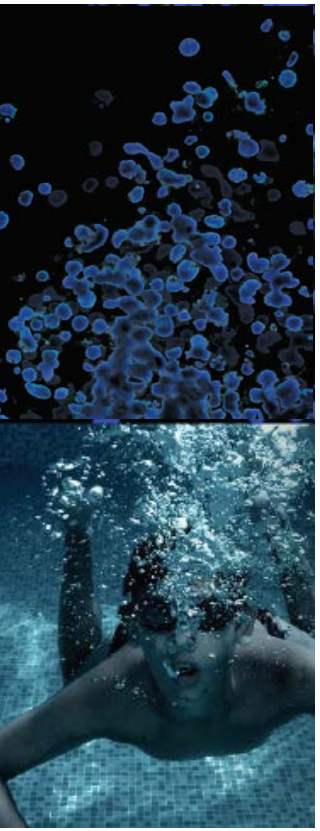
Exploring the Human Virome: The Importance of Standards



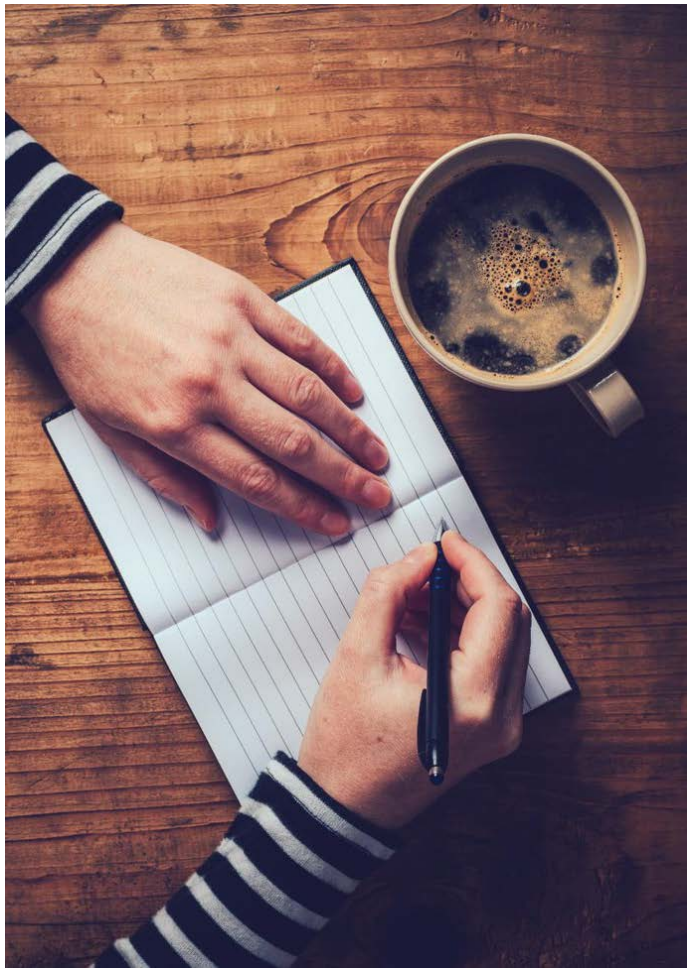
Heather Couch, PhD, PMP®
Program Manager

Credible Leads to Incredible™

Agenda

- ✓ History of ATCC
- ✓ Virome: what is a virome and why are we talking about it
- ✓ Advances in virome research
- ✓ Addressing the need for standards in virome research
- ✓ Development and evaluation of the ATCC Virome Standards
- ✓ Applications of the ATCC Virome Standards

- ✗ Assay development
- ✗ Recommend any specific assay, kit, protocol, or instrument



