host immune system [122]. Moreover, their segmented genomes provide advantages for antigenic drift, antigenic shift, and recombination between strains infecting the same host, thus producing new combinations of genes that potentially favor the emergence of new viruses and increase their ability to infect and spread in new and unexpected host species [122–125]. By contrast, CST events in herpesviruses are extremely rare due to intrinsic genomic properties that lead to a persistent or latency stage rather than CST. Firstly, herpesviruses use specific receptors, such as Class III membrane protein or entry glycoprotein, that are specific to the herpesvirus family to enter host cells [126]. In the case of OsHV-1, only one Class III

membrane protein was predicted in OsHV-1 genome, mainly involved in catalyzing fusion of viral and cellular membranes, suggesting a high specificity of the viral genes for a particular host species. In addition, herpesviruses have co-evolved with their specific host species over a long time, leading to specific adaptations that enhance their ability to infect and persist within their natural host rather than infecting multiple hosts [127–129]. A herpesvirus that is adapted to one species may not be able to effectively be transmitted to or evade the immune system of a different species, reducing its ability to establish an infection [130]. Furthermore, herpesviruses have a low mutation rate, because of error correcting activity of exonuclease-polymerase, which does not allow a rapid evolution and adaptation to the new host species.

In the specific case of OsHV-1, CST events between oysters may be potentially limited by ecological and physiological barriers. In fact, only a few oyster production sites (Brittany, Normandy and Languedoc-Roussillon, Marennes-Oléron and Arcachon) bring the two oyster species together. In addition, given the phylogeny of oyster species and the genetic distance between them, physiological barriers are inevitably present, notably due to the immune defenses of the different host species [131]. In our study, high selective sweep signals were detected within the ORFs coding for transmembrane and membrane proteins, as well as for other functions involved in the maintenance of the replication process and packaging of the virus suggesting adaptation of the different OsHV-1 genotypes to a specific host species. Moreover, a previous study involved injecting flat oysters with OsHV-1 virus isolated from naturally moribund Pacific oysters, resulting in lower viral loads and reduced mortality rates compared to *M. gigas* infections. These findings suggest the presence of specific host-virus interactions and potentially during the immune response [47]. Based on the results, the co-cultivation of the two oyster species *M. gigas* and *O. edulis* does not stand for a major infection hazard, but it could increase genetic recombination and cross-species transmission events opportunities thus potentially promoting the emergence of variants that could be problematic.

Based on our findings, it seems that the French ancestral OsHV-1 genotype did originate from *M. gigas* but rather was differentiated into at least two genotypes that subsequently adapted to different marine bivalve host species such as *O. edulis* and *M. gigas*. The adaptation process continued through

coevolution, which was probably facilitated by host-specific factors such as immunity, living environment (temperature, salinity, light, host density), and food availability, as well as viral genomic properties such as mutation rates, host entry receptor specificity, and low genetic diversity.

Conclusion

In this study, our objective was to explore the diversity of OsHV-1 within oysters, between oysters, and among two oyster species. This study aimed to shed light on the evolutionary history of OsHV-1 by examining potential cross-species transmission events and subsequent genetic differentiation of viral lineages. Our findings align with previous research and emphasize the significant influence of host species on the genetic diversity and differentiation of OsHV-1 clusters. OsHV-1 has evolved into two distinct lineages in France. One lineage, known as the variant µVar, primarily infects *M. gigas*, while the other lineage could emerge after a cross-species transmission event between *M. gigas* and *O. edulis*, specifically infecting *O. edulis*. Our analysis revealed that actually each oyster species is primarily infected by a different OsHV-1 genotype, and no additional cross-species transmission events have been identified thus far.

Considering the obtained results and the differentiation among various viral lineages, it would be prudent to improve the classification of *Malacoherpesviridae* by incorporating the observed genotypes found in different host species.

Furthermore, our analysis identified selection signals within regions responsible for crucial biological functions and domains involved in viral particle binding, DNA synthesis, replication, and packaging. We also observed the presence of transmembrane proteins that are critical for host adaptation. These findings suggest the possibility of OsHV-1 undergoing adaptation and specialization in response to its respective host species.

To deepen our understanding of the evolutionary processes and host-virus interactions, research should focus on investigating the coevolution between OsHV-1 and a wider range of host species using phylogenetic approaches. This would involve sampling from diverse host species across different

geographical regions. By conducting such studies, we can gain valuable insights into the diversification, adaptation, and coevolutionary dynamics of OsHV-1 with its various hosts.

Acknowledgments

The present study was supported by the French Institute for Exploitation of the Sea (Ifremer), by DGAL (French General Directorate for Food) through the National Reference Laboratory for Mollusc Diseases and by the European-Union Reference Laboratory for Mollusc Diseases, Ifremer, La Tremblade. CP was supported by grant from the region Nouvelle-Aquitaine and from the scientific direction of Ifremer. We also thank the SeBiMER from Ifremer for their technical assistance in bioinformatic.

We declare that we have no competing interests.

References

1. Roche B, Benbow ME, Merritt R, Kimbirauskas R, McIntosh M, Small PLC, et al. Identifying the Achilles' heel of multi-host pathogens: The concept of keystone "host" species illustrated by Mycobacterium ulcerans transmission. 8(4):045009. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3925833/

2. Woolhouse MEJ, Taylor LH, Haydon DT. Population Biology of Multihost Pathogens. 292(5519):1109–12. Available from: https://www.science.org/doi/10.1126/science.1059026

3. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. 451(7181):990–3. Available from: https://www.nature.com/articles/nature06536

4. Archie EA, Luikart G, Ezenwa VO. Infecting epidemiology with genetics: a new frontier in disease ecology. 24(1):21–30. Available from: https://www.sciencedirect.com/science/article/pii/S0169534708003303

5. Handel A, Rohani P. Crossing the scale from within-host infection dynamics to between-host transmission fitness: a discussion of current assumptions and knowledge. 370(1675):20140302. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4528500/

6. VanderWaal KL, Ezenwa VO. Heterogeneity in pathogen transmission: mechanisms and methodology. 30(10):1606–22. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2435.12645

7. Geoghegan JL, Senior AM, Holmes EC. Pathogen population bottlenecks and adaptive landscapes: overcoming the barriers to disease emergence. 283(1837):20160727. Available from: https://royalsocietypublishing.org/doi/full/10.1098/rspb.2016.0727

8. Kuiken T, Holmes EC, McCauley J, Rimmelzwaan GF, Williams CS, Grenfell BT. Host Species Barriers to Influenza Virus Infections. 312(5772):394–7. Available from: https://www.science.org/doi/10.1126/science.1122818

9. Park M, Loverdo C, Schreiber SJ, Lloyd-Smith JO. Multiple scales of selection influence the evolutionary emergence of novel pathogens. 368(1614):20120333. Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2012.0333

10. Azab W, Dayaram A, Greenwood AD, Osterrieder N. How Host Specific Are Herpesviruses? Lessons from Herpesviruses Infecting Wild and Endangered Mammals. 5(1):53–68. Available from: https://www.annualreviews.org/doi/10.1146/annurev-virology-092917-043227

11. Bandín I, Dopazo CP. Host range, host specificity and hypothesized host shift events among viruses of lower vertebrates. 42(1):67. Available from: http://www.veterinaryresearch.org/content/42/1/67

12. Davison AJ, Trus BL, Cheng N, Steven AC, Watson MS, Cunningham C, et al. A novel class of herpesvirus with bivalve hosts. 86(1):41–53. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/vir.0.80382-0

13. Huff JL, Barry PA. B-Virus (Cercopithecine herpesvirus 1) Infection in Humans and Macaques: Potential for Zoonotic Disease. 9(2):246–50. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2901951/

14. Arzul I, Corbeil S, Morga B, Renault T. Viruses infecting marine molluscs. 147:118–35.

15. Arzul I, Nicolas JL, Davison AJ, Renault T. French Scallops: A New Host for Ostreid Herpesvirus-1. 290(2):342–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0042682201911869

16. Arzul I, Renault T, Lipart C, Davison AJ. Evidence for interspecies transmission of oyster herpesvirus in marine bivalves. 82(4):865–70. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/0022-1317-82-4-865

17. Carnegie RB, Arzul I, Bushek D. Managing marine mollusc diseases in the context of regional and international commerce: policy issues and emerging concerns. 371(1689):20150215. Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0215

18. Bai CM, Morga B, Rosani U, Shi J, Li C, Xin LS, et al. Long-range PCR and high-throughput sequencing of Ostreid herpesvirus 1 indicate high genetic diversity and complex evolution process. 526:81–90. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0042682218303052

19. Xia J, Bai C, Wang C, Song X, Huang J. Complete genome sequence of Ostreid herpesvirus-1 associated with mortalities of Scapharca broughtonii broodstocks. 12(1):110. Available from: http://www.virologyj.com/content/12/1/110

20. Renault, Arzul. Herpes-like virus infections in hatchery-reared bivalve larvae in Europe: specific viral DNA detection by PCR. 24(3):161–7. Available from: http://doi.wiley.com/10.1046/j.1365-2761.2001.00282.x

21. Renault T. Infection herpétiques chez les invertébrés : détection de virus de type herpès chez les mollusques bivalves marins. (2):401–3. Available from: https://archimer.ifremer.fr/doc/00000/2903/

22. Farley CA, Banfield WG, Kasnic G, Foster WS. Oyster Herpes-Type Virus. 178(4062):759–60. Available from: https://www.sciencemag.org/lookup/doi/10.1126/science.178.4062.759

23. Comps M. Epizootic diseases of oysters associated with viral infections. (18):23–37.

24. Renault T, Lipart C, Arzul I. A herpes-like virus infects a non-ostreid bivalve species: virus replication in Ruditapes philippinarum larvae. 45:1–7. Available from: http://www.int-res.com/abstracts/dao/v45/n1/p1-7/

25. Burioli EAV, Prearo M, Houssin M. Complete genome sequence of Ostreid herpesvirus type 1 µVar isolated during mortality events in the Pacific oyster Crassostrea gigas in France and Ireland. 509:239–51. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0042682217302040

26. Nicolas JL, Comps M, Cochennec N. Herpes-like virus infecting Pacific oyster larvae. 12(1):11.

27. Renault T, Cochennec R, Le Deuff RM, Chollet B. HERPES-LIKE VIRUS INFECTING JAPANESE OYSTER (CRASSOSTREA GIGAS) SPAT. :3.

