Whole-Genome Sequencing Protocols for IBV and Other Coronaviruses Using High-Throughput Sequencing

Graham L. Freimanis and Michael S. Oade

Abstract

This chapter reports the high-throughput sequencing protocol for sequencing Coronaviruses and other positive strand viruses to produce a dataset of significant depth of coverage. The protocol describes sequencing of infectious bronchitis virus propagated in embryonated eggs and harvested in the allantoic fluid. The protocol is composed of three main steps—enrichment of the allantoic fluid using ultracentrifugation, extraction of total RNA from allantoic fluid, and library preparation from total RNA to DNA sequencing libraries. The workflow will be suitable for all coronaviruses using high-throughput sequencing platforms.

Key words Quasispecies, High-throughput sequencing, RNA, Virus, Infectious bronchitis virus, Coronavirus

1 Introduction

High-throughput sequencing (HTS) technologies provide the opportunity to rapidly obtain full genome sequence data of pathogenic organisms. In the recent COVID-19 pandemic, HTS was used to quickly identify the causative agent as a coronavirus, which has subsequently been named SARS-COV-2 [1]. The subsequent dissemination of sequence information online has allowed laboratories worldwide to study the virus and begin the development of vaccines eliminating the requirement of access to live virus samples. HTS and whole-genome sequencing (WGS) are therefore powerful, versatile tools for the development of novel treatments and vaccines, for studying virus evolution and genetic associations to disease or for tracking outbreaks with notable recent publications investigating viral outbreaks in health-care settings [2, 3].

The depth of data and high sequence quality obtained by such HTS also affords the possibility to dissect and study virus populations leading to the identification of low frequency mutations that
would otherwise be undetected by conventional Sanger-based sequencing. This is particularly relevant when working with RNA viruses, like coronaviruses, which have multiple variant viruses within a viral swarm due to factors including rapid replication rates, an error prone RNA-dependent RNA polymerase and large population sizes [4]. Despite the relatively small size of virus genomes, their sequencing is complicated by the atypical low qualities of viral nucleic acid template extracted from samples. Sample preparation therefore is the key to deep sequencing, in that it is important to enrich for viral RNA whilst maintaining the integrity of input material, resulting in a DNA-free high-quality preparation.

The protocol described here can produce high coverage depth datasets for Coronavirus genomes, such as Infectious Bronchitis Virus (IBV) grown in ovo from around 50 ng of total RNA. The growth of IBV in embryonated eggs has been described previously and is not discussed in this protocol [5]. Allantoic fluid must been free of membrane or other solid masses and ideally be free from blood. If the IBV causes hemorrhaging of the blood vessels, it is important to harvest the allantoic fluid before this happens. Due to the low sample matrix content in allantoic fluid and the polyA enrichment component of this protocol, some host mRNAs are also captured during sequencing, however this is dependent of the purity of the original sample.

This protocol is suitable for the capture of both viral gRNA and mRNA (combined) from allantoic fluid, is robust in terms of coverage depth, producing >99% genome recovery.

2 Materials

2.1 Purification of IBV

1. 50 ml centrifuge tubes.
2. 30% sucrose (w/v) in PBS adjusted to pH 7.2 with HCl, filtered through 0.22 μm syringe.
3. Refrigerated benchtop centrifuge.
4. Ultracentrifuge and rotor (e.g., Beckman Coulter SW55Ti rotor and Sorvall WX Ultra 80 ultracentrifuge).
5. Ultracentrifuge tubes compatible with rotor (e.g., Beckman polypropylene (13 × 51 mm) ultracentrifuge tubes).
6. Cannula.

2.2 RNA Extraction

1. RNase cleaning agent (e.g., RNaseZAP).
2. RNA extraction kit (e.g., RNeasy kit) or RNA isolation reagent (e.g., TRIzol).
3. Nuclease-free water.
4. DNA Lobind tubes.
5. 80% ethanol.
2.3 QC of Extracted Total RNA

1. High sensitivity DNA quantitation kit (e.g., Qubit high sensitivity RNA kit or equivalent).
2. Agilent Bioanalyzer 2100 high sensitivity RNA kit (or equivalent).