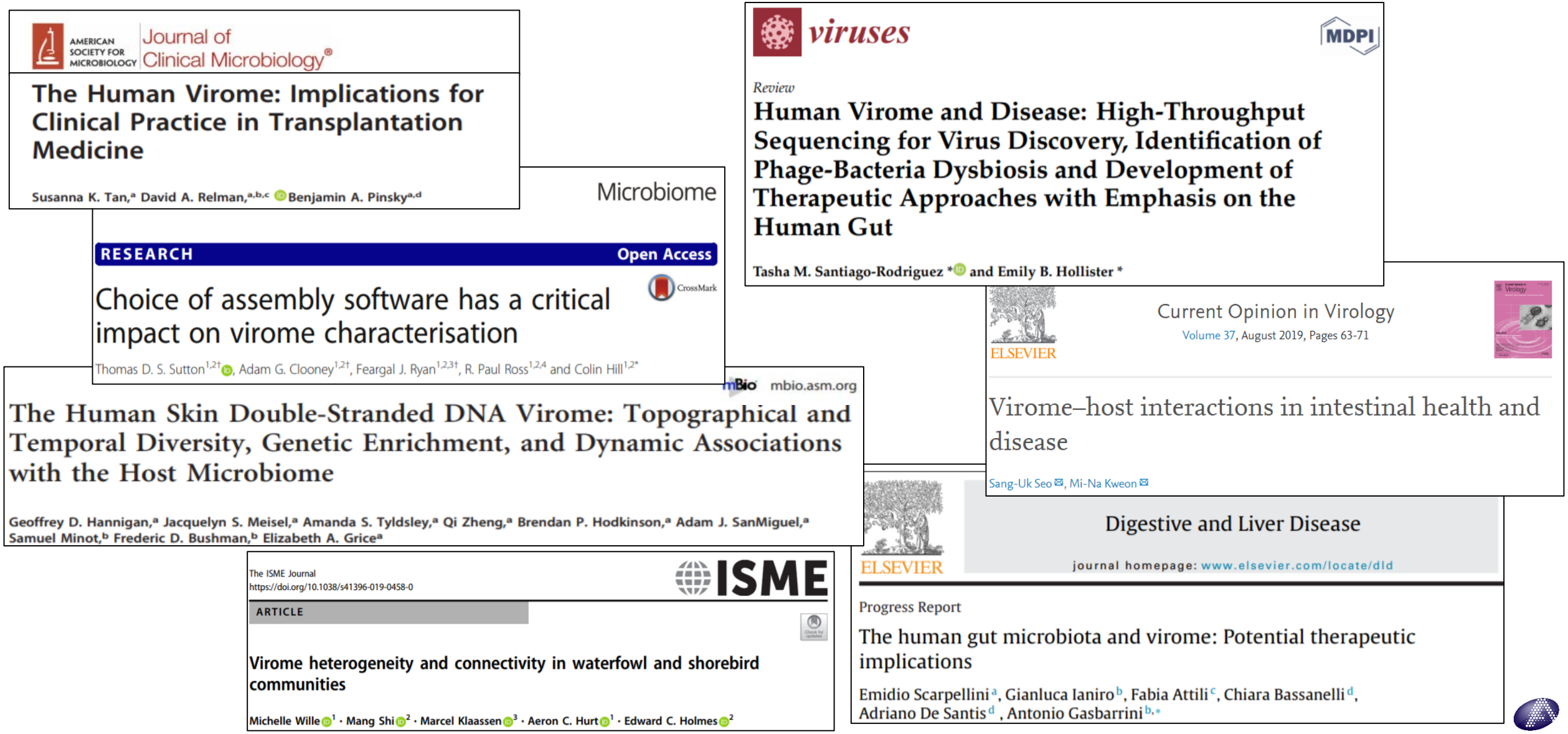
About ATCC

- Founded in 1925, ATCC is a non-profit organization with HQ in Manassas, VA, and an R&D and Services center in Gaithersburg, MD
- World's premier biological materials resource and standards development organization
  - 5,000 cell lines
  - 80,000 microorganisms
  - Genomic & synthetic nucleic acids
  - Media/reagents

- ATCC collaborates with and supports the scientific community with industry-standard biological products and innovative solutions
- Growing portfolio of products and services
- Sales and distribution in 150 countries, 18 international distributors
- Talented team of 450+ employees, over one-third with advanced degrees