28. Grizel H, Heral M. Introduction into France of the Japanese oyster (Crassostrea gigas). (47):399–403. Available from: https://archimer.ifremer.fr/doc/00000/2760/

29. FAO. Production mondiale par source de production de la pêche et aquaculture - Quantité (1950 - 2021) [Internet]. Available from: https://www.fao.org/fishery/statistics-query/fr/global_production/global_production_quantity

30. Burge CA, Griffin FJ, Friedman CS. Mortality and herpesvirus infections of the Pacific oyster Crassostrea gigas in Tomales Bay, California, USA. 72(1):31–43.

31. EFSA Panel on Animal Health and welfare (AHAW). Scientific Opinion on the increased mortality events in Pacific oysters, *Crassostrea gigas*: Oyster mortality. 8(11):1894. Available from: http://doi.wiley.com/10.2903/j.efsa.2010.1894

32. Paul-Pont I, Dhand NK, Whittington RJ. Spatial distribution of mortality in Pacific oysters Crassostrea gigas: reflection on mechanisms of OsHV-1 transmission. 105(2):127–38.

33. Renault T, Moreau P, Faury N, Pepin JF, Segarra A, Webb S. Analysis of Clinical Ostreid Herpesvirus 1 (Malacoherpesviridae) Specimens by Sequencing Amplified Fragments from Three Virus Genome Areas. 86(10):5942–7. Available from: https://journals.asm.org/doi/10.1128/JVI.06534-11

34. Davison AJ. Evolution of the herpesviruses. 86(1–2):69–88. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378113501004928

35. Segarra A, Pépin JF, Arzul I, Morga B, Faury N, Renault T. Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, Crassostrea gigas, in France in 2008. 153(1):92–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168170210002364

36. Arzul I, Renault T, Lipart C. Experimental herpes-like viral infections in marine bivalves: demonstration of interspecies transmission. 46:1–6. Available from: http://www.int-res.com/abstracts/dao/v46/n1/p1-6/

37. Morga B, Jacquot M, Pelletier C, Chevignon G, Dégremont L, Biétry A, et al. Genomic Diversity of the Ostreid Herpesvirus Type 1 Across Time and Location and Among Host Species. 12:711377. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.711377/full

38. Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. 11(10):728–36. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3980634/

39. Sunde J, Yıldırım Y, Tibblin P, Forsman A. Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies. 11. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2020.00218

40. Abbadi M, Zamperin G, Gastaldelli M, Pascoli F, Rosani U, Milani A, et al. Identification of a newly described OsHV-1 µvar from the North Adriatic Sea (Italy). 99(5):693–703. Available from: https://www.microbiologyresearch.org/content/journal/jgv/10.1099/jgv.0.001042

41. Ren W, Chen H, Renault T, Cai Y, Bai C, Wang C, et al. Complete genome sequence of acute viral necrosis virus associated with massive mortality outbreaks in the Chinese scallop, Chlamys farreri. 10(1):110. Available from: https://doi.org/10.1186/1743-422X-10-110

42. Arzul I, Langlade A, Chollet B, Robert M, Ferrand S, Omnes E, et al. Can the protozoan parasite Bonamia ostreae infect larvae of flat oysters Ostrea edulis? 179(1):69–76. Available from: https://www.sciencedirect.com/science/article/pii/S030440171100094X

43. Kuny CV, Szpara ML. Alphaherpesvirus Genomics: Past, Present and Future. :41–80. Available from: https://www.caister.com/cimb/abstracts/v42/41.html

44. Schikorski D, Renault T, Saulnier D, Faury N, Moreau P, Pépin JF. Experimental infection of Pacific oyster Crassostrea gigas spat by ostreid herpesvirus 1: demonstration of oyster spat susceptibility. 42(1):27. Available from: http://www.veterinaryresearch.org/content/42/1/27

45. da Silva P, Renault T, Fuentes J, Villalba A. Herpesvirus infection in European flat oysters Ostrea edulis obtained from brood stocks of various geographic origins and grown in Galicia (NW Spain). 78:181–8. Available from: http://www.int-res.com/abstracts/dao/v78/n3/p181-188/

46. Comps M, Cochennec N. A Herpes-like Virus from the European Oyster Ostrea edulis L. 62(2):201–3. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022201183710980

47. López Sanmartín M, Power DM, de la Herrán R, Navas JI, Batista FM. Experimental infection of European flat oyster Ostrea edulis with ostreid herpesvirus 1 microvar (OsHV-1µvar): Mortality, viral load and detection of viral transcripts by in situ hybridization. 217:55–62. Available from: https://www.sciencedirect.com/science/article/pii/S0168170216300739

48. Howard D W. Histological techniques for marine bivalve mollusks and crustaceans [Internet]. p. 218. Report No.: 5. Available from: https://repository.library.noaa.gov/view/noaa/1790

49. Pepin JF, Riou A, Renault T. Rapid and sensitive detection of ostreid herpesvirus 1 in oyster samples by real-time PCR. 149(2):269–76. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0166093408000396

50. Delmotte J, Pelletier C, Morga B, Galinier R, Petton B, Lamy JB, et al. Genetic diversity and connectivity of the Ostreid herpesvirus 1 populations in France: A first attempt to phylogeographic inference for a marine mollusc disease. 8(1):veac039. Available from: https://doi.org/10.1093/ve/veac039

51. Dotto-Maurel A, Pelletier C, Morga B, Jacquot M, Faury N, Dégremont L, et al. Evaluation of tangential flow filtration coupled to long-read sequencing for ostreid herpesvirus type 1 genome assembly. 8(11). Available from: https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000895

52. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. 34(17):i884–90. Available from: https://academic.oup.com/bioinformatics/article/34/17/i884/5093234

53. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. 9(4):357–9. Available from: http://www.nature.com/articles/nmeth.1923

54. Peñaloza C, Gutierrez AP, Eöry L, Wang S, Guo X, Archibald AL, et al. A chromosome-level genome assembly for the Pacific oyster Crassostrea gigas. 10(3):giab020. Available from: https://doi.org/10.1093/gigascience/giab020

55. Boutet I, Alves Monteiro HJ, Baudry L, Takeuchi T, Bonnivard E, Billoud B, et al. Chromosomal assembly of the flat oyster (Ostrea edulis L.) genome as a new genetic resource for aquaculture. n/a(n/a). Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.13462

56. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. 25(16):2078–9. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btp352

57. Li H. Seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences.

58. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. 19(5):455–77. Available from: http://www.liebertpub.com/doi/10.1089/cmb.2012.0021

59. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. 215(3):403–10. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022283605803602

60. Morse AM, Calabro KR, Fear JM, Bloom DC, McIntyre LM. Reliable Detection of Herpes Simplex Virus Sequence Variation by High-Throughput Resequencing. 9(8):226. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5580483/

61. Katoh K, Misawa K, Kuma K ichi, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. 30(14):3059–66.

62. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. 38(7):3022–7. Available from: https://doi.org/10.1093/molbev/msab120

63. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. 7(1):214. Available from: https://doi.org/10.1186/1471-2148-7-214

64. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. 11(1):119. Available from: https://doi.org/10.1186/1471-2105-11-119

65. González-Tortuero E, Krishnamurthi R, Allison HE, Goodhead IB, James CE. Comparative analysis of gene prediction tools for viral genome annotation [Internet]. Bioinformatics; Available from: http://biorxiv.org/lookup/doi/10.1101/2021.12.11.472104

66. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale [Internet]. bioRxiv; p. 2021.06.03.446934. Available from: https://www.biorxiv.org/content/10.1101/2021.06.03.446934v2

67. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. 40(7):1023–5. Available from: https://www.nature.com/articles/s41587-021-01156-3

68. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks [Internet]. bioRxiv; p. 2022.04.08.487609. Available from: https://www.biorxiv.org/content/10.1101/2022.04.08.487609v1

69. Cameron ES, Schmidt PJ, Tremblay BJM, Emelko MB, Müller KM. Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. 11(1):22302. Available from: https://www.nature.com/articles/s41598-021-01636-1

70. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. Available from: http://arxiv.org/abs/1207.3907

71. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. 31(13):2202–4. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv112

72. R Development Core Team. R: A language and environment for statistical computing. [Internet]. R Foundation for Statistical Computing, Vienna, Austria; Available from: http://www.R-project.org

73. Csardi G, Nepusz T. The igraph software package for complex network research. Complex Systems:1695. Available from: https://igraph.org

74. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. New York, NY: Springer New York; Available from: https://link.springer.com/10.1007/978-0-387-98141-3

75. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. 14(7):1394–403. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC442156/

76. Li H. Minimap2: pairwise alignment for nucleotide sequences. Birol I, editor. 34(18):3094–100. Available from: https://academic.oup.com/bioinformatics/article/34/18/3094/4994778

77. Goel M, Sun H, Jiao WB, Schneeberger K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. 20(1):277. Available from: https://doi.org/10.1186/s13059-019-1911-0

78. Goel M, Schneeberger K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. 38(10):2922–6. Available from: https://doi.org/10.1093/bioinformatics/btac196

79. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. 43(3):e15. Available from: https://doi.org/10.1093/nar/gku1196

80. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. 34(12):3299–302. Available from: https://doi.org/10.1093/molbev/msx248

81. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. 59(3):307–21.

