



**HAL**  
open science

## Evaluation of tangential flow filtration coupled to long-read sequencing for ostreid herpesvirus type 1 genome assembly

Aurélie Dotto-Maurel, Camille Pelletier, Benjamin Morga, Maude Jacquot, Nicole Faury, Lionel Dégremont, Maëlis Bereszczynki, Jean Delmotte, Jean-Michel Escoubas, Germain Chevignon

### ► To cite this version:

Aurélie Dotto-Maurel, Camille Pelletier, Benjamin Morga, Maude Jacquot, Nicole Faury, et al.. Evaluation of tangential flow filtration coupled to long-read sequencing for ostreid herpesvirus type 1 genome assembly. *Microbial Genomics*, 2022, 8 (11), 10.1099/mgen.0.000895 . hal-03863848

**HAL Id: hal-03863848**

**<https://hal.science/hal-03863848>**

Submitted on 21 Nov 2022

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

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

# Evaluation of tangential flow filtration coupled to long-read sequencing for ostreid herpesvirus type 1 genome assembly

Aurélien Dotto-Maurel<sup>1†</sup>, Camille Pelletier<sup>1†</sup>, Benjamin Morga<sup>1</sup>, Maude Jacquot<sup>1</sup>, Nicole Faury<sup>1</sup>, Lionel Dégremont<sup>1</sup>, Maëlis Bereszczynski<sup>2</sup>, Jean Delmotte<sup>2</sup>, Jean-Michel Escoubas<sup>2,\*</sup> and Germain Chevignon<sup>1,\*</sup>

## Abstract

Whole-genome sequencing is widely used to better understand the transmission dynamics, the evolution and the emergence of new variants of viral pathogens. This can bring crucial information to stakeholders for disease management. Unfortunately, aquatic virus genomes are usually difficult to characterize because most of these viruses cannot be easily propagated *in vitro*. Developing methodologies for routine genome sequencing of aquatic viruses is timely given the ongoing threat of disease emergence. This is particularly true for pathogenic viruses infecting species of commercial interest that are widely exchanged between production basins or countries. For example, the ostreid herpesvirus type 1 (OsHV-1) is a Herpesvirus widely associated with mass mortality events of juvenile Pacific oyster *Crassostrea gigas*. Genomes of Herpesviruses are large and complex with long direct and inverted terminal repeats. In addition, OsHV-1 is unculturable. It therefore accumulates several features that make its genome sequencing and assembly challenging. To overcome these difficulties, we developed a tangential flow filtration (TFF) method to enrich OsHV-1 infective particles from infected host tissues. This virus purification allowed us to extract high molecular weight and high-quality viral DNA that was subjected to Illumina short-read and Nanopore long-read sequencing. Dedicated bioinformatic pipelines were developed to assemble complete OsHV-1 genomes with reads from both sequencing technologies. Nanopore sequencing allowed characterization of new structural variations and major viral isomers while having 99,98% of nucleotide identity with the Illumina assembled genome. Our study shows that TFF-based purification method, coupled with Nanopore sequencing, is a promising approach to enable in field sequencing of unculturable aquatic DNA virus.

## DATA SUMMARY

All sequence data have been deposited in NCBI under Bioproject ID PRJNA823871. Nanopore data have been deposited to SRA under accession number SAMN27363714 with base-called fastq reads under accession number SRR19025776 and raw fast5 and slow5 file under accession number SRR19039499; Illumina fastq reads have been deposited to SRA under accession number SAMN27363675. OsHV-1 assembled genome from Nanopore data have been deposited to GenBank under accession number ON411413; OsHV-1 assembled genome from Illumina data have been deposited to GenBank under accession number ON411412.

Received 13 May 2022; Accepted 06 September 2022; Published 10 November 2022

**Author affiliations:** <sup>1</sup>Ifremer, ASIM, F-17390 La Tremblade, France; <sup>2</sup>IHPE, Univ. Montpellier, CNRS, Ifremer, UPVD, F-34095 Montpellier, France.

**\*Correspondence:** Jean-Michel Escoubas, jean.michel.escoubas@ifremer.fr; Germain Chevignon, germain.chevignon@ifremer.fr

**Keywords:** *Crassostrea gigas*; Illumina; ostreid herpesvirus type 1; oxford nanopore technologies; tangential flow filtration; virus.

**Abbreviations:** ASW, artificial seawater; Cp, crossing point; cp, copy number; DP, DNA polymerase; FSW, filtered seawater; GU, genomic units; IRL, inverted repeats long; IRS, inverted repeats small; mNGS, metagenomic next-generation sequencing; ONT, Oxford Nanopore Technologies; ORF, open reading frames; OsHV-1, ostreid herpesvirus type; P, permeate; POMS, pacific oyster mortality syndrome; R, retentate; SNP, single nucleotide polymorphism; SV, structure variations; TFF, tangential flow filtration; TRL, tandem repeats long; TRS, tandem repeats small; UL, unique long; US, unique small; VLP, virus-like particles.

All sequence data have been deposited in NCBI under Bioproject ID PRJNA823871; biosample IDs SAMN27363714 and SAMN27363675; and to GenBank under accession number ON411412 and ON411413.

†These authors contributed equally to this work

**Data statement:** All supporting data, code and protocols have been provided within the article or through supplementary data files. Three supplementary figures and two supplementary tables are available with the online version of this article.

000895 © 2022 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

### Impact Statement

Several aquatic viruses are pathogenic to species of commercial interest, leading to massive value losses during disease outbreaks. To better understand the origin, transmission and propagation of such viruses, more genomic data is required and needs to be analysed. However genomic characterization of viruses that are mainly unculturable is a major challenge. In the present study, we have successfully tested the combination of tangential flow filtration (TFF) and Nanopore sequencing for virus purification and genome sequencing. Our results demonstrate that it is now possible to easily sequence the complete genome of large DNA viruses directly from host infected tissues without prior *in vitro* propagation.

## INTRODUCTION

Viruses found in aquatic environments are often unculturable and their genomic characterization often relies on our ability to enrich samples for viral nucleic acid. To date, several methods have been implemented to this end and different workflows are applied depending on the type of virus studied: small or large virus, RNA or DNA virus, virus with linear or circular genome, single or double strand virus. Enrichment methods range from rapid and simple steps involving low-speed centrifugation, filtration and nuclease treatment, to more complex methods involving ultracentrifugation, tilling PCR, rolling circle amplification and hybridization-based capture. While the former shows higher efficiency at purifying viral nucleic acid, these methods habitually rely on *a priori* knowledge of the targeted virus and are limited at detecting a new variant or new pathogen [1–4]. Recently, metagenomic next-generation sequencing (mNGS) has been successfully tested for broad-spectrum pathogen identification for clinical diagnosis of infectious diseases, outbreak surveillance, and pathogen discovery [5, 6]. However, the cost, sequencing depth and background contamination are true limitations to the accuracy of mNGS-based diagnostics relative to specific targeting testing.

Given that aquaculture is continuously faced with the challenge of identifying novel pathogenic viruses infecting different fish, crustaceans, and shellfish species, there is an urgent need to develop tools enabling those viruses to be sequenced easily and with limited bias to develop timely disease control strategies.

Ostreid herpesvirus type 1 (OsHV-1) is a double strand DNA virus belonging to the Malacoherpesviridae family and is the only species of the genus *Ostreavirus* [7]. OsHV-1 has been associated with mass mortality events of naive Pacific oyster *Crassostrea gigas* [8] from larvae to adults [9, 10]. It has plagued the oyster production worldwide, from Europe to north and south America, Australia and New Zealand, and in a lesser extent in Asia, for more than a decade [11–17]. OsHV-1 mostly affects spat and juvenile (less than 18 months) in oyster farms, and leads to mass mortalities that can reach 100% within days [18–20]. Research efforts have revealed a series of factors contributing to the disease, including seawater temperature, oyster age and genetics [8, 9, 16, 18, 21–24]. Recently, holistic molecular approaches uncovered the pathogens associated to Pacific oyster mortality syndrome (POMS) [25]. These studies showed that an infection by OsHV-1 is the critical step of the infectious process leading to an immune-compromised state followed by a microbiota destabilization that ‘opens the door’ to opportunistic bacterial pathogens (e.g. *Vibrio* spp.) that target haemocytes to induce their lysis [26]. The infectious process is completed with subsequent bacteraemia ultimately leading to oyster death (reviewed in [27]).

Since the characterization of the emergent genotype OsHV-1  $\mu$ Var, associated with the 2008 high mortality events [18], several variants of the  $\mu$ Var genotype have been described (reviewed in [28]). However, most studies on OsHV-1 genetic diversity have been based on PCR molecular markers focusing on a few variable regions of the viral genome, which do not reflect the whole genomic diversity [29, 30]. Indeed, OsHV-1  $\mu$ Var whole-genome sequencing [31] and its comparison with the reference genome published in 2005 [7] showed that both genomes differed at the nucleotide level (single nucleotide polymorphism, SNP) as well as at the structural level by the loss or gain of several genomic regions carrying many open reading frames (ORFs). This observation indicates that whole-genome sequencing is needed to fully reveal viral diversity and better understand the structure and evolution of the virus populations [31].

OsHV-1 particles harbour a linear genome of approximately 207 kilobases (kb) with the majority of the 128 ORFs detected (54.7%), having no known homologies with other organisms [31]. OsHV-1 genomic organization is characteristic of other Herpesviruses consisting of two invertible unique regions ( $U_L$  for ‘Unique long’ 170 kbp;  $U_S$  for ‘Unique small’ 3.5 kbp) each flanked by tandem and inverted repeats ( $TR_L/IR_L$ : 7.5 kbp;  $TR_S/IR_S$ : 9.7 kbp) separated by a third unique region (X: 1.5 kbp) with a global genomic organization  $TR_L-U_L-IR_L-X-IR_S-U_S-TR_S$ . Furthermore, several isomers of the virus cohabit at equimolar levels in the same samples [7].

To date, 50 fully assembled OsHV-1 genomes are available in public databases, those genomes have been sequenced and assembled using various protocol and sequencing technologies (Table S1, available in the online version of this article). Although recent genomes have been obtained from short-read mNGS [31–34], OsHV-1 complete genome assembly remains a complex process. Indeed, a short-read approach is ideal for the identification of small-scale variation like SNPs but this technology struggles to identify variations at a larger scale like structure variations (SV) or large inverted and repeated regions. Due to its intrinsic

Herpesvirus genome architecture with high G+C content regions, long repeated regions ( $TR_L/IR_L$  and  $TR_S/IR_S$ ), and the presence of four major isomers differing in the relative orientations of the two unique regions, it is extremely difficult, if not impossible, to solve the complete structure of OsHV-1 genome using short-read sequencing [35].

These limitations can now be addressed by long-read sequencing approaches of which Oxford Nanopore Technologies (ONT) is one example. Indeed, ONT has already demonstrated its ability to sequence and assemble several types of viruses from RNA [36–42] as well as DNA [43–46] matrices. Nanopore sequencing has also been successfully tested on Herpesviruses like the Human Herpes Virus Type 1 (HHV-1) [47], the Human cytomegalovirus (HCMV) [48], the Equine herpesvirus 1 (EHV-1) [49] and the koi herpesvirus (KHV) [50]. In addition, ONT already allowed researchers to sequence viruses infecting aquaculture species like the White spot syndrome virus (WSSV) infecting shrimp [51] or the Salmonid alphavirus (SAV) infecting Atlantic salmon [52]. Finally, ONT has also demonstrated efficiency to sequence viruses from sea-water samples [53–55]. However, it has never been tested for the characterization of the OsHV-1 genome.

In order to implement ONT long-read sequencing, high molecular weighted DNA is required in a large quantity. However, due to the lack of a cell line, that could be used for the propagation and isolation of viral particles, OsHV-1 must be extracted from infected animal tissues. Therefore, it is extremely challenging to obtain viral DNA in sufficient quantity to perform long-read mNGS.

To overcome these difficulties, we need a new approach for viral particle purification coupled with long-read sequencing to assemble complex OsHV-1 genomes efficiently and accurately.

One approach to purify viral particles is tangential flow filtration (TFF). This method has been widely used for purification of a large variety of virus particles, such as particles of human influenza A virus [56], insect baculovirus [57, 58], HIV-like virus [59], oncolytic measles virus [60] and of several waterborne viruses like norovirus and adenovirus [61]. While TFF has been used to reduce oyster pathogen concentration in seawater to bio-secure shellfish farms [62], it has never been used with the goal of purifying those pathogens. Recently, TFF coupled with Nanopore sequencing, has been used to concentrate a large amount of seawater for the characterization of bacteriophage population [63].