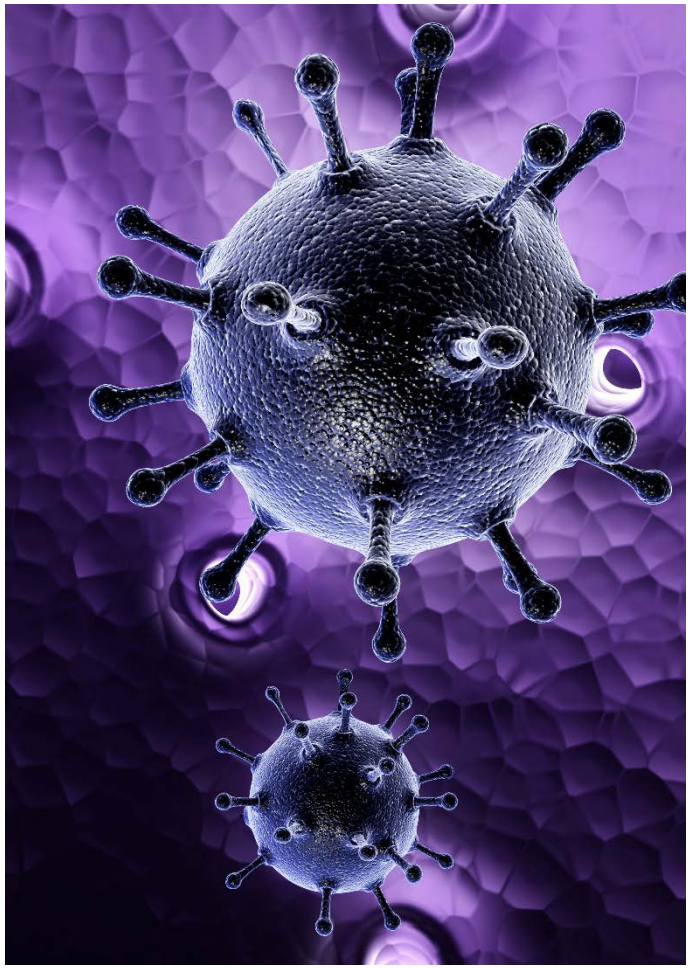
ATCC®

Virome: A growing area of research



ATCC

What is a virome and why are we talking about it?



- The virome is the collection of viruses associated with a particular ecosystem.
- Recent Pubmed search for “virome” revealed ~750 Virome publications.
  - Viral communities and their interaction with hosts
  - Human “virome” composition and its function
  - Impact of the virome on human health and disease
- Methods used for virome research:
  - Shotgun DNA or RNA sequencing
  - ~16 different virome bioinformatics tools
- Availability of controls for the study of viral communities:
  - Several experimental mock communities (prepared for internal use)
  - Only one viral mock community commercially available (ATCC)

Challenges in virome research

Studying viral communities is challenging because viruses do not possess phylogenetically conserved genes like 16S rRNA or ITS. Therefore, shotgun DNA and RNA sequencing is the most used approach to characterize viral communities.

VIROME WORKFLOW

Sample Collection

Nucleic acid extraction
- Storage
- Extraction method
- DNA/RNA recovery
- Quality & quantity

Sequencing

Shotgun
- NGS technology
- NGS library kit

Data Analysis

Bioinformatics
- NGS quality
- Depth & coverage
- Host sequences removal
- Taxonomy & database

Development of mock viral communities

Whole virus (ATCC® MSA-2008™)

ATCC viruses are selected based on genomic characteristics and clinical relevance

Individual viral cultures are quantified by droplet digital™ PCR (ddPCR)

Mixed in even proportions based on genome copy number (2 X 10^3 genome copies/µL per virus in 500 µL)

Six viruses detected by ddPCR and shotgun sequencing

Nucleic acids (ATCC® MSA-1008™)