82. Winter DJ. MMOD: an R library for the calculation of population differentiation statistics. 12(6):1158–60. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2012.03174.x

83. Alachiotis N, Pavlidis P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. 1(1):1–11. Available from: https://www.nature.com/articles/s42003-018-0085-8

84. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). 2(1):vew007. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4989882/

85. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. 67(5):901–4. Available from: https://doi.org/10.1093/sysbio/syy032

86. Burioli EAV, Prearo M, Riina MV, Bona MC, Fioravanti ML, Arcangeli G, et al. Ostreid herpesvirus type 1 genomic diversity in wild populations of Pacific oyster Crassostrea gigas from Italian coasts. 137:71–83. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022201116300519

87. Mahiet C, Ergani A, Huot N, Alende N, Azough A, Salvaire F, et al. Structural Variability of the Herpes Simplex Virus 1 Genome *In Vitro* and *In Vivo*. 86(16):8592–601. Available from: https://jvi.asm.org/content/86/16/8592

88. López-Muñoz AD, Rastrojo A, Kropp KA, Viejo-Borbolla A, Alcamí A. Combination of long- and short-read sequencing fully resolves complex repeats of herpes simplex virus 2 strain MS complete genome. 7(6):000586. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8461477/

89. Chou J, Roizman B. Isomerization of herpes simplex virus 1 genome: identification of the cis-acting and recombination sites within the domain of the a sequence. 41(3):803–11. Available from: https://www.sciencedirect.com/science/article/pii/S0092867485800611

90. Chaitanya KV. Structure and Organization of Virus Genomes. :1–30. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7119911/

91. Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. Recombination in viruses: Mechanisms, methods of study, and evolutionary consequences. 30:296–307. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7106159/

92. Turner PE, Elena SF. Cost of Host Radiation in an RNA Virus. 156(4):1465–70. Available from: https://doi.org/10.1093/genetics/156.4.1465

93. Kolb AW, Brandt CR. Genomic nucleotide-based distance analysis for delimiting old world monkey derived herpes simplex virus species. 21:436. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7318535/

94. Longdon B, Day JP, Alves JM, Smith SCL, Houslay TM, McGonigle JE, et al. Host shifts result in parallel genetic changes when viruses evolve in closely related species. Lauring AS, editor. 14(4):e1006951. Available from: https://dx.plos.org/10.1371/journal.ppat.1006951

95. Villordo SM, Filomatori CV, Sánchez-Vargas I, Blair CD, Gamarnik AV. Dengue Virus RNA Structure Specialization Facilitates Host Adaptation. 11(1):e1004604. Available from: https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1004604

96. Lee K, Kolb AW, Sverchkov Y, Cuellar JA, Craven M, Brandt CR. Recombination Analysis of Herpes Simplex Virus 1 Reveals a Bias toward GC Content and the Inverted Repeat Regions. Longnecker RM, editor. 89(14):7214–23. Available from: https://journals.asm.org/doi/10.1128/JVI.00880-15

97. Renner DW, Szpara ML. Impacts of Genome-Wide Analyses on Our Understanding of Human Herpesvirus Diversity and Evolution. 92(1):e00908-17. Available from: https://journals.asm.org/doi/full/10.1128/JVI.00908-17

98. Burrel S, Boutolleau D, Ryu D, Agut H, Merkel K, Leendertz FH, et al. Ancient Recombination Events between Human Herpes Simplex Viruses. 34(7):1713–21. Available from: https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msx113

99. Koelle DM, Norberg P, Fitzgibbon MP, Russell RM, Greninger AL, Huang ML, et al. Worldwide circulation of HSV-2 × HSV-1 recombinant strains. 7(1):44084. Available from: https://www.nature.com/articles/srep44084

100. Sijmons S, Thys K, Mbong Ngwese M, Van Damme E, Dvorak J, Van Loock M, et al. High-throughput analysis of human cytomegalovirus genome diversity highlights the widespread occurrence of gene-disrupting mutations and pervasive recombination. 89(15):7673–95.

101. Lassalle F, Depledge DP, Reeves MB, Brown AC, Christiansen MT, Tutill HJ, et al. Islands of linkage in an ocean of pervasive recombination reveals two-speed evolution of human cytomegalovirus genomes. 2(1):vew017. Available from: https://academic.oup.com/ve/article-lookup/doi/10.1093/ve/vew017

102. Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, et al. Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases. 72(3):457–70. Available from: https://journals.asm.org/doi/10.1128/MMBR.00004-08

103. Thakur S, Sasi S, Pillai SG, Nag A, Shukla D, Singhal R, et al. SARS-CoV-2 Mutations and Their Impact on Diagnostics, Therapeutics and Vaccines. 9:815389. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8902153/

104. Webby R, Hoffmann E, Webster R. Molecular constraints to interspecies transmission of viral pathogens. 10(S12):S77–81. Available from: http://www.nature.com/articles/nm1151

105. Brito AF, Pinney JW. The evolution of protein domain repertoires: Shedding light on the origins of the Herpesviridae family. 6(1):veaa001. Available from: https://doi.org/10.1093/ve/veaa001

106. Al-Khatib HA, Smatti MK, Ali FH, Zedan HT, Thomas S, Ahmed MN, et al. Comparative analysis of within-host diversity among vaccinated COVID-19 patients infected with different SARS-CoV-2 variants. 25(11):105438. Available from: https://www.sciencedirect.com/science/article/pii/S2589004222017102

107. Gu H, Quadeer AA, Krishnan P, Ng DYM, Chang LDJ, Liu GYZ, et al. Within-host genetic diversity of SARS-CoV-2 lineages in unvaccinated and vaccinated individuals. 14(1):1793. Available from: https://www.nature.com/articles/s41467-023-37468-y

108. Kieleczawa J. Fundamentals of Sequencing of Difficult Templates—An Overview. 17(3):207–17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2291785/

109. Lupo C, Ezanno P, Arzul I, Garcia C, Jadot C, Joly JP, et al. How network analysis of oyster movements can improve surveillance and control programs of infectious diseases? 3. Available from: http://www.frontiersin.org/Community/AbstractDetails.aspx?ABS_DOI=10.3389/conf.FVETS.2016.02.00044

110. Nørgaard LS, Phillips BL, Hall MD. Infection in patchy populations: Contrasting pathogen invasion success and dispersal at varying times since host colonization. 3(5):555–66. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/evl3.141

111. Chikhi L, Sousa VC, Luisi P, Goossens B, Beaumont MA. The Confounding Effects of Population Structure, Genetic Diversity and the Sampling Scheme on the Detection and Quantification of Population Size Changes. 186(3):983–95. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2975287/

112. Sul JH, Martin LS, Eskin E. Population structure in genetic studies: Confounding factors and mixed models. 14(12):e1007309. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007309

113. Liu P, Song Y, Colijn C, MacPherson A. The impact of sampling bias on viral phylogeographic reconstruction. Falcão De Oliveira E, editor. 2(9):e0000577. Available from: https://dx.plos.org/10.1371/journal.pgph.0000577

114. McGeoch DJ, Dolan A, Ralph AC. Toward a Comprehensive Phylogeny for Mammalian and Avian Herpesviruses. 74(22):10401–6. Available from: https://journals.asm.org/doi/10.1128/JVI.74.22.10401-10406.2000

115. McGeoch DJ, Cook S, Dolan A, Jamieson FE, Telford EAR. Molecular Phylogeny and Evolutionary Timescale for the Family of Mammalian Herpesviruses. 247(3):443–58. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022283685701520

116. Donohoe O, Zhang H, Delrez N, Gao Y, Suárez NM, Davison AJ, et al. Genomes of Anguillid Herpesvirus 1 Strains Reveal Evolutionary Disparities and Low Genetic Diversity in the Genus Cyprinivirus. 9(5):998. Available from: https://www.mdpi.com/2076-2607/9/5/998

117. Peck KM, Lauring AS. Complexities of Viral Mutation Rates. 92(14):10.1128/jvi.01031-17. Available from: https://journals.asm.org/doi/10.1128/jvi.01031-17

118. Helmer L, Farrell P, Hendy I, Harding S, Robertson M, Preston J. Active management is required to turn the tide for depleted Ostrea edulis stocks from the effects of overfishing, disease and invasive species. 7:e6431.