2.4 IBV Sequencing Library Preparation

1. NEBNext Ultra II stranded mRNA RNA-Seq (or equivalent).
2. NEBNext Ultra II indices for multiplexing (or equivalent).
3. SPRI paramagnetic beads (e.g., Beckman Coulter AMPure XP beads or equivalent).
4. 80% ethanol.
5. 10 mM Tris–HCl, pH 8.

2.5 Normalization of Sequencing Libraries

1. Agilent Bioanalyzer DNA 1000 kit.
2. DNA quantitation kit (e.g., Qubit DNA BR kit).

2.6 Quantification of Sequencing Library Pool

1. NEBNext Illumina library quantitation kit (or equivalent).
2. Fast PCR thermocycler (e.g., ABI 7500 fast thermocycler).

2.7 Denaturation of Sequencing Pool and Loading of Libraries

1. 0.2 N sodium hydroxide.
2. Control library (e.g., PhiX).
3. MiSeq reagent cartridge v3 600 cycles.
4. Illumina MiSeq benchtop sequencer.

3 Methods

Perform all activities prior to library preparation in a clean RNase free environment with dedicated RNA pipettes, RNase-free wipes, and other consumables. Always clean workspace with RNase cleaning agents prior to starting protocol. A schematic of the whole protocol is shown in Fig. 1.

3.1 Purification of IBV (2 h)

1. Place 3–10 ml of IBV infected allantoic fluid into a 15 ml tube and centrifuge for 10 min, $1150 \times g$ at 4 °C in a benchtop centrifuge in order to remove large solid matter from the IBV infected allantoic fluid.
2. Take the supernatant and add it to a 5 ml ultracentrifuge tube. A cushion of 0.5 ml 30% Sucrose is placed at the base of the tube under the allantoic fluid using a canula being careful not to disturb the allantoic fluid. Tubes must be balanced carefully.
3. Ultracentrifuge for 60 min, $236,880 \times g$ at 4 °C in an ultracentrifuge.
4. Remove supernatant in layers, taking care not to disturb the pellet (see Note 1).
5. Resuspend the pellet in 350 μl of RLT lysis buffer or TRIzol.
6. Either store at −20 °C or progress directly onto total RNA extraction.

3.2 RNA Extraction (1 h)

In our experience although TRIzol produced higher-quality RNA, RNeasy extractions by comparison yielded more RNA. A greater RNA yield and the easier scalability of RNeasy versus TRIzol resulted in RNeasy being our preferred method of RNA extraction for this protocol. Other RNA extraction kits and methods may also be used.

1. Extract total RNA from the 350 μl of nucleic acid dissolved in lysis buffer (RLT) using the Qiagen RNeasy RNA extract kit (or equivalent) according to the manufacturer’s protocol, following the optional additional spin to remove excess RPE. Elute in a total of 30 μl of elution buffer. Sometimes running elution buffer, prewarmed at 37 °C, through the column twice can improve the yield of the kit.

3.3 Quality Control (QC) of Extracted Total RNA (1 h)

1. Quantify total RNA content using Qubit and the Qubit High sensitivity RNA kit (or equivalent) following the manufacturer’s instructions with each sample measured in duplicate.
2. Identify size distribution of RNA using the Bioanalyzer 2100 and Bioanalyzer RNA nano chip (or equivalent) following the manufacturer’s protocol (see Note 2).

3.4 Library Preparation (1–2 Days Dependent Upon Kit)

1. Calculate a minimum input of 50–200 ng of total RNA and dilute into a total volume of 50 μl to satisfy input requirements for the NEBNext Ultra II stranded mRNA RNA-Seq kit (or equivalent) (see Note 3).

2. The remainder of the protocol is carried out following the manufacturer’s instructions, fragmenting for 200–300 bp fragments in terms of fragmentation and size selection protocols. However, if the amount of input material is low (<50 ng) then 200 bp is recommended.

3. Elute the final libraries in 20 μl of 10 mM Tris–HCl, pH 8. Store at −20 °C for a maximum of 4 weeks (see Note 4).

3.5 Library QC and Normalization of Sequencing Libraries (3 h)

1. Run generated libraries on a Bioanalyzer 2100 using Bioanalyzer DNA 1000 kit (or equivalent) following the manufacturer’s instructions (see Note 5). A typical library profile is as below (Fig. 2). Record the average library size for each library using smear analysis to increase accuracy of average size.

