

1. Title

Evaluation of long-read sequencing for Ostreid herpesvirus type 1 genome characterization from *Magallana gigas* infected tissues

1.1. Author names

Aur lie Dotto-Maurel¹, Camille Pelletier¹, Lionel Degremont¹, Serge Heurtebise¹, Isabelle Arzul Benjamin Morga¹, Germain Chevignon^{1#}

1.2. Affiliation

¹Ifremer, ASIM, F-17390 La Tremblade, France

1.3. Corresponding author

Germain.Chevignon@ifremer.fr

1.4. Keyword

Virus; Ostreid herpesvirus type 1; Long reads; genome; Oxford Nanopore Technologies; *Magallana gigas*.

2. Abstract

Since the 1990s, the Pacific oyster *Magallana gigas* has faced significant mortality, which has been associated with the detection of the Ostreid Herpesvirus type 1 (OsHV-1). Due to the complex genomic architecture and the presence of multiple genomic isomers, short-read sequencing using Illumina method struggles to accurately assemble tandem and repeat regions and to identify and characterize large structural variations in the OsHV-1 genome. Third-generation sequencing technologies, as long-read real-time nanopore sequencing from Oxford Nanopore Technologies (ONT), offer new possibilities for OsHV-1 whole genome analysis. Identification of the best method for extraction of high molecular weight (HMW) DNA and development of accurate bioinformatic pipelines for its characterization are now required. To this end, we evaluated and compared six HWM methods and one conventional DNA extraction kit for their ability to extract OsHV-1 DNA from *M. gigas*-infected tissues. We then evaluated the ability of ONT sequencing to produce an accurate OsHV-1 genome from both whole genome and “adaptive sampling” (AS) sequencing approaches. Finally, we evaluated the efficiency of bioinformatics tools for *de novo* assembly and consensus calling to generate accurate OsHV-1 genomes. The HMW DNA extraction kit coupled with ONT sequencing and dedicated bioinformatics tools allowed us to produce accurate OsHV-1 genomes compared to those assembled using Illumina technology. The AS approach allowed up to 60% enrichment for viral data, and the long reads generated by ONT allowed the characterization of OsHV-1 isomers. Together with its portability, this sequencing shows great promise as a diagnostic tool for the characterization of unculturable aquatic viruses directly from host tissues.

3. Importance

Many aquatic viruses threaten commercially valuable species and cause significant economic losses during outbreaks. To improve our understanding of the origin, transmission patterns and spread of these viruses, additional genomic data are essential. However, genomic characterization of unculturable large DNA viruses is a major challenge. In the present study, we have successfully evaluated the ability of ONT sequencing and adaptive sequencing (AS) to sequence and assemble the complete OsHV-1 genome. Our results show that it is now possible to sequence the whole genome of large DNA viruses directly from infected host tissue, without the need for prior *in vitro* propagation or prior laboratory steps for virus enrichment.

4. Introduction

Ostreid herpesvirus type 1 (OsHV-1) is a double-stranded DNA virus of Malacoherpesviridae family that has been associated with mass mortality events of Pacific oyster *Magallana gigas* (formerly named *Crassostrea gigas*), the most widely produced mollusk in the world (1, 2). OsHV-1 has impacted oyster production in many production basin worldwide for fifteen years (3–9). It can cause mass mortalities of up to 100% within days (10, 11) and several factors trigger disease development including seawater temperature, oyster age and genetics background (8, 11–16). While this virus cannot be cultured *in vitro*, OsHV-1 infection can be propagated *in vivo*. Therefore, its genomic characterisation relies on our ability to enrich samples for viral nucleic acids.

OsHV-1 has a linear genome of approximately 207 kilobases (kb) and its genomic organization is similar to other Herpesviruses. This genome consists of two invertible unique regions U_L for “Unique long” (170 kbp) and U_S for “Unique short” (3.5 kbp) each flanked by tandem and inverted repeats TR_L/IR_L (7.5 kbp each) around the U_L and TR_S/IR_S (9.7 kbp each) around the U_S and separated by a third unique region named X (1.5 kbp) for a global genomic organization TR_L - U_L - IR_L - X - IR_S - U_S - TR_S. In addition, as characterized previously (1) and similarly to other herpes virus (17), 4 majors isomers of the virus coexist at equimolar levels in the same samples: isomers P for Prototype with U_L and U_S in the forward orientation, I_L for Inverted U_L, I_S for Inverted U_S and I_{LS} for Inverted U_L and U_S. Furthermore, each of these 4 majors isomers have 3 minors isomers: D_X where the X region is Duplicated (i.e. X - TR_L - U_L - IR_L - X - IR_S - U_S - TR_S), T_X where the X region is Translocated at the left termini of the genome (i.e. X - TR_L - U_L - IR_L - IR_S - U_S - TR_S) and N_X where the X region is absent (i.e. TR_L - U_L - IR_L - IR_S - U_S - TR_S) (1, 17, 18).

To date, detection and characterization of OsHV-1 virus in infected oysters are mainly performed by PCR and its genome is regularly sequenced using Illumina short-read sequencing method (11, 19). However, Illumina sequencing does not allow for accurate assembly of tandem and inverted repeats (i.e. TR_L/IR_L and TR_S/IR_S) resulting in the assembly of a Non-Redundant (NR) genome version (i.e. U_L - IR_L - X - IR_S - U_S) (19).

While the OsHV-1 life cycle is not well characterized, research perspectives for its characterization include better identification of the genomic modification of the virus during the cycle. Those modifications could include the production of noncanonical genome with large structural variation or could involve variation in the number of genomic isomers. Currently, the characterization of such genomic alteration can only be envisioned thanks to long read approaches.

In the past decade, new sequencing technologies, allowing long DNA fragments sequencing have been developed by Oxford Nanopore Technologies (ONT). Recently, we showed that ONT sequencing allowed us to use long read technology to assemble a complete OsHV-1 genome and accurately characterize viral isomers from TFF-filtered viral particles (18).

While the combination of TFF and ONT sequencing has shown promising results in improving the genomic characterization of OsHV-1, the implementation of this approach is complex and requires a large amount of biological material, making it difficult to use for OsHV-1 genomic characterization at the individual host level. To achieve the goal of real-time sequencing of the OsHV-1 genome at the individual host level, it is necessary to simplify the procedure and reduce the number of laboratory steps involved in sample processing.

ONT has recently introduced a selective sequencing approach named "adaptive sampling" (AS), which enables real-time enrichment of specific sequences of interest. This method is software-driven and provides the ability to accept or reject DNA molecules within individual nanopore sequencing units. By rapidly analyzing the first few hundred base pairs of a DNA molecule and comparing them to a reference sequence, it determines whether the molecule matches the target or not. If a match is found, the molecule is fully sequenced. If not, the current in the pore is reversed, ejecting the DNA molecule and allowing the pore to capture a new one (20).

This approach has already been successfully tested with a number of different organisms and has been able to enrich for viruses, bacteria or mammals by a factor of up to 7 (21–24).

The AS approach has been implemented in several third-party software with varying levels of efficiency for targeted sequences (25–28). Recently, ONT implemented AS as a beta version in the Mk1C control software, a simple user-selectable option. Users can provide a file of reference sequences and choose an enrichment or depletion mode.

This study aims to evaluate the ONT sequencing technology to characterize OsHV-1 genome directly from infected oyster tissues. To achieve it, we used infected moribund oysters to first evaluate a number of high molecular weight (HMW) DNA extraction kits for extracting OsHV-1 DNA directly from infected host tissues. We then tested the ability of ONT to sequence the OsHV-1 genome from DNA extracted from a single host individual with and without AS. We then tested the ability of ONT to sequence the OsHV-1 genome from DNA extracted from a single host individual with and without AS. To determine the level of accuracy, the ONT genome was compared to a genome generated using Illumina technology, and we determined the minimum number of ONT reads required to obtain an ONT

OsHV-1 genome with 99.99% accuracy. Finally, we evaluated the quality of different consensus callers for producing accurate OsHV-1 genomes without the time-consuming de novo assembly steps.

5. Results

5.1. Yield, quality metrics and OsHV-1 copy number of extracted DNA

Our results showed that the majority of the kits tested on 20 ± 2 mg of tissues, yielded similar amounts of DNA ranging from 10.6 to 14.9 μg , with the exception of the MagAttract and Zymo kits, which yielded an average of 8.7 and 4.6 μg respectively (Figure 3A). Most of the kits tested produced a consistent yield across different samples, with the exception of the Circulomics kit, which had a yield range from 3.7 μg to 27.5 μg .

The extracted DNA was then migrated onto an agarose gel to assess the DNA size distribution of the fragments for each extraction method. Samples with the longest DNA fragment sizes were those extracted using the Monarch, Circulomics and Macherey Nagel kits, while the Qiagen kit produced predominantly fragmented DNA, as expected. Promega, Zymo and MagAttract kits produced similar mixed DNA fragments, with some long fragments and some fragmented DNA (Figure 3B).

The DNA yields and qualities claimed by the manufacturers are shown in Table 3. Overall, we obtained DNA yields that were in line with the claims, except for the Promega kit which yielded ten times more DNA than expected. On the other hand, we extracted less DNA than expected from the MagAttract and the Macherey Nagel kits. As expected, based on the manufacturer's claims, the Monarch, Circulomics and Macherey Nagel kits provided the longest apparent DNA fragments (Table 3).

The number of OsHV-1 genome copies per ng of DNA was similar among kits except for the Macherey Nagel kit ranging from 115 000 to 7 520 000 copies per ng of DNA (Figure 3C), indicating that all kits have similar abilities to extract OsHV-1 DNA from infected oyster tissue.

5.2. ONT allows sequencing of OsHV-1 from infected oyster tissue

5.2.1. Adaptive sampling enables enrichment of OsHV-1 reads

The six ONT sequencing runs produced a total of 13.4 M reads and 18.6 G bases, ranging from 0.59 M reads for run 4 to 5 M reads for run 2, and from 0.3 G bases for run 4 to 8.4 G bases for run 2 (Figure S1).

By mapping the reads against the *M. gigas* and OsHV-1 μVarA reference genomes, we were able to determine the origin of the reads for each DNA sample in each sequencing run (Figure 4). Run 1, which used DNA samples with the best ratio of OsHV-1 per ng of extracted DNA, had the highest proportions of OsHV-1 reads among the "whole sequencing" runs (runs 1, 2, and 3). For the run 1, the proportion of OsHV-1 reads ranged from 4.4% for the MagAttract kit to 17.8% for the Macherey Nagel kit, with an average of 9.7% (Figure 4). Sequencing runs 2 and 3, which used samples selected for their

higher apparent fragment size on agarose gels, had generally a lower proportion of OsHV-1 compared to run 1 (Figure 4). In addition, the proportions of OsHV-1 reads and bases were very similar for each sample within the three “whole sequencing” runs. Conversely, the three sequencing runs where AS was enabled for OsHV-1 enrichment, showed at least a 2-fold increase in the percentage of OsHV-1 bases sequenced compared to the percentage of reads sequenced (Figure 4). For the run 4, where DNA fragments were size selected, there was an even more pronounced enrichment of OsHV-1 bases compared to the number of reads, with Monarch, Macherey Nagel and Circulomics DNA samples containing up to 68.1%, 56.3% and 38.9% of OsHV-1 sequenced bases respectively (Figure 4).

M. gigas reads had higher N50 values for most of the sequenced samples in the three “whole sequencing” runs, while reads that did not align to *M. gigas* or OsHV-1 had the lowest N50 value. OsHV-1 N50 values were always close to those of *M. gigas*, but rarely higher (Macherey Nagel and QiAgen samples from run 1) (Figure 5). However, in the AS runs, non-OsHV-1 reads (i.e. *M. gigas* and other reads) showed a significant decrease in N50 values with values around 500 bp.

Furthermore, OsHV-1 reads in runs 5 (OsHV-1 N50 = 2 065 bp) and 6 (OsHV-1 N50 = 2 302 bp) did not show an increase in N50 compared to runs without AS (run 1 OsHV-1 N50 = 4 436 bp, run 2 OsHV-1 N50 = 2 627 bp and run 3 OsHV-1 N50 = 2 267 bp), but OsHV-1 N50 values benefited greatly from the combination of the AS and size selection. Indeed, in run 4, N50 values ranged from 4 534 bp for the MagAttract sample to 8 350 bp for the Monarch sample, with an average OsHV-1 N50 of 5 820 bp (Figure 5).

5.3. ONT sequencing allows accurate characterization of the complete OsHV-1 genome

5.3.1. Illumina OsHV-1 genome assembly

The Illumina sequencing run yielded more than 102 M of 150 bp paired-end reads, distributed as follows: 0.90 % of OsHV-1 reads, 93.82 % of *M. gigas* reads and 5.29 % of unaligned reads (Figure 6 top panel). Our pipeline, based on the *de novo* assembler SPAdes, produced five OsHV-1 contigs ranging from 2 484 bp to 164 393 bp (Figure 6, top panel). These 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 Materials and Methods for more details). By re-aligning the Illumina reads on the new OsHV-1 NR genome, we were able to correct the junction between contigs C7709 and C3973 (addition of one A/T bp) and to confirm that the other metaSPAdes contigs were correctly assembled. As expected, since the NR genome contains only one copy of these duplicated regions, we observed that the read coverage within IR_L and IR_S was twice that of the unique regions U_L , U_S and X (Figure 7, top panel).

5.3.2. ONT OsHV-1 genome assembly

Shasta produced the smallest contigs (N50: 3 kb) and the highest number of OsHV-1 and non-OsHV-1 contigs with a total of 494/1 314 and 494/1 454 contigs for all reads and the top 10% longest read dataset, respectively. Smartdenovo struggled to assemble a complete OsHV-1 genome for all datasets except with the top 0.1% longest reads (N50: 237 kb and 1/1 contig). RA required a low coverage to assemble the OsHV-1 genome and generated an extended OsHV-1 contig of 233 kb for the top 0.1% longest reads dataset only. Raven generated an extended OsHV-1 contig of more than 200 kb for all but the “all read” dataset however, like Shasta, never fully assembled the U_L region. Flye produced the highest number of long contigs in both the OsHV-1 and the non-OsHV-1 data sets, with a total of 58/32, 28/4, 3/4 and 6 contigs and N50 of 44 kb, 32 kb, 179 kb and 186 kb, respectively. Canu produced 1 to 4 contigs among the datasets and seemed to be the best tool for OsHV-1 assembly. In addition, Flye and Canu consistently produced one OsHV-1 contig that was longer than 170 kb and 228 kb respectively across all datasets. This contig is always shorter than expected for Flye whereas it is always longer than expected for Canu.