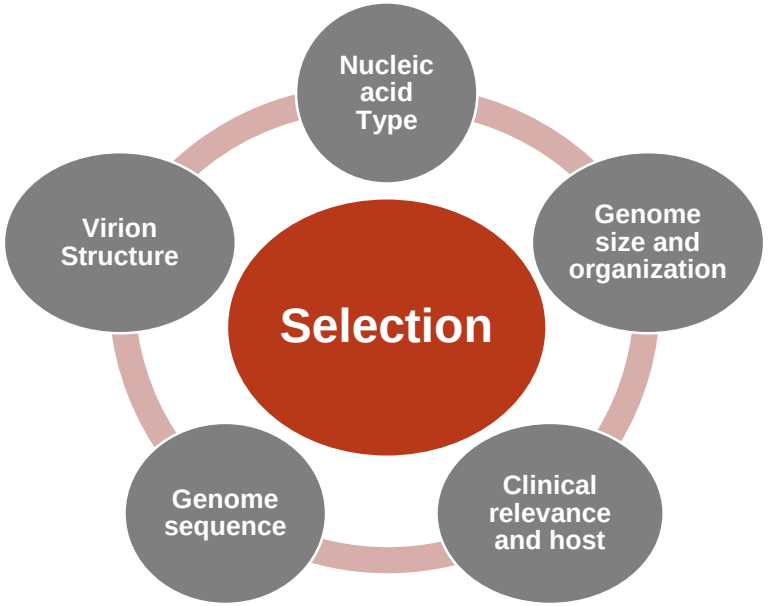
ATCC viruses are selected based on genomic characteristics and clinical relevance

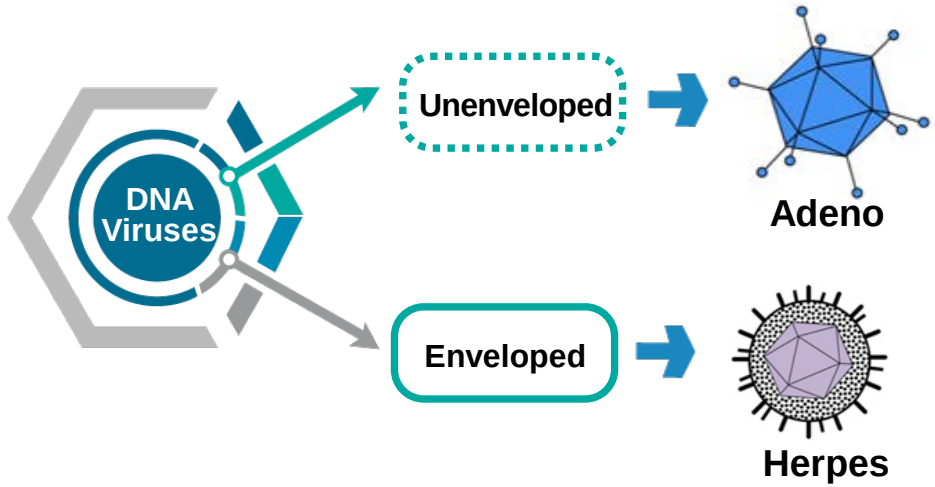
Individual viral nucleic acids are quantified by droplet digital™ PCR (ddPCR)

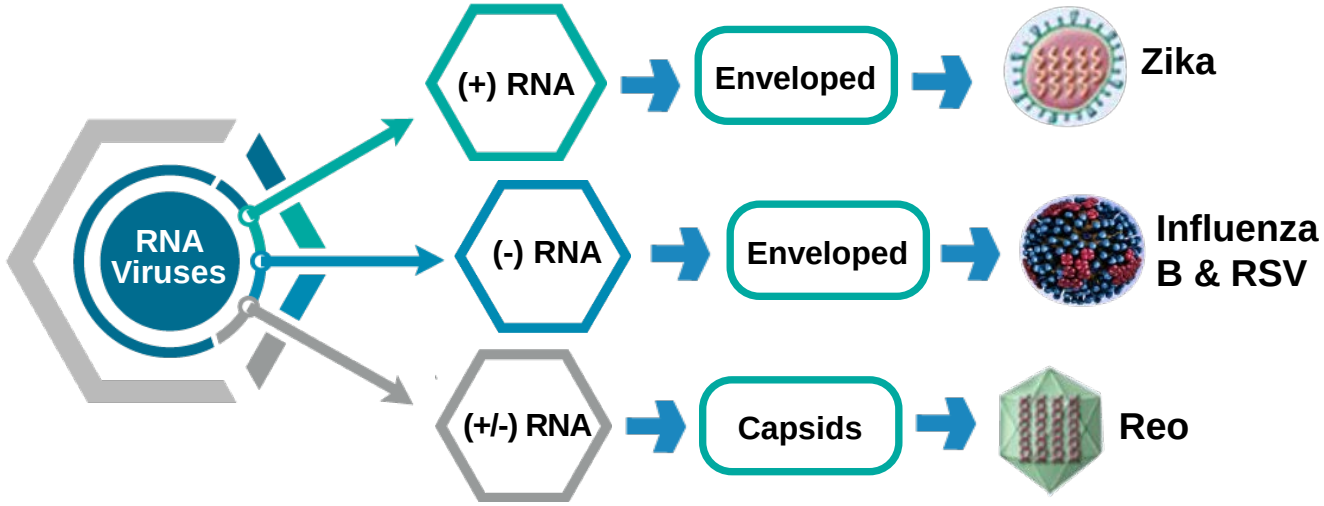
Mixed in even proportions based on genome copy number (2 X 10^4 genome copies/µL per virus in 50 µL)