2. Manually quantitate each sample using the Qubit high sensitivity kit measuring each sample in duplicate to get an accurate reading.

3. Convert the average Qubit reading from ng/μl to nM using the average library size from the Bioanalyzer results using the formula below;

$$\text{concentration in } \text{ng/μl} = \frac{660 \, \text{g/mol} \times \text{average library size (in bp)}}{\times 10^6} = \text{concentration in nM}$$

4. For longer term storage, dilute in 10 mM Tris–HCl, pH 8.5. Do not pipette volumes <2 μl (see Note 6). Otherwise, proceed to step 5.

5. Dilute libraries to a concentration of 5 nM using the formula below. A minimum of 4 nM is required before addition to the pool.

$$C_1 V_1 = C_2 V_2$$

where

$C_1 = \text{initiation concentration of solution.}$

$V_2 = \text{volume of the initiation solution that will be diluted.}$

$C_2 = \text{concentration of final solution.}$

$V_2 = \text{desired volume of final solution.}$
6. Take 5 µl from all samples and pool in a single 1.5 ml LoBind tube and store at −20 °C. If you do not intend to progress to the next step immediately then ensure sequencing is performed within 2–3 weeks for optimal cluster densities where possible.

3.6 Quantitation of Sequencing Pool Using qPCR, Sample Denaturation and Loading onto Illumina MiSeq (1 Day)

1. Quantitate the pooling using the NEBNext Quantitation for Illumina Libraries kit (or equivalent) using the manufacturer’s protocol.

2. On confirmation that the sequencing pool is at a concentration of approximately 4–5 nM, proceed to denaturing libraries with NaOH, for loading on the MiSeq.

3. Denature libraries with 0.2 N NaOH and dilute according to the manufacturer’s guidelines. Samples are then loaded at a concentration of 15 pM onto a 2 × 300 cycle v3 MiSeq reagent cartridge, using a 1% PhiX spike (see Note 7).

4 Notes

1. This protocol can be very sensitive to input material. Depending on the yield of virus grown in eggs, the viral pellet may not be prominent after ultracentrifugation. Insufficient yield of virus at this stage will greatly hamper the quality of obtained reads. It may therefore be beneficial to repeat this stage as necessary until a greater yield of virus is achieved.
2. The purpose of running the Bioanalyzer is to check that RNA is present within the sample. The RNA Integrity Number (RIN) will not be informative to the final library yield in this case. An example of a typical trace is shown in Fig. 3.

3. Using the poly-A enrichment module will directly enrich for both genomic vRNA and viral mRNAs due to the presence of a poly-A tail at the 3’ end of the IBV genome.

4. Libraries should not be left for longer than 4 weeks at $-20\,^\circ\text{C}$ or this can result in a reduction in clustering and sequencing quality.

5. Performing library QC is important for indicating inefficient library preparation/low library yield.

6. Where possible, do not dilute values less than 2 μl. Pipetting smaller volumes is highly inaccurate and will unduly bias the proportion of this library on the flow cell.

7. Loading concentrations may need to be optimized for individual machines.

References

1. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579 (7798):270–273. https://doi.org/10.1038/s41586-020-2012-7

2. Garvey MJ, Bradley CW, Holden KL, Hewins P, Ngui SL, Tedder R, Jumaa P, Smit E (2017) Use of genome sequencing to identify hepatitis C virus transmission in a renal healthcare setting.
3. Houlihan CF, Frampton D, Ferns RB, Raffle J, Grant P, Reidy M, Hail L, Thomson K, Mattes F, Kozlakidis Z, Pillay D, Hayward A, Nastouli E (2018) Use of whole-genome sequencing in the investigation of a nosocomial influenza virus outbreak. J Infect Dis 218(9):1485–1489. https://doi.org/10.1093/infdis/jiy335

4. Andino R, Domingo E (2015) Viral quasispecies. Virology 479-480C:46–51. https://doi.org/10.1016/j.virol.2015.03.022

5. Guy JS (2015) Isolation and propagation of coronaviruses in embryonated eggs. In: Maier HJ, Bickerton E, Britton P (eds) Coronaviruses: methods and protocols. Springer New York, New York, NY, pp 63–71. https://doi.org/10.1007/978-1-4939-2438-7_7