119. Feng C, Wang J, Liston A, Kang M. Recombination Variation Shapes Phylogeny and Introgression in Wild Diploid Strawberries. 40(3):msad049. Available from: https://doi.org/10.1093/molbev/msad049

120. Tian J, Sun J, Li D, Wang N, Wang L, Zhang C, et al. Emerging viruses: Cross-species transmission of coronaviruses, filoviruses, henipaviruses, and rotaviruses from bats. 39(11):110969. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2211124722007550

121. Long JS, Mistry B, Haslam SM, Barclay WS. Host and viral determinants of influenza A virus species specificity. 17(2):67–81. Available from: https://www.nature.com/articles/s41579-018-0115-z

122. Shao W, Li X, Goraya MU, Wang S, Chen JL. Evolution of Influenza A Virus by Mutation and Re-Assortment. 18(8):1650. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5578040/

123. Kim H, Webster RG, Webby RJ. Influenza Virus: Dealing with a Drifting and Shifting Pathogen. 31(2):174–83. Available from: https://www.liebertpub.com/doi/10.1089/vim.2017.0141

124. Both GW, Sleigh MJ, Cox NJ, Kendal AP. Antigenic drift in influenza virus H3 hemagglutinin from 1968 to 1980: multiple evolutionary pathways and sequential amino acid changes at key antigenic sites. 48(1):52–60. Available from: https://journals.asm.org/doi/10.1128/jvi.48.1.52-60.1983

125. Webster RG, Laver WG, Air GM, Schild GC. Molecular mechanisms of variation in influenza viruses. 296(5853):115–21. Available from: http://www.nature.com/articles/296115a0

126. Krummenacher C, Carfí A, Eisenberg RJ, Cohen GH. Entry of Herpesviruses into Cells: The Enigma Variations. In: Pöhlmann S, Simmons G, editors. New York, NY: Springer New York; p. 178–95. (Advances in Experimental Medicine and Biology; vol. 790). Available from: http://link.springer.com/10.1007/978-1-4614-7651-1_10

127. McGeoch DJ, Rixon FJ, Davison AJ. Topics in herpesvirus genomics and evolution. 117(1):90–104. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168170206000219

128. Kaján GL, Doszpoly A, Tarján ZL, Vidovszky MZ, Papp T. Virus–Host Coevolution with a Focus on Animal and Human DNA Viruses. 88(1):41–56. Available from: https://doi.org/10.1007/s00239-019-09913-4

129. Waltzek T, Kelley G, Alfaro M, Kurobe T, Davison A, Hedrick R. Phylogenetic relationships in the family Alloherpesviridae. 84:179–94. Available from: http://www.int-res.com/abstracts/dao/v84/n3/p179-194/

130. O'Connor CM, Sen GC. Innate Immune Responses to Herpesvirus Infection. 10(8):2122. Available from: https://www.mdpi.com/2073-4409/10/8/2122

131. Danic-Tchaleu G, Heurtebise S, Morga B, Lapègue S. Complete mitochondrial DNA sequence of the European flat oyster Ostrea edulis confirms Ostreidae classification. 4(1):400. Available from: https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-4-400

Figure captions

Fig 1. Schematic representation of the experimental sampling design. (A-B) Larvae, spat, and adults naturally infected with the virus were collected from *M. gigas* and *O. edulis* in France and Sweden. (C) DNA was extracted from individual tissues and the viral copy number was assessed by qPCR. D) DNA samples were sequenced using the Illumina short-read technology. *M. gigas* and *O. edulis* samples and treatments were represented respectively in green and purple.

Fig 2. distribution of SNPs within the OsHV-1 genome. (A) Distribution of single nucleotide polymorphisms (SNPs) along the alignment of OsHV-1 de novo assembled genomes and (B) zoom on the repeated regions of OsHV-1 genomes. SNPs detected in the *M. gigas* population are represented in green, SNPs detected in the *O. edulis* population are represented in purple and SNPs detected in both populations are represented in blue. The intensity of the point varies according to the number of alleles detected for each position. Locations of SNPs are plotted against regions and positions along the OsHV-1 multiple alignment. Variations in GC content are represented in grey and when greater than 50% highlighted in red.

Fig 3. From intra sample population diversity to host specialization. A) Multidimensional scaling (MDS) plot showing allelic frequencies distribution at polymorphic sites. Points are colored according to the host species alleles were detected in. B) MDS representing oyster individual's similarity according to allelic frequencies of polymorphisms they harbored. The MDS analysis provides a visual representation of this relationship in a two-dimensional space and the genotype association graph shows the co-occurrence of alleles within individuals. Samples from O. edulis are shown in purple, while those from M. gigas are shown in green and are divided into two clusters. Links between samples/genotypes individuals are represented by lines, which vary in thickness based on the similarity between the two genotypes, and the color of the lines indicates if the individuals are from the same species or not.

Fig 4. Comparison of the six structural variants detected in the dataset. On the right side of the structural variation graph, the phylogenetic relationships between the six representative individuals are plotted (estimated by a maximum likelihood approach [80]. The description of genomic structural

variations was based on a pairwise comparison to the OsHV-1 reference genome (NC_005881.2, Davison et al. 2005). In each pairwise comparison, the reference genome is represented by a dark blue bar. Syntenic regions are represented in light grey, inversions in orange, and inverted translocations in light blue. Below the graph of structural variations is the architecture and position in the OsHV-1 reference genome.

Fig 5. Recombination hotspots. Linkage disequilibrium (LD) heatmap between pairs of SNPs (site 1 and 2) along the alignment of 40 de novo assembled OsHV-1 genomes and the three OsHV-1 reference genomes: OsHV-1 Ref. (NC_005881.2; Davison et al. 2005), µVarA (KY242785; Burioli et al. 2017) and µVarB (KY271630; Burioli et al. 2017). LD values vary between 0 and 1, consequently, the color scale varies respectively from beige to black. At the bottom, the architecture of OsHV-1 genomes and position along the multiple alignment are represented.

Fig 6. Consensus-based population genomics of OsHV-1. (A) Genetic differentiation between samples isolates from both species computed using the mean GST values within 1-kb windows along the multiple alignment of OsHV-1 genomes. The red line at y=0.545 represent the mean global GST, which gives an overall estimate of the genetic differentiation of OsHV-1 between species. Red dots on the curve represent windows having a mean GST higher than 0.6. (B) Distribution of Tajima's D values obtained from 1-kb windows for all genomes included in this study (in blue), for OsHV-1 genomes sampled from M. gigas (in green) and O.edulis (in purple) If values of Tajima's D were significant, dots were filled, while if not significant, dots are empty. The blue, green and purple lines represent the averaged Tajima's D value for each genome data-set. For a better comparison of the neutrality and differentiation measures, significant Tajima's D values are linked by dotted lines to GST values. Below the graph is represented the architecture and position in the OsHV-1 genome multiple alignment. Genomic regions are reported on the two graphs above. Additionally, regions present only in M. gigas and O. edulis OsHV-1 subpopulations are represented with green and purple backgrounds respectively.

Fig 7. Detection of selection patterns. (A) Detection of selective sweep signatures along the multiple alignment of OsHV-1 genomes in the global group (blue points), in the *O. edulis* group (purple points) and in the *M. gigas* group (green points). Below those graphs are represented the ORFs predicted on

ancestral sequence of each group and the architecture and position in the OsHV-1 genomes multiple alignment. (B) Pie charts representing the proportion of selective sweep signals detected within ORFs coding for specific function or domain for both species.

Fig 8. Evolution and demographic history of OsHV-1 within both host species. A) Time-scaled phylogeny of OsHV-1 genomes and ancestral state reconstruction of the discrete host species trait. Demographical reconstruction of OsHV-1 population size and observed population within B) *M. gigas* and C) *O. edulis*. Similarly, to other figures, color of the branches of the Maximum clade credibility (MCC) tree represents the host species from which the virus was isolated (green for samples from *M. gigas* and purple for samples from O. edulis). Numbers at branch nodes represent the posterior probabilities of the host species inference.

Supporting information

S1 Fig. Viral copies and reads number in the libraries used in the analysis. Bars represent the viral copies number per ng of DNA quantified by qPCR and points represent the viral reads number in each library. The color code is consistent with other figures: purple for samples from *O. edulis* and green for samples from *M. gigas*.

S2 Fig. Histological haematoxylin eosin tissue sections of *Ostrea edulis* larvae showing healthy larvae (A) and larvae presenting different lesions degrees associated with the detection of OsHV-1 (B to E). (A) Healthy flat oyster larvae with section of the digestive gland (DG), velum (Vel) and mantle (Mn). (B) larvae presenting a light haemocyte infiltration in the mantle (arrow); other tissues as digestive gland (DG) and posterior abductor muscle (Pam) presented no lesion. (C) Apparition of necrosis (arrow) and nuclear abnormalities (arrowhead) in the mantle and velum; digestive gland (DG) presented no lesion. (D) Larvae presenting a necrosis of all tissues. (E) Dead larvae in decomposition state. Scale-bars = 10µm.

S3 Fig. Rarefaction curves showing the number of polymorphisms detected by Freebayes according to the read depth in (A) libraries that contain more reads than the expected number needed to accurately detect polymorphisms and (B) libraries that does not contain enough reads than the expected number needed to accurately detect polymorphisms. The red dotted lines correspond to the value at which number of polymorphisms detected reach a plateau corresponding to the minimum number of reads required for effective detection of polymorphisms. This threshold has been set at 150k reads.

S4 Fig. Graphical representation of the global OsHV-1 genomes multiple alignment. The alignment contains one genome per line. Each colored block represents orthologous genomic region. White space represents gaps in the alignment. When a block lies above the center line the aligned region is in the forward orientation relative to the first genome sequence.

S5 Fig. Recombination events detected along OsHV-1 genomes multiple alignment. A) Phylogenetic relationships between the 43 individuals estimated by a maximum likelihood approach. B) Metadata corresponding to OsHV-1 samples, respectively, the host species, the origin country, the origin basin, the sampling year and the host developmental stage. C) Recombination plot showing syntenic regions in grey, recombination occurring on internal branches in red and recombination occurring on terminal branches in blue. Below the graph are represented the recombination hotspots and the architecture and position in the OsHV-1 genome multiple alignment.

S6 Fig. Heatmap of pairwise sequence identities comparison. Color scale varies according to identity percentages, from black for most closely related pairs of sequences to beige for less similar sequences. On the left of the heatmap, the phylogenetic relationships between the 43 individuals are plotted. Those were estimated by a maximum likelihood approach.

S7 Fig. Clock-likeness of OsHV-1 genomes. Regression of root-to-tip genetic distance against sampling time were computed for each host species based on ML tree using TempEst and the best fitting root method. Dots and lines are colored according to clusters defined on the ML tree: purple for OsHV-

1 group sampled from *O. edulis*, light green and dark green for OsHV-1 group sampled from *M. gigas* reference genotypes and µVar genotypes respectively.