5.3.3. Illumina short-read and ONT long-read sequencing produce nearly identical genomes

Pairwise alignment of the Illumina and ONT assembled genomes revealed only 15 bp differences between the two genomes, corresponding to 99.992 % average nucleotide identity (ANI). Interestingly, all of these differences occurred only in G/C homopolymers, which is a known limitation of long-read sequencing and more specifically of ONT sequencing (Figure 8). We then evaluated the minimum ONT read depth required to assemble and polish an OsHV-1 genome as accurately as possible (i.e. to achieve the highest ANI compared to the Illumina genome). The polishing step increased the ANI from 99.947% for the Canu assembly alone to 99.994% for the Medaka-polished genome (Figure 9). Finally, we found that when the coverage of the OsHV-1 genome reached 50 reads, the quality of the Canu assembly polished with a run of Medaka was close to the Illumina genome (99.994% accuracy) (Figure 9).

5.4. ONT consensus callers show contrasting results depending on read depth

To ensure a rapid and accurate capture of OsHV-1 genomes, we can also call a consensus sequence directly after aligning the reads to a reference genome. This approach bypasses the need for *de novo* assembly and polishing steps, which can be time consuming.

To evaluate this approach, we tested reference mapping with reads coming from all 42 libraries individually, with pooled reads from each of the 6 runs, and with a pool of all the sequenced reads on the new *de novo* assembled and polished ONT OsHV-1 genome. These 49 libraries were consensus called using five different tools (See M&M for more details). For each consensus call, we obtained 49 consensus genomes and compared them pairwise using the alignment *all versus all* (Figure S2). Since all the 42 ONT sequenced libraries were prepared with DNA extracted from tissues infected by the same

viral inoculum, the observed ANI is mainly the result of ONT sequencing error combined with the ability of the different consensus callers to correct this error. Our results show that Clair3 and Pepper DeepVariant and to a lesser extent Medaka were the most consistent for producing accurate genomes (i.e. producing the genome closest to the one generated from Illumina) (Figure S2). When comparing the 49 ONT consensus calls to the Illumina assembled genome, Medaka, Clair3 and Pepper DeepVariant had an ANI of 99.991% with the Illumina genome (Figure 10). However, Pepper DeepVariant and Clair3 were more consistent than Medaka to produce accurate genomes across all coverage levels tested (Figure 10). Longshot performed best at generating accurate genomes with coverage greater than 100 (Figure S2 and 10), achieving an ANI greater than 99.990% for read depths ranging from 104 to 6 697 (Figure 10). Finally, Nanocaller excelled at producing accurate genomes with coverage below 200 (Figure 10 and S2), achieving an ANI greater than 99.990% for a coverage between 11 and 244 (Figure 10).

5.5. ONT sequencing allows detection of OsHV-1 structural variation and the quantification of the four major isomers

The Canu assembly of the top 0.1% longest reads provided the most reliable genome structure and produced 3 additional contigs corresponding to the other 3 major OsHV-1 isomers (Figure 6 bottom panel) (1). Similar to our previous results (18), we quantified the relative abundance of the four isomers by screening all the reads long enough to cover at least a part of both the U_L and the U_S regions. Thus, we identified 137 long reads divided into 32 I_{LS} isomers, 36 P isomers, 39 I_L isomers and 30 I_S isomers (Figure 11).

Furthermore, thanks to the ability of ONT to generate long reads, we use Sniffle to analyse the structural variations (SV) directly within the OsHV-1 reads. We detected 18 SVs, six of which were likely to be related to the U_L inversion. As the reads were shorter than the U_L length (< 164 kb), we could not determine whether they were true SVs or artefacts due to the OsHV-1 isomers, so we decided to filter them out. Finally, we obtained 12 SVs (Figure 12), including a deletion of 1 510 bp corresponding to the absence of the X region supported by 91 reads, an inversion of 3 371 bp corresponding to the inversion of the U_S region supported by 46 reads and a duplication of 1 515 bp located in the X region, corresponding to the duplication of at least two copies of the X region supported by 12 reads. We also detected 9 SVs spread across the U_L , IR_L , X, IR_S , U_S and TR_S regions, each supported by only 2 reads. Due to the low number of reads supporting these SVs, they were considered as spurious. Six of them were located in the U_L region; one overlapped the U_L and IR_L -X regions; one was located over two parts of the U_L region and in the U_S region; and the last SV corresponds to a deletion of approximately 8 200 bp that includes the entire U_S region and part of the TR_S and IR_S regions (data not shown).

To determine the distribution of reads supporting the different SV at the individual library level, we analyzed each DNA extraction from each kit individually for each sequencing run, specifically for the X deletion, U_S inversion and X duplication identified from the whole dataset (Figure S3). The U_S

inversion was identified in 35 reads for run 1; two for run 2, eight for run 4 and one for run 6. Of these, 22 were generated from the Monarch kit, two from the Promega kit, one from the QiAgen kit, five from the Circulomics kit and 16 from the Macherey Nagel kit. The X deletion was identified in 32 reads for run 1; 37 for run 2; six for run 3; five for runs 4 and 5 and six for run 6. Of these, two were generated by the Promega kit, 33 by the Monarch kit, 10 by the QiAgen kit, 19 by the Zymo kit, 5 by the MagAttract kit, 13 by the Circulomics kit and 9 by the Macherey Nagel kit. Finally, the X duplication was identified in 6 reads for run 1, 4 reads for run 2 and 4 reads for run 6. Among them, 1 was generated by the Promega and the QiAgen kits, 5 by the Monarch kit, 1 by the Circulomics kit and 4 by the Macherey Nagel kit.

Most of these SV-containing reads came from the first two runs (37 of the 46 reads identified as Us inversion, 69 of the 91 reads identified as X deletion and 10 of the 12 reads identified as X duplication). Regarding the different kits, 40% of the 149 SV-bearing reads were from the Monarch extraction, 20% from the Macherey Nagel extraction and 13% from the Circulomics extraction.

6. Discussion

Recently, whole genome sequencing has become a standard method to characterize and study viruses (29–31). However, this approach is still challenging when working with DNA viruses with large and complex genomes such as Herpesviruses. The inverted repeat regions in the OsHV-1 genome increase the difficulty of obtaining a properly assembled genome. In addition, several isomers have been characterized and recently observed and quantified from purified OsHV-1 particles (1, 18). Thus, these features further complicate genome assembly using a short-read sequencing approach. Finally, because OsHV-1 is unculturable, obtaining sufficient viral material to analyze and assemble an accurate OsHV-1 genome is a real challenge.

In a previous study, we showed that ONT sequencing of viral OsHV-1 particles purified by TFF allowed us to overcome these difficulties (18). Indeed, we were able to generate up to 60% of viral bases, which represents a significant enrichment compared to previous work where the ratio of OsHV-1 bases from a short read sequencing of infected tissues ranged from 0.1% to 13.1% (19, 32, 33). However, although TFF is a powerful tool for purifying unculturable viruses from clinical and environmental isolates, it still requires additional experiments and increases cost, which slows down and complicates viral pathogen identification. Therefore, TFF cannot be used routinely.

In our study, seven DNA extraction kits were tested. Monarch, Circulomics, and Macherey Nagel kits were those that extracted the largest DNA fragment sizes. Overall, the amount of the OsHV-1 genomes were similar among the kits tested and they were also similar to those observed in previous studies (34, 35) kits have similar abilities to extract OsHV-1 DNA from infected oyster tissues.

Two types of sequencing were then performed: Whole Genome (WG) sequencing without enrichment, and Adaptive Sampling (AS) sequencing with real-time enrichment for OsHV-1 sequences. While WG sequencing resulted in an average OsHV-1 base count of 9.7% under all conditions, similar to previous studies where no enrichment was performed (19, 32, 33), the AS sequencing enriched our data in viral bases by at least 2-fold, reaching up to 60% of viral bases for some samples, approaching the results obtained with our TFF-based OsHV-1 purification (18). The levels of enrichment observed with AS were also similar to those obtained for enrichment of the Human Adenovirus (HAdV) where 0.24 to 7.08-fold enrichment were obtained (22). Although the availability of a reference genome is required to perform AS, this approach could overcome the limitations associated with traditional sequencing methods, such as the need for greater sequencing depth, thus reducing the cost of sequencing and data storage.

Improvements in the extraction of viral data from infected tissues represent a significant advance. However, optimal use of this method requires the ability to construct high-quality viral genomes. To date, complex viral genomes, such as herpesviruses, require a combination of long- and short-read sequencing for accuracy and resolution of inverted repeat regions (36, 37). Our recent validation of the accuracy of ONT sequencing in generating OsHV-1 genomes without hybrid *de novo* assembly encouraged us to evaluate its ability in *de novo* assembly directly from infected *M. gigas* tissues. (18).

To our knowledge, only Canu and Flye have been tested to assemble OsHV-1 genomes from ONT reads (18). Here, we have also evaluated RA, Shasta, Smartdenovo and Raven to identify the best tool for OsHV-1 genome assembly. Interestingly, none of the *de novo* assemblers/data subset combinations produced a perfect OsHV-1 genome with a structure similar to the OsHV-1 reference genome (1). All tools struggled with the placement of the inverted repeat regions and with the orientation of the U_L / U_S isomer inversions. However, similar to our previous results (18), Canu produced the best assembly by combining the top 0.1% longest reads. Thanks to a limited manual curation (the removal of an extra TR_S - U_S - IR_S - X region at the 5' end of the genome and an extra X - TR_L region at the 3' end), we obtained a complete OsHV-1 genome with an architecture similar to the OsHV-1 reference genome (1). In addition, Canu was able to assemble three other contigs corresponding to the three other major OsHV-1 isomers (1, 18).

In our study, we compared the *de novo* ONT assembly with the Illumina short-read assembly to assess the accuracy of the ONT genome. Due to limitations in the resolution of inverted repeats and viral isomerisation, the short-read approach produced five contigs that required extensive manual scaffolding to obtain a NR OsHV-1 genome structure similar to the reference genome (1, 18). While a previous comparison identified 187 bp differences between Illumina and ONT sequencing approaches (18), our present study found only 15 bp differences between the two sequencing methods. Discrepancy between studies could be related to the use of different sequencing library kits, indeed in the previous comparison, the Rapid sequencing kit (SQK-RAD004) was used for ONT library preparation, while in this

comparison, we used the Ligation sequencing kit (LSK109). The LSK109 kit includes a DNA repair step using NEB's NEBNext® FFPE DNA Repair Mix, which potentially corrects for sequence differences observed with the Rapid kit. Although the Rapid sequencing kit reduces handling time, the ligation sequencing approach appears to offer a lower final error rate.

In order to maximize sequencing costs and processing time, it is of paramount importance to evaluate the minimum read depth required to assemble an accurate OsHV-1 genome. Using the approach tested here, we found that a coverage of 50 OsHV-1 ONT long reads was sufficient to obtain a complete OsHV-1 genome with more than 99.99% accuracy compared to the Illumina genome. In other words, only 6,061 OsHV-1 long reads with an N50 of 3 kb were required to assemble a complete and accurate OsHV-1 genome. Assuming that 5% of the data sequenced from infected tissues is from OsHV-1, it would take approximately 3 hours to generate an accurate OsHV-1 genome.

An alternative approach to rapidly obtain accurate genomes is to bypass time-consuming *de novo* assembly and polishing steps by directly calling consensus sequences after aligning reads to a reference genome. In our study, we assessed the performance of five consensus callers (NanoCaller, Clair3, Longshot, Medaka, and Pepper DeepVariant) in calling an accurate OsHV-1 genome using ONT reads from different DNA extraction kits and sequencing runs.

Since the same viral inoculum was used for all ONT samples, and both the Illumina and ONT assembled genomes showed minimal differences (15 bp), the observed percentage of identity between the consensus sequences and the Illumina OsHV-1 genome primarily reflects ONT sequencing errors and the ability of consensus callers to generate accurate consensus. Notably, we found that the accuracy of certain consensus callers was directly influenced by read depth.

NanoCaller demonstrated higher accuracy with low read depth samples, enabling the generation of accurate consensus sequences from limited data, which could be advantageous for real-time genome calling of individuals with low levels of viral infection. Conversely, Longshot performed exceptionally well in high read depth scenarios, generating accurate consensus sequences from extensive data, particularly in the later stages of infection. Clair3, Pepper DeepVariant and, to a lesser extent, Medaka, showed consistent performance, generating accurate consensus sequences from all samples regardless of sequencing depth. These results are consistent with previous studies that have observed similar results regarding read depth and tool performance (38–41).

Pairwise comparisons between the Illumina and ONT genomes, along with results from minimum sequencing depth and consensus calling, strongly suggest that ONT is an efficient method for assembling a complete and accurate OsHV-1 genome. This is achieved without the need for a hybrid approach (i.e. without the need of both long and short-read sequencings) and with minimal bias from sequencing errors.

In addition to its efficiency at the base level, ONT sequencing allowed us to detect structural variations within the OsHV-1 genome and to characterize and quantify the different isomers. While four major OsHV-1 isomers are characterized by inversions of the U_L and U_S regions, three minor isomers

are characterized by the absence of the X region or its displacement at the 5' end of the TR_L region and between the IR_L and IR_S regions (1). Recently, we have reconfirmed the presence of the four major isomers in equimolar proportions within viral particles (18). Here, similar results were obtained from infected tissues. In addition, few reads were also missing in the X region in this new dataset. However, this observation is more likely due to a higher sequencing depth rather than phenotypic differences between genomes sequenced from infected tissues and viral particles.