In the present study, we have developed a TFF-based method allowing purification and concentration of OsHV-1 infectious particles from infected oyster tissues and we have evaluated the relative benefits of long-read (Nanopore) and short-read (Illumina) sequencing for genome assembly. Our results show that TFF purification of OsHV-1 viral particles coupled with ONT sequencing is a promising strategy for OsHV-1 whole-genome reconstruction and to decipher genetic diversity and evolution of virus populations.

## METHODS

### Oyster production

Experimental *C. gigas* spat was produced at the Ifremer hatchery in La Tremblade, (Marennes-Oléron Bay, Charente-Maritime, France). Oysters belonged to a line that had been characterized for its higher susceptibility to OsHV-1 [64]. Thirty adult oysters were conditioned in seawater at 20 °C for 2 months to favour gametogenesis. Then, ripe oysters were triggered for a mass spawning event in March 2015. Larvae and spat were cultured as described by [64, 65] and they were always maintained in our controlled facilities using filtered and UV-treated seawater to protect from pathogens naturally present in the Marennes-Oléron Bay. Prior to the experiment, spat (2 g and 4 cm) were acclimatized at 19 °C for 2 weeks in flow-through 120 l tanks using filtered and UV-treated seawater enriched in phytoplankton (*Skeletonema costatum*, *Isochrysis galbana* and *Tetraselmis suecica*).

### Viral inoculum preparation and OsHV-1 quantification

OsHV-1 inoculum were prepared from tissue of infected oysters following protocols of Schikorsky *et al.* [66]. Briefly, gills and mantle were dissected from moribund oysters naturally or experimentally infected with OsHV-1 in previous trials and frozen at –80 °C. Then, ten volumes of 0.2 µm filtered and artificial seawater (ASW) were added and tissues were crushed on ice using an Ultraturax mixer (3×5 s). The tissue mash was clarified by centrifugation (1000 g, 5 min, 4 °C), and then filtered consecutively in sterile conditions using syringe filters at 8, 0.45 and 0.22 µm pore sizes (Millipore; Billerica, USA).

Quantification of OsHV-1 was carried out using quantitative PCR (qPCR) using primers amplifying a 197 bp fragment of the OsHV-1 DNA polymerase (DP) gene. DNA was extracted from viral inoculum using the QiaAmp DNA Mini Kit (Qiagen) according to the manufacturer's protocol. DNA purity ( $1.8 \leq A_{260} \text{ nm}/A_{280} \text{ nm} \leq 2.0$ ;  $2.0 \leq A_{260} \text{ nm}/A_{230} \text{ nm} \leq 2.20$ ) and concentration were checked using a Nano-Drop ND-1000 spectrometer (Thermo Scientific) and Qubit® dsDNA HS assay kits (Molecular Probes Life Technologies), respectively.

qPCRs were performed in duplicate using a Mx3005 P Thermocycler (Agilent). Amplification reactions were each performed in a total volume of 20 µl. Each reaction contained 5 µl of DNA extracted from viral inoculum or from TFF fractions (readjusted to initial inoculum volume) or 5 µl of Milli-Q water for the negative control, 10 µl of Brilliant III Ultra-Fast SYBR® Green PCR

Master Mix (Agilent), 2 µl of primer OsHVDP For (forward) 5'-ATTGATGATGTGGATAATCTGTG-3' and 2 µl of primer OsHVDP Rev (reverse) 5'-GGTAAATACCATTGGTCTTGTTC-3' [67] at the final concentration of 550 nM each, and 1 µl of Milli-Q water. qPCR cycling conditions were as follows: 3 min at 95 °C followed by 40 cycles of amplification at 95 °C for 5 s and 60 °C for 20 s. For absolute quantification, the amplification product of the DP gene was cloned into the pCR4-Topo vector and replicated in *Escherichia coli* DH5α (Invitrogen). Plasmid was extracted using the Wizard Plus SV miniprep DNA purification system (Promega) and standard curves of known concentration of plasmid were generated according to the Applied Biosystems manual for absolute real-time PCR quantification. Absolute quantification of OsHV-1 genome copies in oyster samples was then estimated by comparing the observed crossing point (Cp) values to known plasmid standards from 10<sup>3</sup> to 10<sup>9</sup> copies of the DP product. The results were expressed as the log-transformed copy number of viral DNA per micro liter of DNA extracts (cp µl<sup>-1</sup>).

### Virus-like particle quantification by epifluorescence microscopy

Virus-like particles (VLP) were quantified using epifluorescence microscopy after staining with nucleic acid dye [68, 69]. Briefly, viral inoculum were first fixed with glutaraldehyde (Sigma-Aldrich G6257, final concentration 0.5%) and diluted in 1% potassium citrate solution (pH 7) to prevent aggregation of viral particles. Inoculum were vacuum filtrated to capture viruses on a 0.02 µm aluminium oxide filter (Whatman 6809–6002 Anodisc 25 inorganic membrane filter), before subsequent staining (SYBR Gold nucleic acid gel stain Invitrogen S11494, diluted at 0.2% in Tris-EDTA) and washing by 4 ml of 1% potassium citrate solution. Filters were then mounted on glass slides and VLP were observed and enumerated under epifluorescence microscopy (Olympus BH2-RFCA).

### Experimental infections

For experimental infections, oysters were previously myorelaxed in hexahydrate MgCl<sub>2</sub> (ACROS, 50 g l<sup>-1</sup>) [70]. Oysters were injected using 26-gauge needle attached to a multi-dispensing hand pipette, with 100 µl of inoculum (diluted in ASW to reach the desired viral concentration) in the abductor muscle. For each experiment, a group of oysters was myorelaxed and injected with ASW as the control. Oysters were replaced in seawater in plastic tanks at 20 °C and mortality was monitored daily.

### Virus purification and concentration by tangential flow filtration (TFF)

To develop a TFF-based process for the concentration of OsHV-1 particles, we used MicroKros® hollow fibre filter modules (Spectrum®) with three different pore sizes: 750 kDa (CO2-E7500-05-N), 500 kDa (CO2-E500-05-N) and 300 kDa (CO2-E300-05-N).

The first step of the process was to determine the pore size of the appropriate membrane to retain the OsHV-1 particles. To this end, we have filtered 10 ml of OsHV-1 inoculum containing 5.9×10<sup>4</sup> genomic units (GU)/µl with the 750 kDa TFF-device until the volume reached 1 ml. This 1 ml retentate (R), consisting of particles larger than the pore size of the hollow fibre, was diluted to the initial volume (10 ml) with filtered (0.02 µm) seawater and stored at 4 °C. The permeate (P), composed of particles smaller than the pore size of the hollow fibre, was collected, the volume adjusted to 10 ml with filtered (0.02 µm) seawater and was then filtered on the 500 kDa TFF device. The 500 kDa retentate was then processed similarly to the 750 kDa retentate and the 500 kDa permeate was then filtered on the 300 kDa TFF-device. At the end of this filtering process, we obtained three retentates (R-750; R-500 and R-300) and one permeate (P-300) (Fig. 1a).

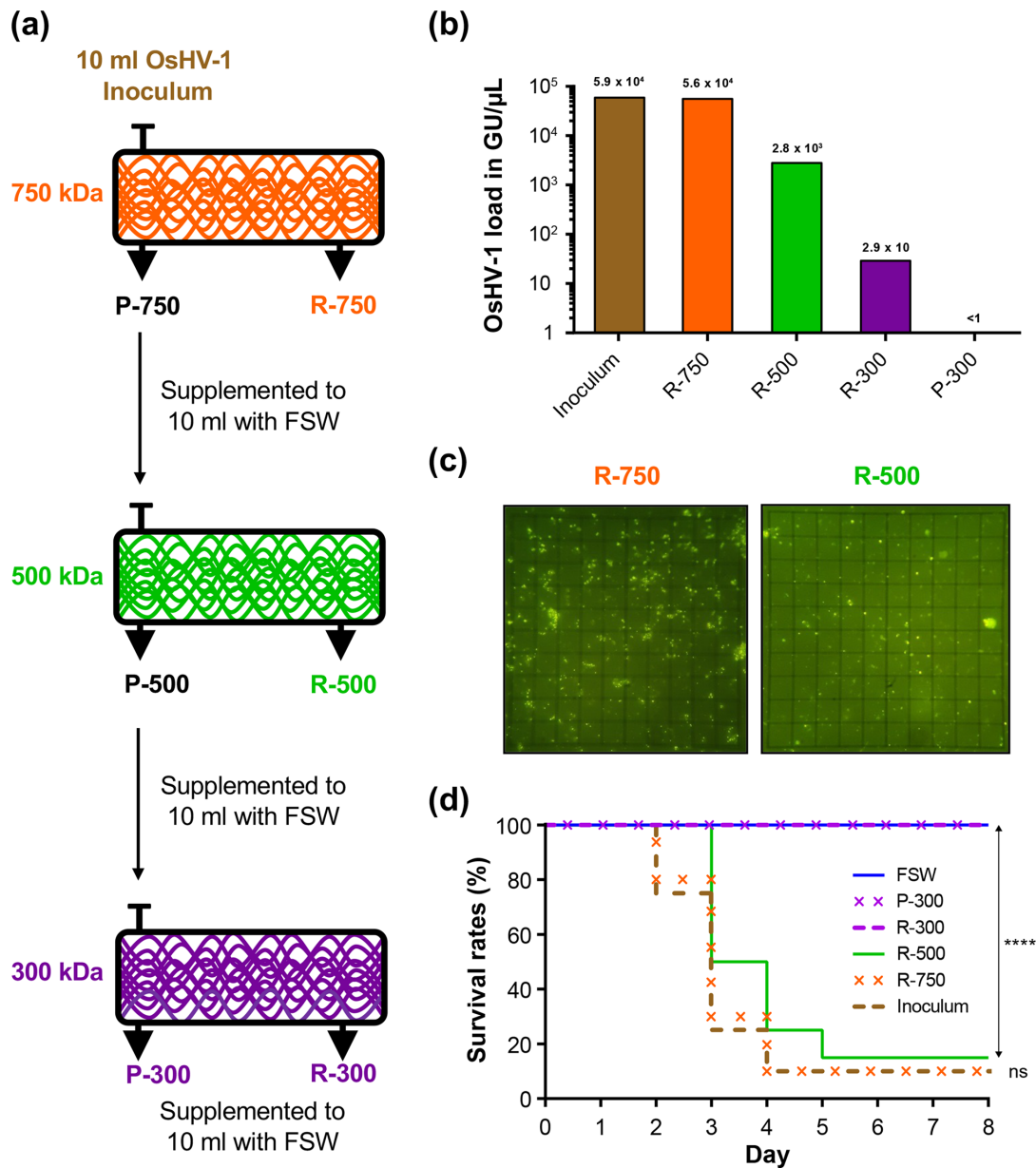
The second step consisted of purifying and concentrating OsHV-1 particles to extract enough DNA for both short- and long-read sequencing. For this purpose, we prepared a viral inoculum (see Viral inoculum preparation and OsHV-1 quantification for details) from 31.12 g of tissue obtained from 28 susceptible oysters (family F15) naturally infected by OsHV-1 in the Thau Lagoon in 2016 and stored at –80 °C (for more details on oyster origin and infection process see [71]). We obtained 462 ml of 0.22 µm filtrated inoculum, which were processed on a 500 kDa hollow fibre filter. Concentrated viral particles contained in 82 ml of the 500 kDa retentate were recovered by centrifugation (2 h at 18 000 r.p.m. at 4 °C). Pellets were then re-suspended in 1.2 ml of PBS and DNA was extracted and quantified as described above. We finally obtained 17.9 µg of viral DNA.

### Illumina short-read sequencing

Illumina DNA-Seq library preparation and sequencing were performed by the Genome Quebec Company (Genome Quebec Innovation Center, McGill University, Montreal, Canada) using 120 ng of DNA extracted from the OsHV-1 inoculum concentrated on 500 kDa TFF-device [as described at the end of Virus purification and concentration by tangential flow filtration (TFF)], the NxSeq AmpFREE Low DNA Library Kit (Lucigen) and the NovaSeq 6000 Sequencing system (Illumina®) (paired ends, 150 bp).