S1 Table. Summary of sequencing data and reads alignments statistics on OsHV-1 reference genome (NC_005881). The 40 samples come from two host species: *Crassostrea gigas* and Ostrea edulis collected in France and Sweden between 2005 and 2012. Sequencing has been realized by different sequencing platform: [a]Ligan (CNRS UMR 8199, Lille, France); [b]Genome Quebec Company (Genome Quebec Innovation Center, McGill University, Montreal, Canada), [c]Fasteris Life Science (Plan-les-Ouates, Swiss) and [d]GenoToul (Toulouse, France).

S2 Table. Frequencies of the 985 minor polymorphisms identified by variant calling analyses.

S3 Table. Putative ORFs and associated functions impacted by intra-samples polymorphisms in OsHV-1 population collected from both *O. edulis* and *M. gigas*.

S4 Table. Ratio between number of polymorphisms and OsHV-1 genomic regions length.

S5 Table. Start, end and length of recombination breakpoints detected by Gubbins and number of SNPs removed for subsequent analyses with the recombinant region. Red values correspond to breakpoints also detected by linkage disequilibrium values.

S6 Table. Percentage of identity and number of differences between the 44 NR-genomes based on pairwise distances computing using MAFFT. Cells are colored with a gradient between green for highly similar genomes (~99,99 % identity or less than 100 bp differences) to red for less similar genomes (~91 % identity or more than 17 kbp differences).

S7 Table. Significant genetic differentiation among OsHV-1 genomes collected in both *O. edulis* and *M. gigas*. Initially, the mean GST values were calculated for 1-kbp windows along the OsHV-1 genomes. Subsequently, 100-bp windows were computed within each 1-kbp region that displayed substantial genetic differentiation (GST > 0.6). The table exclusively presents windows that exhibit significant GST values. For every distinct region of differentiation, information regarding the affected ORF as well as its associated function or domain are provided.

S8 Table. Summary of selective sweep events position and score detected using RaISD in both species. Additionally, information about ORF and function or domain impacted by these events are provided.

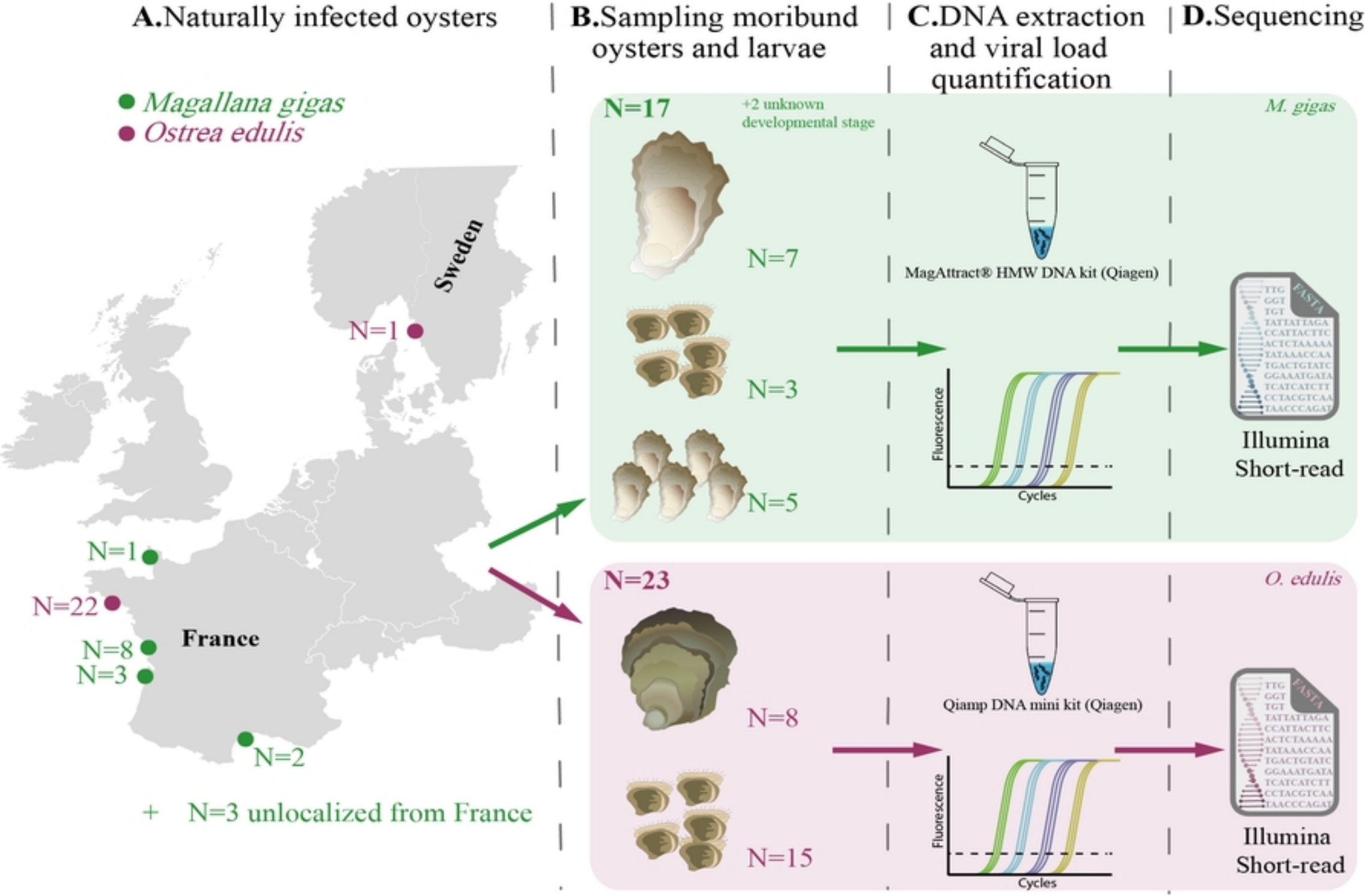
A.Naturally infected oysters
B.Sampling moribund oysters and larvae
C.DNA extraction and viral load quantification
D.Sequencing
Magallana gigas
Ostrea edulis
Sweden
N=1
France
N=1
N=22
N=8
N=3
N=2
+ N=3 unlocalized from France
N=17
+2 unknown developmental stage
M. gigas
N=7
N=3
N=5
MagAttract® HMW DNA kit (Qiagen)
Fluorescence
Cycles
FASTA
Illumina Short-read
N=23
O. edulis
N=8
N=15
Qiamp DNA mini kit (Qiagen)
Fluorescence
Cycles
FASTA
Illumina Short-read

Figure1

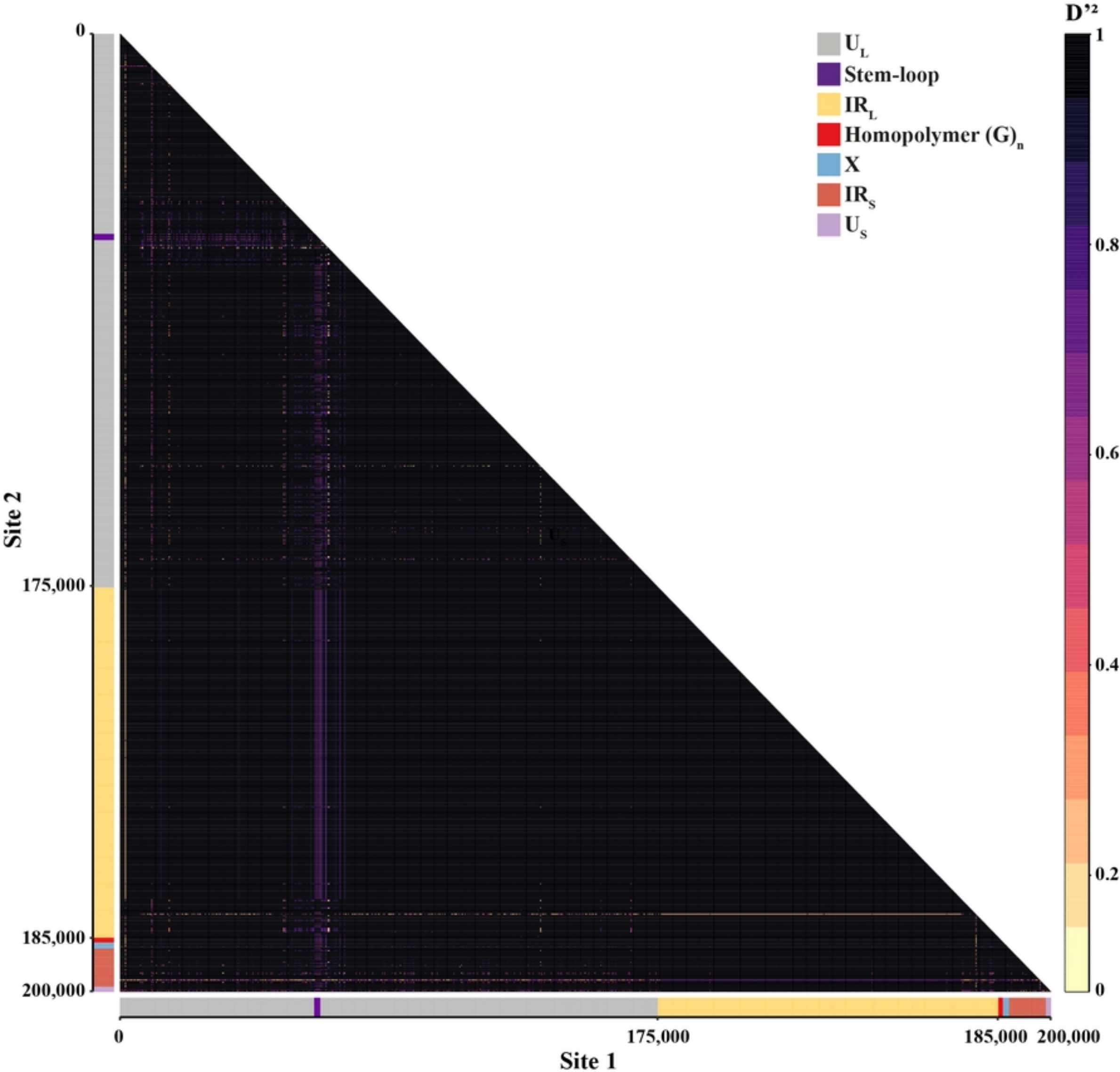
D'²
1
0.8
0.6
0.4
0.2
0
U_L
Stem-loop
IR_L
Homopolymer (G)_n
X
IR_S
U_S
Site 2
0
175,000
185,000
200,000
Site 1
0
175,000
185,000
200,000

