

Whole-Genome Sequencing of Porcine Epidemic Diarrhea Virus by Illumina MiSeq Platform

Leyi Wang, Tod Stuber, Patrick Camp, Suelee Robbe-Austerman, and Yan Zhang

Abstract

Porcine epidemic diarrhea virus (PEDV) belongs to the genus *Alphacoronavirus* of the family *Coronaviridae*. PEDV was identified as an emerging pathogen in US pig populations in 2013. Since then, this virus has been detected in at least 31 states in the USA and has caused significant economic loss to the swine industry. Active surveillance and characterization of PEDV are essential for monitoring the virus. Obtaining comprehensive information about the PEDV genome can improve our understanding of the evolution of PEDV viruses, and the emergence of new strains, and can enhance vaccine designs. In this chapter, both a targeted amplification method and a random-priming method are described to amplify the complete genome of PEDV for sequencing using the MiSeq platform. Overall, this protocol provides a useful two-pronged approach to complete whole-genome sequences of PEDV depending on the amount of virus in the clinical samples.

Key words Porcine epidemic diarrhea virus, Targeted amplification, SISPA, Whole-genome sequencing, MiSeq

1 Introduction

Complete genome sequencing and genetic analysis significantly improved our understanding of the evolution and relationship of porcine epidemic diarrhea virus (PEDV) strains worldwide. The first PEDV whole-genome sequence was completed for the prototype strain CV777 in 2001 [1]. Since then, several PEDV strains have been sequenced and now over 170 whole-genome sequences have been deposited in GenBank. Based on the phylogenetic analysis of the whole-genome sequence, PEDV has been classified into two Genogroups—1 and 2—which the variant and classical strains of US PEDV belong to, respectively [2].

Since the first 454 FLX pyrosequencing platform was introduced to the market in 2005, next-generation sequencing (NGS) has significantly advanced research in diverse fields. NGS has the

advantages of high-throughput and cost-effectiveness. Currently, there are several platforms available, including the Genome Analyser developed by Illumina/Solexa, and the Personal Genome Machine (PGM) by Ion Torrent. One of Illumina NGS platforms—MiSeq—is commonly used in diagnostic laboratories.

In general, sequencing viruses directly from fecal samples for PEDV are technically challenging without prior amplification with specific primers. In this chapter, we describe a useful two-pronged approach where the random-priming method can be used to sequence the complete PEDV genome from samples with Ct values of less than 15, whereas the targeted amplification method is recommended to be used to sequence clinical fecal samples with higher Ct values (low viral loads).

2 Materials

2.1 RNA Extraction from Feces or Intestinal Contents

1. MagMAX Pathogen RNA Kit (Life Technologies) for viral RNA extraction from fecal or intestinal contents (*see Note 1*).
2. HBSS (GIBCO).

2.2 Real-Time Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) Reaction

1. One-Step RT-PCR kit (Qiagen).
2. Smart Cycler II (Cepheid, Sunnyvale CA).
3. 10 pmol Primers and probes [3] (*see Note 2*).

Forward primer: 5'-CATGGGCTAGCTTTCAGGTC-3'.

Reverse primer: 5'-CGGCCCATCACAGAAGTAGT-3'.

Probe: 5'/56-FAM/CATTCTTGGTGGTCT TTCAAT CCTGA/ZEN 3IABkFQ/3'.

2.3 Targeted Amplification One-Step RT-PCR

1. One-Step RT-PCR kit (Qiagen) (*see Note 3*).
2. Oligonucleotide primers dissolved in nuclease-free water to a stock concentration of 100 pmol/μl and a working concentration of 20 pmol/μl. The sequences for 19 pairs of primers are listed in Table 1.
3. Qiagen gel purification kit.
4. Qubit 2.0 Fluorometer (Life Technologies).

2.4 SISPA Method (Sequence-Independent, Single-Primer Amplification)

1. SuperScript III Reverse Transcriptase kit (Invitrogen).
2. RNase H treatment (NEB).
3. Klenow amplification (NEB).
4. Advantage 2 PCR kit (Clontech).
5. 10 mM dNTP mix (NEB).
6. RNase Inhibitor (Promega).