This is the first time such an observation has been made from viral DNA isolated from infected tissues since the original OsHV-1 genome was sequenced (1). However, longer reads are needed to better quantify and characterize these major and minor isomers. In this study, none of the isomer identification methods used were able to confirm the presence of all minor isomers in the OsHV-1 read population because none of the reads included the entire U_L region, making it impossible to determine whether the X region located at the 5' end of the TR_L region is a true minor isomer or just a result of the inverted U_L isomer.

The study of structural variation among herpesviruses is critical to understanding their complete viral life cycles. While OsHV-1 and HSV-2 show equimolar fractions of major isomers in both viral particles and infected tissues (42), HSV-1 shows variable isomer proportions depending on the cell type and strain infected (43), whereas VZV shows differences in the proportions of inverted short and long region isomers (44). Understanding these structural variations is essential for the comprehensive characterization of herpesvirus life cycles.

The use of ONT sequencing coupled with AS allowed us to sequence OsHV-1 genomes directly from host tissues. Thanks to the availability of an OsHV-1 reference genome, this approach allowed an enrichment of viral genomes without the constraints of laborious laboratory steps required to get enough viral particles isolated from the host tissues.

This advanced sequencing technology is a promising solution to rapidly and cost-effectively detect OsHV-1 in infected oyster tissues, to better characterize its life cycle, and to provide an efficient diagnostic tool to detect unculturable aquatic viruses directly from infected host tissues.

7. Materials and Methods

7.1. Short-read Illumina sequencing

7.1.1. OsHV-1 natural infection

During a mortality outbreak in August 2020, highly susceptible standardized oyster spat were deployed in the field at “La Floride” in the Marennes-Oléron Bay, France (Lat 45.80° and Long -1.15°). When mortality began, spat were transferred to the laboratory and checked twice a day to sample fresh moribund individuals.

7.1.2. DNA extraction and library preparation

About 35 mg of mantle were collected from one moribund oyster. Nucleic acid extraction was performed using the Magattract HMW kit (Qiagen) according to the manufacturer's protocol. DNA purity and concentration were measured using the Nanodrop ND-1000 spectrometer (Thermo Scientific) and Qubit® dsDNA BR assay kits (Molecular Probes Life Technologies). Before sequencing, DNA was tested for viral copy number by qPCR (see below). Illumina DNA-Seq library preparation and sequencing were performed by Genome Quebec Company (Genome Quebec Innovation Center, McGill University, Montreal, Canada) as previously described (Dotto-Maurel et al, 2022), using 120 ng of DNA extracted from the infected oyster, using the Shotgun PCR Free Library Kit (Lucigen) and the NovaSeq™ 6000 Sequencing device (Illumina®) (paired ends, 150 bp) (Figure 2).

7.2. Long-read ONT sequencing

7.2.1. Oyster production

Highly susceptible *M. gigas* spat were produced at the Ifremer hatchery in Bouin in June 2020, (Vendée, France) as described in (45). Oysters were always protected in our facilities using filtered and UV-treated seawater, and so they never experienced mortality outbreaks. Prior the experiment, spat (9 months old) 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*).

7.2.2. Viral inoculum and experimental OsHV-1 infection

From the same cohort as the one sampled for the Illumina sequencing (see above), the OsHV-1 strain used for the experimental infection was primarily isolated from moribund infected oysters collected on August 5th 2020, and was then successively experimentally inoculated twice in juvenile oysters. The virus was isolated from the gills and mantle of moribund oysters infected with OsHV-1 on January 25th, 2021. Viral inoculum was prepared following previously described protocols (35) and, for each oyster, 100 µL aliquot of OsHV-1 inoculum ($9.3 \cdot 10^5$ OsHV-1 DNA copies µL⁻¹) was injected into the adductor muscle. In total, 50 juvenile oysters were used.

7.2.3. Sampling and tissue dissection

Oysters were checked 48 hrs and 72 hrs post OsHV-1 injection, and each moribund oyster (47 in total) was dissected for gills and mantle isolation. Dissections were then subjected to manual homogenization with plastic pestles. Homogenates were then carefully weighed and stored on ice until DNA extraction.

7.2.4. DNA extraction

For each sample, between 18 mg and 25 mg of homogenized tissues were subjected to DNA extraction with one or two of the seven commercial DNA extraction kits. Six of these kits were specifically designed for the extraction of high molecular weight (HMW) DNA: the Promega Wizard® HMW DNA Extraction kit (hereafter referred as Promega, P) which uses centrifugation for the DNA cleaning step; the New England Biolabs Monarch® HMW DNA Extraction kit for Tissue (hereafter referred as Monarch, M), which uses glass beads to retain DNA during the cleaning steps; the Zymo Research Quick-DNA HMW MagBead kit (hereafter referred as Zymo, Z) and the Qiagen MagAttract® HMW DNA kit (hereafter referred as MagAttract, Ma), both of which use magnetic beads for the DNA cleaning steps; the Circulomics Nanobind Tissue Big DNA kit (hereafter referred as Circulomics, C), which uses magnetic disk to retain DNA during the cleaning steps; and the Macherey Nagel NucleoBond AXG 20 kit (hereafter referred as Macherey Nagel, Mn) which uses a gravity column to extract DNA. The last kit used was the Qiagen QIAmp® DNA kit (hereafter referred as Qiagen, Q and Q'), which is designed for the recovery of regular weight DNA and uses silica columns for DNA purification. Each extraction kit was tested using 5 to 7 moribund oysters (Table 1).

Each DNA extraction procedure was carried out according to the manufacturer's protocol. In order to preserve HMW DNA as much as possible, the main modifications involved replacing pipette/vortex mixing with gentle flicking. For the Circulomics and Monarch kits, we reduced the agitation to 250 rpm during the lysis and RNase treatment steps, instead of 900 rpm and 300 rpm respectively. Following DNA extraction with the different kits, all purified DNA were stored at 4°C until further processing, avoiding unnecessary freeze-thaw cycles.

7.2.5. DNA quality control

The purity of freshly extracted DNA was evaluated by Nanodrop spectrophotometer, and the DNA concentrations were measured using the Qubit™ dsDNA BR Assay kit. The distribution size of the DNA fragments was then evaluated by electrophoresis with the migration of 300 ng of purified DNA on a 0.6% agarose gel containing ethidium bromide for 2.5h at 112 V. The agarose gels were also loaded with Quick-Load® 1kb Extend DNA Ladder, which has sizes ranging from 0.5 kpb to 48.5 kpb. These agarose gels were then visualized under UV light and the images were captured with the ImageLab 3.0 software.

7.2.6. OsHV-1 genome copy number estimation by qPCR

Quantification of the OsHV-1 genome copy number was performed using quantitative PCR (qPCR) using the primers OsHVDP For (forward) 5'-ATTGATGATGTGGATAATCTGTG-3' and OsHVDP Rev (reverse) 5'-GGTAAATACCATTGGTCTTGTTC-3' (46) (18). The absolute quantification of viral genome copy number was estimated by comparing the observed Ct value with a standard curve obtained using a serial dilution of plasmidic DNA (pUC57) including the target region (equivalent to 1.10^5 to 1.10^1 OsHV-1 DNA copies μL^{-1}). Each qPCR run included a complete standard

curve and a non-template control (NTC). All reactions were performed with three technical replicates. Data were then collected on the MxPro software, and the dissociation curves were verified. Reaction efficiency was confirmed to be between 90% and 110% with $R^2 > 90$. Finally, technical replicates with Ct differences over 0.5 were rejected. The OsHV-1 genome copy number was then calculated for all DNA-extracted samples for each extraction kit.

7.2.7. Library preparation and ONT sequencing

ONT sequencing libraries were prepared using the Ligation Sequencing Kit (SQK-LSK109) coupled with the PCR-free Native Barcoding Expansion Kit (EXP-NBD104 and EXP-NBD114) for multiplexing following the ONT protocol “Native barcoding genomic DNA” (Version: NBE_9065_v109_revAA_14Aug2019). The libraries were loaded on MinION flow cells (FLO-MIN106D, R9.4) and sequenced on a MinION Mk1C device using the MinKNOW software version 21.02.2.

To determine the ability of ONT to sequence OsHV-1 from infected oyster tissues, we performed a total of six ONT sequencing runs, each with seven multiplexed samples corresponding to DNA extracted with the seven different extraction kits. Three ONT sequencing runs were performed on the whole DNA samples (hereafter named “Whole sequencing”) and three ONT sequencing runs were performed with AS option activated with enrichment on the OsHV-1 μ VarA reference genome (GenBank: KY242785.1) in order to enrich for OsHV-1 DNA molecules in real time during the sequencing process.

Flow cells used for two different runs were washed using the Flow Cell Wash kit (EXP-WSH004), allowing them to be re-used in subsequent runs. Additionally, libraries loaded consecutively on the same washed flow cell were built with different barcode set. Each sequencing run was setup as described in Table 2. Run 1 was performed on samples with the best ratio of the OsHV-1 genome copies per ng of extracted DNA while the five others runs were performed with the best apparent migration profile on agarose. Additionally, a short fragment elimination step, with the Circulomics Short Read Eliminator kit (10kb) was performed for the run 4 (Table 2, Figure 3).

After sequencing, all fast5 data generated by the Mk1C with the MinKNOW software were transferred to the Ifremer HPC cluster Datarmor for data analysis (Figure 2).

7.3. Data Analysis

7.3.1. Illumina assembly

Illumina OsHV-1 genome assembly has been performed using a previously developed bioinformatic pipeline (Dotto-Maurel et al, 2022). This pipeline is summarized in Figure 2.

7.3.2. ONT basecalling, cleaning, quality control and reads sorting

The resulting fast5 reads were basecalled with Guppy GPU version 5.0.7 with the following parameters: `--config dna_r9.4.1_450bps_hac.cfg --detect_mid_strand_barcodes, --min_score_mid_barcodes 50, --min_score 50, --trim_threshold 50, -- trim_barcodes and --barcode_kits "EXP-NBD104 EXP-NBD114"`. The fastq reads obtained after basecalling were cleaned with Porechop v0.2.4 (<https://github.com/rrwick/Porechop>) to remove remaining adapters and barcodes. To this end, the adapter.py file containing reference adapters and barcode sequences used by Porechop was modified to include the sequences of the adapters from the SQS-LSK109 kit. Porechop was run with the following parameters: `--adapter_threshold 78, --middle_threshold 78 and --extra_middle_trim_bad_side 10 --check_reads 100000`. The resulting clean fastq reads were then subject to quality control analysis with MinIONQC v1.4.2 (47). To determine the origins of the read (*M. gigas* versus OsHV-1), we performed a read classification by mapping the ONT reads on the *M. gigas* reference genome (GenBank: GCA_902806645.1) and on the OsHV-1 μ Var A reference genome (GenBank: KY242785.1) using Minimap2 v2.19 (48) with the parameter: `-x map-ont`. The alignment files were then compressed, sorted and indexed using Samtools v1.11 (49). Custom homemade scripts were then used to extract non-*M. gigas* reads, which were used for OsHV-1 *de novo* assembly (Figure 2).

7.3.3. ONT OsHV-1 *de novo* assembly

Six long-read *de novo* assemblers were tested: Canu version 2.1.1 (50) with standard parameters; Flye version 2.8.3 16 (51) with the parameter `--meta`; Shasta version 0.6.0 (<https://github.com/paoloshasta/shasta>) with standard parameters; Smartdenovo (<https://github.com/ruanjue/smartdenovo>) with parameter: `-J 200`; Raven version 1.5 (<https://github.com/lcb-science/raven>) with standard parameters and RA version 0.2.1 (<https://github.com/lcb-science/ra>) with standard parameters. Each assembler was run: (i) on a pool of reads corresponding to all the non-*M. gigas* reads from all the tested kits and from all the sequenced runs, (ii) on the top 10% longest reads, (iii) on the top 1% longest reads and (iv) on the top 0.1% longest reads. OsHV-1 *de novo* assembled contigs were then identified with BlastN (52) against the OsHV-1 μ Var A genome (GenBank: KY242785.1) and curated in Geneious version 2020.2.4 (<https://www.geneious.com>) The curated OsHV-1 genome were then consensus called with all non-*M. gigas* reads using Medaka version 1.4.3 with parameters `medaka_consensus` and `-m r103_sup_g507` (<https://github.com/nanoporetech/medaka>) (Figure 2).

To evaluate the minimum ONT reads needed to obtain a complete and accurate OsHV-1 genome, we generated several fastq datasets, containing non-*M. gigas* reads from all runs, each corresponding to a targeted coverage on the OsHV-1 genome. We used a combination of Canu *de novo* assembly version 2.1.1 and Medaka consensus calling version 1.4.3 (<https://github.com/nanoporetech/medaka>) on i) 99 fastq datasets with coverage ranging from 20 to 1000 and ii) 56 fastq datasets with coverage ranging from 1100 to 6700. Each assembly was separately

compared to the OsHV-1 genome assembled from Illumina to evaluate its quality through the ANI between the two genomes.

7.3.4. Comparison of Illumina and ONT OsHV-1 genome sequencing

Before whole genome alignment, both genomes were modified to fit the same isomers version (i.e. P isomer). To evaluate the accuracy of ONT sequencing, the manually constructed Illumina OsHV-1 Non-Redundant (NR) genome was aligned with the manually curated ONT OsHV-1 complete genome using Mafft version 1.4.0 with default parameters (53) implemented in Geneious version 2020.2.4 (<https://www.geneious.com>).