Figure5

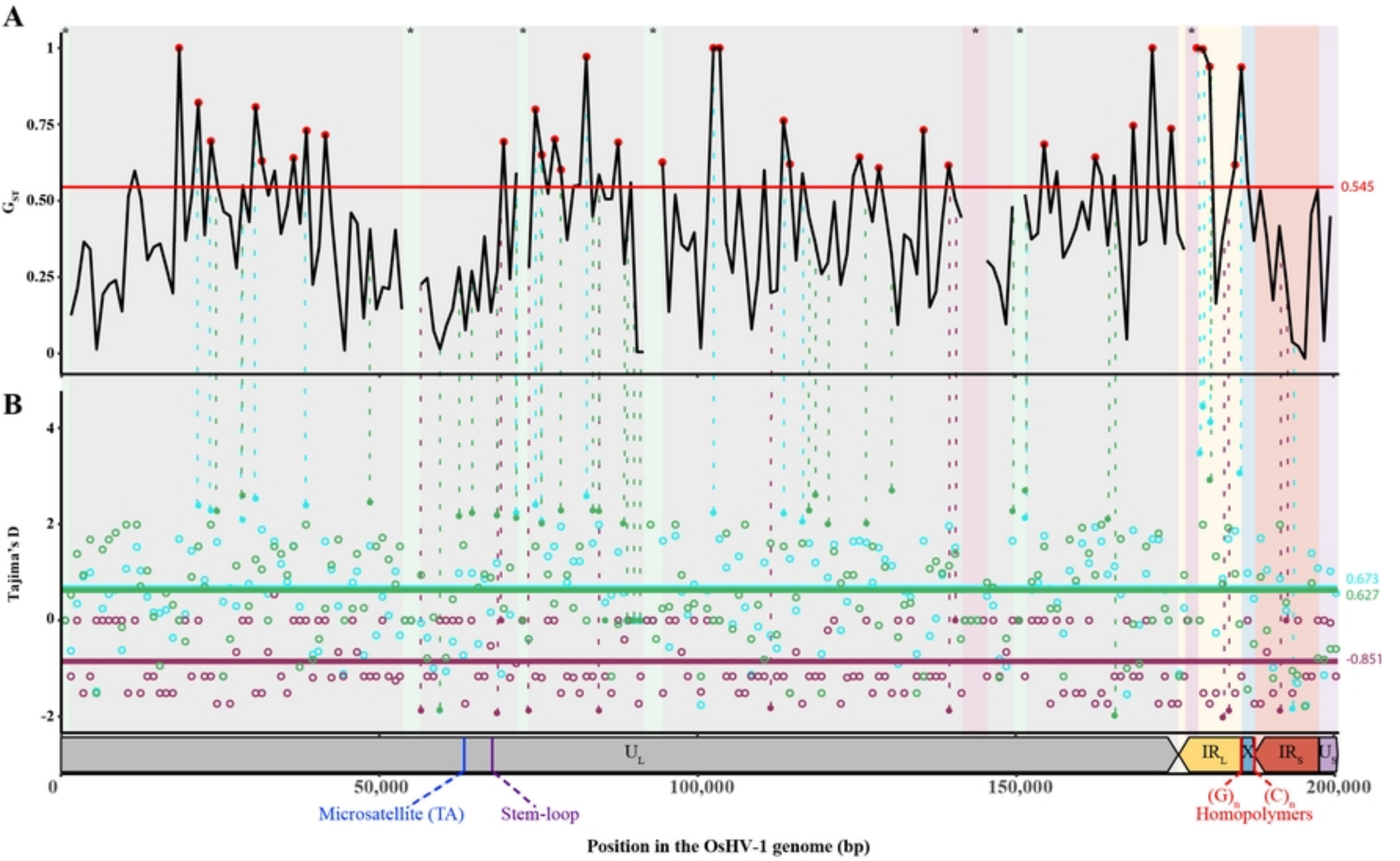
A
B
G_{ST}
Tajima's D
0.545
0.673
0.627
-0.851
U_L
IR_L
X
IR_S
U_S
0
50,000
100,000
150,000
200,000
Microsatellite (TA)
Stem-loop
$(G)_n$
$(C)_n$
Homopolymers
Position in the OsHV-1 genome (bp)

Figure6

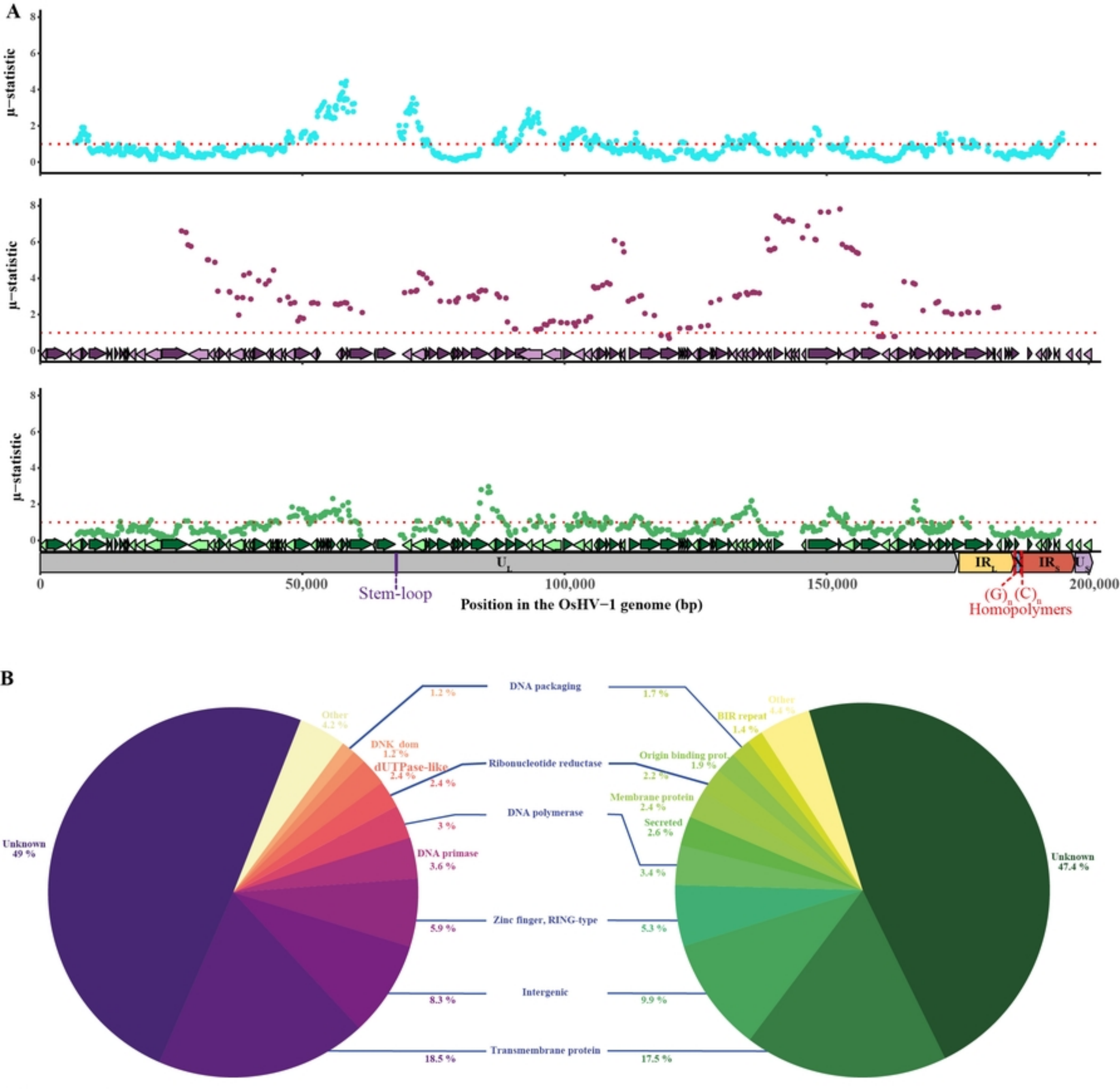
A
μ−statistic
0
2
4
6
8
U_L
IR_L
IR_S
U_S
0
50,000
100,000
150,000
200,000
Stem-loop
Position in the OsHV−1 genome (bp)
(G)n (C)n
Homopolymers
B
DNA packaging
1.2 %
1.7 %
Other
4.2 %
Other
4.4 %
BIR repeat
1.4 %
DNK dom
1.2 %
dUTPase-like
2.4 %
Ribonucleotide reductase
2.4 %
2.2 %
Origin binding prot.
1.9 %
Membrane protein
2.4 %
DNA polymerase
3 %
Secreted
2.6 %
3.4 %
Unknown
49 %
DNA primase
3.6 %
Unknown
47.4 %
Zinc finger, RING-type
5.9 %
5.3 %
Intergenic
8.3 %
9.9 %
Transmembrane protein
18.5 %
17.5 %