Table 1
Nineteen pairs of primers for whole-genome amplification of PEDV

Fragment no.	Sequence	Sense	Size
F1	ACTTAAAAAGATTTTCTATCTAC CGTTAACGATACTAAGAGTGGC	Forward Reverse	1622
F2	TGGTGACCTTGCAAGTGCAGC ATTACCAACAGCCTTATTAAGC	Forward Reverse	1603
F3	ACCATTGACCCAGTTTATAAGG ACAAAAGCACTTACAGTGGC	Forward Reverse	1587
F4	TACACCTTTGATTAGTGTTGG TTTGTAGCGTCTAACTCTAC	Forward Reverse	1614
F5	GTACCAGGTGATCTCAATGTG ACGTGGCAATGTCATGGACG	Forward Reverse	1615
F6	ATGCTGCTGTTGCTGAGGCTC TCAGTTGAGATAGAGTTGGC	Forward Reverse	1600
F7	GTGACAAGTTCGTAGGCTC TAAGTGACAGAACTCACAGG	Forward Reverse	1597
F8	TGCACAAGGTCTTGTTAACATC TCTGTGCACCATTAGGAGAATC	Forward Reverse	1601
F9	ACCTGCGTGTAGTCAAGTGG GTTACCAGTGGAACACCATC	Forward Reverse	1599
F10	ACTGTGCCAACTTCAATACG TCATCAACAAACACACCTGC	Forward Reverse	1611
F11	TGCTCGCAGCATACTATGCAG GTGGTGCAGGCAGCTGTTGAG	Forward Reverse	1588
F12	TCTATGTGCACTAATTATGAC TGATTGCACAATTCGGCCGC	Forward Reverse	1599
F13	CCATACATGATTGCTTTGTC ATCGTCAAGCAGGAGATCC	Forward Reverse	1595
F14	TGTCTAGTAATGATAGCAGC TTATCCCATGTTATGCCGAC	Forward Reverse	1647
F15	TAATGATGTTACAACAGGTCG AAGCCATAGATAGTATACTTG	Forward Reverse	1554
F16	TGAGTTGATTACTGGCACGCC GTACTGTATGTAAAAACAGCAG	Forward Reverse	1598

(continued)

Table 1
(continued)

Fragment no.	Sequence	Sense	Size
F17	ATCGCAATCTCAGCGTTATG	Forward	1596
	GTGTAAACTGCGCTATTACAC	Reverse	
F18	CTGCTTATTATAAGCATTAC	Forward	1603
	GCTTCTGCTGTTGCTTAAGC	Reverse	
F19	AGTCTCGTAACCAGTCCAAG	Forward	1065
	TTTTTTTTTTTTTGTGTATCCAT	Reverse	

- Oligonucleotide primers [4] dissolved in nuclease-free water to a concentration of 50 μM , 1 μM , and 10 μM for P1, P2, and P3, respectively.

P1: GAC CAT CTA GCG ACC TCC ACN NNN NNN N.

P2: GAC CAT CTA GCG ACC TCC AC TTT TTTTTTTT
TTTTTTTTT TT.

P3: GAC CAT CTA GCG ACC TCC AC.

- QIAquick PCR Purification Kit (Qiagen).

2.5 Detection of PCR Products

- Agarose.
- Distilled water.
- Ethidium bromide.
- 1 \times TAE buffer: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA.

2.6 Illumina Nextera DNA Library Preparation

- Nextera XT Library Prep Kit 96 samples (Box 1 of 2) (Illumina).
- Nextera XT Library Prep Kit 96 samples (Box 2 of 2) (Illumina).
- Nextera XT Index Kit 96 indexes-192 samples.
- Agencourt AMPure XP beads (Beckman Coulter).
- 96-well PCR plate (Scientific Inc.).
- 96 Deep Well Block (Invitrogen).
- Microseal "B" adhesive seals (BioRad).
- Magnetic plate stand-96 (Life, Technologies).
- Ethanol, 200 proof (Sigma-Aldrich).

