Development and Evaluation of Standards for Virome Research

Poster: 351
Date: June 8, 2018

Phone: 800.638.6597
Email: Tech@atcc.org

Briana Benton, B.S., Juan Lopera, Ph.D., Stephen King, M.S., Dev Mittar, Ph.D., and Reed Shabman, Ph.D.
ATCC, Manassas, VA 20110

Background

Though there is an abundance of studies, applications, and publications on the human bacterial microbiome, there are a limited number of reagents and publications focused specifically on the "virome". Next-generation sequencing (NGS) has enabled virus sequencing on a large scale at an affordable cost. However, the complexities involved in the NGS methodology and the diversity of viral genomes pose a significant challenge to assay standardization. Therefore, there is a critical need for standardized reference materials across the research and diagnostics communities to serve as controls in assay development. To support this need, we are developing a viral panel comprising both quantified virus and virus nucleic acids prepared from diverse RNA and DNA virus families.

Strategy for Virome Mix

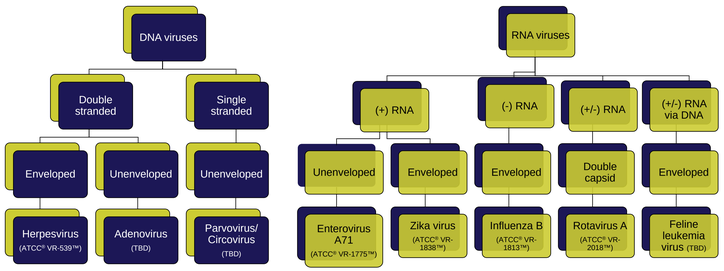
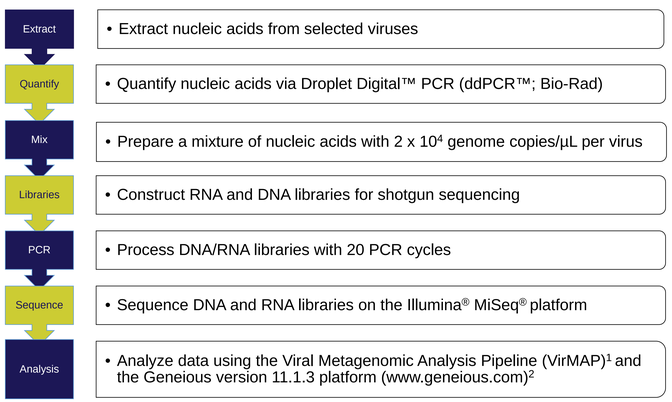


Figure 1. Relevant genotypic attributes that were evaluated for the development of the virome standards. Ideally, a comprehensive NGS virus standard would include a representative virus from each category.

Work Flow



Development of Virome Mix

Table 1. Viruses selected for proof-of-concept experiments

Virus	Strain	ATCC® No.	Host
Human herpesvirus 1	MacIntyre	VR-539™	Vero cells
Enterovirus A71 (EV-A71)	BrCr	VR-1775™	Vero cells
Influenza B virus	B/Massachusetts/2/2012	VR-1813™	Embryonated chicken eggs
Zika virus	MR-766	VR-1838™	Vero cells
Rotavirus A	WA (TC adapted)	VR-2018™	MA104 Clone 1 cells

ddPCR Analysis

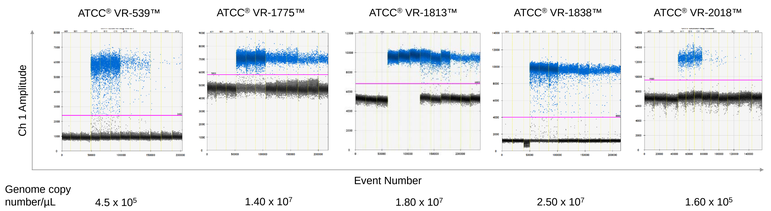


Figure 2. Nucleic acids from each virus were quantitated via ddPCR using proprietary molecular probe-based assays that target a single gene copy. Graphs depict no template (black droplets) and template (blue droplets), which serve as quantitation controls. The viruses ranged from 1.60 x 10^5 genome copies/µL to 2.5 x 10^7 genome copies/µL.