7.3.5. Evaluation of long-read consensus callers

For consensus caller evaluation, reads from each sample, each kit and each run were mapped (i) individually (42 alignments); (ii) by run (sum of seven samples: 6 alignments) and (iii) all sequenced reads (sum of the six runs: 1 alignment) on the new *de novo* assembled and consensus called ONT OsHV-1 genome. A total of 49 alignment files were consensus called with five long reads consensus callers: Pepper-MARGIN-DeepVariant version r0.5 (40) with parameter: --ont; Nanocaller version 0.4.0 (38) with parameter: -seq ont; Medaka version 1.4.3 (<https://nanoporetech.github.io/medaka>) with parameter: medaka_haploid_variant; Longshot version 0.4.3 (54) with parameter --no_haps; and Clair3 version 0.1 (55) with parameters: --chunk_size=205000, --no_phasing_for_fa, --include_all_ctgs and --haploid_precise. For each consensus caller, 49 consensus sequences were then produced using Bcftools version 1.10.2 (49) with standard parameters of bcftools consensus. Mafft version 7.313 with standard parameters was used to create an alignment of the 49 consensus sequences along with five other OsHV-1 genome sequences (i) the OsHV-1 genome from 2005 (GenBank: NC005881.2), (ii) the OsHV-1 μ Var A sequence (GenBank: KY242785.1), (iii) the OsHV-1 genome assembled with Illumina reads and both (iv) non-polished and (v) polished OsHV-1 genomes assembled with ONT reads. Five pairwise percentage identity matrices (one for each consensus caller) of the 54 OsHV-1 sequences were then extracted with Geneious v2020.2.4 (<https://www.geneious.com>) (Figure 2).

7.3.6. OsHV-1 structural variation and isomers quantification

For structural variation detection and OsHV-1 isomers characterization, we evaluated two methods. First, to quantify OsHV-1 reads corresponding to different isomers, we used the Canu *de novo* assembly using the top 0.1% of the longest reads, which produced four contigs each representing one of the four major isomers (P: $U_L \rightarrow / U_S \rightarrow$; I_S : $U_L \rightarrow / U_S \leftarrow$; I_{LS} : $U_L \leftarrow / U_S \leftarrow$ and I_L : $U_L \leftarrow / U_S \rightarrow$). They were trimmed on the U_L side to retain a 10 kb U_L fragment with Geneious version 2020.2.4 (<https://www.geneious.com>). The four major isomers were then polished with Medaka version 1.4.3 (<https://github.com/nanoporetech/medaka>) using all the non-*M. gigas* reads and were used as reference for mapping all reads aligning on the ONT assembled genome using Minimap2 version 2.19 with

parameters -ax map-ont, --sam-hit-only and --secondary=no. Samtools version 2.11 was then used to filter out all the shortest aligned reads and all alignments with MapQ < 40, retaining only the reads spanning at least a fragment of the U_L and a fragment of the U_S (i.e., 28 kb long). Second, to evaluate the detection of structural variations (SV) within the OsHV-1 genome, we used Sniffles2 version 2.0.6 (56) with the parameters: --minsupport 1 and --no-qc, to detect all SV regardless of the number of reads supporting the SV or the quality of the SV. We used two types of alignments as input for Sniffles 2: (i) all ONT reads of all runs on the polished ONT OsHV-1 genome and (ii) ONT reads from each DNA extraction kits of each run on the polished ONT OsHV-1 genome. We then filtered the SVs, keeping only those that were flagged as “PRECISE”, longer than 500 bp and supported by at least 2 reads (Figure 2).

7.4. Visualization

We generated all the figures using R (<https://www.r-project.org>) with ggplot2 (<https://ggplot2.tidyverse.org>) or using Geneious. We prepared the figures using Inkscape (<https://inkscape.org/fr>).

7.5. Data availability

All sequence data have been deposited in ENA ...
ONT data are available here <https://doi.org/10.12770/1a301465-ae7-42c2-a950-7708f422a6aa> and
Illumina data are accessible via <https://doi.org/10.12770/e12a5d05-cc95-407f-b0f7-c86e79cc5b3a>
Analysis scripts used for ONT data are publicly available at <https://gitlab.ifremer.fr/lgpmm/ont-bioinformatic-tools-for-oshv-1-genome-characterization>

8. Acknowledgments

8.1. Authors and contributors

GC, BM, CP, SH and ADM designed the study; LD, BM and CP provided the biological materials; ADM, GC and SH processed the samples. ADM and CG performed ONT sequencing. CP and GC performed the bioinformatic analysis of the Illumina data, and ADM and GC performed the bioinformatic analysis of the ONT data. ADM, GC and CP drafted the manuscript and BM and IA corrected it. All authors read and approved the final version of the manuscript.

8.1. Conflicts of interests

The authors declare that there are no conflicts of interest.

8.2. Funding information

CP was financially supported by grant from the Ifremer Scientific Board and from the Nouvelle-Aquitaine region. ADM 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. 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. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

8.4. Acknowledgments

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 the SEBIMER team for maintaining bioinformatics tools.

9. Appendices

Figure Supplementary 1: Statistics of sequencing runs

Pie chart showing percentages reads sequenced and percentages of bases sequenced for each DNA samples sequenced from seven different HMW DNA extraction kits, for six ONT sequencing runs, along with some run characteristics: Flow cell batch number, number of flow cell washes, number of pores available before sequencing, run duration, and total number of reads and bases sequenced.

Figure Supplementary 2: Consensus caller evaluation. Each consensus caller was tested on 49 alignment files made by mapping reads coming from each sample and from each sequencing run, plus all reads pooled together from each run, plus all reads pooled together from all runs against the OsHV-1 ONT genome produced from all reads from all run and polished with medaka. The 49 OsHV-1 consensus sequences were aligned with the OsHV-1 reference sequence (GenBank: NC005881.2); OsHV-1 μ VarA (GenBank: KY242785.1), Illumina NR OsHV-1 (GenBank: XXX); and both polished and non-polished ONT OsHV-1 genomes (GenBank of the polished ONT OsHV-1 genome: XXX). Pairwise percentage of identity and base number differences are color-coded from red (100% identity) to green. Average reads coverage for each alignment is color-coded from light grey (low coverage) to dark grey (high coverage).

Figure Supplementary 3: Structural variation of ONT reads from each library. Identification of the major structural variations (X deletion / duplication and U_s inversion) detected in each of the 42

ONT libraries. The OsHV-1 structure is shown with the specific regions: TR_L/IR_L (turquoise); U_L (white); X (blue); TR_S/IR_S (grey); U_S (dark grey).

10. References

1. Davison AJ, Trus BL, Cheng N, Steven AC, Watson MS, Cunningham C, Deuff R-ML, Renault T. 2005. A novel class of herpesvirus with bivalve hosts. *Journal of General Virology* 86:41–53.
2. EFSA Panel on Animal Health and Welfare (AHAW). 2015. Oyster mortality. *EFSA Journal* 13:4122.
3. Xia J, Bai C, Wang C, Song X, Huang J. 2015. Complete genome sequence of Ostreid herpesvirus-1 associated with mortalities of *Scapharca broughtonii* broodstocks. *Virology Journal* 12:110.
4. Cáceres-Martínez J, Vásquez-Yeomans R, Danigo P, Reyes-Roel C. 2018. Histological Alterations in Pacific Oysters *Crassostrea gigas* that Survived a Summer Mortality Event in Baja California, Mexico. *J Aquat Anim Health* 30:31–38.
5. Hwang JY, Park JJ, Yu HJ, Hur YB, Arzul I, Couraleau Y, Park MA. 2013. Ostreid herpesvirus 1 infection in farmed Pacific oyster larvae *Crassostrea gigas* (Thunberg) in Korea. *Journal of Fish Diseases* 36:969–972.
6. Jenkins C, Hick P, Gabor M, Spiers Z, Fell S, Gu X, Read A, Go J, Dove M, O'Connor W, Kirkland P, Frances J. 2013. Identification and characterisation of an ostreid herpesvirus-1 microvariant (OsHV-1 μ -var) in *Crassostrea gigas* (Pacific oysters) in Australia. *Dis Aquat Org* 105:109–126.
7. Mortensen S, Strand Å, Bodvin T, Alfjorden A, Skår CK, Jelmert A, Aspán A, Sælemyr L, Naustvoll L-J, Albretsen J. 2016. Summer mortalities and detection of ostreid herpesvirus microvariant in Pacific oyster *Crassostrea gigas* in Sweden and Norway. *Dis Aquat Organ* 117:171–176.
8. Pernet F, Barret J, Le Gall P, Corporeau C, Dégremont L, Lagarde F, Pépin J, Keck N. 2012. Mass mortalities of Pacific oysters *Crassostrea gigas* reflect infectious diseases and vary with farming practices in the Mediterranean Thau lagoon, France. *Aquacult Environ Interact* 2:215–237.
9. Prado-Alvarez M, Darmody G, Hutton S, O'Reilly A, Lynch SA, Culloty SC. 2016. Occurrence of OsHV-1 in *Crassostrea gigas* Cultured in Ireland during an Exceptionally Warm Summer. Selection of Less Susceptible Oysters. *Front Physiol* 7.
10. Garcia C, Thébault A, Dégremont L, Arzul I, Miossec L, Robert M, Chollet B, François C, Joly J-P, Ferrand S, Kerdudou N, Renault T. 2011. Ostreid herpesvirus 1 detection and relationship with *Crassostrea gigas* spat mortality in France between 1998 and 2006. *Veterinary Research* 42:73.
11. Segarra A, Pépin JF, Arzul I, Morga B, Faury N, Renault T. 2010. Detection and description of a particular Ostreid herpesvirus 1 genotype associated with massive mortality outbreaks of Pacific oysters, *Crassostrea gigas*, in France in 2008. *Virus Research* 153:92–99.
12. Azema P, Lamy J-B, Boudry P, Renault T, Travers M-A, Degremont L. 2017. 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* 49:1–16.

13. Petton B, Pernet F, Robert R, Boudry P. 2013. Temperature influence on pathogen transmission and subsequent mortalities in juvenile Pacific oysters *Crassostrea gigas*. *Aquacult Environ Interact* 3:257–273.
14. Petton B, Boudry P, Alunno-Bruscia M, Pernet F. 2015. Factors influencing disease-induced mortality of Pacific oysters *Crassostrea gigas*. *Aquacult Environ Interact* 6:205–222.
15. Le Roux F, Wegner KM, Polz MF. 2016. Oysters and *Vibrios* as a Model for Disease Dynamics in Wild Animals. *Trends Microbiol* 24:568–580.
16. Delisle L, Pauletto M, Vidal-Dupiol J, Petton B, Bargelloni L, Montagnani C, Pernet F, Corporeau C, Fleury E. 2020. High temperature induces transcriptomic changes in *Crassostrea gigas* that hinder progress of ostreid herpesvirus (OsHV-1) and promote survival. *Journal of Experimental Biology* 223:jeb226233.
17. Hayward GS, Jacob RJ, Wadsworth SC, Roizman B. 1975. Anatomy of herpes simplex virus DNA: evidence for four populations of molecules that differ in the relative orientations of their long and short components. <https://www.pnas.org/doi/10.1073/pnas.72.11.4243>. Retrieved 3 September 2023.
18. Dotto-Maurel A, Pelletier C, Morga B, Jacquot M, Faury N, Dégremont L, Bereszczynski M, Delmotte J, Escoubas J-M, Chevignon G. 2022. Evaluation of tangential flow filtration coupled to long-read sequencing for ostreid herpesvirus type 1 genome assembly. *Microbial Genomics* 8.
19. Delmotte J, Pelletier C, Morga B, Galinier R, Petton B, Lamy J-B, Kaltz O, Avarre J-C, Jacquot M, Montagnani C, Escoubas J-M. 2022. Genetic diversity and connectivity of the Ostreid herpesvirus 1 populations in France: A first attempt to phylogeographic inference for a marine mollusc disease. *Virus Evolution* 8:veac039.
20. Loose M, Malla S, Stout M. 2016. Real-time selective sequencing using nanopore technology. *Nat Methods* 13:751–754.
21. Kipp EJ, Lindsey LL, Khoo BS, Faulk C, Oliver JD, Larsen PA. 2021. Enabling metagenomic surveillance for bacterial tick-borne pathogens using nanopore sequencing with adaptive sampling. preprint. *Microbiology*.
22. Lin Y, Dai Y, Liu Y, Ren Z, Guo H, Li Z, Li J, Wang K, Yang L, Zhang S, Liu H, Jia L, Ni M, Li P, Song H. 2022. Rapid PCR-Based Nanopore Adaptive Sequencing Improves Sensitivity and Timeliness of Viral Clinical Detection and Genome Surveillance. *Frontiers in Microbiology* 13.
23. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. 2022. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biology* 23:11.
24. Wanner N, Larsen PA, McLain A, Faulk C. 2021. The mitochondrial genome and Epigenome of the Golden lion Tamarin from fecal DNA using Nanopore adaptive sequencing. *BMC Genomics* 22:726.
25. Weilguny L, De Maio N, Munro R, Manser C, Birney E, Loose M, Goldman N. 2023. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nat Biotechnol* 41:1018–1025.
26. Ulrich J-U, Lutfi A, Rutzen K, Renard BY. 2022. ReadBouncer: precise and scalable adaptive sampling for nanopore sequencing. *Bioinformatics* 38:i153–i160.