### Nanopore long-read sequencing

For long-read sequencing, we used the Oxford Nanopore Technologies (ONT) MinION platform (Mk1b). Briefly, 10 µg of DNA extracted from the OsHV-1 inoculum concentrated on 500 kDa TFF-device [as described at the end of Virus purification and concentration by tangential flow filtration (TFF)] were fractionated on Agencourt AMPure XP beads (Beckmann Coulter, 0.5 × final concentration) to eliminate small DNA molecules (< 300 bp). DNA size range analysis on Agilent Fragment Analyzer

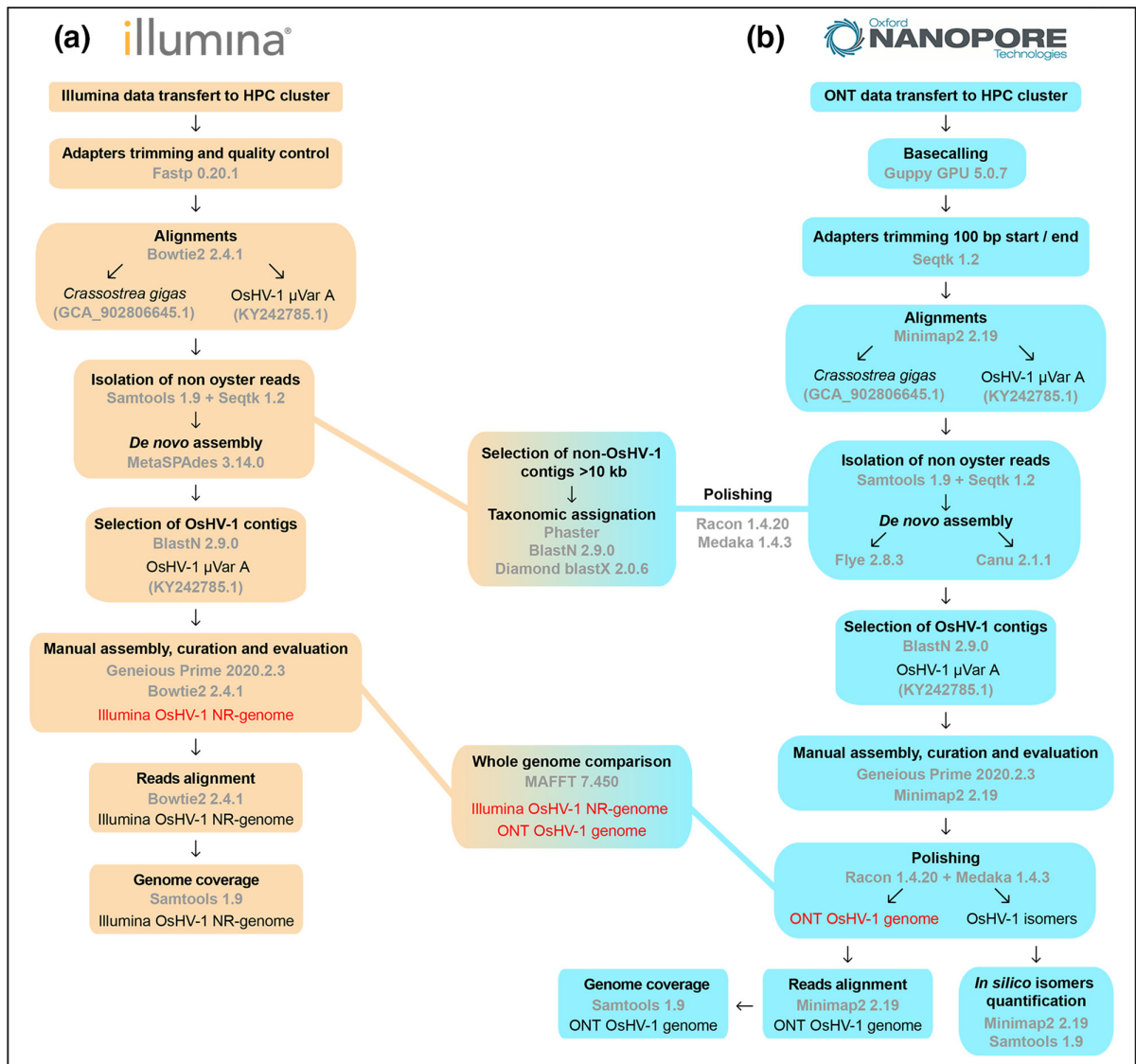


**Fig. 1.** (a) Workflow for the processing of OsHV-1 inoculum on tangential flow filtration (TFF) devices with three membrane pore sizes (750, 500 and 300 kDa). After filtration on each TFF-device, the volumes of retentates (R) and permeates (P) were supplemented to 10 ml with filtered seawater (FSW, 0,02 μm). (b) qPCR monitoring of OsHV-1 loads (in genomic units per μl) at the different steps of TFF. (c) Visualization and quantification of virus-like particles (VLP) by epifluorescence microscopy in retentates of TFF devices of 750 kDa (R-750) and 500 kDa (R-500). An eyepiece with a 10×10 grid reticle was used for counting VLP (note in our setup the grid reticle is 100×100 μm). (d) Kaplan-Meier survival curves of oysters injected with 100 μl of OsHV-1 inoculum, or retentates of TFF-devices of 750, 500, 300 kDa (R-750, R-500 and R-300, respectively), or permeate of TFF device 300 kDa (P-300) or FSW as negative control. \*\*\*\**P* < 0.0001; ns not significant (Mantel-Cox Log-rank test).

(Genomic DNA analysis kit DNF-467–33) revealed that the average size was over 60 kb with a genomic quality number (GQN) of 8.9 (threshold 10 kb). Then, 400 ng of AMPure purified DNA were used to construct the sequencing library using SQK-RAD0004 tagmentation kit and sequencing was performed as per manufacturer's guidelines using R9.4 flowcells (FLO-MIN106) and run for 20 h.

### Illumina genome assembly

Illumina raw reads files were quality filtered (PHRED score >31) and adapters trimmed with Fastp version 0.20.1 [72]. To evaluate the origin of reads, fastq reads were aligned to the oyster *C. gigas* genome (accession: GCA\_902806645.1) and to OsHV-1 μVar



**Fig. 2.** Flow diagram to illustrate the bioinformatic pipelines used to assemble OsHV-1 genome from (a) Illumina short-read or (b) Nanopore long-read.

A genome (accession: KY242785.1) with Bowtie2 version 2.4.1 [73]. All the non-oyster reads were then selected with Samtools version 1.9 [74] and Seqtk version 1.2 (<https://github.com/lh3/seqtk>). Those reads were then used for *de novo* assembly with SPAdes version 3.14.0 with parameter `-meta` [75]. The *de novo* OsHV-1 contigs generated by SPAdes were then identified with BLASTN version 2.6.0 [76] using OsHV-1  $\mu$ Var A sequence as a query (accession number: KY242785.1). The identified OsHV-1 contigs were then manually scaffolded in Geneious Prime (version 2020.2.3) to reconstruct the common virus architecture ( $U_L$ - $IR_L$ -X- $IR_S$ - $U_S$ ) [7]. To overcome the problem of inverted repeated regions that cannot be confidently positioned with short-read data, we constructed non-redundant genomes (NR genomes) that contain only one copy of each repeated region ( $U_L$ - $IR_L$ -X- $IR_S$ - $U_S$ ) an approach commonly used for Herpesvirus genome analyses [34, 77, 78]. Quality filtered reads were then aligned to the new constructed OsHV-1 NR genome to evaluate coverage uniformity. This pipeline is summarized in Fig. 2a.

### Nanopore genome assembly

Nanopore fast5 file were GPU base called with Guppy version 5.0.7 with standard parameters except `-config dna_r9.4.1_450bps_hac.cfg` (<https://community.nanoporetech.com/downloads>). Fastq reads were then 100 bp trimmed at their beginning and at their end to ensure proper adapter removal with SeqTk version 1.3 with parameters `-b 100` and `-e 100` (<https://github.com/lh3/seqtk>). To assign reads to either host or viral origin, fastq reads were aligned to the oyster *C. gigas* genome (accession: GCA\_902806645.1) and to OsHV-1  $\mu$ Var A genome (accession: KY242785.1) with Minimap2 version 2.22 with standard parameters [79]. Then two

types of *de novo* assemblies were performed with non-oyster reads with two different *de novo* assemblers. All fastq reads or only the top 10% of the longest reads were used as input for *de novo* assembly with Flye version 2.8.3 and parameters -meta [80] or with Canu version 2.1.1 and parameters -nanopore [81]. OsHV-1 *de novo* assembled contigs were then identified with BLASTN [76] against the OsHV-1  $\mu$ Var A genome and were manually trimmed in Geneious version 2020.2.4 (<https://www.geneious.com>). All *de novo* contigs plus the new assembled OsHV-1 genome were then polished using Nanopore reads with three run of Racon version 1.4.20 with parameters --no-trimming -m 8x-6g -8w 500 [82] and one run of Medaka version 1.4.3 with parameters medaka\_consensus (<https://github.com/nanoporetech/medaka>).

For OsHV-1 isoform coverage evaluation, the four contigs produced by Canu with the top 10% of the longest reads were trimmed on the  $U_L$  side to keep a 10 kb  $U_L$  fragment with Geneious version 2020.2.4 (<https://www.geneious.com>). The four major isomers were then used as a reference for reads mapping with Minimap2 version 2.22 with standard parameters -axe map-on (<https://github.com/lh3/minimap2>). Samtools version 1.10 [83] were then used to filter out all the shortest aligned read and all the alignment with MapQ <40, keeping only the reads spanning at least a fragment of the  $U_L$  and a fragment of the  $U_S$  (i.e. 28 kb long). A pipeline summary can be found in Fig. 2b.

### Non-OsHV-1 contigs' classification

Polished non-OsHV-1 contigs with a size larger than 10 kb were then subjected to taxonomic identification and classification with PHASTER [84] accessed in September 2021, with BLASTN version 2.0.9 with parameters -max\_hsps 1 -max\_target\_seqs 100 -evalue 0.0001 [76] or with DIAMOND BLASTX [85] version 2.0.6 with parameters --sensitive --max-hsps 1 --max-target-seqs 0 --range-culling -F 15 --evalue 0.0001 against the refseq database accessed in September 2021.

### Comparison of Illumina and Nanopore OsHV-1 genomes

Before whole-genome alignment evaluation, both genomes were modified to fit the same isomers' version of the genome (i.e.  $TR_L/U_L \rightarrow IR_L/X/IR_S/U_S \rightarrow TR_S$ ). To evaluate the capacity of Nanopore sequencing to produce an accurate OsHV-1 genome, the manually constructed Illumina OsHV-1 non-redundant (NR) genome was aligned with the manually curated Nanopore OsHV-1 complete genome with MAFFT version 1.4.0 using default parameters [86] implemented in Geneious version 2020.2.4 (<https://www.geneious.com>).

### OsHV-1 phylogeny

In order to illustrate the global phylogenetic relationships among our and other available genomes, 13 OsHV-1 NR-genomes (11 available in GenBank: AY509253, GQ153938, KP412538, KY242785, KY271630, MG56175, MF509813, OM811579, OM811584, OM811586, OM811594 and the two from this study) were aligned with progressiveMauve algorithm [87] implemented in Geneious version 2020.2.4 to local collinear blocks. Reordered genomes block were then aligned with MAFFT version 7.4 using default parameters [86] implemented in NGPhylogeny.fr [88]. Alignment was cleaned with BMGE with default parameters [89] implemented in NGPhylogeny.fr. The best fitting model of nucleotide substitution was chosen according to the Akaike Information Criterion with jModelTest version 2.1.10 with default parameter [90]. Tree was constructed with phyML version 3.3 [91] with default parameters except '-b 100 m HKY85 -f m -c 8 s BEST -a e -v e -o tlr' implemented in NGPhylogeny.fr. Phylogenetic tree was midpoint rooted and arranged in FigTree version 1.4.4 (<https://github.com/rambaut/figtree>).

