Viral RNA Genome Characterization via Illumina MiSeq Technology

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

August 2016

Revised by: Corina Valencia, Research Specialist, and Shelby O'Connor, Professor

May 2022

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

Viral RNA Genome Characterization with Illumina MiSeq Technology

Introduction

The Pathogen Sequencing Unit of the WNPRC Genomics Services specializes in viral RNA genome sequencing on a variety of pathogens including simian immunodeficiency virus (SIV), SARS-CoV-2 virus, and Zika virus. As medicine, community health needs, and research priorities are evolving in response to the COVID-19 pandemic, Zika virus outbreaks, and other emerging infectious diseases, we are continuously developing and adapting the tools to sequence RNA viruses that are the cause of diverse infectious diseases.

The techniques for viral sequencing have advanced significantly during the last decade. Illumina sequencing technologies offer many advantages when characterizing a population of viruses including, but not limited to, advanced insight on viral mechanics and pathology, mutation identification, and variant tracking. We use Illumina technology to sequence the entire coding sequence of RNA viruses, shorter amplicons, and plasmids. We can produce this data from various starting materials, including cell culture, plasma, PCR products (amplicons), RNA, and DNA.

Much of our sequencing is performed with an in-house Illumina MiSeq, although we have access to higher throughput instruments, such as the NovaSeq. The Illumina MiSeq can generate up to 15 Gb of output with up to 25 million sequencing reads. A variety of cartridges can be run on this instrument up to a maximum of 2 x 300 bp read lengths. There is flexibility in sequencing using the MiSeq Reagent Kit with diverse cartridge styles and flow cell sizing. Further details on costs for sequencing RNA viruses as a fee-for-service work on the Illumina MiSeq are listed below.

Viral RNA Genome Sequencing with Illumina

General Workflow for Sequencing

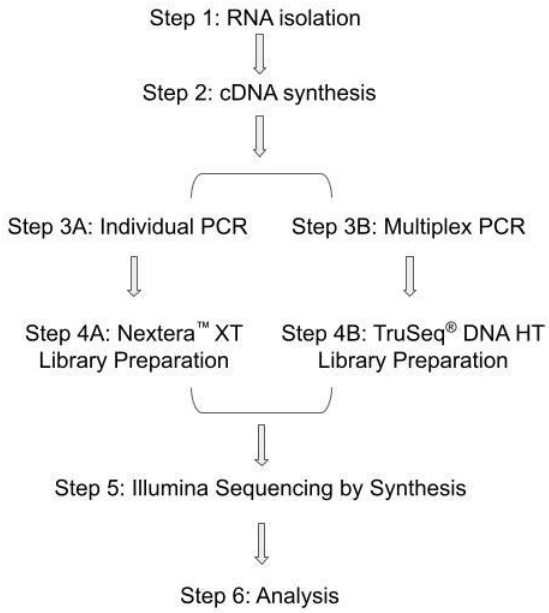

Figure 1. General workflow for viral RNA Illumina sequencing. Viral RNA is isolated then processed into cDNA. Then PCR amplification occurs using either individual PCR or multiplex PCR. Following the generation of PCR products, Nextera™ or TruSeq® library preparation is performed. Library-prepared samples are pooled and normalized for sequencing on the MiSeq. Analysis can then be performed with the MiSeq data.

Starting materials for viral RNA sequencing

Our lab is able to process samples for sequencing from different starting materials. We can receive cell culture fluid, plasma, plasmids, RNA, DNA, and PCR products (amplicons). 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.

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 own RNA, DNA, or PCR amplicons yourself, please ship it on dry ice as well, and include information about each sample (tube labeling, dates, sample description, concentrations of DNA, etc.).

