Viral RNA & DNA Genome Characterization via Newly Developed Oxford Nanopore Sequencing Protocol

Prepared by: William Gardner, Research Specialist, and Shelby O'Connor, Professor

October 2024

UW - Madison WNPRC 555 Science Drive Madison, WI 53711 T (608) 890-0843 F (608) 265-8084
soconnor@primate.wisc.edu

Introduction

The Viral Pathogen Sequencing Unit offers a complimentary service designed to enhance research rigor and reproducibility. [1] Our viral stock verification process allows researchers to confidently proceed with infection studies, ensuring the identity and integrity of both established and newly prepared virus samples. To sequence such a wide diversity of viruses, we use an unbiased amplification approach. This method employs random hexamers to create double stranded viral cDNA (dsDNA) to sequence virus inoculum without any prior knowledge of the virus sequence.

Our firm belief is that sequencing viral inoculum for use in animal studies is essential. By agreeing to this free service, we will automatically publish the inoculum sequences in NCBI's Sequence Read Archive (SRA). By making this sequence available, we enhance the quality of viral research across the scientific community, enabling researchers to conduct more informed and precise studies of viral evolution and pathogenesis in their animal models.

To sequence such a wide diversity of viruses we use an unbiased amplification approach. This method employs random hexamers to create double stranded viral cDNA (dsDNA) to sequence virus inoculum without any prior knowledge of the virus sequence. These cDNAs can be sequenced by different methods. We have used Illumina's Nextera tagmentation method for its exceptional quality scores and depth of sequencing. We recently upgraded this service to take advantage of Oxford Nanopore Technology's (ONT) long-read sequencing. This advancement brings multiple benefits:

- Faster results
- Lower operational costs
- Improved detection of structural variants

Our Illumina Nextera service will continue to be offered, but please note that free viral inoculum sequencing with Illumina methods will experience a delay of up to 10 weeks due to multiplexing needs for cost efficiency. In direct comparison, ONT long-read sequencing provides faster results with lower costs for small sample sequencing and ensures a quick turnaround time for inoculum sequencing. We remain committed to offering both options depending on your research needs.

Viral RNA extraction

At our lab we can process samples from varying starting materials. We can isolate viral RNA and DNA from plasma samples or cell supernatants using a Maxwell® RSC 48 Instrument and Viral Nucleic Acid Isolation kit. If you wish to isolate your own RNA or prepare your own DNA amplicons, you may send those samples to us, and we will proceed from there.

Starting material for viral RNA isolation

If you wish to send us plasma or cell supernatant, we require 0.2-1.0 mL of sample to be shipped to Genomics Services. Please ship samples on dry ice to minimize the degradation of RNA from the starting material during shipping. Best results are obtained if the sample is promptly stored at -80°C upon isolation and never thawed.

How we perform “Unbiased” Amplification

We use an approach called Sequence Independent Single Primer Amplification (SISPA) that uses 2 steps of priming, first a set of random hexamers with a 5’ tag binds to many regions across the DNA/RNA strand. Then a second set of primers complementary to the tag will amplify the segments of tagged DNA/RNA. This provides a way to acquire DNA in an unbiased way and eliminates the need for target primers. SISPA requires an input that is single stranded so ssDNA viruses may also be amplified using this technique such as parvoviruses. Our amplified product can then be used with ONT or Illumina based sequencing protocols.

Oxford Nanopore MinION Sequencing Protocol

ONT Sample Preparation

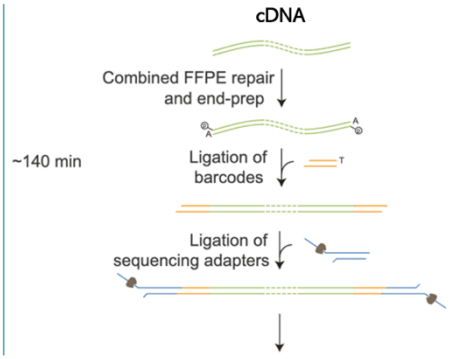

Figure 1. Native Barcoding workflow. Credit: Oxford Nanopore Technologies

Proceeding from the unbiased amplification step we use Oxford Nanopore’s Native Barcoding Kit 24 V14 (Figure 1) to prep our DNA for nanopore sequencing. DNA input sizes for nanopore sequencing are not strict and allow us to process DNA fragments of all sizes. We first perform DNA repair with New England Biolab’s NEBNext® FFPE DNA Repair Mix to fix any damage to the DNA that took place during freeze thaw cycles. We then perform end prep with the NEBNext® Ultra™ II End Repair/dA-Tailing Module. Barcodes from the Native Barcoding kit are ligated onto each sample Finally, adapters are added to the ends of each DNA fragment to allow tethering of DNA molecules to the nanopore.