27. Kovaka S, Fan Y, Ni B, Timp W, Schatz MC. 2021. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. 4. *Nat Biotechnol* 39:431–441.
28. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. 2021. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. 4. *Nat Biotechnol* 39:442–450.
29. Al-Eitan LN, Wu G, Golding M, Tang Y, Goharriz H, Marston DA, Fooks AR, McElhinney LM. 2021. Whole-genome sequencing and phylogenetic analysis of rabies viruses from Jordan. *PLoS Negl Trop Dis* 15:e0009431.
30. Kobayashi H, Uchida Y, Fujino K, Horie M, Umezawa E, Aihara N, Kamiie J, Shimoda H, Maeda K, Une Y, Taharaguchi S. 2022. Isolation and whole-genome sequencing of a novel aviadenovirus from owls in Japan. *Arch Virol* 167:829–838.
31. Wei T, Wang C, Ma F, Guo J, Chen A, Huang Y, Xie Z, Zheng L. 2023. Whole genome sequencing and evolution analyses of Human metapneumovirus. *Virus Genes* 59:524–531.
32. Morga B, Jacquot M, Pelletier C, Chevignon G, Dégremont L, Biétry A, Pepin J-F, Heurtebise S, Escoubas J-M, Bean TP, Rosani U, Bai C-M, Renault T, Lamy J-B. 2021. Genomic Diversity of the Ostreid Herpesvirus Type 1 Across Time and Location and Among Host Species. *Front Microbiol* 12:711377.
33. Burioli EA V, Prearo M, Houssin M. 2017. Complete genome sequence of Ostreid herpesvirus type 1 μ Var isolated during mortality events in the Pacific oyster *Crassostrea gigas* in France and Ireland. *Virology* 509:239–251.
34. Green TJ, Rolland J-L, Vergnes A, Raftos D, Montagnani C. 2015. OsHV-1 countermeasures to the Pacific oyster's anti-viral response. *Fish & Shellfish Immunology* 47:435–443.
35. Schikorski D, Renault T, Saulnier D, Faury N, Moreau P, Pépin J-F. 2011. Experimental infection of Pacific oyster *Crassostrea gigas* spat by ostreid herpesvirus 1: demonstration of oyster spat susceptibility. *Veterinary Research* 42:27.
36. Karamitros T, van Wilgenburg B, Wills M, Klenerman P, Magiorkinis G. 2018. Nanopore sequencing and full genome de novo assembly of human cytomegalovirus TB40/E reveals clonal diversity and structural variations. *BMC Genomics* 19:577.
37. López-Muñoz AD, Rastrojo A, Kropp KA, Viejo-Borbolla A, Alcamí A. 2021. Combination of long- and short-read sequencing fully resolves complex repeats of herpes simplex virus 2 strain MS complete genome. *Microb Genom* 7:000586.
38. Ahsan MU, Liu Q, Fang L, Wang K. 2021. NanoCaller for accurate detection of SNPs and indels in difficult-to-map regions from long-read sequencing by haplotype-aware deep neural networks. *Genome Biology* 22:261.
39. Helal AA, Saad BT, Saad MT, Mosaad GS, Aboshanab KM. 2022. Evaluation of the Available Variant Calling Tools for Oxford Nanopore Sequencing in Breast Cancer. 9. *Genes* 13:1583.
40. Shafin K, Pesout T, Chang P-C, Nattestad M, Kolesnikov A, Goel S, Baid G, Kolmogorov M, Eizenga JM, Miga KH, Carnevali P, Jain M, Carroll A, Paten B. 2021. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods* 18:1322–1332.
41. Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. 2022. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling <https://doi.org/10.1101/2021.12.29.474431>.

42. Chang W, Jiao X, Sui H, Goswami S, Sherman BT, Fromont C, Caravaca JM, Tran B, Imamichi T. 2022. Complete Genome Sequence of Herpes Simplex Virus 2 Strain G. 3. *Viruses* 14:536.
43. Mahiet C, Ergani A, Huot N, Alende N, Azough A, Salvaire F, Bensimon A, Conseiller E, Wain-Hobson S, Labetoulle M, Barradeau S. 2012. Structural Variability of the Herpes Simplex Virus 1 Genome In Vitro and In Vivo. *Journal of Virology* 86:8592–8601.
44. Kinchington PR, Reinhold WC, Casey TA, Straus SE, Hay J, Ruyechan WT. 1985. Inversion and circularization of the varicella-zoster virus genome. *Journal of Virology* 56:194–200.
45. Dégremont L, Nourry M, Maurouard E. 2015. Mass selection for survival and resistance to OsHV-1 infection in *Crassostrea gigas* spat in field conditions: response to selection after four generations. *Aquaculture* 446:111–121.
46. Webb SC, Fidler A, Renault T. 2007. Primers for PCR-based detection of ostreid herpes virus-1 (OsHV-1): Application in a survey of New Zealand molluscs. *Aquaculture* 272:126–139.
47. Lanfear R, Schalamun M, Kainer D, Wang W, Schwessinger B. 2019. MinIONQC: fast and simple quality control for MinION sequencing data. *Bioinformatics* 35:523–525.
48. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.
49. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008.
50. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736.
51. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan J, Pevzner PA. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat Methods* 17:1103–1110.
52. Altschul SF. 2014. BLAST Algorithm Encyclopedia of Life Sciences. John Wiley & Sons, Ltd.
53. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780.
54. Edge P, Bansal V. 2019. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. 1. *Nat Commun* 10:4660.
55. Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, Luo R. 2022. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. 12. *Nat Comput Sci* 2:797–803.
56. Smolka M, Paulin LF, Grochowski CM, Mahmoud M, Behera S, Gandhi M, Hong K, Pehlivan D, Scholz SW, Carvalho CMB, Proukakis C, Sedlazeck FJ. 2022. Comprehensive Structural Variant Detection: From Mosaic to Population-Level. *bioRxiv* <https://doi.org/10.1101/2022.04.04.487055>.

11. Legends

Figure 1: Schematic representation of adaptive sampling with OsHV-1 enrichment

Figure 2: Experimental design and bioinformatic pipeline implemented in this study

Figure 3: DNA extractions characteristic. A. Total extracted DNA in ng obtained with the seven different HMW DNA extraction kits. The X-axis represents the replicates of the extraction kits (5 replicates for Promega and QiAgen, 6 replicates for QiAgen', Circulomics and Macherey Nagel and 7 replicates for Monarch, Zymo and MagAttract) while the Y-axis represents the total DNA extracted (ng). The horizontal black lines correspond to the mean of extracted DNA for each extraction kit while the vertical black lines represent its standard deviation. B. Agarose gel electrophoresis of genomic DNA isolated from OsHV-1-infected oyster tissue using seven different HMW DNA extraction kits. First line of each gel corresponds to DNA ladders ranging from 0.5 kb to 48.5kb. C. OsHV-1 genome copy number per ng of extracted DNA for seven different HMW DNA extraction kits. The X-axis represents the replicates of the extraction kits (5 replicates for Promega and QiAgen, 6 replicates for QiAgen', Circulomics and Macherey Nagel and 7 replicates for Monarch, Zymo and MagAttract) while the Y-axis represents the OsHV-1 copy number per ng of DNA. The horizontal black lines correspond to the mean of extracted DNA for each extraction kit while the vertical black lines represent its standard deviation.

Figure 4: Classification of sequenced reads and bases for six ONT sequencing runs and seven different HMW DNA extraction kits. Histograms showing the proportion of reads (blue) and bases (green) for OsHV-1 (dark colors), *M. gigas* (mid-dark colors), "OsHV-1 + *M. gigas*" (mid-light colors) and "Unaligned" (light colors) for six sequencing run and seven different HMW DNA extraction kits. "Unaligned" corresponds to reads that did not align to either OsHV-1 μ VarA or *M. gigas*. "OsHV-1 + *M. gigas*" corresponds to reads that align on OsHV-1 μ VarA and on *M. gigas*.

Figure 5: Raincloud plot of read length distribution for six ONT sequencing runs and seven different HMW DNA extraction kits. Each of the 42 graphs show on x-axis: the reads density distribution, the reads length distribution as dot plot and bar plot representing \log_{10} (read length) in bp. Different colours correspond to OsHV-1 reads (dark blue), *M. gigas* reads (mid-dark blue), "OsHV-1 + *M. gigas*" reads (mid-light blue) and "Unaligned" reads (light blue). Coloured dots on bar plot represent N50 values. "Unaligned" corresponds to reads that did not align to either OsHV-1 μ VarA or *M. gigas*. "OsHV-1 + *M. gigas*" corresponds to reads that align on OsHV-1 μ VarA and on *M. gigas*.

Figure 6: Evaluation of *de novo* assembly tools for short and long read sequences. Illumina assembly: Distribution of the reads used for the assembly is shown on the pie chart with *M. gigas* reads in darkblue, Unaligned reads in light blue and OsHV-1 reads in black. MetaSPADES was used to assemble an OsHV-1 reference genome using Illumina short reads. Five contigs were generated and manually scaffolded to produce a non-redundant OsHV-1 genome. ONT assembly: Six different long read *de novo* assemblers (Flye RA, Raven, Smartdenovo, Shasta and Canu) were evaluated with four

different read sets: all non-*M. gigas* reads from all sequenced samples from all six ONT sequencing runs; the top 10%, 1%, and 0.1% longest non-*M. gigas* reads from all sequenced samples from all six ONT sequencing runs. This evaluation produced 24 panels showing the contigs produced by each *de novo* assembler and each dataset. Only the OsHV-1 contigs are visualized and each contig has been annotated for specific OsHV-1 regions: TR_L/IR_L (turquoise); U_L (white); X (blue); TR_S/IR_S (grey); U_S (dark grey).

Figure 7: Distribution of read coverage on the Illumina and ONT assembled genome. Top panel: Distribution of Illumina reads (turquoise) on Illumina assembled contigs (yellow). Bottom panel: Distribution of ONT reads (turquoise) on the ONT-assembled contig (yellow). The genomic architecture of OsHV-1 is shown with the specific regions: TR_L/IR_L (turquoise); U_L (white); X (blue); TR_S/IR_S (grey); U_S (dark grey). The Y-axis represents the reads coverage while the X-axis represents the position of the OsHV-1 genome.

Figure 8: Pairwise alignment of ONT and Illumina assembled genomes. Alignment of the Illumina NR OsHV-1 genome generated by MetaSPADES with the non-*M. gigas* Illumina reads against the ONT OsHV-1 genome generated by Canu with the 0.1% longest non-*M. gigas* ONT reads. These two genomes differ by a total of 15 SNPs compared to a consensus of 186 kb, all located in homopolymers (red triangles). Yellow bars represent the OsHV-1 genome position of these SNPs. Light green bars represent contigs used to generate references and black bars represent nucleic sequences. The OsHV-1 structure is shown with the specific regions: TR_L/IR_L (turquoise); U_L (white); X (blue); TR_S/IR_S (grey); U_S (dark grey).

Figure 9: Evaluation of the minimum ONT read depth required to assemble an accurate OsHV-1 genome. Dot plots of nucleotide identity between the assembled genomes and the Illumina OsHV-1 genome according to coverage. Each blue dot represents an ONT OsHV-1 genome assembled with Canu but not polished. Each green dot represents an ONT OsHV-1 genome assembled with Canu and polished with Medaka. Nucleotide identity (%) and the error per bp between ONT and Illumina OsHV-1 genomes are shown on Y-axes. The number of reads used for the assembly and the corresponding coverage on the Illumina genome are shown on X-axes. The centre panel is an extension of low coverage from 0 to 300. The orange rectangle frames the first ONT assembled and polished genome which is accurate (99.994% nucleotide identity between ONT and Illumina genome) and was generated with 6 061 OsHV-1 reads.

Figure 10: Consensus caller scoring. Pairwise comparison of OsHV-1 consensus sequences generated by five different long reads consensus callers: Clair3; Pepper DeepVariant; Medaka; Nanocaller; and

Longshot. Red lines represent the mean percentage of identity for each of the consensus caller tested. The Y-axis represents percentage of nucleotide identity and the X-axis represents read coverage.

Figure 11: OsHV-1 isomer characterization. Coverage distribution of long reads (> 42kb) overlapping at least part of U_L and U_S with a mapping quality score > 40 against the 4 major isomers (P, I_L, I_S, I_{LS}). Reads are color coded at the base level by similarity from dark blue (100% identity) to red (less than 60% identity).

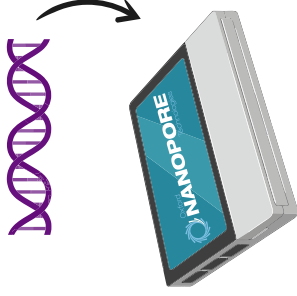
Figure 12: Structural variation. Representation of the 12 structural variations (the X deletion (red), the U_S inversion (purple), the X duplication (green) and the spurious structural variations (blue)) identified by Sniffles2 with all non-*M. gigas* reads. The genomic architecture of OsHV-1 is shown with the specific regions: TR_L/IR_L (turquoise); U_L (white); X (blue); TR_S/IR_S (grey); U_S (dark grey). The Y-axis represents the number of reads supporting SV and the X-axis represents the position of the OsHV-1 genome.

