Viral RNA Genome Characterization via Illumina MiSeq Technology

Prepared by: Dane Gellerup, Research Specialist, and Shelby O'Connor, Assistant Professor

May 2021
Viral RNA Genome Characterization with Illumina MiSeq Technology

Introduction

The Viral Sequencing Unit of the WNPRC Genomics Services specializes in RNA viral genome sequencing with an emphasis on simian immunodeficiency viruses. Although we specialize in RNA viral genome sequencing, we are continuously developing the tools to sequence additional pathogens.

The techniques for viral sequencing have advanced significantly during the last decade. Illumina sequencing technologies offer many advantages when characterizing a population of viruses. In our lab we house an Illumina MiSeq and have the tools to sequence the entire coding sequence of RNA viruses, shorter amplicons, and plasmids. We can produce this data from various starting materials, including plasma, cell culture, RNA, and DNA.

On the Illumina MiSeq we can multiplex sequences from an average of 25 different RNA virus populations. The ability to multiplex this amount of genomes on one run is what keeps the cost of deep sequencing virus populations low. A single run of the Illumina MiSeq generates an average of 25,000,000 reads, such that we obtain about 1,000,000 reads per virus population. Our requirements for sequencing RNA viruses as fee-for-service work are listed below.

Illumina viral RNA genome sequencing process

At our lab we are able to process samples from varying starting materials. We can isolate viral RNA from plasma samples and then perform RT-PCR to get cDNA. Alternately, we can use random hexamers to produce dsDNA from viral RNA. Also, 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. Depending on what type of starting material you send us, the cost of sequencing may be less expensive. A breakdown of costs is available upon request.

Starting material for viral RNA isolation

If you wish to send us plasma we require 0.2-1.0 mL plasma/EDTA samples to be shipped to Genomics Services, ideally containing >1000 copies of virus template. Please ship plasma samples on dry ice to minimize degradation of RNA from the starting material during shipping. Best results are obtained if the plasma was promptly stored at -80°C upon
Isolation and never thawed. If you choose to produce your RNA or DNA yourself, please ship it on dry ice as well, and include information about each sample (concentrations of DNA, etc.).

Isolation of RNA from plasma and RT-PCR amplification of SIV cDNA

We isolate viral RNA using the QiAamp MinElute Virus Spin Kit (Qiagen). For SIV sequencing, we generate viral cDNA using the Superscript® III One-Step RT-PCR System with Platinum® Taq High Fidelity by Invitrogen. With sequence specific primers, we generate four overlapping ~2.5 kb amplicons that span the entire SIV coding sequence. The amplified products are run on a 1% agarose gel and purified using Qiagen's MinElute Gel Extraction Kit. For samples with particularly low viral loads (<1000 copies/mL), we may opt to use the Agencourt Ampure Xp (Beckman Coulter) system to purify amplicons. Samples are quantified using Invitrogen’s Qubit™ dsDNA HS Assay Kit.

Isolation of RNA from plasma and unbiased synthesis of SIV dsDNA with random hexamers

Oftentimes researchers infect animals with virus inocula without knowing the sequence of the input virus. We can sequence virus inocula without any a priori knowledge of the virus sequence. We do this using a method that employs random hexamers to create double stranded viral cDNA (dsDNA). This provides a way to acquire DNA in an unbiased way and eliminates the need for specific primers. The RNA is isolated in the same way as described above with the QiAamp MinElute Virus Spin Kit (Qiagen), but then we perform various steps using random hexamers, first- and second-strand dsDNA reagents, and limited runs on a thermocycler to create dsDNA.

Nextera™ XT DNA Sample Preparation

Using the transposon-based Nextera™ XT DNA Sample Prep Kit, the purified RT-PCR/random hexamer products are fragmented. This reaction (termed “tagmentation”) requires 1 ng of DNA and yields approximately 400-600 bp fragments. To ensure balanced coverage of the viral genome, we pool approximately 0.25 ng of each RT-PCR amplicon into one tube prior to adding the Tagment DNA Enzyme. (For further information, refer to Illumina’s website.) Following completion of a limited-cycle PCR that attaches Illumina- compatible sequencing primer sequences as well as index sequences, samples are purified using the Agencourt AMPure XP system. Samples are quantified again 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
Each pool is denatured and sequenced as detailed on the Illumina website: (http://www.illumina.com/systems/miseq.ilmn). Generally, we multiplex 20-30 viral genomes per sequencing run on either a 2x250 or 2x300 cartridge. This yields approximately 25,000,000 paired-end sequencing reads (600 bp in length) per sequencing run, which results in about 1,000,000 reads per genome. We reliably obtain complete coverage of the RNA genome by virtue of paired-end reads and large volumes of data.

### Average yields

<table>
<thead>
<tr>
<th>Average Total Reads</th>
<th>Avg Total Reads/ Virus (Pool of 25)</th>
<th>Read Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>~25,000,000</td>
<td>~1,000,000</td>
<td>~400-600 bp</td>
</tr>
</tbody>
</table>

The averages listed here reflect a typical sequencing run; not all samples are equally successful, but our average of 25 genomes per run ensures that there is adequate coverage for all genomes.

### Analysis of data from a MiSeq run

After images are processed and bases are called, we bin the sequences by sample indices. In all cases, we can provide fastq files for each read. If a reference genome is available, we can also generate BAM or SAM alignments for you. Additional data analyses may be performed after discussions with Genomics Services.

### Costs for SIVmac239 viral sequencing

Sequencing costs per animal can be found on the WNPRC Genomics Services Home Page. Direct cost recovery is used to set the base rate for Wisconsin National Primate Research Center clients (Tier 1). For clients outside of the WNPRC, we must also recover indirect costs (facilities and administrative overhead, etc.). As of May 1st, 2021 the rates for these indirect costs are 31% for federally funded clients (Tier 2); 55.5% for foreign universities and non-profits (such as foreign NIH equivalents, e.g. the UK’s NHS) and US for-profit entities using federal funds (Tier 2a); and 72.5% for non-federally funded clients (Tier 3). A more detailed breakdown of costs is available upon request.