Methods

Methods: We normalized input viral nucleic acid genome copy numbers via ddPCR to a concentration of 2 x 10^4 genome copies/µL per virus. The virome mix was then used to prepare DNA and RNA libraries for NGS using the Illumina MiSeq sequencing platform.

Data analysis: Raw sequence data was analyzed in two independent pipelines. First, VirMAP[3] (Baylor College of Medicine, Houston, Texas) was used to successfully identify viral species present in the virome mix, independent of mapping to a known reference. Second, we used Geneious[2] software to map raw data to the published reference genomes of each virus. While we normalized input genomic copies by ddPCR, abundance by NGS varied significantly (Figure 4). Additionally, the number of reads per virus varied significantly between analysis platforms (Figure 4). This proof-of-concept study highlights the need and utility of such virome standards to allow researchers to optimize their sequencing methods to recover sequences from diverse viral families.

NGS Results

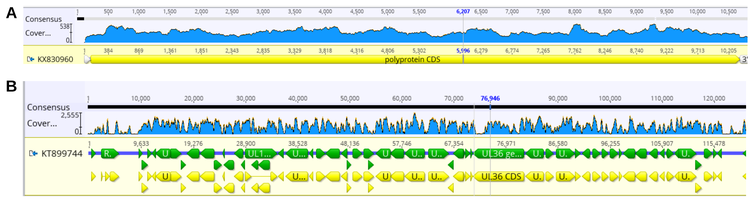


Figure 3. MiSeq alignment data. (A) Zika virus and (B) Herpesvirus sequence data were aligned to the respective reference genome. Data were generated with Geneious[2] v11.1.3. The depth of the sequence coverage (blue) and the alignment to the annotated reference genome (yellow & green) were analyzed.

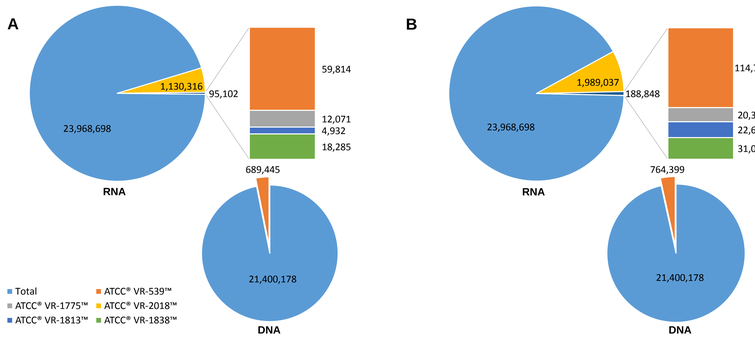


Figure 4. DNA and RNA library data sets were analyzed using two different data analysis platforms. Using (A) VirMAP[3] and (B) Geneious[2], we were able to align the paired end data to all 5 viruses with an average confidence value of 99.5% and 99.4%, respectively.

Conclusions

- The standardized concentration of 2 x 10^4 genome copies/µL per virus in our virome standard is sufficient for NGS library construction and data analysis.
- On average, viruses were identified with a greater than 99% confidence value using the Viral Metagenomic Analysis Pipeline[3] and Geneious[2] software.
- Approximately 7.1% of reads from the DNA library and 3.4% of reads from the RNA library mapped to spiked viral samples. Host reads were far more abundant, which was expected based on the purity of the starting material.

Acknowledgements/References

1. We would like to thank Nadim Ajami and Matthew Wong, Baylor College of Medicine, Houston, Texas for their analysis of our MiSeq data using the Viral Metagenomic Analysis Pipeline (VirMAP).
2. Kearse M, *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics 28*(12): 1647-1649, 2012.
3. Ajami NJ, *et al*. VirMAP, towards maximal viral information recovery from non-targeted sequencing. *In-review.*