Nanopore Sequencing on ONT MinION

Barcoded and adapter ligated samples are loaded onto our MinION sequencer on a Flongle flow cell. The relatively inexpensive cost of an individual Flongle flow cell allows us to sequence individual samples and yield very high sequencing depth. One Flongle can output a maximum of 2.6 Gbases of read data in 24 hours which can generate 260,000x coverage of a Zika virus genome (10kb). Nanopores are embedded in a polymer membrane with high electrical resistance. As DNA fragments pass through these nanopore channels, the nucleotides disrupt the ionic flow, causing measurable changes in electrical current. These disruptions are recorded as a continuous signal, which is then processed by ONT’s MinKNOW™ software and stored in POD5 signal files. ONT’s base calling program Dorado is

used to transform these POD5 files into useable fastq files and greatly improves the default “fast” base calling accuracy. From an example run with three samples of Zika virus stocks multiplexed on a Flongle, we yielded 217K reads with an average read length of 724 bp. ONT’s Super Accurate base caller greatly improves the quality of sequencing reads. Our average read quality improved from 13 to 35 (Figure 2). We achieved over 1000X sequencing depth with relatively equal coverage on our samples allowing for low frequency variant calling. (Figure 3).

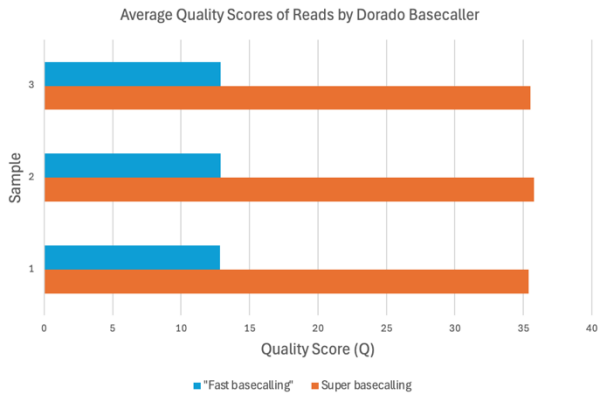

Figure 2. Example unbiased amplification and ONT Native Barcoded sequencing of three Zika virus samples’ run quality scores between fast and Super Accurate base calling models.

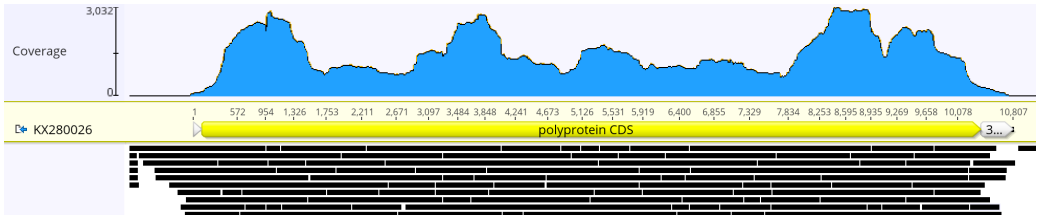

Figure 3. Sequencing Alignment Map of 1 Zika Virus sample mapped to Zika virus Paraiba isolate KX280026.1.

Analysis of data from an ONT Flongle run

First, we take the raw ONT signal files (POD5) and use ONT's advanced machine learning algorithm, Dorado, for enhanced base-calling accuracy. Using chopper (https://github.com/wdecoster/chopper) we then remove low quality reads and reads smaller than 100 base pairs. Unbiased sequencing results in a lot of host read contaminants (Figure 4) that we remove using minimap2 (https://github.com/lh3/minimap2) and taxon identification of remaining reads is performed using Kraken2 (https://github.com/DerrickWood/kraken2). If no reference sequence is provided, we will return multiple alignments to the highest taxonomic matches. BAM alignments will be sent back to researchers and submitted to NCBI’s Sequence Read Archive (SRA). Our custom metagenomic pipeline can take raw POD5 files and convert them into useable alignment files in a matter of minutes (Figure 5).

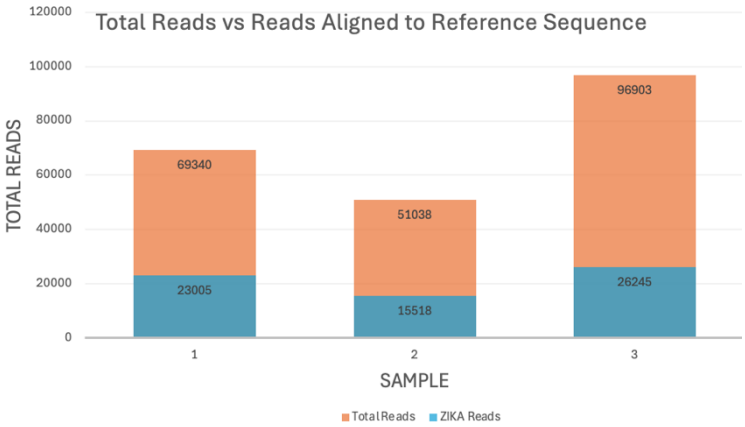

Figure 4. Total reads from each sample vs the number of reads aligned to the reference genome.