Table 1: Sample description

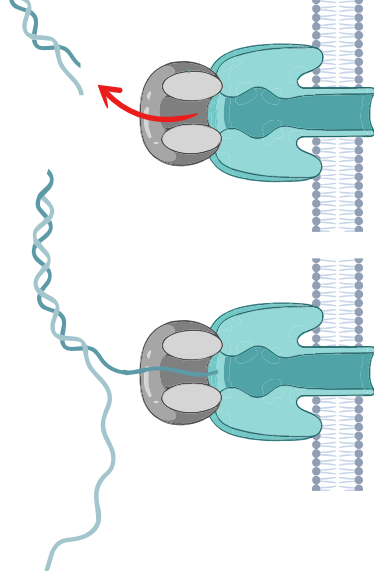
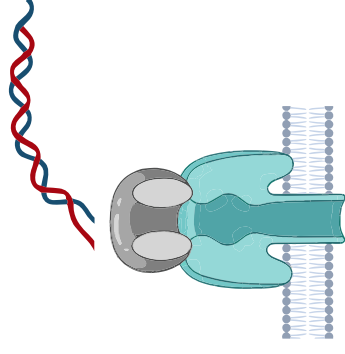
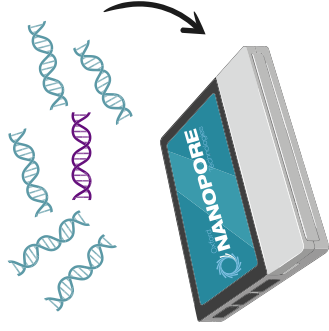
Table 2: Description of the six sequencing runs

Table 3: Expected and obtained DNA yield and fragment size by the different DNA extraction kits

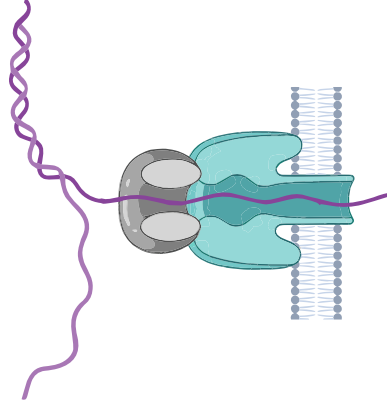
OsHV-1 reference genome



DNA library



Non-OsHV-1 DNA



OsHV-1 DNA

Figure 1: Schematic representation of adaptive sampling with OsHV-1 enrichment

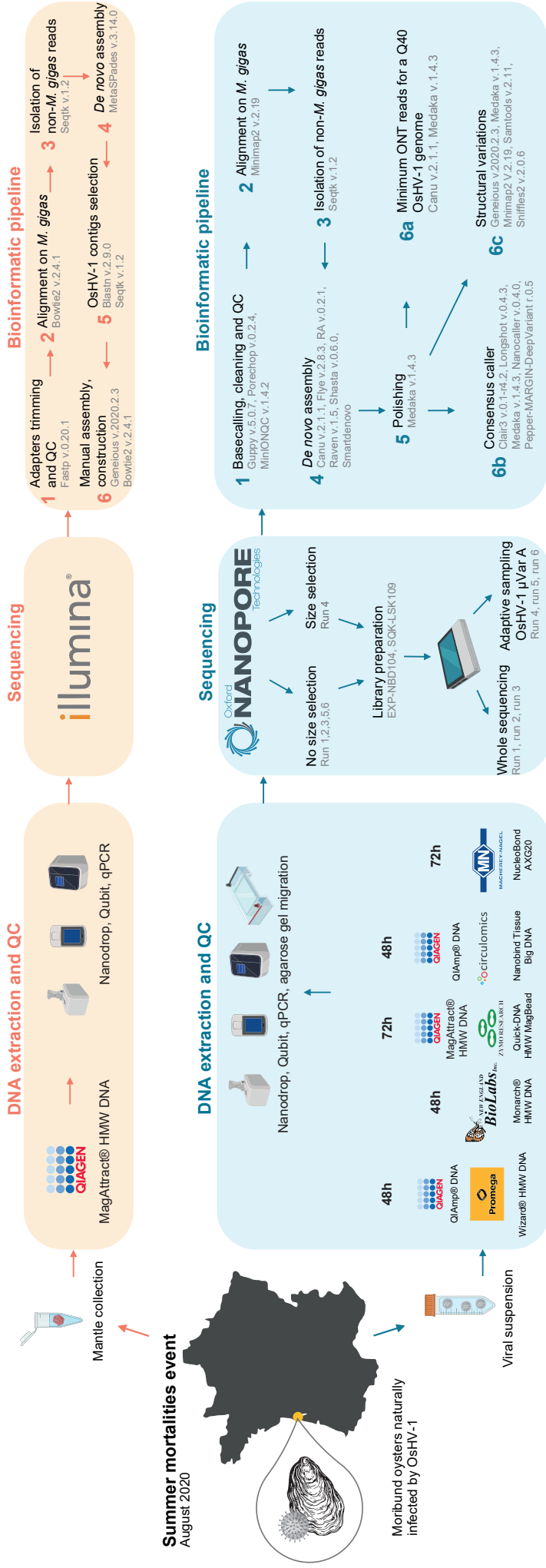


Figure 2: Experimental design and bioinformatic pipeline implemented in this study

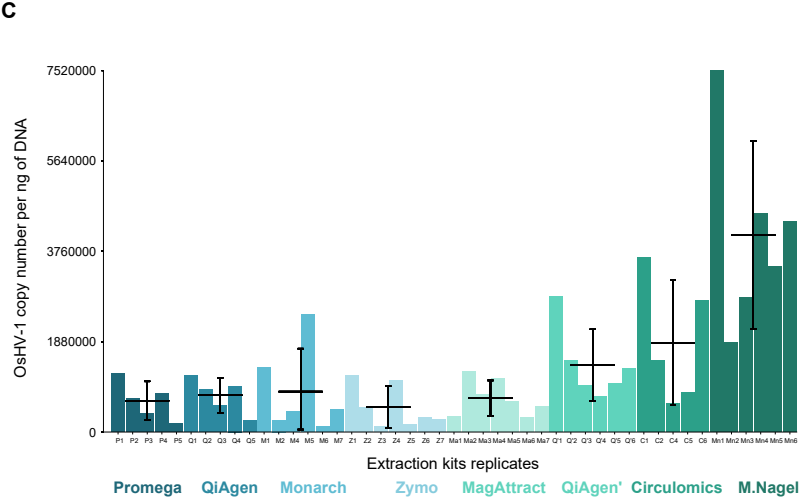
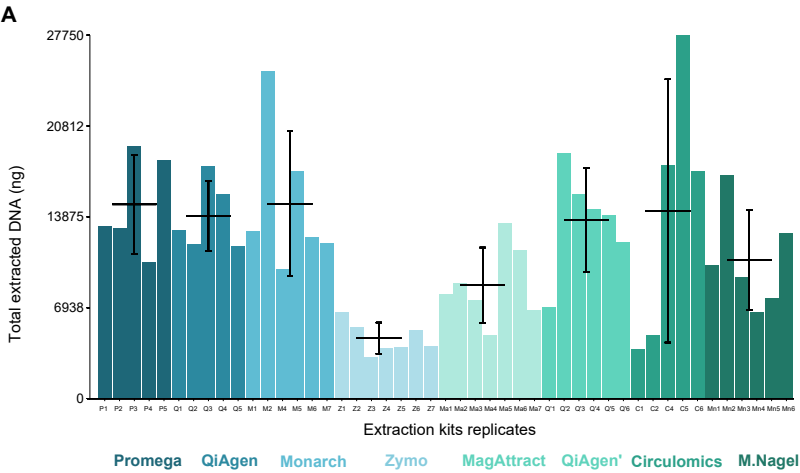


Figure 3: DNA extractions characteristic. A. Total extracted DNA in ng obtained with the seven different HMW DNA extraction kits. The X-axis represents the replicates of the extraction kits (5 replicates for Promega and Qiagen, 6 replicates for Qiagen', Circulomics and Macherey Nagel and 7 replicates for Monarch, Zymo and MagAttract) while the Y-axis represents the total DNA extracted (ng). The horizontal black lines correspond to the mean of extracted DNA for each extraction kit while the vertical black lines represent its standard deviation. B. Agarose gel electrophoresis of genomic DNA isolated from OshV-1-infected oyster tissue using seven different HMW DNA extraction kits. First line of each gel corresponds to DNA ladders ranging from 0.5 kb to 48.5kb. C. OshV-1 genome copy number per ng of extracted DNA for seven different HMW DNA extraction kits. The X-axis represents the replicates of the extraction kits (5 replicates for Promega and Qiagen, 6 replicates for Qiagen', Circulomics and Macherey Nagel and 7 replicates for Monarch, Zymo and MagAttract) while the Y-axis represents the OshV-1 copy number per ng of DNA. The horizontal black lines correspond to the mean of extracted DNA for each extraction kit while the vertical black lines represent its standard deviation.

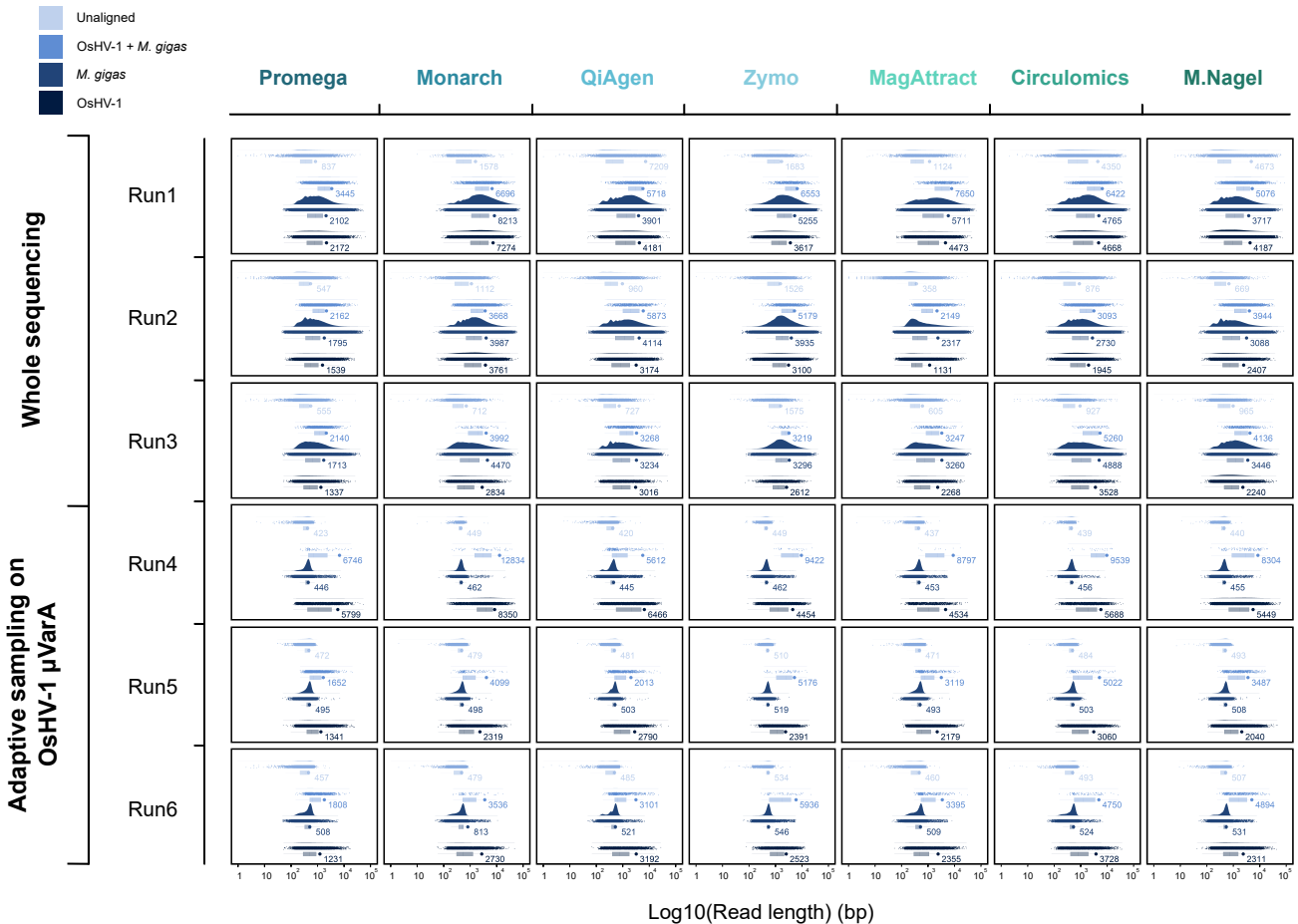


Figure 5: Raincloud plot of read length distribution for six ONT sequencing runs and seven different HMW DNA extraction kits. Each of the 42 graphs show on x-axis: the reads density distribution, the reads length distribution as dot plot and bar plot representing log10 (read length) in bp. Different colours correspond to OsHV-1 reads (dark blue), *M. gigas* reads (mid-dark blue), “OsHV-1 + *M. gigas*” reads (mid-light blue) and “Unaligned” reads (light blue). Coloured dots on bar plot represent N50 values. “Unaligned” corresponds to reads that did not align to either OsHV-1 μ VarA or *M. gigas*. “OsHV-1 + *M. gigas*” corresponds to reads that align on OsHV-1 μ VarA and on *M. gigas*.

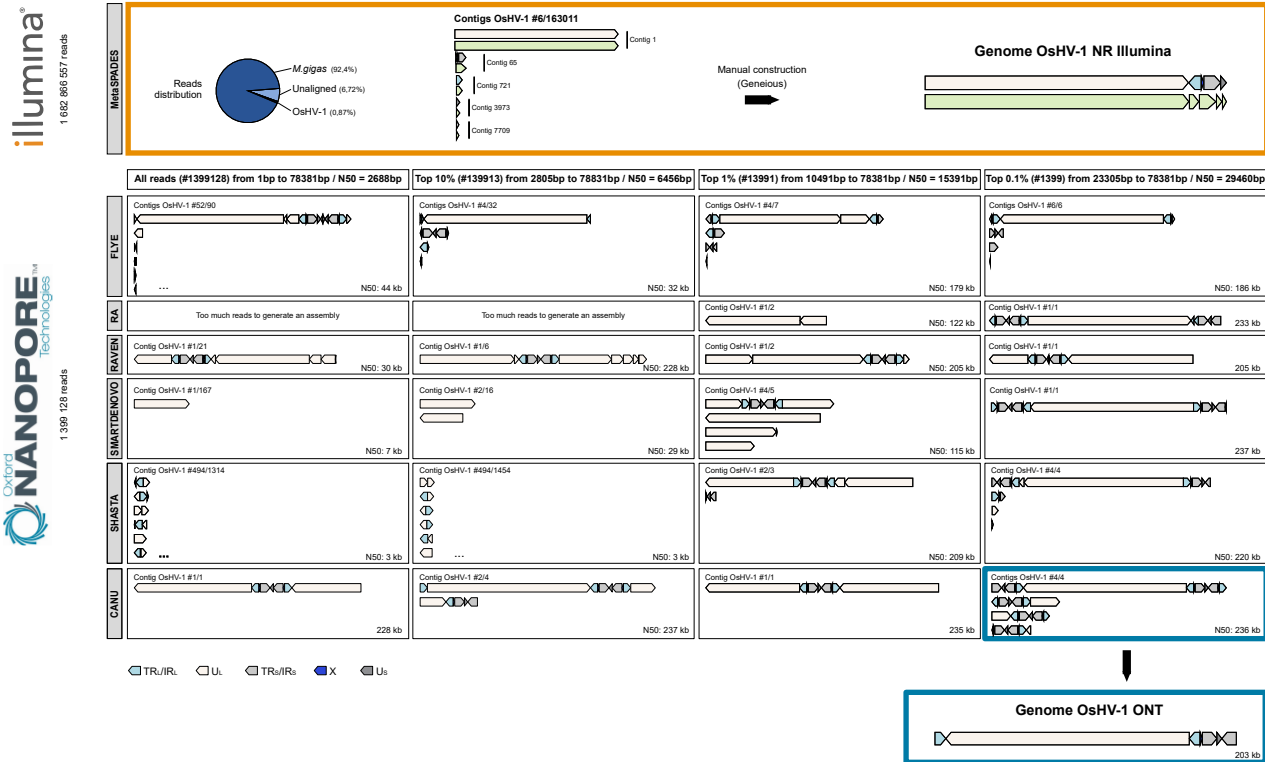


Figure 6: Evaluation of de novo assembly tools for short and long read sequences. Illumina assembly: Distribution of the reads used for the assembly is shown on the pie chart with *M. gigas* reads in darkblue, Unaligned reads in light blue and OshV-1 reads in black. MetaSPADES was used to assemble an OshV-1 reference genome using Illumina short reads. Five contigs were generated and manually scaffolded to produce a non-redundant OshV-1 genome. ONT assembly: Six different long read *de novo* assemblers (Flye, RA, Raven, Smartdenovo, Shasta and Canu) were evaluated with four different read sets: all non-*M. gigas* reads from all sequenced samples from all six ONT sequencing runs; the top 10%, 1%, and 0,1% longest non-*M. gigas* reads from all sequenced samples from all six ONT sequencing runs. This evaluation produced 24 panels showing the contigs produced by each de novo assembler and each dataset. Only the OshV-1 contigs are visualized and each contig has been annotated for specific OshV-1 regions: TRL/IRL (turquoise); UL (white); X (blue); TRS/IRS (grey); US (dark grey).