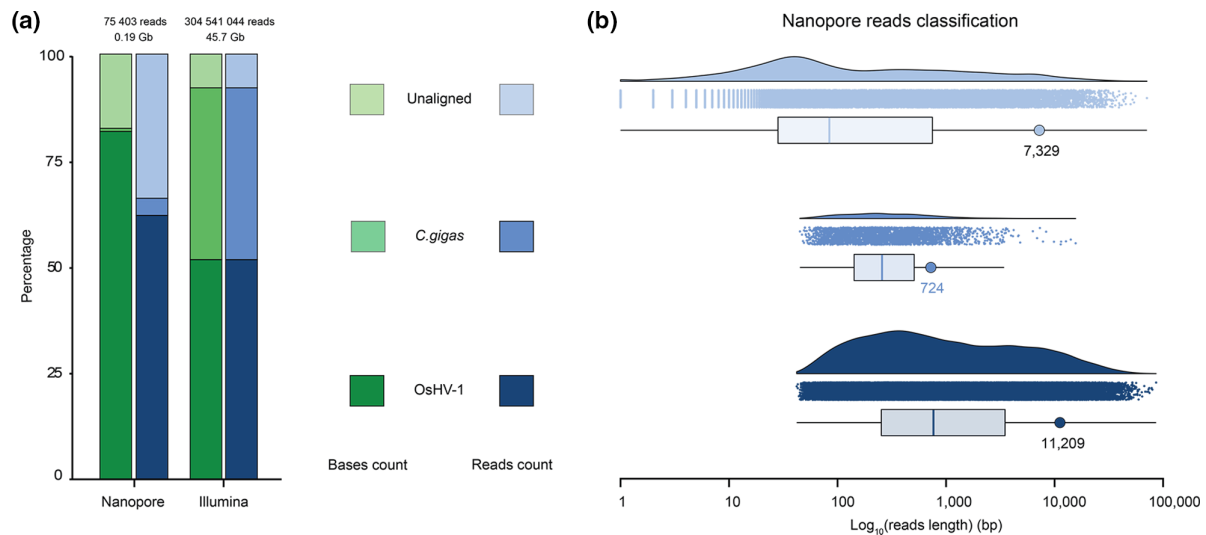
## RESULTS

### Calibration of tangential flow filtration (TFF) device to purify and concentrate OsHV-1 particles

To develop a tangential flow filtration-based process for the purification and concentration of OsHV-1, we tested TFF-devices with three membrane pore sizes [molecular weight cut-off (MWCO) of 750, 500 and 300 kDa]. Then, we used an OsHV-1 inoculum as starting material and successively filtered it through the three TFF-devices (Fig. 1a; see Methods for further detail). Finally, we have obtained three retentates (R-750, R-500 and R-300) and one permeate (P-300), which were then used to analyse the OsHV-1 load and infectivity.

To evaluate the ability of the three TFF-devices to retain OsHV-1, we have quantified OsHV-1 load by qPCR in the initial inoculum, in the three retentates and in the permeate (Fig. 1b). The number of OsHV-1 genomic units (GU) in the R-750 fraction ( $5.6 \times 10^4$  GU) was of the same order of magnitude as it was in the initial inoculum ( $5.9 \times 10^4$  GU), which suggests that most of the viruses (95%), were retained by 750 kDa MWCO. However, a small fraction (4.8%) of GUs present in the initial inoculum were also found in the R-500 retentate ( $2.8 \times 10^3$  GU) and traces of amplification (below standard curve) were also observed in the R-300 ( $2.9 \times 10$  GU). Epifluorescence microscopy did not reveal virus-like particles (VLP) in the R-300 and P-300 fractions (data not shown) while in the R-750 and R-500 fractions we were able to enumerate  $2.71 \times 10^4$  ( $7.05 \times 10^3 \pm \text{SD}$ ) and  $4.88 \times 10^3$  ( $2.79 \times 10^3 \pm \text{SD}$ ) VLP  $\mu\text{l}^{-1}$ , respectively (Fig. 1c). Epifluorescence microscopy gave similar results to qPCR and confirms that most of the viral particles present in the starting inoculum were preserved in the R-750 fraction.





**Fig. 3.** Global sequencing statistics. (a) Histogram representing percentage of sequenced bases (green) and reads (blue) that aligned to OsHV-1 (dark) or to *C. gigas* (regular) genomes or that did not align to these genomes (light). (b) Raincloud plot representing the size distribution of the Nanopore reads that aligned to OsHV-1 (dark blue) or to *C. gigas* (blue) genomes or that did not align to these genomes (light blue). Each panel is made of a density plot, a dot plot where each dot corresponds to one Nanopore read and a box plot showing average read size (vertical coloured bar) and N50 (coloured circle).

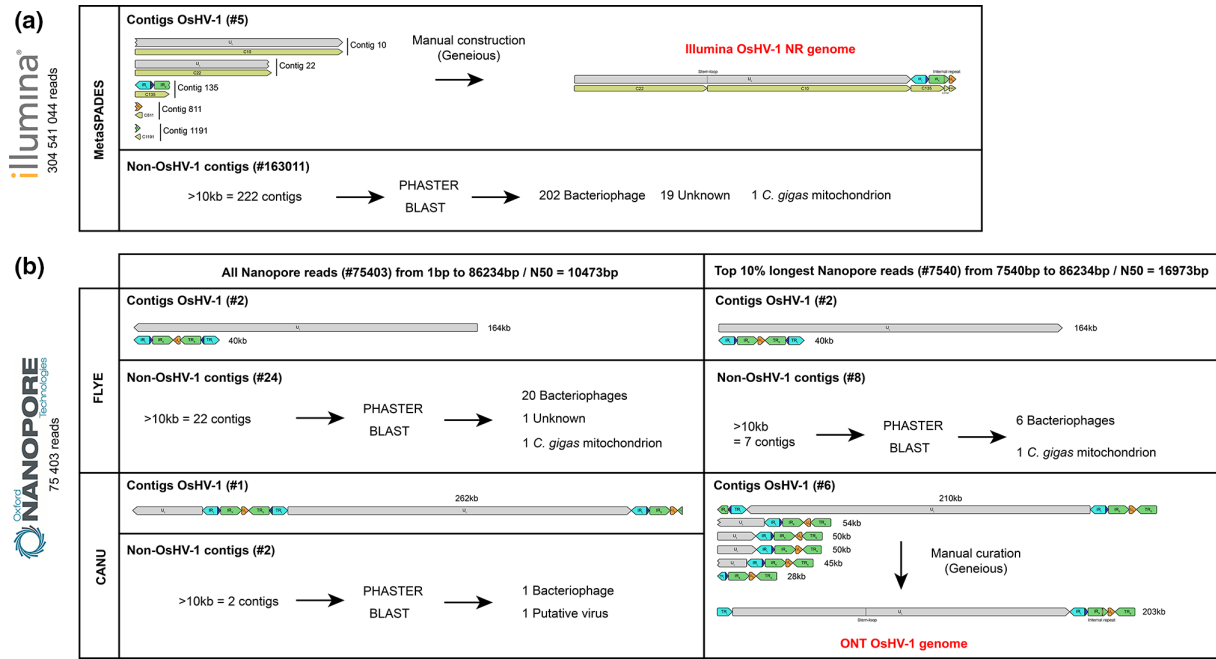
Finally, we compared the infectious capacity of the different fractions with the initial inoculum by injecting them into oysters. No mortality was observed for oysters injected with fractions R-300 and P-300 as well as for oysters injected with filtered sea water (Fig. 1d). However, oysters that had been injected with R-750 and R-500 fractions exhibit survival curves, which are not significantly different from those of oysters injected with the starting inoculum. Interestingly, survival curves of R-750 and R-500 fractions are not significantly different although R-500 fraction contained around 20 times less OsHV-1 GU than the R-750 fraction. These results showed that filtration on 500 kDa TFF-device allowed the purification and concentration of fully infective OsHV-1 viral particles. Consequently, we used the 500 kDa TFF-device to purify and concentrate viral particles from an inoculum prepared from 28 oysters associated from mortality event, which have occurred in 2016. We have been able to extract ~18 µg of viral DNA, an amount sufficient to perform both short- and long-read sequencing (see Methods for further detail).

### General statistics of the two sequencing technologies

Viral DNA extracted from TFF purified viral particles was sequenced using two different sequencing technologies: (i) a short-read approach based on the Illumina platform and (ii) a long-read approach based on the Nanopore platform.

The emergence of mNGS, has allowed researchers to put aside complex techniques such as ultracentrifugation coupled to sucrose gradient for OsHV-1 genome sequencing. The direct consequence of the lack of viral purification is the production of a large amount of non-viral data, resulting in the waste of a significant portion of the sequencing run. In recent studies, sequencing of the OsHV-1 genome from DNA extracted directly from infected tissues has led to the acquisition of only 0.01–13.21% OsHV-1 reads (Table S1).

As expected TFF purification of OsHV-1 particles allowed both sequencing technologies to obtain a large enrichment in OsHV-1 reads compared to host reads, leading to a mean coverage of 740 reads/bp and 213 790 reads/bp for Nanopore and Illumina sequencing, respectively. In more details, the Illumina sequencing run yielded more than 304 million paired-end reads of 150 base pairs (bp) corresponding to 45.68 Gb and distributed as follows: 52% of OsHV-1 reads, 40% of *C. gigas* reads and 8% of unaligned reads with a similar proportion of sequenced base number (Fig. 3a). The Nanopore sequencing run yielded 75 403 reads corresponding to 0.19 Gb and distributed as follows: 62 and 82% of OsHV-1 reads and bases, respectively, 4 and 0.7% of *C. gigas* reads and bases, respectively, and 33 and 17% of unaligned reads and bases, respectively (Fig. 3a). Since the Nanopore technology sequence native DNA molecules present in the DNA library, a large range of reads' size has been sequenced. In detail, OsHV-1 reads ranged from 42 to 86 234 bp with a N50 of 11 209 bp, *C. gigas* reads ranged from 45 to 15 635 bp with a N50 of 724 bp and unaligned reads ranged from 1 to 70 966 bp with a N50 7 329 bp (Fig. 3b). The difference in N50 between host and viral DNA (11 209 vs 724 bp) was probably due to oyster DNA fragmentation during purification steps while viral DNA was protected by the viral envelope and capsid. This observation has probably led to the observed difference in host reads' percentage between both sequencing approaches. Indeed, before and during the Nanopore libraries' construction several Ampure XP bead-based DNA purification were performed, and this type of DNA purification is known to induce a loss of small size DNA molecules.



**Fig. 4.** Schematic representation of OsHV-1 whole-genome assembly outputs for (a) Illumina short-read sequencing and (b) Nanopore long-read sequencing.

### OsHV-1 genome assembly using Illumina reads

Our short-read pipeline based on the *de novo* assembler metaSPAdes produced five OsHV-1 contigs ranging from 2 429 to 99 197 bp (Fig. 4a). Those five contigs were then manually assembled to construct an OsHV-1 NR genome in which only one copy of each repeat was conserved ( $U_L$ - $IR_L$ - $X$ - $IR_S$ - $U_S$ , see Methods for further details). The re-alignment of the Illumina reads on the new OsHV-1 NR genome allowed us to correct the junction between contigs C1191 and C811 (addition of one A/T bp) and to confirm that the other metaSPAdes contigs were correctly assembled.

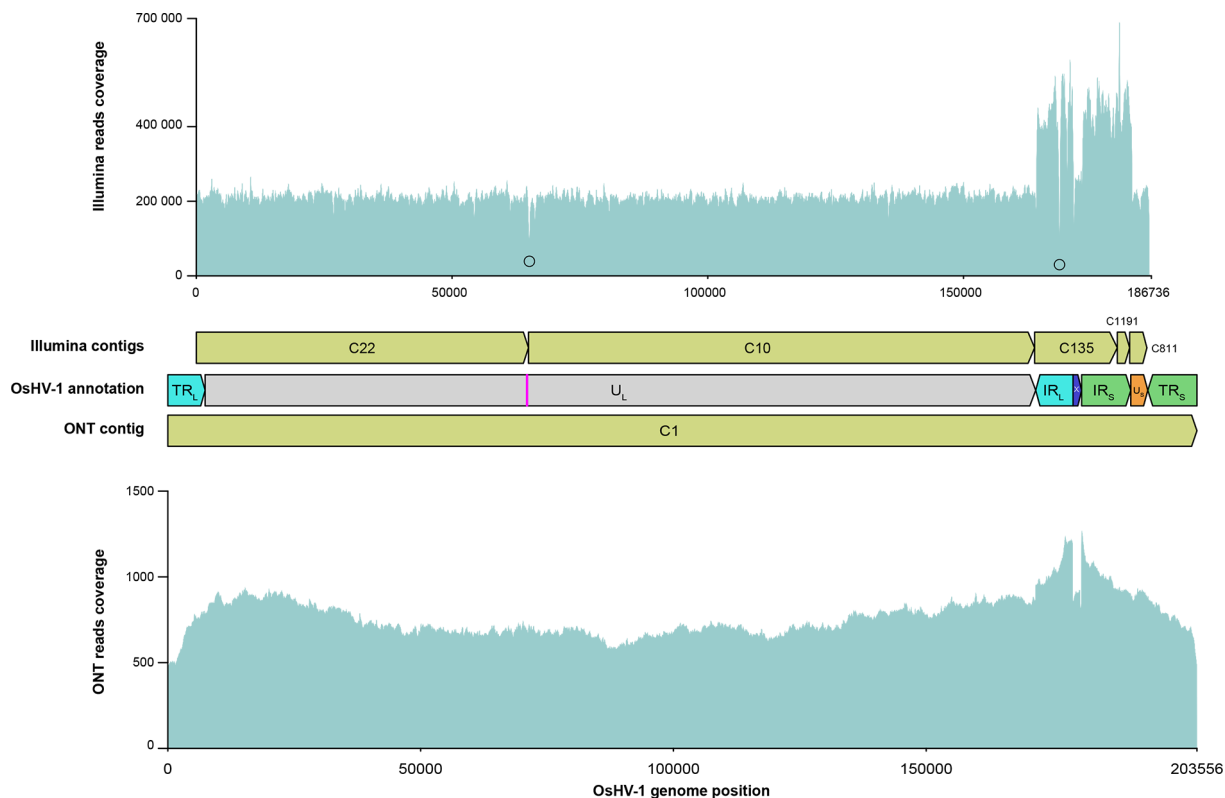
As expected, we observed that the reads' coverage within  $IR_L$  and the  $IR_S$  is twice the reads' coverage for unique regions  $U_L$ ,  $U_S$  and  $X$  given that the NR genome contains only one copy of these duplicated regions (Fig. 5 upper panel). Interestingly, two decreases of reads' coverage were observed (Fig. 5 upper panel). The first decrease of coverage was at the junction of the C22 and C10 contigs and takes place in a stem-loop suspected to be the viral origin of replication [7]. The presence of this stem-loop could have induced a difficulty to access the DNA in this region by the sequencing polymerase leading to a decrease of sequencing yield around this palindromic region and the production of two distinct contigs by the *de novo* assembler (Fig. S1). The second decrease of reads' coverage takes place within the C135 contig, in a long C22 homopolymer of the  $IR_L$  region. This type of bias in read coverage has already been observed for long homopolymer sequenced by Illumina [92] (Fig. 5 upper panel and Fig. S2).