2.7 Next-Generation Sequencing

- MiSeqv2 Reagent Kit 500 cycles PE-Box 1 of 2 (Illumina).
- MiSeqv2 Reagent Kit Box 2 of 2 (Illumina).
- MiSeq (Illumina).

2.8 Sequence Assembly and Analysis

1. Kraken.
2. Krona.
3. BWA—Burrows-Wheeler Alignment Tool.
4. SAMtools.
5. Picard.
6. BLAST.
7. BioPython.
8. GATK—Genome Analysis Toolkit.
9. R.
10. IGV.

3 Methods

3.1 Viral RNA Extraction

1. Fecal or intestinal contents were diluted in HBSS to a final concentration of 20 % and were homogenized by five stainless steel balls followed by a centrifuge step at 2000 RCF at 4 °C for 5 min.
2. The supernatant was used for RNA extraction by using the MagMAX Pathogen RNA/DNA Kit (Life Technologies) (*see Note 2*).

3.2 Real-Time RT-PCR Reaction

1. Real-time RT-PCR with a 25 µl reaction volume was completed using QIAGEN one-step RT-PCR kit: 5 µl 5× RT-PCR buffer, 0.5 µl forward primer (10 pmol), 0.5 µl reverse primer (10 pmol), 0.5 µl probe (10 pmol), 1 µl dNTP, 1 µl enzyme mix, 0.2 µl RNasin inhibitor (40 Unit/µl, Promega), and RNA temple: 2.5 µl.
2. The amplification conditions were 50 °C for 30 min; 95 °C for 15 min; and 45 cycles of 94 °C, 15 s, and 60 °C, 45 s.

3.3 One-Step RT-PCR Reaction

1. RT-PCR with a 25 µl reaction volume was completed using QIAGEN one-step RT-PCR kit: 5 µl 5× RT-PCR buffer, 0.8 µl forward primer (20 pmol), 0.8 µl reverse primer (20 pmol) (Table 1), 1 µl dNTP, 1 µl enzyme mix, 0.2 µl RNasin inhibitor (40 Unit/µl, Promega), and RNA temple: 2.5 µl.
2. The amplification conditions were 50 °C for 30 min; 95 °C for 15 min; and 45 cycles of 94 °C, 30 s, 54 °C, 30 s, and 72 °C, 1 min 30 s.
3. Analyze the PCR products on a 1 % agarose gel and migrate for 1 h at 90 V.
4. Excise the correct size bands and perform gel purification with a Qiagen gel purification kit (*see Note 4*).

5. Quantify the DNA generated by a fluorescence-based method (Qubit 2.0 Fluorometer) and final amount of DNA input as 1 μg (*see Note 5*).

3.4 SISPA Method (See Note 6)

3.4.1 First-Strand Synthesis

1. 1 μl of 50 μM random primer P1; 1 μl of 1 μM oligo dT primer P2; 1 μl 10 mM dNTP mix; 10 pg–5 μg of RNA template. Add water up to 13 μl total volume.
2. Incubate the reaction at 65 $^{\circ}\text{C}$ for 5 min and incubate on ice for at least 1 min.
3. Add 4 μl 5 \times first-strand buffer; 1 μl 0.1 M DTT; 1 μl RNase inhibitor; 1 μl of SuperScript III Reverse Transcriptase.
4. Incubate the reaction at 25 $^{\circ}\text{C}$ for 5 min, 50 $^{\circ}\text{C}$ for 30–60 min, and 70 $^{\circ}\text{C}$ for 15 min.
5. Add 1 μl RNase H (NEB) to the reaction.
6. Incubate at 37 $^{\circ}\text{C}$ for 20 min.