Figure7

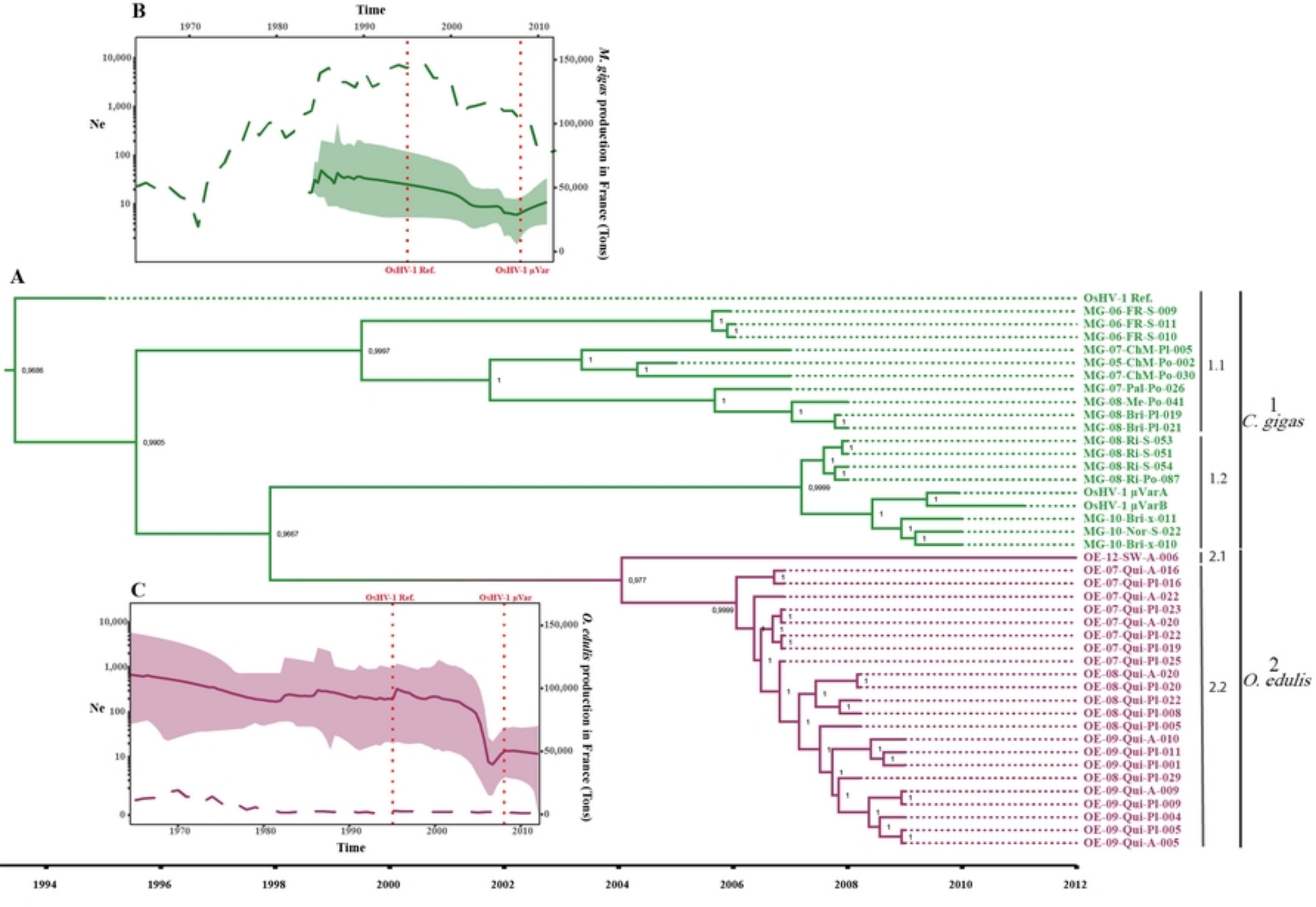
B
Time
1970
1980
1990
2000
2010
Ne
10,000
1,000
100
10
150,000
100,000
50,000
0
M. gigas production in France (Tons)
OsHV-1 Ref.
OsHV-1 µVar
A
OsHV-1 Ref.
MG-06-FR-S-009
MG-06-FR-S-011
MG-06-FR-S-010
MG-07-ChM-Pl-005
MG-05-ChM-Po-002
MG-07-ChM-Po-030
MG-07-Pal-Po-026
MG-08-Me-Po-041
MG-08-Bri-Pl-019
MG-08-Bri-Pl-021
MG-08-Ri-S-053
MG-08-Ri-S-051
MG-08-Ri-S-054
MG-08-Ri-Po-087
OsHV-1 µVarA
OsHV-1 µVarB
MG-10-Bri-x-011
MG-10-Nor-S-022
MG-10-Bri-x-010
OE-12-SW-A-006
OE-07-Qui-A-016
OE-07-Qui-Pl-016
OE-07-Qui-A-022
OE-07-Qui-Pl-023
OE-07-Qui-A-020
OE-07-Qui-Pl-022
OE-07-Qui-Pl-019
OE-07-Qui-Pl-025
OE-08-Qui-A-020
OE-08-Qui-Pl-020
OE-08-Qui-Pl-022
OE-08-Qui-Pl-008
OE-08-Qui-Pl-005
OE-09-Qui-A-010
OE-09-Qui-Pl-011
OE-09-Qui-Pl-001
OE-08-Qui-Pl-029
OE-09-Qui-A-009
OE-09-Qui-Pl-009
OE-09-Qui-Pl-004
OE-09-Qui-Pl-005
OE-09-Qui-A-005
1.1
1.2
2.1
2.2
1
C. gigas
2
O. edulis
0,9686
0,9997
0,9905
0,9999
0,9667
0,977
0,9999
C
OsHV-1 Ref.
OsHV-1 µVar
Ne
10,000
1,000
100
10
0
150,000
100,000
50,000
0
O. edulis production in France (Tons)
1970
1980
1990
2000
2010
Time
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012

Figure8

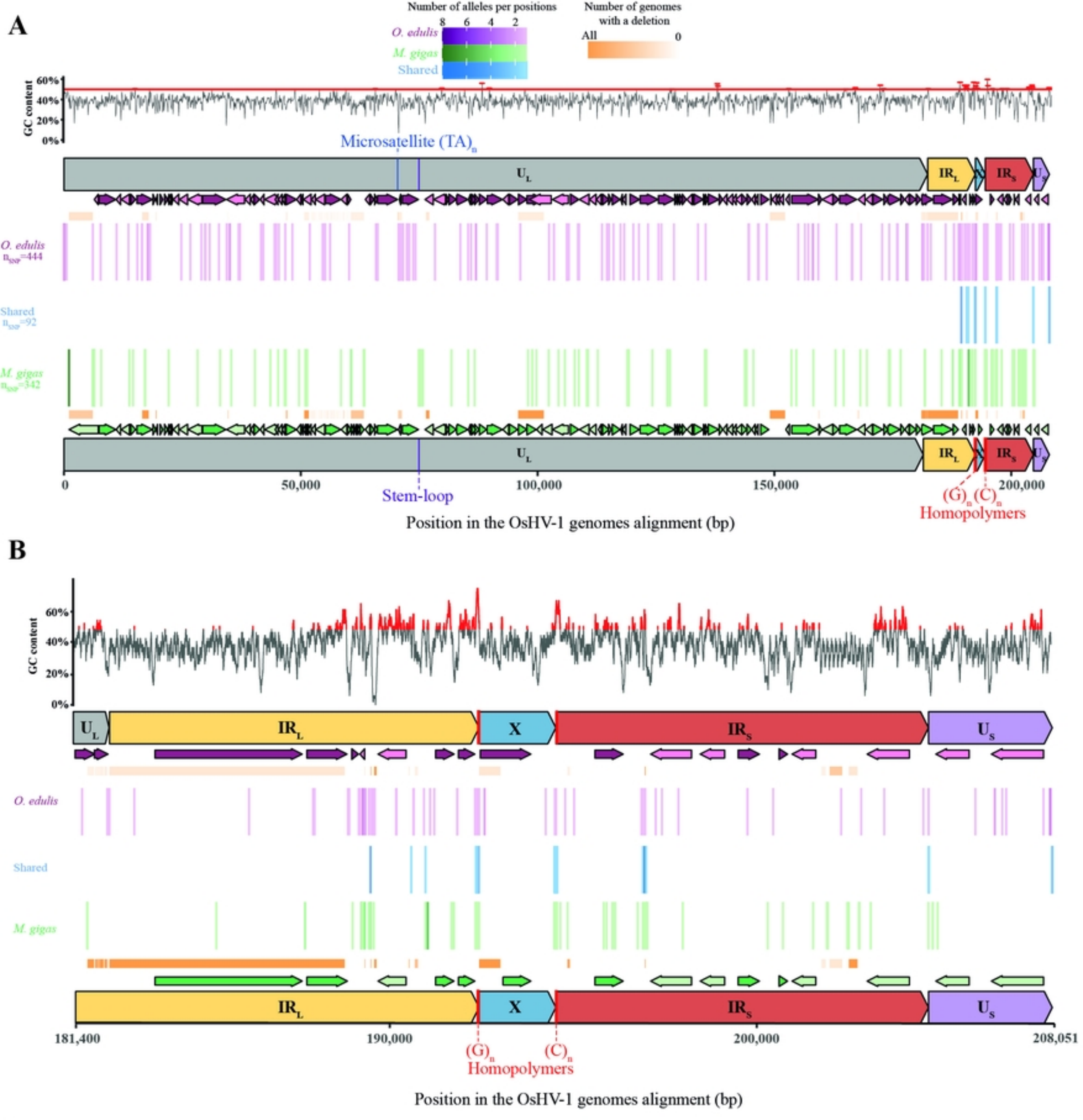
A
Number of alleles per positions
8 6 4 2
O. edulis
M. gigas
Shared
Number of genomes with a deletion
All
0
GC content
60%
40%
20%
0%
Microsatellite (TA)n
UL
IRL
IRS
US
O. edulis
nSNP=444
Shared
nSNP=92
M. gigas
nSNP=342
0
50,000
100,000
150,000
200,000
Stem-loop
(G)n (C)n
Homopolymers
Position in the OsHV-1 genomes alignment (bp)
B
GC content
60%
40%
20%
0%
UL
IRL
X
IRS
US
O. edulis
Shared
M. gigas
181,400
190,000
200,000
208,051
(G)n
(C)n
Homopolymers
Position in the OsHV-1 genomes alignment (bp)