### OsHV-1 genome assembly using Nanopore reads

Since this is the first time that an OsHV-1 genome has been assembled from ONT long-read, we decided to evaluate the two most used long-read *de novo* assembler Flye and Canu with two datasets: (i) all the Nanopore reads and (ii) only the top 10% longest Nanopore reads (see Methods for further detail) (Fig. 4b).

Whatever the dataset used, Flye yielded the same two contigs; a long one (164 kb) corresponding to the complete  $U_L$  and a short one (40 kb) made of the  $U_S$  surrounded by one copy of  $IR_L$ - $X$ - $IR_L$  on its left side and one copy of  $TR_S$ - $X$ - $TR_L$  on its right side (Fig. 4b). With the dataset including all Nanopore reads, Canu yielded one large OsHV-1 contig (262 kb). This contig is made of one fragment of the  $U_L$  (32 kb) followed by a fragment made of the  $U_S$  surrounded by one copy of  $IR_L$ - $X$ - $IR_S$  on its left side and one copy of  $TR_S$ - $X$ - $TR_L$  on its right side (40 kb), followed by the complete  $U_L$  (164 kb) and followed by another fragment made of  $U_S$  surrounded by one copy of  $IR_L$ - $X$ - $IR_S$  on the left side and one copy of the  $TR_L$  (24 kb) on its right side (Fig. 4b).

With the dataset composed of only the top 10% longest Nanopore reads, Canu yielded six OsHV-1 contigs ranging from 28 to 210 kb (Fig. 4b). The 28 kb contig consisted of the  $U_S$  surrounded on its left side by one copy of  $TR_L$ - $X$ - $IR_S$  and on its right side by one copy of  $TR_S$ . The four contigs ranging from 45 to 54 kb share the same general structure; that is to say, a fragment of the  $U_L$  region followed by a copy of  $IR_L$ - $X$ - $IR_S$ - $U_S$ - $TR_S$ . These four contigs are differentiated by the relative

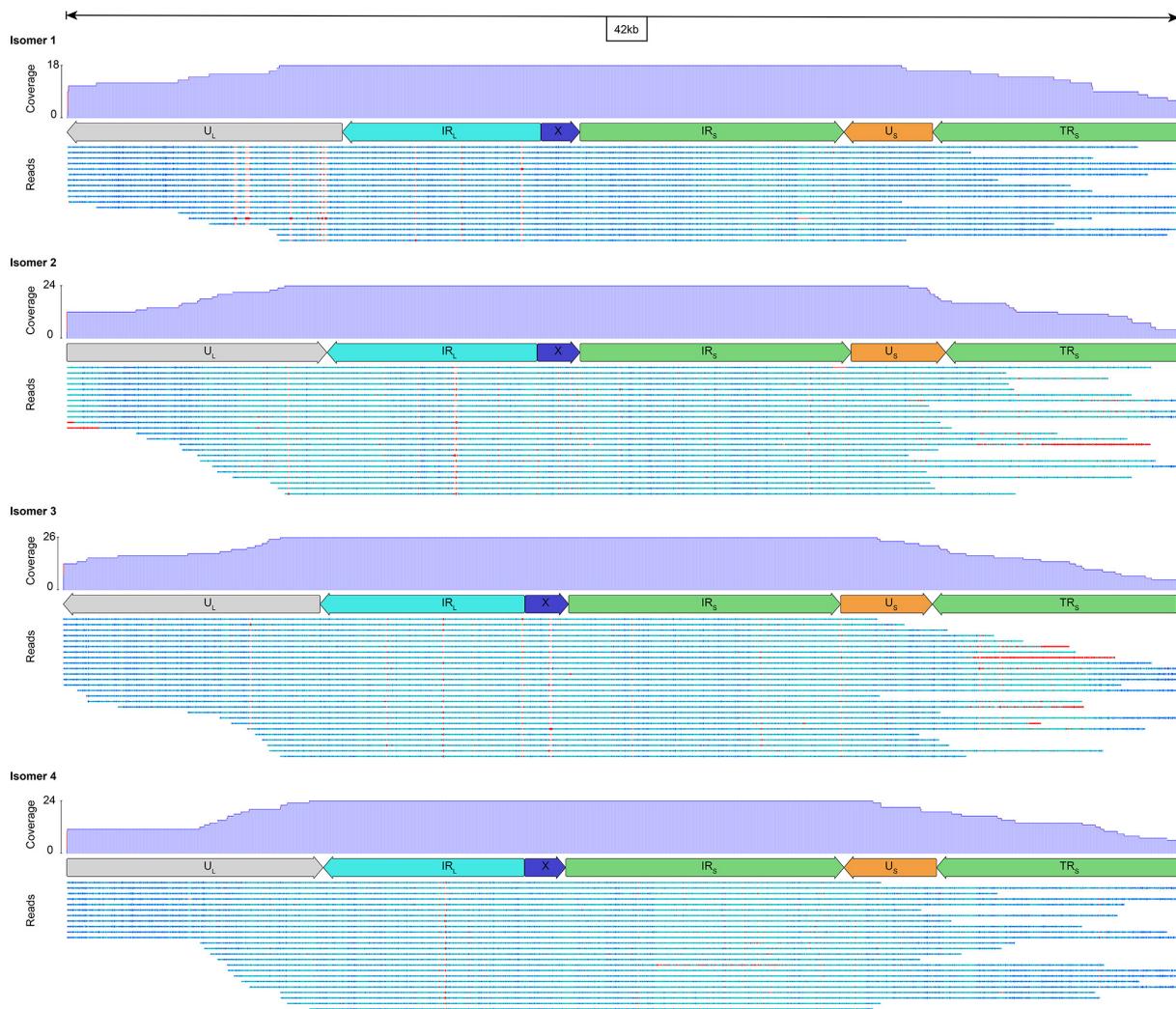


**Fig. 5.** OshV-1 whole-genome assemblies' evaluation by read mapping. Upper panel represent Illumina short-read coverage on the Illumina OshV-1 NR genome. Black circles indicate bottom line of the two losses of coverage identified in the Illumina OshV-1 NR genome. Middle panel represent Illumina *de novo* assembled contigs (green), OshV-1 genomic region annotation with the pink bar representing the stem-loop and the ONT *de novo* assembled contig (green). Bottom panel represent ONT long-read coverage on the ONT OshV-1 genome. C: Contig.

orientation of the  $U_L$  and  $U_S$  fragments (same or inverted orientations) and their orientation with respect to the  $IR_L$ -X- $IR_S$  fragment (Fig. 4b). It was likely that these four contigs correspond to the four major OshV-1 isomers described by Davison *et al.* [7]. In order to verify this observation and to quantify the relative abundance of the four isomers, we have screened all the reads long enough to cover at least a part of both  $U_L$  and  $U_S$  regions. We identified 92 long-read divided into 18  $\langle -U_L / \langle -U_S \rangle$ , 24  $U_L \rightarrow / U_S \rightarrow$ , 26  $\langle -U_L / U_S \rightarrow$  and 24  $U_L \rightarrow / \langle -U_S \rangle$  (Fig. 6). This result confirms the OshV-1 genome isomerization and shows that the viral population is composed of approximately equimolar amounts of the four isomers. Finally, Canu assembled one large contig (210 kb) made of the complete  $U_L$  surrounded by one copy of  $IR_S$ -X- $TR_L$  on its left side and one copy of  $IR_L$ -X- $IR_S$ - $U_S$ - $TR_S$  on its right side (Fig. 4b). This former contig have been manually curated to create the complete ONT OshV-1 genome by removing the  $IR_S$ -X fragment on the left side of the  $U_L$  and by reverse complement the  $U_L$  to fit with the most used representation of OshV-1 genomes isomer [7].

To assess the robustness of the manual assembly, we aligned all ONT reads back to the ONT OshV-1 genome and observed a continuity of coverage throughout the genome, which validates the assembly of the ONT OshV-1 genome with our pipeline (Fig. 5 lower panel). Interestingly, we can see a small decrease of coverage at both extremities of the genome and an increase of coverage on  $IR_L$  and  $IR_S$  at the junction surrounding the X region (Fig. 5 lower panel). This observation reflects the use of one of the 16 minor isomers described by Davison *et al.* to represent our ONT OshV-1 genome [7]. Indeed, in addition to the four isomers described previously, Davison *et al.* showed that some of these isomers contained an additional copy of the X region at the left or right terminus whereas some might have lacked the X region internally. Since only one copy of the X region is presented in the ONT OshV-1 genome, the reads corresponding to the minor variant of the X region might tend to align around the X region between  $IR_L$  and  $IR_S$  instead of aligning at the extremities of the genome ( $TR_L$  and  $TR_S$ ) leading to a higher coverage of the region surrounding X compared to the rest of the genome (Fig. 5 lower panel).

Taken together these results validates the assembly of the ONT OshV-1 genome, confirms for the first time the OshV-1 genome isomerization and shows that the viral population is composed of approximately equimolar amounts of the four major isomers.

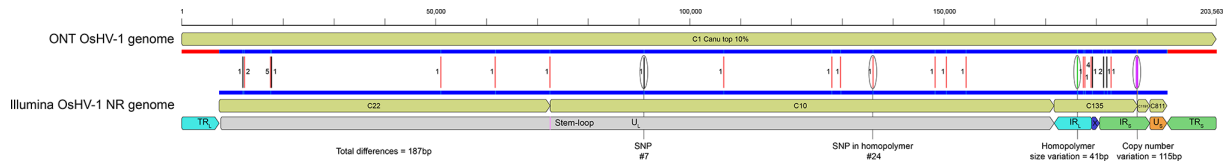


**Fig. 6.** OsHV-1 major isomer validation. Each panel corresponds to a 42 kb long OsHV-1 genomic region made of  $U_L$  fragment plus  $IR_L$ ,  $X$ ,  $IR_S$ ,  $U_S$  and  $TR_S$ . The different panels correspond to the different OsHV-1 isomers made by permutation of  $U_L$  and  $U_S$  regions. Each panel is made of a diagram of reads coverage, a diagram of OsHV-1 region annotation and a diagram representing long-read alignment that span at least one fragment of the  $U_L$  and one fragment of the  $U_S$ . Read bases are color-coded from conserved (dark blue), to less conserved (light blue) with gaps in sequence colored in red.

### Comparison of Illumina OsHV-1 NR genome and ONT OsHV-1 genome

Since it is almost impossible to assemble a complete OsHV-1 genome from short-read technology due to the presence of large inverted repeats, and because long-read technology tends to produce longer OsHV-1 genome than expected, both technologies still requires manual editing to join the different fragments or to remove extra repeats to obtain a genome structure as accurate as possible. Furthermore, ONT is known to produce a high error rate on raw sequences compared to the Illumina technology. Here, rather than looking at the raw sequence error rate, we attempted to quantify the residual error by comparing to each other, the polished ONT OsHV-1 genome, and the Illumina OsHV-1 NR genome. To do so, we used a pairwise alignment and were able to identify a total of 187 bp difference between the two genomes (Fig. 7).