3.4.2 Klenow Amplification

1. Add 3 μl 10 \times Klenow reaction buffer; 1 μl of 25 μmol dNTP; and 1 μl of 1 μM random primer P1 to the reaction in Sect. 3.4.1.
2. Incubate at 95 $^{\circ}\text{C}$ for 2 min and cool to 4 $^{\circ}\text{C}$.
3. Add 1 μl Klenow fragment (NEB).
4. Incubate at 37 $^{\circ}\text{C}$ for 60 min and 75 $^{\circ}\text{C}$ for 20 min.

3.4.3 PCR Amplification

1. 5 μl 10 \times Advantage 2 PCR buffer; 1 μl 50 \times dNTP mix; 2 μl 10 μM barcode primer P3; 1 μl 50 \times Advantage 2 Polymerase Mix; DNA template from Klenow amplification. Add water up to 50 μl total volume.
2. Incubate the reaction using the following PCR program: 1 cycle: 95 $^{\circ}\text{C}$ 5 min; 5 cycles: 95 $^{\circ}\text{C}$ 1 min; 59 $^{\circ}\text{C}$ 1 min; 68 $^{\circ}\text{C}$ 1 min 10 s; 33 cycles: 95 $^{\circ}\text{C}$ 20 s; 59 $^{\circ}\text{C}$ 20 s; 68 $^{\circ}\text{C}$ 1 min 30 s; 1 cycle: 68 $^{\circ}\text{C}$ 10 min.
3. Use 5 μl to analyze the PCR products on a 1 % agarose gel and migrate for 1 h at 90 V (*see Note 7*).
4. Use QIAquick PCR Purification Kit to purify the remaining 45 μl .
5. Quantify the DNA generated by a fluorescence-based method (Qubit 2.0 Fluorometer) and final amount of DNA input as 1 μg .

3.5 Library Preparation Using Nextera XT Kit

1. Perform the library preparation based on the Illumina company manual, which includes tagmentation of genomic DNA, PCR amplification, PCR cleanup, library normalization, and final library pooling for MiSeq sequencing.

3.6 Sequence Assembly and Analysis

1. Kraken is used to initially identify raw reads and provide a graphical representation of the reads using Krona. A custom Kraken database is used. It was built using the standard database containing all Ref Seq bacteria and virus genomes along with all complete swine enteric coronavirus disease (SECD) genomes available at NCBI and a pig genome.
2. Raw reads are run through an in-house custom shell script. In brief, 18 complete genomes from NCBI representing the 4 SECD virus species (TGEV, PRCV, PEDV, and PDCoV) are used as references to align raw reads. A function is looped 3x. This function aligns and removes duplicates, creates a VCF, updates reference with VCF information, and performs a BLAST search against the nt database using the updated reference. Eighteen complete genomes are used to start the initial loop. From this first loop the top hit returned is used as the reference for the next loop. A total of three loops are performed to find the best reference.
3. After the best reference has been found, alignment metrics including read counts, mean depth of coverage, and percent of genome with coverage are collected.
4. Reports summarizing the alignment metrics (Fig. 1) along with Kraken identification interactive Krona HTML file, a FASTA of assembled genome, and depth of coverage profile graph (Fig. 2) are e-mailed to concerned individuals.
5. The assembled FASTA file can be visually verified in IGV using the BAM and VCF output from the script. If necessary the FASTA can be corrected in program of choice.
6. Script details are provided on GitHub (<https://github.com/USDA-VS/public/blob/master/secd/idvirus.sh>).

```

Sample: OH851-RP-Virus
Reference_Set: SECD
R1 file size: 451M, read count: 914188
R2 file size: 451M, read count: 914188
223,955 virus reads --> 24.4977% of total reads

Alignment stats (reference guided):
reference used          read count percent cov  ave depth
Porcine_epidemic_diarrhea_virus-KJ399978  346,926      98.36%      2,764X
Porcine_Resp_Corona_Virus-DQ811787       68,978       50.83%       713X

-----
*** NT Database ***
query ID              qlength slength % id  mis value bscore Description
Porcine_epidemic_diarrhea_virus-KJ399978  28025  28029  99.98  1  0.0  51729  Porcine epidemic diarrhea virus strain OH851, complete genome KJ399978.1
Porcine_Resp_Corona_Virus-DQ811787       26617  27550  93.00  1654 0.0  38778  PRCV ISU-1, complete genome DQ811787.1