Six viruses detected by ddPCR and shotgun sequencing

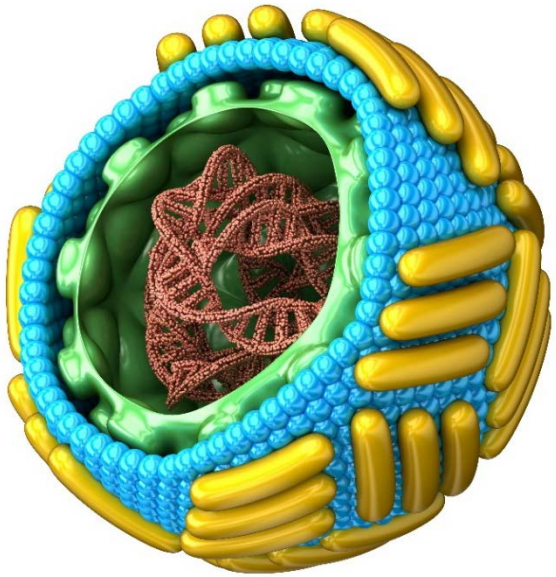
Selection of viruses







ATCC virome standards



Composition of Virome Standards
Human herpesvirus 5 strain AD169 (ATCC® VR-538™)
Human mastadenovirus strain F (ATCC® VR-931™)
Influenza B virus strain B/Florida/4/2006 (ATCC® VR-1804™)
Zika virus strain MR 766 (ATCC® VR-1838™)
Reovirus 3 strain Dearing (ATCC® VR-824™)
Human respiratory syncytial virus strain A2 (ATCC® VR-1540™)

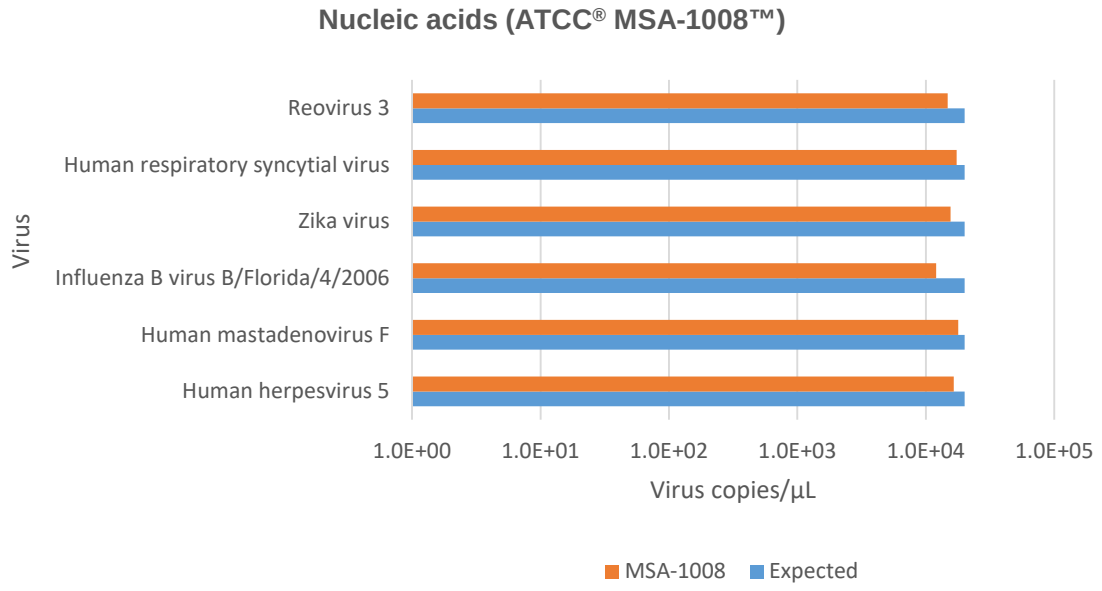
Standard	Preparation	ATCC® Catalog No.	Number of Viruses	Specification	Applications
Virome	Virus Mix	MSA-2008™	6	2 x 10^3 genome copies/µL per virus	Standards for virome assay development, optimization, verification, and validation; evaluating reproducibility; and use as a daily run quality control
	Nucleic Acid Mix	MSA-1008™	6	2 x 10^4 genome copies/µL per virus	