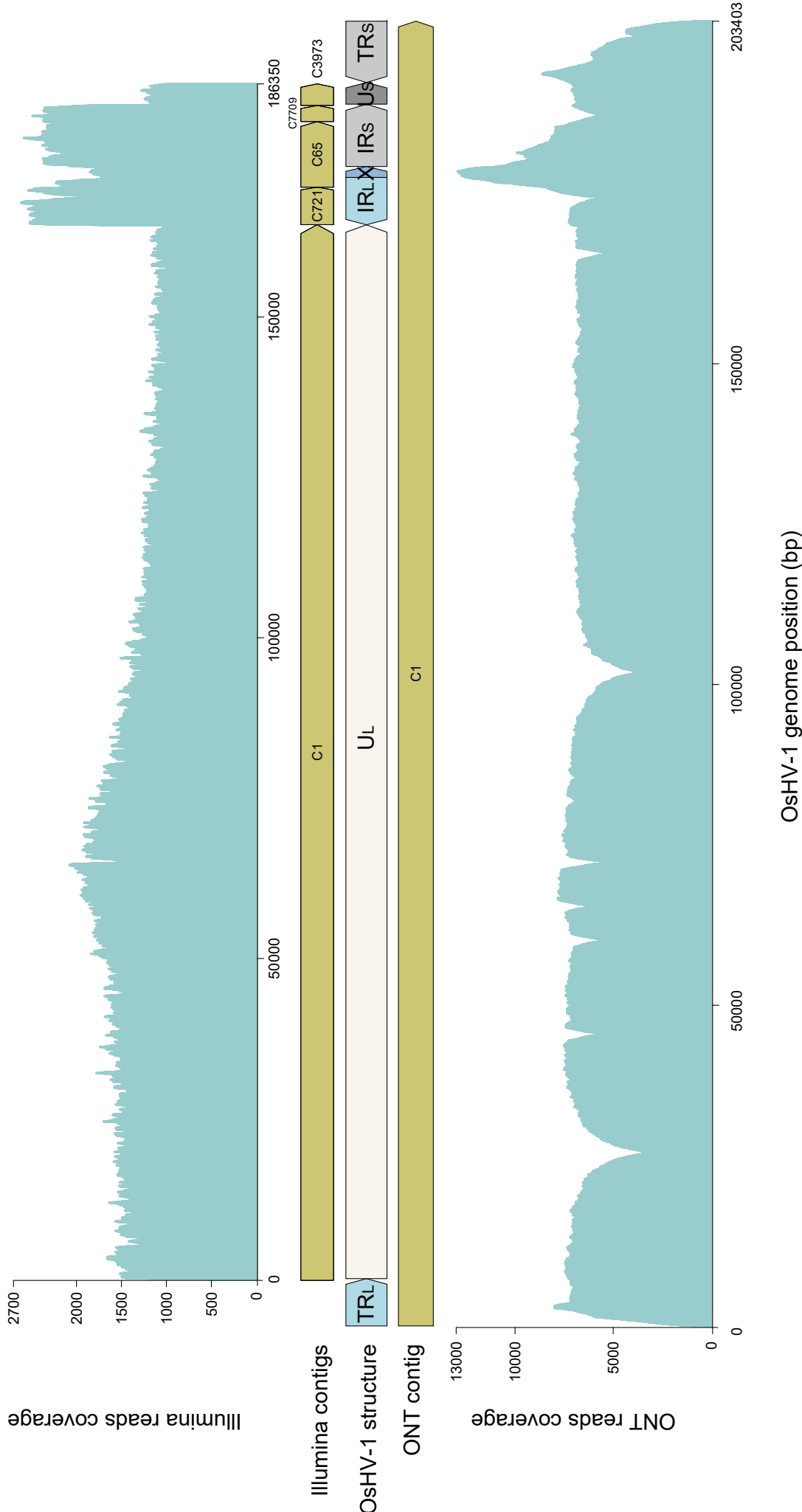


Figure 7: Distribution of read coverage on the Illumina and ONT assembled genome. Top panel: Distribution of Illumina reads (turquoise) on Illumina assembled contigs (yellow). Bottom panel: Distribution of ONT reads (turquoise) on the ONT-assembled contig (yellow). The genomic architecture of OsHVV-1 is shown with the specific regions: TRL/IRL (turquoise); UL (white); X (blue); TRS/IRS (grey); US (dark grey). The y-axis represents the reads coverage while the x-axis represents the position of the OsHVV-1 genome.

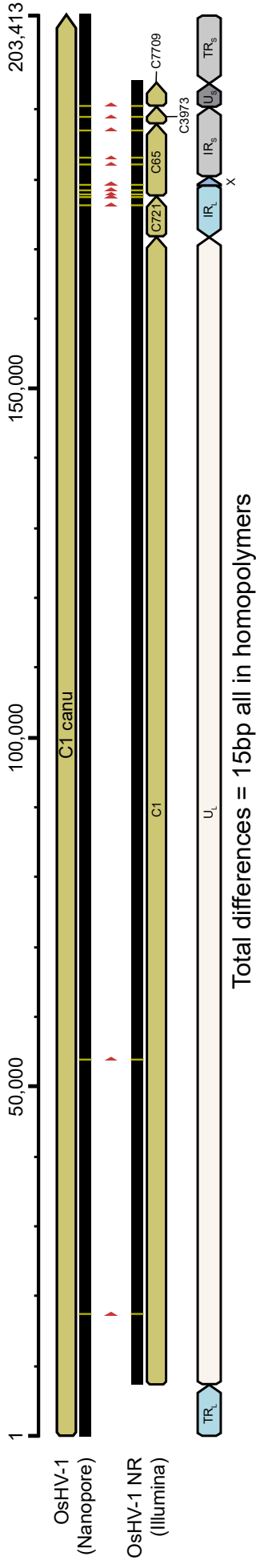


Figure 8: Pairwise alignment of ONT and Illumina assembled genomes. Alignment of the Illumina NR OshV-1 genome generated by MetaSPADES with the non-*M. gigas* Illumina reads against the ONT OshV-1 genome generated by Canu with the 0.1% longest non-*M. gigas* ONT reads. These two genomes differ by a total of 15 SNPs compared to a consensus of 186 kb, all located in homopolymers (red triangles). Yellow bars represent the OshV-1 genome position of these SNPs. Light green bars represent contigs used to generate references and black bars represent nucleic sequences. The OshV-1 structure is shown with the specific regions: TRL/IRL (turquoise); UL (white); X (blue); TRS/IRS (grey); US (dark grey).

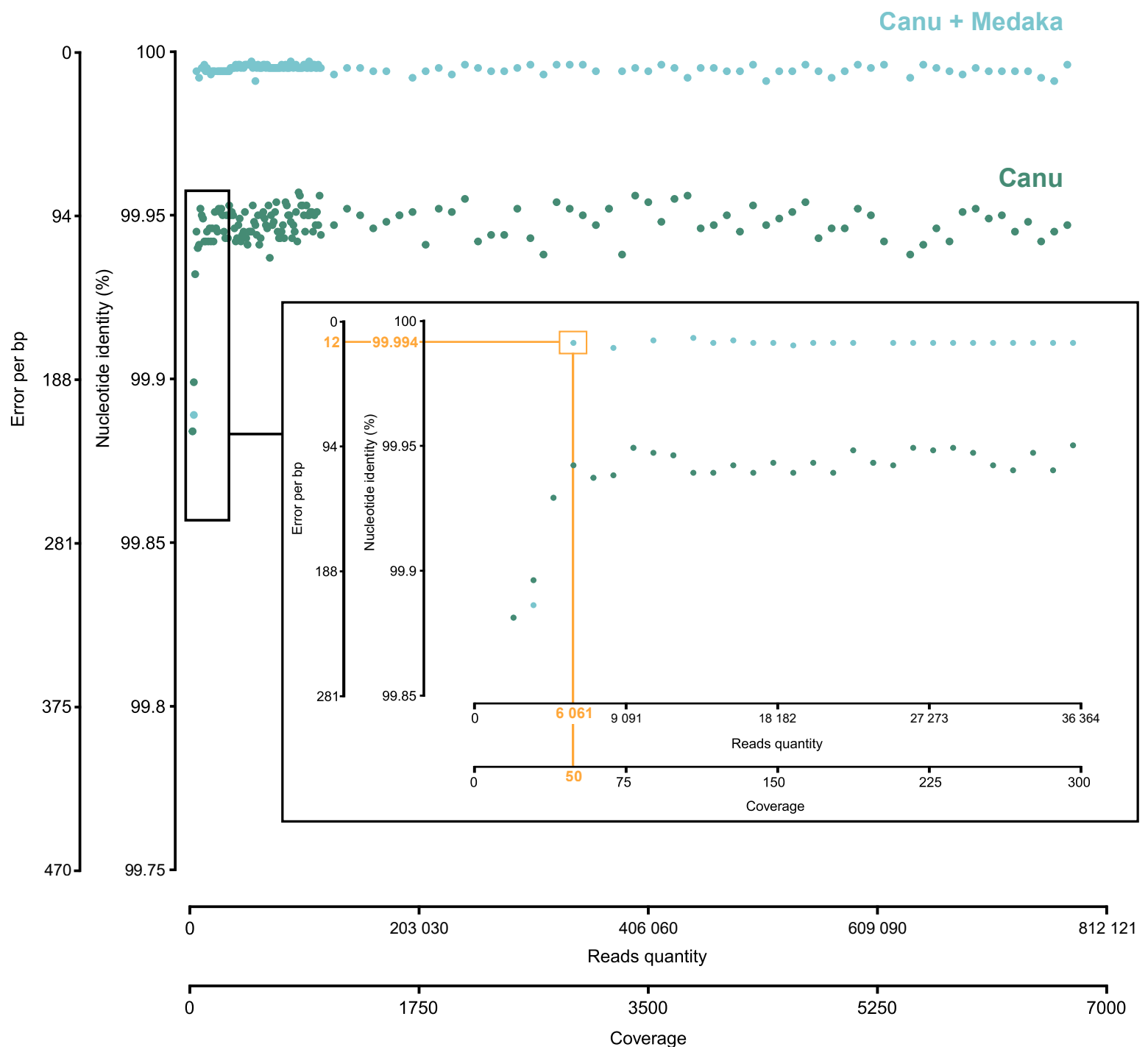


Figure 9: Evaluation of the minimum ONT read depth required to assemble an accurate OshV-1 genome. Dot plots of nucleotide identity between the assembled genomes and the Illumina OshV-1 genome according to coverage. Each blue dot represents an ONT OshV-1 genome assembled with Canu but not polished. Each green dot represents an ONT OshV-1 genome assembled with Canu and polished with Medaka. Nucleotide identity (%) and the error per bp between ONT and Illumina OshV-1 genomes are shown on Y-axes. The number of reads used for the assembly and the corresponding coverage on the Illumina genome are shown on X-axes. The centre panel is an extension of low coverage from 0 to 300. The orange rectangle frames the first ONT assembled and polished genome which is accurate (99.994% nucleotide identity between ONT and Illumina genome) and was generated with 6 061 OshV-1 reads.