```

Fig. 1 Report summary for sample OH851-RP-Virus. The reference set used to initiate the shell script was SECD (swine enteric coronavirus diseases). File size and read counts for each fastq file are shown. Provided by Kraken, 223,955 virus reads were identified. “Reference used” is the closest finding in the NCBI nt database. The read count shows the number of raw reads shown to match the reference. “Percent cov” shows the percent of reference having coverage. A coverage of 98.36 % for PEDV indicates a true find relative to the sporadic <51 % coverage seen from PRCV, although in this case, the presence of PRCV cannot be ruled out. There were no reads matching TGEV and PDCoV, which were not shown. Because of the high percent of genome coverage, the completed reference-guided assembly for PEDV was BLAST against the nt database to provide mismatches, *e*-value, and bit score against the most closely related publicly available genome

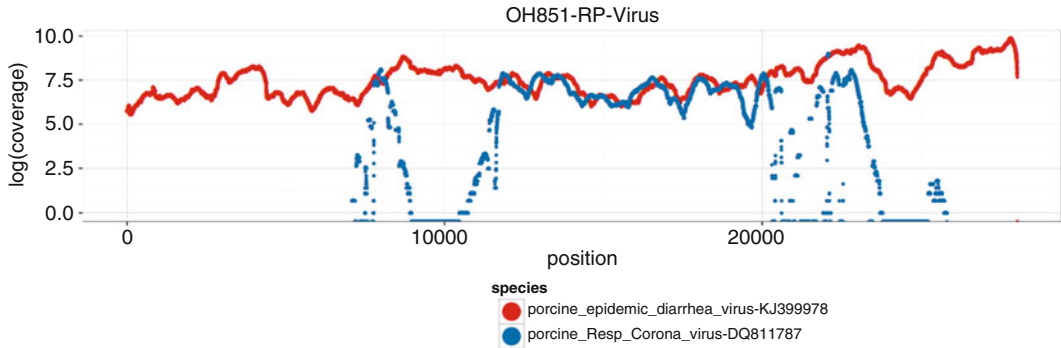


Fig. 2 The depth of coverage profiles for sample OH851-RP-Virus. The x-axis is the genome position. The y-axis is the log depth of coverage. Reads matching any of the four SECD target viral species are shown

4 Notes

1. The MagMAX Pathogen RNA/DNA Kit was used for the extraction of nucleic acid for pathogen detection—including the detection of TGEV, PEDV, and PDCoV—from pig feces or intestinal contents.
2. The real-time RT-PCR assay was developed by our laboratory and the primers and probes target the M gene of the virus.
3. In our laboratory, the QIAGEN one-step RT-PCR kit has been used to amplify RT-PCR products between 100 bp and 1.8 kb in length.
4. Alternatively, 5 μ l out of 25 μ l could be loaded on the gel to confirm that each target band is amplified and then the remaining 20 μ l can be purified by QIAquick PCR Purification Kit.
5. A smaller amount of input DNA than the required 1 μ g for targeted amplification method could be used to avoid overwhelming sequence reads.
6. The SISPA method is recommended when the Ct value of real-time RT-PCR is below 15.
7. When 5 μ l was applied to the gel, a smear bank can be observed.

References

1. Kocherhans R, Bridgen A, Ackermann M, Tobler K (2001) Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* 23:137–144
2. Wang L, Byrum B, Zhang Y (2014) New variant of porcine epidemic diarrhoea virus, United States, 2014. *Emerg Infect Dis* 20:917–919
3. Wang L, Zhang Y, Byrum B (2014) Development and evaluation of a duplex real-time RT-PCR for detection and differentiation of virulent and variant strains of porcine epidemic diarrhoea viruses from the United States. *J Virol Methods* 207:154–157
4. Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL (2008) Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathog* 4:e1000163