Figure2

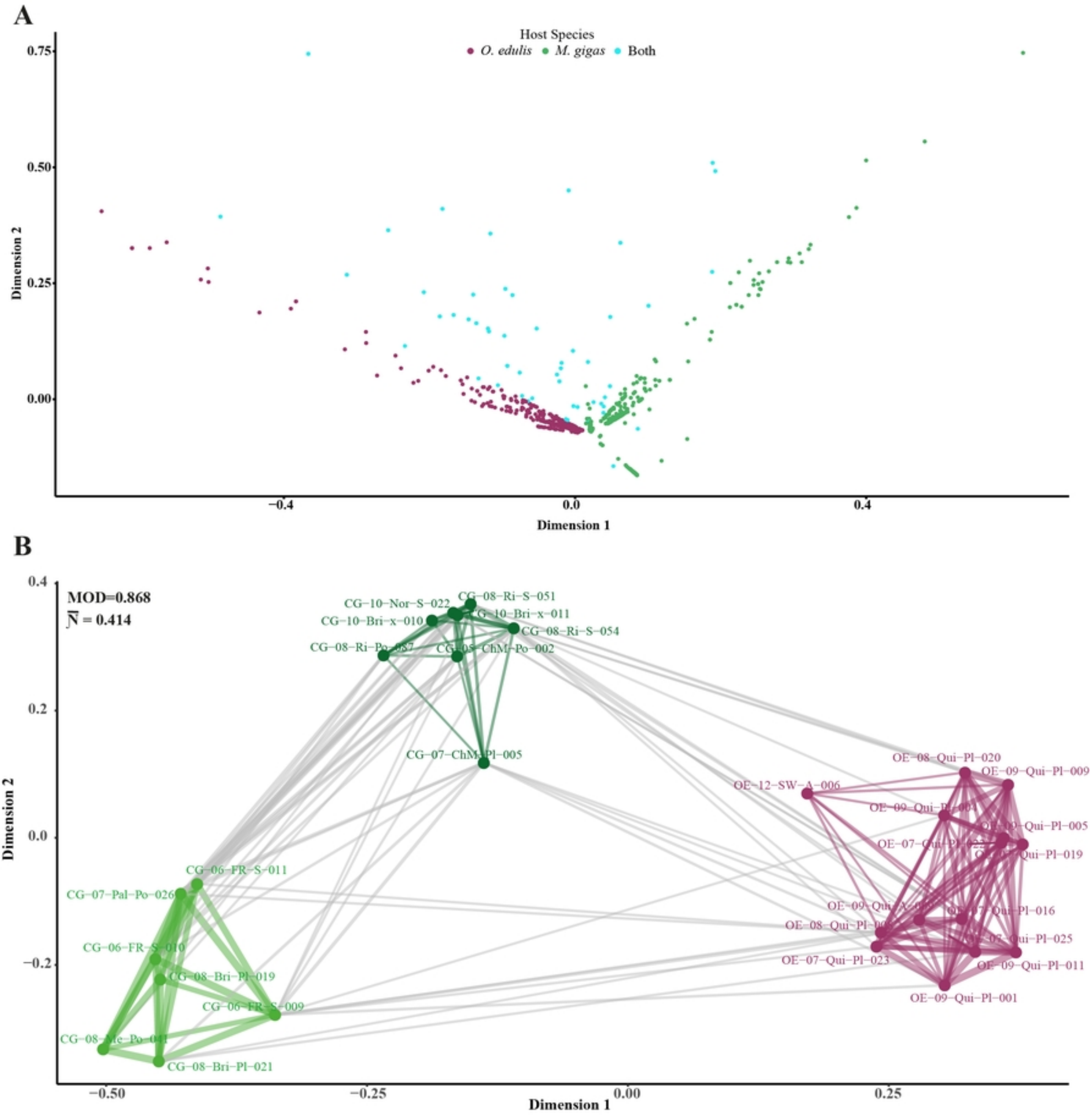
A
Host Species
O. edulis
M. gigas
Both
0.75
0.50
0.25
0.00
Dimension 2
−0.4
0.0
0.4
Dimension 1
B
MOD=0.868
N̄ = 0.414
CG−10−Nor−S−022
CG−08−Ri−S−051
CG−10−Bri−x−011
CG−10−Bri−x−010
CG−08−Ri−S−054
CG−08−Ri−Po−037
CG−05−ChM−Po−002
CG−07−ChM−Pl−005
OE−08−Qui−Pl−020
OE−09−Qui−Pl−009
OE−12−SW−A−006
OE−09−Qui−Pl−005
OE−07−Qui−Pl−019
CG−06−FR−S−011
CG−07−Pal−Po−026
OE−07−Qui−Pl−016
OE−07−Qui−Pl−025
CG−06−FR−S−010
OE−07−Qui−Pl−023
OE−09−Qui−Pl−011
CG−08−Bri−Pl−019
CG−06−FR−S−009
OE−09−Qui−Pl−001
CG−08−Bri−Pl−021
0.4
0.2
0.0
−0.2
Dimension 2
−0.50
−0.25
0.00
0.25
Dimension 1

Figure3

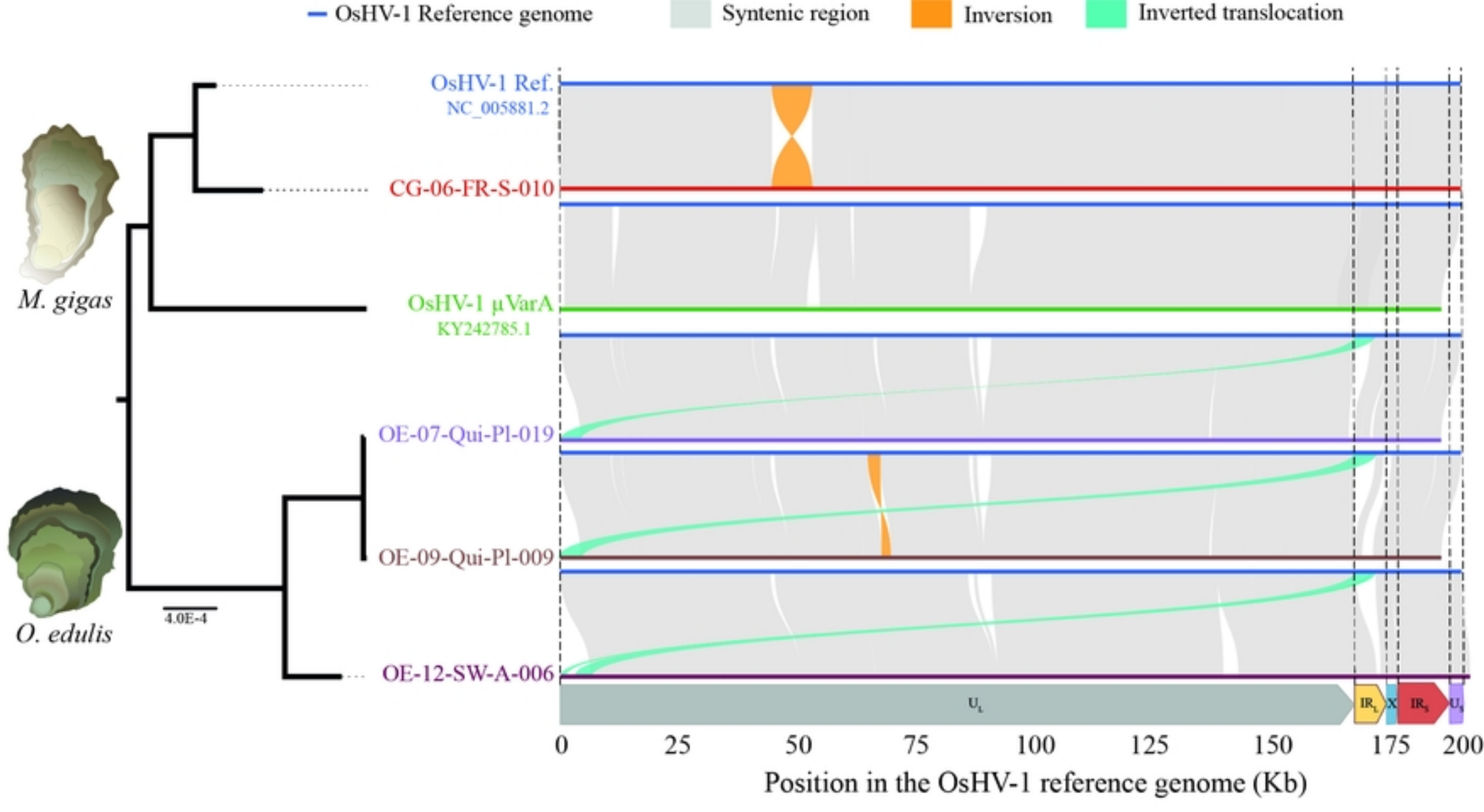
OsHV-1 Reference genome
Syntenic region
Inversion
Inverted translocation
OsHV-1 Ref.
NC_005881.2
CG-06-FR-S-010
OsHV-1 µVarA
KY242785.1
OE-07-Qui-Pl-019
OE-09-Qui-Pl-009
OE-12-SW-A-006
M. gigas
O. edulis
4.0E-4
0
25
50
75
100
125
150
175
200
Position in the OsHV-1 reference genome (Kb)

Figure4