Those differences could be classified in three categories: (i) SNPs, (ii) copy number variation of a 115 bp repeat and (iii) large homopolymer size variation. We observed a total of 31 SNPs between the two genomes, 24 SNPs correspond to InDel, which take place at the extremity of small homopolymers (from 5 to 11 bp), and 7 SNPs were distributed elsewhere in the genome. The 115 bp copy number variation takes place into the  $IR_S$  region. While the Illumina OsHV-1 genome contains two copies of the 115 bp sequence, the ONT OsHV-1 genome harbour three copies of this repetition (Fig. S3). Finally, we observed an homopolymer size variation in the  $IR_L$  region corresponding to 41 bp insertion in the ONT OsHV-1 genome compared to Illumina OsHV-1 genome. Since sequencing error in homopolymer is a well-known caveat of Nanopore sequencing [93], and because long homopolymer are known to be error prone also in Illumina sequencing [92, 94], we have performed a careful analysis of the read alignment for



**Fig. 7.** ONT and Illumina OsHV-1 whole-genome comparison. A total of 187 bp differences are identified between both genomes. Overall, 7 bp correspond to SNPs (black bar), 24 bp correspond to SNPs in small homopolymer (red bar), 41 bp correspond to a homopolymer size variation (green bar), and 115 bp correspond to a copy number variation (purple bar). Numbers next to the bars correspond to the number of variable bases at that position.

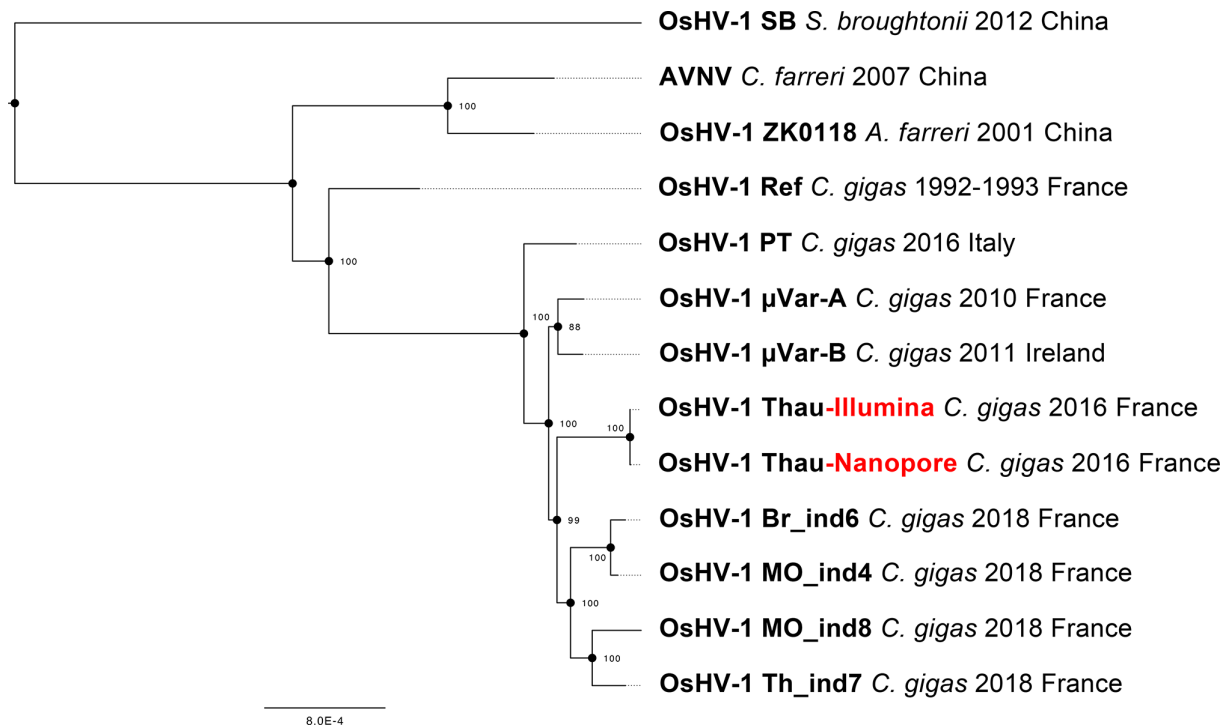
both technologies in this specific region. Interestingly, alignment observation in this large homopolymer identify heterogeneity in homopolymer size at the level of individual reads for both sequencing technologies. While size variation for Illumina reads ranged from 7 to 22 bp, size variation for Nanopore reads ranged from 5 to 98 bp (Fig. S2A and B respectively).

Because of the impossibility to accurately distinguish error in this large homopolymer for both Illumina and Nanopore sequencing, and because we cannot accurately assign the number of Cs inside this large homopolymer, we have decided to keep the most represented number of Cs for each of the sequencing techniques. We therefore did not consider this large homopolymer in our final pairwise homology comparison of both sequencing technologies.

Finally, if we disregard the additional 115 bp repeat highlighted by ONT sequencing, the two *de novo* assembled genomes only differ by a total of 31 bp and show 99.98% of nucleotide identity.

### OsHV-1 phylogenetic analysis

In order to determine the relationships of the OsHV-1 genomes from this study with those known so far, we constructed a phylogenetic tree from the alignment of the two genomes we generated in the present study (viruses collected in 2016), seven OsHV-1 genomes available in public databases (viruses collected before 2016) and four representatives of the 21 genomes acquired by Delmotte-Pelletier *et al.* [34] (viruses collected in 2018). This analysis confirms that the OsHV-1 virus sampled in Thau lagoon



**Fig. 8.** Phylogenetic tree of the 11 OsHV-1 NR version of the genomes available in public database, plus the Nanopore and Illumina NR genomes generated in the present study (red). The tree was mid-point rooted. Bootstrap values are shown at nodes (100 replicates).

(France) in 2016 belongs to  $\mu$ Var clade and that genomes reconstructed from the same sample with either the Nanopore or the Illumina sequencing data are really close to each other (Fig. 8).

### Non-OsHV-1 contigs' classification

Interestingly, our TFF-based purification process led to the characterization of other viral populations present in oyster tissues. To obtain a global overview of this viral metagenome, we made an extensive taxonomic classification of the non-OsHV-1 contigs produced by both ONT and Illumina sequencing. While the Nanopore sequencing produced 22 non-OsHV-1 contigs (from 14 374 to 163 592 bp) (Fig. 4b, Table S2A), Illumina sequencing produced 222 non-OsHV-1 contigs with size greater than 10 kb (from 10 017 bp to 172 478 bp) (Fig. 4a, Table S2B). Among the 22 Nanopore contigs 14 were also detected in the Illumina non-OsHV-1 contigs (Table S2A). Nanopore identified 20 contigs coding for putative bacteriophages, 1 contig coding for a CRESS DNA virus and 1 contig coding for the *C. gigas* mitochondrial DNA. Illumina identified 202 contigs coding for putative bacteriophages, 19 contigs with no hit and 1 contig coding for the *C. gigas* mitochondrial DNA (Table S2B).

## DISCUSSION

Viral whole-genome sequencing has recently become a standard in molecular epidemiology and evolutionary genomics, however this approach can be challenging when implemented on DNA viruses with large and complex genomes such as Herpesviruses [95, 96]. Indeed, it is often difficult to purify these types of viruses in sufficient quantity to allow whole-genome sequencing and reconstruction. Moreover, whole-genome assembly is a real burden considering the viral complexity due to the presence of long inverted repeated regions and structural variants. This is even more challenging to achieve for uncultured viruses derived from clinical samples or from livestock.

We implemented TFF to purify and to concentrate OsHV-1. To this end, we tested TFF devices with different pore sizes and quantified the virus load and infectious capacity of each fraction. Our results showed that TFF device with a MWCO of 500 kDa was able to retain and to concentrate most the OsHV-1 infectious particles contained in the inoculum. Interestingly, viral load quantification by real-time PCR and epifluorescence microscopy were of the same order of magnitude suggesting that TFF not only concentrates the virus but also eliminates most of the free DNA in the starting inoculum as previously suggested by Loewe *et al.* [60]. This is supported by the tremendous increase of the signal (OsHV-1 reads) to noise (oyster reads) ratio obtained in the sequencing data. Indeed, more than 60,01% of the sequenced reads belonged to OsHV-1 whereas, in previous work, the ratio obtained from infected tissues ranged from 0,01% to 13,21% [31, 34]. While other virus enrichment methods like PCR or capture-based approaches can reach up to 100% of viral data, the TFF approach does not suffer from amplification or hybridization bias caused by sequence variation over time due to genomic divergence [97]. We believe this method will be very helpful to improve the quantification of infectious viral particles in samples from organisms for which there is no cell culture available. Indeed, for tissue samples, the quantification of viruses is mainly based on qPCR approaches, which quantify the genome copy number but not the number of infectious viral particles. Since infected tissues could contain viral genomes undergoing replication or non-canonical genomes [48], the quantification of genome packaged in viral particle reflect a more robust estimate of the true quantification of viable virion in the sample. In addition, we have shown that TFF is an alternative particle enrichment method to centrifugation-based purification. TFF has many advantages including low cost, transportability, rapid purification, and enrichment in infectious particles (i.e. 500 kDa for OsHV-1). We therefore demonstrate, that TFF is a powerful tool to purify unculturable viruses from clinical and environmental isolates.

Then, we performed a high-quality DNA extraction that allowed us to compare the capabilities of long- and short-read sequencing technologies (Illumina and Oxford Nanopore Technologies) to *de novo* assemble OsHV-1 whole genome.

To date and to our knowledge, there is no maintained bioinformatic pipeline for reference-free and *de novo* assembly of long DNA viruses with large inverted-repeat like OsHV-1 for either short- or long-read sequence. Usually, long- and short-read sequencing technologies are combined to *de novo* assemble Herpesvirus whole genomes [48, 98]. Conversely to what is usually done, we have successfully tested two bioinformatic pipelines to independently analyse the long- and short-read datasets, and to reference-free and *de novo* assemble two OsHV-1 whole genomes. The *de novo* assembly of the short-read pipeline produced five OsHV-1 contigs that have been manually scaffolded to produce a complete OsHV-1 genome. The *de novo* assembly of the long-read pipeline produced one large contig that has been manually curated to remove one extra duplicated region. Because of uncertainty with the short-read assembly to position duplicated regions relative to unique region, we have depicted the Illumina assembly as a Non-Redundant (NR) genome, with only one copy of each  $R_L$  and  $R_S$ , as frequently done in the literature [34, 77, 78, 99]. In contrast, the long-read assembly resulted in a genome with both inverted repeats with the help of a minimal manual edit.

Pairwise comparison of the two assembled OsHV-1 genomes showed that the Nanopore genome was 99,98% similar to its Illumina counterpart. Our results demonstrate that Nanopore sequencing is now a mature sequencing technology for the small-scale genomic characterization of difficult long DNA virus with large inverted repeat. This breakthrough sequencing technology coupled with TFF virus purification is now a promising solution to drastically increase the number of sequenced OsHV-1 genomes allowing in-depth molecular evolution and epidemiology characterization.

In addition to being accurate at the base level, ONT sequencing also allowed to resolve large-scale genomic variation. For the first time here, the Nanopore sequencing allowed the identification of one extra copy of a 115 bp multi-copy repetition inside the IR<sub>S</sub> region that was not previously resolved by short-read approach. This long-read sequencing approach also confirmed the presence of the four major OsHV-1 isomers in viral particles for the first time since the acquisition of the first OsHV-1 genome [7]. Isomers are characterized by permutation of the unique regions U<sub>L</sub> and U<sub>S</sub> with respect to the IR<sub>L</sub>-X-IR<sub>S</sub> fragment. Some of the long-read obtained with Nanopore sequencing were large enough to cover one part of the two unique regions allowing us to quantify and confirm that the different OsHV-1 isomers were present in a stoichiometric proportion among viral particles. Interestingly, this observation is different from the one made in herpes simplex virus 1 (HSV-1), where four isomers have also been characterized, but with variable tropisms specific to cell lines and tissues [100]. Since the presence of isomeric genome in Herpesvirus is tightly linked to its replication and recombination process [98, 101, 102], having access to isomer frequency along with full genome characterization will allow in the near future a better understanding of the complex life cycle as well as of the evolutionary processes of this type of virus.