Sequencing SIV genomes with RT-PCR and Nextera™ XT Library Prep

We begin sequencing SIV genomes by isolating viral RNA using the Maxwell® RSC Total Viral Total Nucleic Acid Purification Kit or QIAamp MinElute Virus Spin Kit (Qiagen). With the viral RNA isolate we synthesize viral cDNA using the Superscript® III One-Step RT-PCR System with Platinum® Taq High Fidelity by Invitrogen. Following cDNA synthesis, we generate four overlapping ~2.5 kb amplicons that span the entire SIV coding sequence by RT-PCR using sequence-specific primers. The amplified products are run on a 1% agarose gel and purified using AmPure XP Beads. Samples are quantified using Invitrogen's Qubit™ dsDNA HS Assay Kit.

The purified and quantified PCR products are prepared for sequencing using the transposon-based Nextera™ XT DNA Sample Prep Kit. During this process, the purified RT-PCR 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, normalized, and loaded onto the MiSeq.

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. Samples are filtered via centrifugation prior to viral isolation to remove any remaining host cells. The RNA is isolated in a similar way to that described above with the QIAamp MinElute Virus Spin Kit (Qiagen). Then we use random hexamers for first-strand cDNA synthesis, followed by second-strand dsDNA synthesis. These dsDNA products are purified, quantified, and library prepped with the Nextera™ XT DNA Sample Prep Kit as described above.

Sequencing SARS-CoV-2 and Zika viral genomes with multiplex PCR and TruSeq® DNA HT Library Prep

We sequence SARS-CoV-2 and Zika genomes with similar cDNA synthesis and multiplex PCR methods. Viral RNA is isolated from plasma using the Maxwell® RSC Viral Total Nucleic Acid Purification Kit or QIAamp MinElute Virus Spin Kit (Qiagen). We generate viral cDNA from viral RNA using the SuperScript® IV One-Step RT-PCR System with SuperScript® IV VILO Master Mix (which contains the random hexamers) by Invitrogen. Multiplex PCR is performed using genome specific primer pools and Q5 High-Fidelity 2X master mix to generate ~400 bp amplicon products. Following the generation of short multiplex amplicons, sample-associated primer pools are combined and purified using AmPure XP Beads. To confirm a successful PCR, we examine the PCR products on a 1% agarose gel. Samples are quantified using Invitrogen's Qubit™ dsDNA HS Assay Kit.

Amplicons generated by multiplex PCR are prepared for a sequencing library using the TruSeq® DNA HT Library Prep kit. The purified amplicons are end repaired and removed of large and small DNA fragments. Following this step are a series of limited cycle PCR steps to perform adenylation of 3' ends (A-tailing), adaptor ligation, and library amplification. The yielded fragments are approximately 400-600 bp in size and are tagmented with uniquely paired indices and Illumina-compatible sequencing oligos (For further information, refer to Illumina's website). Samples are purified using the Agencourt AMPure XP system and quantified again using the Qubit™ dsDNA HS Assay Kit. Average fragment size of library prepared sample is determined using the Agilent High Sensitivity DNA Kit. All uniquely tagged samples are then pooled, normalized, 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 5-30 viral genomes per sequencing run on either a 2x300 (600 cycles), 2x250 (500 cycles), or 2x150 (300 cycles) cartridge depending on genome size and the needs of the client. This yields up to approximately 25,000,000 paired-end sequencing reads (for a 2x300 cartridge) per sequencing run. There are both regular sized and small sized flow cell kits. The appropriate selection of kit depends on the number of samples you have and the length of sequences you need. With the Illumina MiSeq we reliably obtain complete coverage of the RNA genome by virtue of paired-end reads and large volumes of data. Further detail on costs for sequencing RNA viruses as a fee-for service work are listed below.

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 alignments for you. Additional data analyses may be performed after discussions with our staff.

Costs for SIVmac239, SARS-CoV-2, and Zika viral sequencing

Sequencing costs per genome characterization 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 rate). For clients outside of the WNPRC, we recover indirect costs (facilities and administrative overhead, etc.). As of May 1st, 2022 the rates for these indirect costs are 38.5% for federally funded or not-for-profit clients (Tier 2 rate) and 72.5% for non-federally funded or for-profit clients (Tier 3 rate). A more detailed breakdown of costs is available upon request.