NATIONAL CENTER FOR IMMUNIZATION AND RESPIRATORY DISEASES

Reverse transcription methods impact genome sequencing coverage for RNA viruses

R.L. Marine[1], N. Momin[2], D.D. Wagner[2], G. Nabakooza[3], and M.S. Oberste[1]

[1]Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA; [2]Eagle Global Scientific LLC, Atlanta, Georgia, USA; [3]Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA

Reverse transcription methods generate distinct and reproducible coverage profiles, with traditional two-step methods resulting in the most even coverage patterns for small linear RNA virus genomes

BACKGROUND

Sequencing of viral pathogens has become indispensable for tracking the evolution and spread of disease. For sequencing of RNA viruses, sample preparation requires converting extracted RNA to cDNA, as most of the existing next-generation sequencing platforms require DNA as template. Previous studies have evaluated the impact of different library preparation methods and sequencing platforms on the quality and completeness of assembled genomes, but the effects of RNA to cDNA conversion strategies are underexplored.

METHODS

In this study, we prepared a panel of four enterovirus (EV) D68 and three EV-A71 isolates for sequencing using four random reverse transcription (RT) procedures (Figure 1):

- a laboratory-developed one-step (**OS**) RT protocol
- a sequence-independent single-primer amplification (**SISPA**)
- two traditional two-step RT procedures leveraging enzymes/protocols available through **Thermo** Fisher and New England BioLabs (**NEB**)

STEP	One-step*	SISPA	NEB	Thermo
First Strand RT	qScript Flex, random hexamer and oligo dT primer (1:100)	SuperScript IV, N1-8N primer	NEB Ultra II RNA First Strand Synthesis Module, random primer and oligo dT primer (1:100)	SuperScript IV, , random hexamer and oligo dT primer (1:100)
Second Strand RT	Vent exo-	Klenow 3'->5' exo-	NEBNext® Ultra™ II Non-Directional RNA Second Strand Synthesis Module	Second strand cDNA Synthesis Kit#
Additional Steps		N1 PCR		

*For the one-step procedure, first strand and second strand are performed as one reaction/step

Figure 1. Enzymes/kits used for each step of the four RT procedures

After reverse transcribing all samples using each procedure, samples were purified using SPRI bead cleanup (1.8X) and sequencing libraries were prepared using the Illumina DNA library preparation kit. Prepared libraries were sequenced on an Illumina MiSeq 300 cycle paired-end run.

DATA ANALYSIS

Raw fastq data was processed using Trimmomatic v0.39 to removed low quality bases, adapters and reads less than 50 bp after trimming. Filtered fastq data was then mapped to respective reference genomes using Geneious Prime v 2023.1.1 ("Low Sensitivity/Fastest" setting and no fine tuning) to determine the number/percentage of reads mapped and general coverage statistics. For the coverage pattern analysis, filtered datasets were normalized per sample by random subsampling using seqtk v1.3.

RESULTS

- The SISPA method generated a higher proportion (9.9–63.5%) of target virus reads per sample compared to the one-step (0–15.5%) and two-step RT procedures (0.5–20.4%) (Figure 2)
- The one-step and two-step RT procedures generated greater breadth and evenness of read coverage
  - Read coverage CV was 28.4–34.2% and 24.5–35.0% for Thermo Fisher and NEB two-step RT procedure samples, respectively, compared to 39.2–70.3% with the one-step RT procedure and 73.7–94.5% with the SISPA procedure (Table 1)
- The SISPA method generated the most distinct read coverage patterns compared to other RT methods (Pearson's r values of 0.54–0.92 for intra-method comparisons vs 0.22–0.59 for inter-method comparisons, Table 2)
- Higher read coverage at the 3' ends of EV-D68 and EV-A71 genomes was obtained for the NEB and Thermo two-step RT procedure (Figure 3)

CONCLUSIONS

For whole genome sequencing of viruses, RT methods which produce more even coverage profiles (traditional two-step) are preferable for generating complete assemblies and maximizing the number of samples per sequencing run. Conversely, because of the high proportion of viral reads obtained using the SISPA method, this procedure may be favored when viral detection is the main goal of sequencing.

FIGURES/TABLES

Table 1. Mean read coverage and coefficient of variation (CV) for EV-D68 and EV-A71 samples.