Finally, and even if this was not the primarily goal of this study, TFF allowed to co-purify multiple DNA viruses along with OsHV-1. For this specific application of viral metagenomics, Illumina allowed, due to the huge sequencing depth, to taxonomically identify ten times more virus than Nanopore sequencing for contigs with size larger than 10 kb. The great majority of identified DNA virus belong to bacteriophage. It is interesting to note that some of them are phages infecting *Vibrio*, bacteria, which have been described as colonizing moribund oysters during POMS outbreak [25, 103], others are *Pseudoalteromonas* and *Aeromonas* phages, pathogens of bacteria considered as beneficial for oysters [104, 105]. Of course, our data does not disentangle the potential role of these phages in POMS but raised questions about the necessity of investigating the virome of healthy and diseased oysters as it has been done for the bacteriome [25].

## CONCLUSION

This first evaluation of Nanopore sequencing for OsHV-1 genome characterization is extremely promising. Coupled with the appropriate bioinformatics pipeline this long-read sequencing approach enables identification of structural variation not observed with the short-read approach and allows characterization and quantification of genomic isomers. As shown in the present study, Nanopore sequencing still struggles to accurately sequence homopolymeric region. However, Nanopore is currently updating software for base calling as well as flow cell and chemistry to improve the resolution of the homopolymeric region [93, 106]. Illumina as well as Nanopore present difficulties to accurately resolve very long homopolymer (>20 bp) in GC-rich regions. It is thus of prime importance to focus efforts on methodological development for a better resolution of those regions that likely play a role for the diversity, replication and the stability of the viral genome [35, 107, 108].

The next challenge lies in combining TFF purification of virus populations harboured by a single individual and long-read sequencing to resolve viral haplotype phasing and better understand intra individual viral dynamics. Indeed, most of the whole-genome studies conducted on Herpesviruses to date only generate a consensus view of the viral population rather than a reconstruction of the viral haplotypes [96]. However, recent studies have shown that viral haplotype phasing could prove to be crucial in understanding the genetic determinants and viral evolution underpinning virulence, pathogenesis and drug resistance [109–111].

### Funding information

J.D. was supported by a grant from the University of Montpellier. C.P. was supported by a grant from the Ifremer Scientific Board and from the Nouvelle-Aquitaine region. A.D.M. was financially supported by Ifremer, by EU DG SANTE through the European Reference Laboratory for mollusc diseases, and by DGAL through the National Reference Laboratory for Mollusc Diseases. This work benefited from support from the 'Laboratoire d'excellence' (LabEx) CeMEB, through the exploratory research project HaploFit and the use of the 'environmental genomics' facility (<http://www.labex-cemeb.org/fr/genomique-environnementale-1>). This work was also supported by the Ifremer Scientific Board, through the HemoVir and GT-oyster projects and by the 'Fond Européen pour les Affaires Maritimes et la Pêche' (FEAMP, GESTINNOV project n°PFEA470020FA1000007). The authors acknowledge the Pôle de Calcul et de Données Marines (PCDM; <https://www.ifremer.fr/en/Research-Technology/Research-Infrastructures/Digital-infrastructures/Computation-Centre>) for providing DATARMOR computing and storage resources. Some data were produced through the GenSeq technical facilities of the « Institut des Sciences de l'Evolution de Montpellier » with the support of LabEx CeMEB, an ANR 'Investissements d'avenir' programme (ANR-10-LABX-04-01). This study is set within the framework of the 'Laboratoires d'Excellences (LABEX)' TULIP (ANR-10-LABX-41). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Acknowledgements

We thank the staff of the Ifremer station at La Tremblade (ASIM), Frédéric Girardin and his team (Plateforme des Mollusques Marins de La Tremblade PMMLT) and Christophe Stavrakakis and his team (Plateforme des Mollusques Marins de Bouin PMMB). We also thank Louis Boismorand and Marc Leroy for technical assistance. We are also grateful to Jean-Christophe Avarre and Marie-ka Tilak (ISEM) for fruitful discussions and advice concerning Nanopore sequencing technology.

### Author contributions

J.M.E. and B.M. conceptualized the study; B.M., N.F., M.B. and J.M.E. collected and processed the samples, J.D. and J.M.E. performed Nanopore sequencing, A.D.M. and G.C. did the bioinformatic analysis for Nanopore data, C.P. and G.C. did the bioinformatic analysis of the Illumina data, J.M.E. and G.C. drafted the manuscript and M.J. and B.M. corrected it. All authors read and approved the final manuscript.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**Ethical statement**

Not applicable, experimental works with bivalves are not currently subjected to ethical approval by Ifremer.