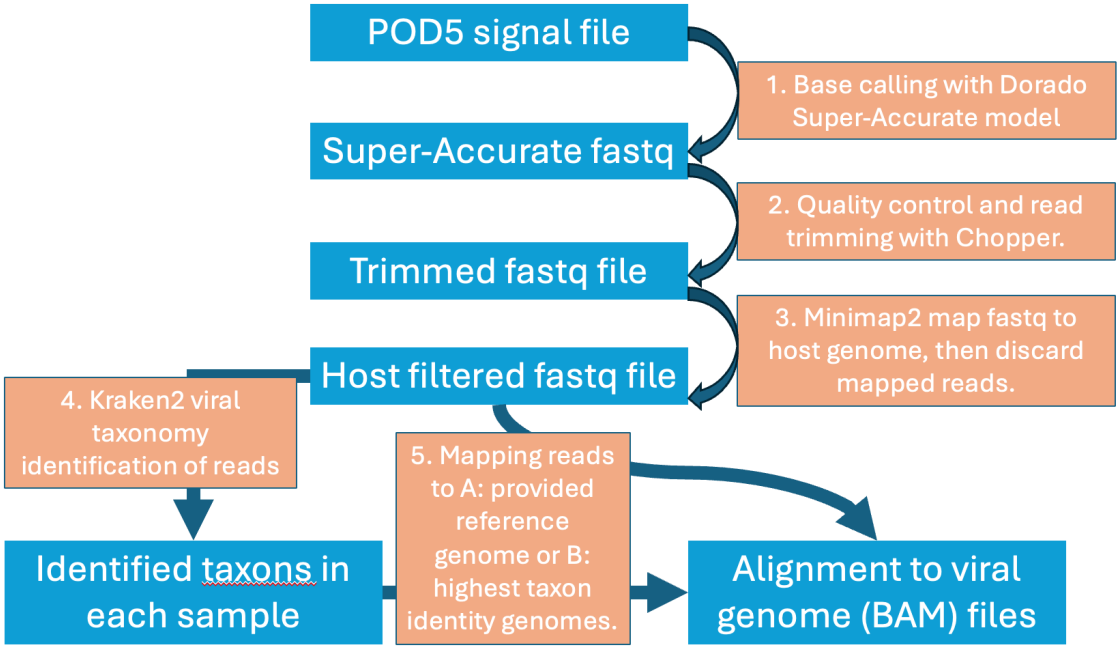

Figure 5. ONT output alignment generation pipeline.

Illumina MiSeq Sequencing Protocol

Nextera™ XT DNA Sample Preparation

Using the transposon-based Nextera™ XT DNA Sample Prep Kit, the purified SISPA cDNA is fragmented. This reaction (termed “tagmentation”) requires 1 ng of DNA and is used to create DNA fragments of varying sizes. Samples are diluted to 5 ng of total DNA prior to adding the Tagment DNA Enzyme. By customizing enzyme concentrations and reaction times we can achieve around 400-600 bp fragments from libraries of varying input sizes (For further information, refer to Illumina's website.) Following completion of a limited-cycle PCR that attaches Illumina-compatible sequencing primers as well as indexes, samples are purified using the Agencourt AMPure XP system. Samples are quantified using the Qubit™ dsDNA HS Assay Kit, and average fragment size is determined using the Agilent High Sensitivity DNA Kit. All uniquely tagged samples are then pooled and loaded onto the MiSeq.

Sequencing-by-synthesis on the Illumina MiSeq

The MiSeq performs sequencing by using fluorescently labeled nucleotides that sequentially build complementary strands of DNA, with each base incorporation triggering the release of a unique fluorescent signal that is detected and recorded to determine the sequence. Our preferred sequencing kit on the MiSeq platform for inoculum sequencing is the MiSeq Reagent Nano Kit v2. The kit offers a maximum output of 500Mb and has the option of either 2x250bp or 2x150 bp paired reads. Generally, we multiplex a minimum of 10 viral genomes per sequencing cartridge to minimize total costs. If our viruses of interest are 10kb in length and we obtain the maximum output we would have 50,000X genome coverage. We reliably obtain very deep coverage of viral genomes when using this process.

Analysis of data from a MiSeq run

After images are processed and bases are called, we bin the sequences by sample indices. We use an almost identical analysis pipeline to our ONT metagenomic pipeline, except that read merging and short read specific quality control uses bbtools. The final BAM files are sent back to researchers and submitted to the NCBI SRA.

References

1. National Institutes of Health. (2024). *Guidance: Rigor and Reproducibility in Grant Applications.* https://grants.nih.gov/policy-and-compliance/policy-topics/reproducibility/guidance

2. Oxford Nanopore Technologies. (2024). *How nanopore sequencing works.* https://nanoporetech.com/platform/technology

3. Oxford Nanopore Technologies. (2024). *Dorado*. https://github.com/nanoporetech/dorado

4. Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken2. Genome Biol 20, 257 (2019). https://doi.org/10.1186/s13059-019-1891-0

5. Illumina Inc. (2024) *Nextera XT DNA Library Prep Kit Support Resources.* https://support.illumina.com/sequencing/sequencing_kits/nextera_xt_dna_kit.html