	One-step		SISPA		NEB		Thermo	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV
EV-D68 Fermon	1	*	2124	73.7	28	35.0	20	34.2
EV-D68 USA/MO/14-18949	126	42.2	3954	83.8	1125	27.6	822	28.4
EV-D68 USA/MO/18954	1319	42.5	3203	85.4	1528	28.9	711	30.8
EV-D68 USA/IL/14-18952	1078	39.2	11744	75.7	1782	24.5	118	29.0
EV-A71 USA/AK/16-19516	368	52.2	4388	91.5	1297	26.1	455	30.0
EV-A71 USA/CT/16-19519	218	40.3	5825	94.5	1804	26.3	1068	29.7
EV-A71 USA/WA/16-19522	309	70.3	7785	68.3	1073	27.1	900	33.0

*CV not considered due to low number of reads for samples

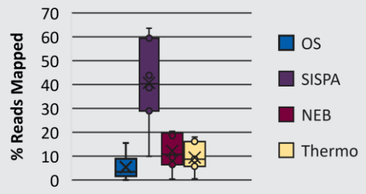

Figure 2. The percent viral reads mapped per sample, grouped by RT procedure
OS- One-step

Table 2. Pearson correlation coefficient (r) values for comparisons of EV-D68 and EV-A71 coverage profiles generated using the same RT method ("Intra") or using differing RT methods ("Inter")

	EV-D68		EV-A71	
	Intra	Inter	Intra	Inter
One-step	0.88-0.94	0.33-0.74	0.57-0.83	0.22-0.85
SISPA	0.76-0.92	0.23-0.59	0.54-0.89	0.22-0.45
NEB	0.77-0.82	0.23-0.78	0.66-0.82	0.25-0.81
Thermo	0.86-0.91	0.48-0.78	0.75-0.84	0.36-0.85

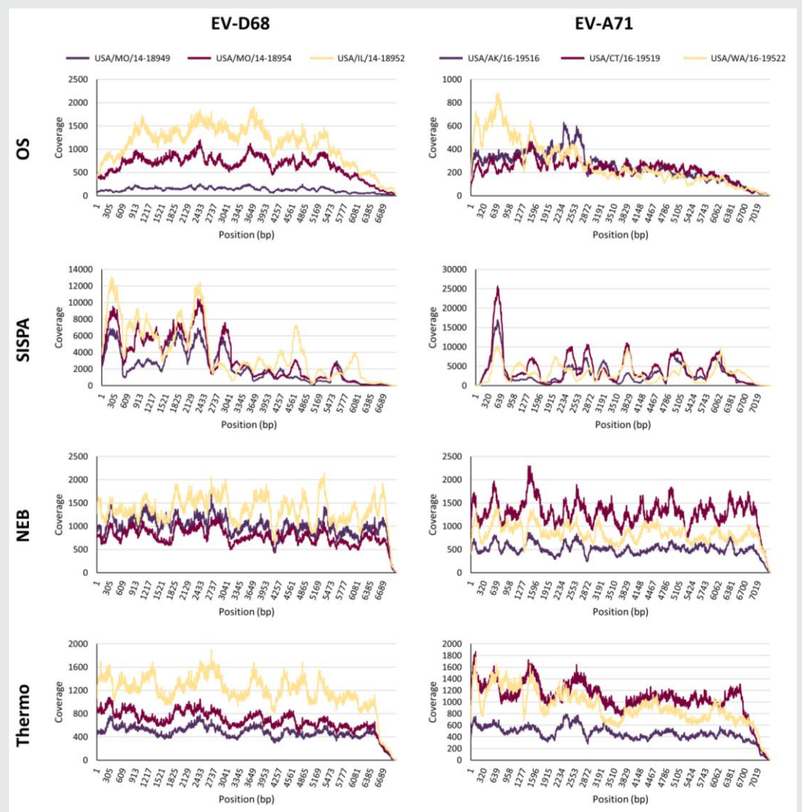

Figure 3. The depth of read coverage across the length of EV-D68 genomes (left panels) and EV-A71 genomes (right panels). Each row depicts the coverage patterns for a particular RT procedure: One-step (OS), SISPA, NEB and Thermo. The EV-D68 Fermon sample was not included in this figure/analysis due to low read coverage.

SCAN HERE FOR MORE INFORMATION

CONTACT INFO

Rachel Marine
rmarine@cdc.gov