**References**

- Hall RJ, Wang J, Todd AK, Bissielo AB, Yen S, et al. Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery. *J Virol Methods* 2014;195:194–204.
- Kumar A, Murthy S, Kapoor A. Evolution of selective-sequencing approaches for virus discovery and virome analysis. *Virus Res* 2017;239:172–179.
- Gaafar YZA, Ziebell H, Pappu HR. Comparative study on three viral enrichment approaches based on RNA extraction for plant virus/viroid detection using high-throughput sequencing. *PLoS ONE* 2020;15:e0237951.
- Deng X, Achari A, Federman S, Yu G, Somasekar S, et al. Metagenomic sequencing with spiked primer enrichment for viral diagnostics and genomic surveillance. *Nat Microbiol* 2020;5:443–454.
- Simner PJ, Miller S, Carroll KC. Understanding the promises and hurdles of metagenomic next-generation sequencing as a diagnostic tool for infectious diseases. *Clin Infect Dis* 2018;66:778–788.
- Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet* 2018;19:9–20.
- Davison AJ, Trus BL, Cheng N, Steven AC, Watson MS, et al. A novel class of herpesvirus with bivalve hosts. *J Gen Virol* 2005;86:41–53.
- EFSA. Oyster mortality. *EFSA Journal* 2015;1.
- Azéma P, Lamy J-B, Boudry P, Renault T, Travers M-A, et al. Genetic parameters of resistance to vibrio aestuarianus, and oshv-1 infections in the pacific oyster, *Crassostrea gigas*, at three different life stages. genetics selection evolution. *BioMed Central* 2017;1–16.
- Dégremont L, Morga B, Maurouard E, Travers M-A. Susceptibility variation to the main pathogens of *Crassostrea gigas* at the larval, spat and juvenile stages using unselected and selected oysters to OsHV-1 and/or *V. aestuarianus*. *J Invertebr Pathol* 2021;183:107601.
- Xia J, Bai C, Wang C, Song X, Huang J. Complete genome sequence of *Ostreid herpesvirus-1* associated with mortalities of *Scapharca broughtonii* broodstocks. *Virol J BioMed Central* 2015;12:110–119.
- Cáceres-Martínez J, Vásquez-Yeomans R, Danigo P, Reyes-Roel C. Histological alterations in pacific oysters *Crassostrea gigas* that survived a summer mortality event in Baja California, Mexico. *J Aquat Anim Health* 2018;30:31–38.
- Hwang JY, Park JJ, Yu HJ, Hur YB, Arzul I, et al. Ostreid herpesvirus 1 infection in farmed Pacific oyster larvae *Crassostrea gigas* (Thunberg) in Korea. *J Fish Dis* 2013;36:969–972.
- Jenkins C, Hick P, Gabor M, Spiers Z, Fell SA, et al. Identification and characterisation of an ostreid herpesvirus-1 microvariant (OsHV-1  $\mu$ -var) in *Crassostrea gigas* (Pacific oysters) in Australia. *Dis Aquat Organ* 2013;105:109–126.
- Mortensen S, Strand Å, Bodvin T, Alfjorden A, Skår CK, et al. Summer mortalities and detection of ostreid herpesvirus microvariant in Pacific oyster *Crassostrea gigas* in Sweden and Norway. *Dis Aquat Organ* 2016;117:171–176.
- Pernet F, Barret J, Le Gall P, Corporeau C, Dégremont L, et al. Mass mortalities of pacific oysters *Crassostrea gigas* reflect infectious diseases and vary with farming practices in the Mediterranean Thau lagoon, France. *Aquacult Environ Interact* 2012;2:215–237.
- Prado-Alvarez M, Darmody G, Hutton S, O'Reilly A, Lynch SA, et al. Occurrence of oshv-1 in *Crassostrea gigas* cultured in Ireland during an exceptionally warm summer. *Selection of Less Susceptible Oysters Front Physiol* 2016;7:492.
- Segarra A, Pépin JF, Arzul I, Morga B, Faury N, et al. Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. *Virus Res* 2010;153:92–99.
- Garcia C, Thébault A, Dégremont L, Arzul I, Miossec L, et al. Ostreid herpesvirus 1 detection and relationship with *Crassostrea gigas* spat mortality in France between 1998 and 2006. *Vet Res* 2011;42:73.
- Renault T, Faury N, Barbosa-Solomieu V, Moreau K. Suppression subtractive hybridisation (SSH) and real time PCR reveal differential gene expression in the Pacific cupped oyster, *Crassostrea gigas*, challenged with *Ostreid herpesvirus 1*. *Dev Comp Immunol* 2011;35:725–735.
- Petton B, Pernet F, Robert R, Boudry P. Temperature influence on pathogen transmission and subsequent mortalities in juvenile Pacific oysters *Crassostrea gigas*. *Aquacult Environ Interact* 2013;3:257–273.
- Petton B, Boudry P, Alunno-Bruscia M, Pernet F. Factors influencing disease-induced mortality of Pacific oysters *Crassostrea gigas*. *Aquacult Environ Interact* 2015;6:205–222.
- Le Roux F, Wegner KM, MPTI. Oysters and vibrios as a model for disease dynamics in wild animals. *Trends Microbiol* 2016;24:568–580.
- Delisle L, Pauletto M, Vidal-Dupiol J, Petton B, Bargelloni L, et al. High temperature induces transcriptomic changes in *Crassostrea gigas* that hinders progress of Ostreid herpesvirus (OsHV-1) and promotes survival. *J Exp Biol* 2020.
- de Lorgeril J, Lucasson A, Petton B, Toulza E, Montagnani C, et al. Immune-suppression by OsHV-1 viral infection causes fatal bacteraemia in Pacific oysters. *Nat Commun* 2018;9:1–14.
- Rubio T, Oyanedel D, Labreuche Y, Toulza E, Luo X, et al. Species-specific mechanisms of cytotoxicity toward immune cells determine the successful outcome of *Vibrio* infections. *Proc Natl Acad Sci U S A* 2019;116:14238–14247.
- Petton B, Destoumieux-Garzón D, Pernet F, Toulza E, de Lorgeril J, et al. The Pacific Oyster Mortality Syndrome, a Polymicrobial and Multifactorial Disease: State of Knowledge and Future Directions. *Front Immunol* 2021;12:630343.
- Barbosa Solomieu V, Renault T, Travers M-A. Mass mortality in bivalves and the intricate case of the Pacific oyster, *Crassostrea gigas*. *J Invertebr Pathol* 2015;131:2–10.
- Renault T, Moreau P, Faury N, Pépin J-F, Segarra A, et al. Analysis of clinical ostreid herpesvirus 1 (malacoherpesviridae) specimens by sequencing amplified fragments from three virus genome areas. 2nd ed. *J Virol* 2012;86(10):5942–5947.
- Renault T, Tchaleu G, Faury N, Moreau P, Segarra A, et al. Genotyping of a microsatellite locus to differentiate clinical *Ostreid herpesvirus 1* specimens. *Vet Res* 2014;45:3.
- Burioli EAV, Prearo M, Houssin M. Complete genome sequence of *Ostreid herpesvirus* type 1  $\mu$ Var isolated during mortality events in the Pacific oyster *Crassostrea gigas* in France and Ireland. *Virology* 2017;509:239–251.
- Abbadì M, Zamperin G, Gastaldelli M, Pascoli F, Rosani U, et al. Identification of a newly described OsHV-1  $\mu$ var from the North Adriatic Sea (Italy). *J Gen Virol* 2018;99:693–703.
- Morga B, Jacquot M, Pelletier C, Chevignon G, Dégremont L, et al. Genomic diversity of the ostreid herpesvirus type 1 across time and location and among host species. *Front Microbiol* 2021;12:711377.
- Delmotte J, Pelletier C, Morga B, Galinier R, Petton B, 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. *Virus Evol* 2022;8:veac039.
35. Kuny CV, Szpara ML. Alphaherpesvirus genomics: past, present and future. *Curr Issues Mol Biol* 2021;42:41–80.
  36. Riaz N, Leung P, Barton K, Smith MA, Carswell S, et al. Adaptation of Oxford Nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. *BMC Genomics* 2021;22:148–12.
  37. Rios G, Lacoux C, Leclercq V, Diamant A, Lebrigand K, et al. Monitoring SARS-CoV-2 variants alterations in Nice neighborhoods by wastewater nanopore sequencing. *Lancet Reg Health Eur* 2021;10:100202.
  38. Toh X, Wang Y, Rajapakse MP, Lee B, Songkasupa T, et al. Use of nanopore sequencing to characterize african horse sickness virus (AHSV) from the African horse sickness outbreak in thailand in 2020. *Transbound Emerg Dis* 2022;69:1010–1019.
  39. Chan WS, Au CH, Lam HY, Wang CLN, Ho DN-Y, et al. Evaluation on the use of Nanopore sequencing for direct characterization of coronaviruses from respiratory specimens, and a study on emerging missense mutations in partial RdRP gene of SARS-CoV-2. *Virol J* 2020;17:183–13.
  40. King J, Harder T, Beer M, Pohlmann A. Rapid multiplex MinION nanopore sequencing workflow for Influenza A viruses. *BMC Infect Dis* 2020;20:648–8.
  41. Batista FM, Stapleton T, Lowther JA, Fonseca VG, Shaw R, et al. Whole genome sequencing of hepatitis A Virus Using A PCR-Free single-molecule nanopore sequencing approach. *Front Microbiol* 2020;11:874.
  42. Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature* 2017;546:406–410.
  43. Eltom KH, Althoff AC, Hansen S, Böhlken-Fascher S, Yousif A, et al. Differentiation of capripox viruses by nanopore sequencing. *Vaccines* 2021;9:351.
  44. Radukic MT, Brandt D, Haak M, Müller KM, Kalinowski J. Nanopore sequencing of native adeno-associated virus (AAV) single-stranded DNA using a transposase-based rapid protocol. *NAR Genomics and Bioinformatics* 2020;2.
  45. Kulikov EE, Golomidova AK, Babenko VV, Letarov AV. A simple method for extraction of the horse feces virome DNA, suitable for oxford nanopore sequencing. *Microbiology* 2020;89:246–249.
  46. Ben Chehida S, Filloux D, Fernandez E, Moubset O, Hoareau M, et al. Nanopore sequencing is a credible alternative to recover complete genomes of geminiviruses. *Microorganisms* 2021;9:903.
  47. Karamitros T, Harrison I, Piorkowska R, Katzourakis A, Magiorkinis G, et al. De novo assembly of human herpes virus type 1 (HHV-1) genome, mining of non-canonical structures and detection of novel drug-resistance mutations using short- and long-read next generation sequencing technologies. *PLoS One* 2016;11:e0157600.
  48. Karamitros T, van Wilgenburg B, Wills M, Klenerman P, Magiorkinis G. Nanopore sequencing and full genome de novo assembly of human cytomegalovirus TB40/E reveals clonal diversity and structural variations. *BMC Genomics* 2018;19:577–12.
  49. Vereecke N, Carnet F, Pronost S, Vanschandevijl K, Theuns S, et al. Genome sequences of equine herpesvirus 1 strains from a European outbreak of neurological disorders linked to a horse gathering in Valencia, Spain, in 2021. *Microbiol Resour Announc* 2021;10:e00333-21.
  50. Fuandila NN, Gosselin-Grenet A-S, Tilak M-K, Bergmann SM, Escoubas J-M, et al. Structural variation turnovers and defective genomes: key drivers for the in vitro evolution of the large double-stranded DNA koi herpesvirus (KHV). Cold Spring Harbor Laboratory. *bioRxiv* 2022;2.
  51. Vinaya Kumar K, Shekhar MS, Otta SK, Karthic K, Ashok Kumar J, et al. First report of a complete genome sequence of White spot syndrome virus from India. *Genome Announc* 2018;6:e00055-18.
  52. Gallagher MD, Matejusova I, Nguyen L, Ruane NM, Falk K, et al. Nanopore sequencing for rapid diagnostics of salmonid RNA viruses. *Sci Rep* 2018;8:16307–16309.
  53. Liem M, Regensburg-Tuink T, Henkel C, Jansen H, Spaink H. Microbial diversity characterization of seawater in a pilot study using Oxford Nanopore Technologies long-read sequencing. *BMC Res Notes* 2021;14:42–47.
  54. Beaulaurier J, Luo E, Eppley JM, Uyl PD, Dai X, et al. Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res* 2020;30:437–446.
  55. Urban L, Holzer A, Baronas JJ, Hall MB, Braeuninger-Weimer P, et al. Freshwater monitoring by nanopore sequencing. *Elife* 2021;10:e61504.
  56. Wickramasinghe SR, Kalbfuss B, Zimmermann A, Thom V, Reichl U. Tangential flow microfiltration and ultrafiltration for human influenza A virus concentration and purification. *Biotechnol Bioeng* 2005;92:199–208.
  57. Michalsky R, Passarelli AL, Pfromm PH, Czermak P. Purification of the baculovirus *Autographa californica* M nucleopolyhedrovirus by tangential flow ultrafiltration. *Desalination* 2009;245:694–700.
  58. Eilts F, Harnischfeger J, Loewe D, Wolff MW, Salzig D, et al. Production of *Baculovirus* and Stem Cells for Baculovirus-Mediated Gene Transfer into Human Mesenchymal Stem Cells. *Methods Mol Biol* 2021;2183:367–390.
  59. Negrete A, Pai A, Shiloach J. Use of hollow fiber tangential flow filtration for the recovery and concentration of HIV virus-like particles produced in insect cells. *J Virol Methods* 2014;195:240–246.
  60. Loewe D, Grein TA, Dieken H, Weidner T, Salzig D, et al. Tangential Flow Filtration for the Concentration of Oncolytic Measles Virus: The Influence of Filter Properties and the Cell Culture Medium. *Membranes* 2019;9:E160.
  61. Shi H, Pasco EV, Tarabara VV. Membrane-based methods of virus concentration from water: a review of process parameters and their effects on virus recovery. *Environ Sci: Water Res Technol* 2017;3:778–792.
  62. Cordier C, Stavrakakis C, Morga B, Degrémont L, Voulgaris A, et al. Removal of pathogens by ultrafiltration from sea water. *Environ Int* 2020;142:105809.
  63. Beaulaurier J, Luo E, Eppley JM, Uyl PD, Dai X, et al. Assembly-free single-molecule sequencing recovers complete virus genomes from natural microbial communities. *Genome Res* 2020;30:437–446.
  64. Azéma P, Lamy J-B, Boudry P, Renault T, Travers M-A, et al. Genetic parameters of resistance to *Vibrio aestuarianus*, and OsHV-1 infections in the Pacific oyster, *Crassostrea gigas*, at three different life stages. *Genet Sel Evol* 2017;49:23.
  65. Dégremont L, Bédier E, Soletchnik P, Ropert M, Huvet A, et al. Relative importance of family, site, and field placement timing on survival, growth, and yield of hatchery-produced Pacific oyster spat (*Crassostrea gigas*). *Aquaculture* 2005;249:213–229.
  66. Schikorski D, Renault T, Sautnier D, Fauray N, Moreau P, et al. Experimental infection of Pacific oyster *Crassostrea gigas* spat by ostreid herpesvirus 1: demonstration of oyster spat susceptibility. *Vet Res* 2011;42:13–27.
  67. Webb SC, Fidler A, Renault T. Primers for PCR-based detection of ostreid herpes virus-1 (OsHV-1): Application in a survey of New Zealand molluscs. *Aquaculture* 2007;272:126–139.
  68. Leruste A, Bouvier T, Bettarel Y. Enumerating viruses in coral mucus. *Appl Environ Microbiol* 2012;78:6377–6379.
  69. Patel A, Noble RT, Steele JA, Schwalbach MS, Hewson I, Fuhrman JA. Virus and prokaryote enumeration from planktonic aquatic environments by epifluorescence microscopy with SYBR Green I. *Nat Protoc*. Nature Publishing Group; 2007;2(2):269–76.
  70. Suquet M, de Kermoyan G, Araya RG, Queau I, Lebrun L, et al. Anesthesia in pacific oyster, *crassostrea gigas*. *Aquat Living Resour EDP Sciences* 2009;22(1):29–34.

71. Dupont S, Lokmer A, Corre E, Auguet J-C, Petton B, et al. Oyster hemolymph is a complex and dynamic ecosystem hosting bacteria, protists and viruses. *Anim Microbiome* 2020;2:12.
72. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
73. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–359.
74. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–2079.
75. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;27:824–834.
76. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
77. Cunningham C, Gatherer D, Hilfrich B, Baluchova K, Dargan DJ, et al. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. *J Gen Virol* 2010;91:605–615.
78. Morse AM, Calabro KR, Fear JM, Bloom DC, McIntyre LM. Reliable detection of herpes simplex virus sequence variation by high-throughput resequencing. *Viruses* 2017;9:E226.
79. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
80. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 2020;17:1103–1110.
81. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, et al. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 2017;27:722–736.
82. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 2017;27:737–746.
83. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, et al. Twelve years of SAMtools and BCftools. *Gigascience* 2021;10:giab008.
84. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016;44:W16–21.
85. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021;18:366–368.
86. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
87. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004;14:1394–1403.
88. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* 2019;47:W260–W265.
89. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 2010;10:210–221.
90. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 2012;9:772–2.
91. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010;59:307–321.
92. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al. Characterizing and measuring bias in sequence data. *Genome Biol* 2013;14:20–R51.
93. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS One* 2021;16:e0257521.
94. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genom Bioinform* 2021;3:lqab019.
95. Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. *Nat Rev Microbiol* 2017;15:183–192.
96. Renner DW, Szpara ML. Impacts of genome-wide analyses on our understanding of human herpesvirus diversity and evolution. *J Virol* 2018;92:126–13.
97. Sozhamannan S, Holland MY, Hall AT, Negrón DA, Ivancich M, et al. Evaluation of signature erosion in ebola virus due to genomic drift and its impact on the performance of diagnostic assays. *Viruses* 2015;7:3130–3154.
98. 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. *Microb Genom* 2021;7.
99. Hage E, Wilkie GS, Linnenweber-Held S, Dhingra A, Suárez NM, et al. Characterization of human cytomegalovirus genome diversity in immunocompromised hosts by whole-genome sequencing directly from clinical specimens. *J Infect Dis* 2017;215:1673–1683.
100. Mahiet C, Ergani A, Huot N, Alende N, Azough A, et al. Structural variability of the herpes simplex virus 1 genome in vitro and in vivo. *J Virol* 2012;86:8592–8601.
101. 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. *Cell* 1985;41:803–811.
102. McVoy MA, Ramnarain D. Machinery to support genome segment invertible elements. *J Virol* 2000;74:4882–4887.
103. Le Roux F, Wegner KM, Polz MF. Oysters and Vibrios as a Model for Disease Dynamics in Wild Animals. *Trends Microbiol* 2016;24:568–580.
104. Desriac F, El Harras A, Simon M, Bondon A, Brillet B, et al. Alterins Produced by Oyster-Associated *Pseudoalteromonas* Are Antibacterial Cyclolipopeptides with LPS-Binding Activity. *Mar Drugs* 2020;18:E630.
105. Gibson LF, Woodworth J, George AM. Probiotic activity of *Aeromonas media* on the Pacific oyster, *Crassostrea gigas*, when challenged with *Vibrio tubiashii*. *Aquaculture* 1998;169:111–120.
106. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 2021;39:1348–1365.
107. Olson AT, Wang Z, Rico AB, Wiebe MS. A poxvirus pseudokinase represses viral DNA replication via A pathway antagonized by its paralog kinase. *PLoS Pathog* 2019;15:e1007608.
108. Castillo-Lizardo M, Henneke G, Viguera E. Replication slippage of the thermophilic DNA polymerases B and D from the Euryarchaeota *Pyrococcus abyssi*. *Front Microbiol* 2014;5:403.
109. Wu NC, De La Cruz J, Al-Mawsawi LQ, Olson CA, Qi H, et al. HIV-1 quasispecies delineation by tag linkage deep sequencing. *PLoS One* 2014;9:e97505.
110. Monaco DC, Zapata L, Hunter E, Salomon H, Dilernia DA. Resistance profile of HIV-1 quasispecies in patients under treatment failure using single molecule, real-time sequencing. *AIDS* 2020;34:2201–2210.
111. Della Bartola M, Byrne S, Mullins E. Characterization of potato virus Y isolates and assessment of nanopore sequencing to detect and genotype potato viruses. *Viruses* 2020;12:E478.