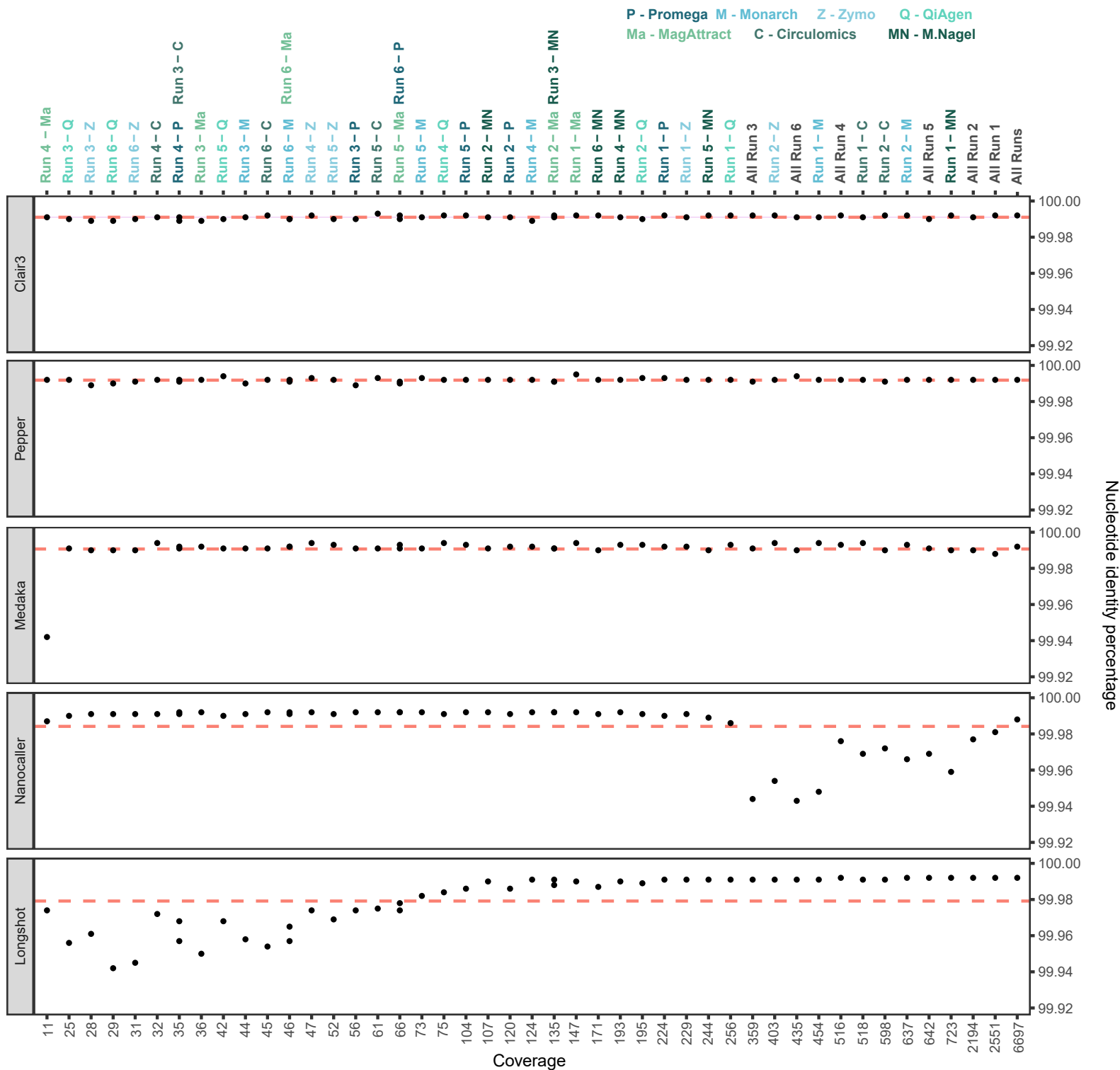


Figure 10: Consensus caller scoring. Pairwise comparison of OsHV-1 consensus sequences generated by five different long reads consensus callers: Clair3; Pepper DeepVariant; Medaka; Nanocaller; and Longshot. Red lines represent the mean percentage of identity for each of the consensus caller tested. The Y-axis represents percentage of nucleotide identity and the X-axis represents read coverage.

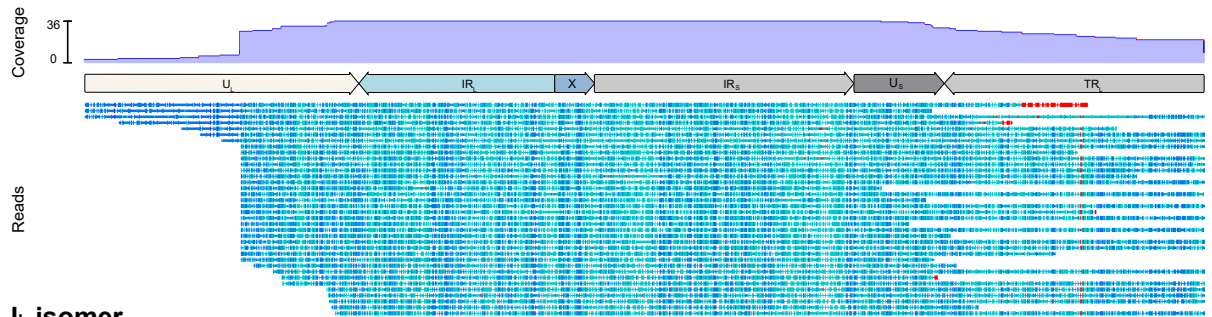
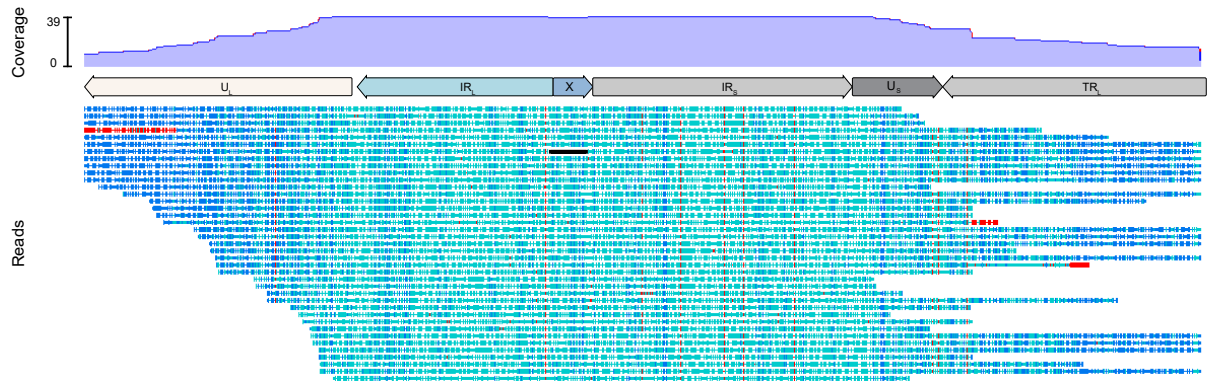
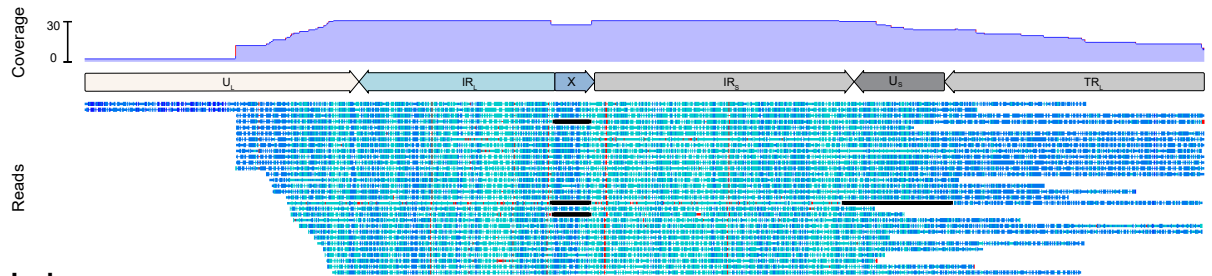
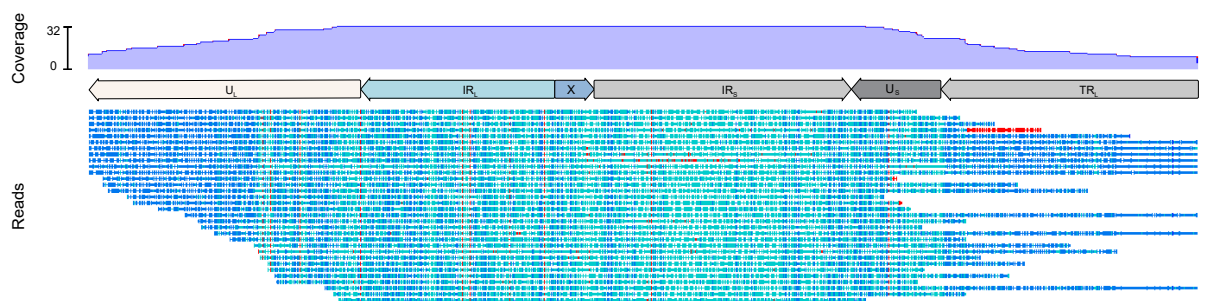
P isomer**IL isomer****IS isomer****ILs isomer**

Figure 11: OsHV-1 isomer characterization. Coverage distribution of long reads (> 42kb) overlapping at least part of UL and US with a mapping quality score > 40 against the 4 major isomers (P, IL, IS, ILs). Reads are color coded at the base level by similarity from dark blue (100% identity) to red (less than 60% identity).

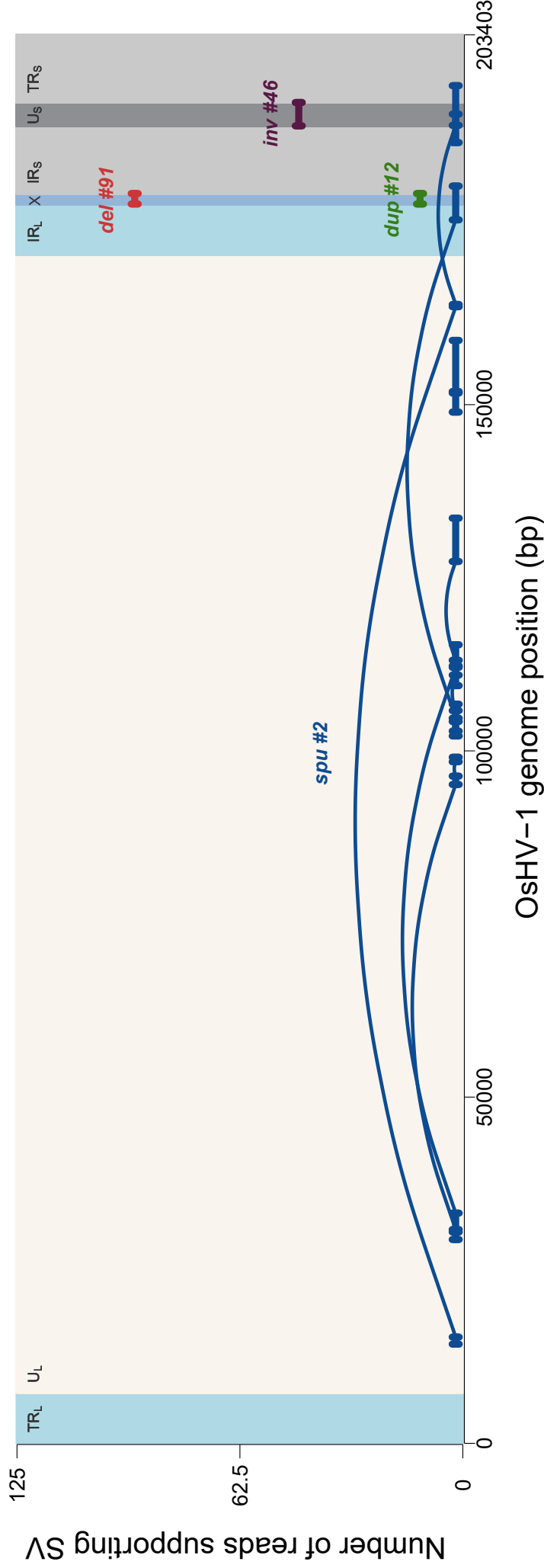


Figure 12: Structural variation. Representation of the 12 structural variations (the X deletion (red), the US inversion (purple), the X duplication (green) and the spurious structural variations (blue)) identified by Sniffles2 with all non-*M. gigas* reads. The genomic architecture of OsHV-1 is shown with the specific regions: TRL/IRL (turquoise); UL (white); X (blue); TRS/IRS (grey); US (dark grey). The Y-axis represents the number of reads supporting SV and the X-axis represents the position of the OsHV-1 genome.

Table 1: Sample description

Biological material	OsHV-1 infection	Nb. of oysters sampled	DNA extraction kit	Sequencing technology
highly susceptible standardized oyster spat	natural	1	MagAttract	Illumina
naive oysters	injection	5	Promega (P) & QiAgen (Q)	ONT
naive oysters	injection	7	Monarch (M)	ONT
naive oysters	injection	7	Zymo (Z)	ONT
naive oysters	injection	7	MagAttract (Ma)	ONT
naive oysters	injection	6	Circulomics (C) & QiAgen (Q')	ONT
naive oysters	injection	6	Macherey Nagel (Mn)	ONT

Table 2: Description of the runs performed in this study

	Duration (h)	Flowcell	Available Pores	Samples	Adaptive Sampling
Run1	48	used	990	P1, M5, Q'1, Z1, Ma2, C1, Mn1	no
Run 2	39	new	1488	P3, M5, Q'5, Z4, Ma7, C6, Mn5	no
Run 3	26	used	189	P2, M4, Q3, Z7, Ma5, C4, Mn3	no
Run 4	16	new	1352	P3, M5, Q3, pool of Z1 and Z2, Ma6, C6, Mn6	yes
Run 5	40	used	341	P2, M4, Q3, Z7, Ma5, C4, Mn3	yes
Run 6	4	used	1054	P2, M4, Q3, Z7, Ma5, C4, Mn3	yes

Table 3: Expected and obtained DNA yield and fragment size by the different DNA extraction kits

Expected			<i>M. gigas</i> tissue	
Kits	DNA yield	DNA fragments size	DNA yield	DNA fragments estimation
Promega	1,7 μ g for Spinach Tissue	64% > 50 kpb	10 to 20 μ g	> 48.5 kb --
Monarch	8,9 to 20,8 μ g	50 to > 500 kpb	9 to 25 μ g	> 48.5 kb ++
QiAgen	5 to 30 μ g	up to 50 kpb	7 to 19 μ g	> 48.5 kb --
Zymo	10 μ g per 50 μ L of beads	up to 150 kpb	3 to 7 μ g	> 48.5 kb +
MagAttract	12 to 70 μ g	up to 150 kpb	4 to 13 μ g	> 48.5 kb +
Circulomics	5 to 100 μ g	50 to > 300 kpb	3 to 28 μ g	> 48.5 kb ++
Macherey Nagel	20 μ g	0,5 to > 300 kpb	6 to 17 μ g	> 48.5 kb ++