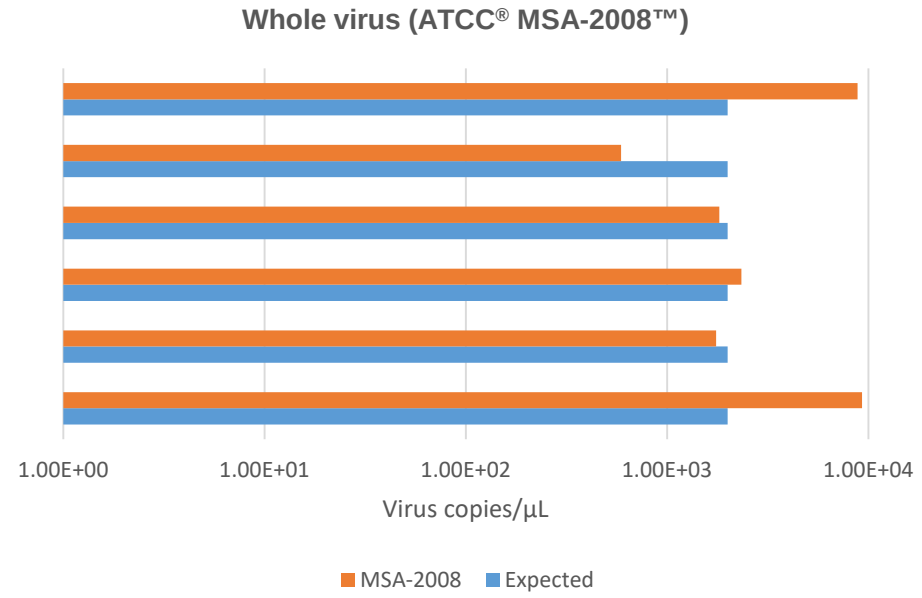
Quantification of virome standards

Genome copy number for each virus in the virome nucleic acid and whole virus mixtures was determined using individual droplet digital™ PCR assays

Genome Type	Virus Name (ATCC® Catalog No.)	MSA-1008™ (Virus copies/µL*)	MSA-2008™ (Virus copies/µL*)
dsDNA	Human herpesvirus 5 (VR-538™)	1.65E+04	9.31E+03
dsDNA	Human mastadenovirus F (VR-931™)	1.79E+04	1.75E+03
ss(-)RNA (8 segments)	Influenza B virus B/Florida/4/2006 (VR-1804™)	1.20E+04	2.34E+03
ss(+)RNA	Zika virus (VR-1838™)	1.56E+04	1.82E+03
ss(-)RNA	Human respiratory syncytial virus (VR-1540™)	1.74E+04	5.92E+02
dsRNA (10 segments)	Reovirus 3 (VR-824™)	1.48E+04	8.83E+03

* Mean copies/µL





Next-generation sequencing analysis

Percent of reads and genome coverage for each virus in the virome mixture were determined by mapping DNA or RNA shotgun sequence to individual reference genomes.

DNA Sequencing Analysis: 34.9% of reads (n= 9.2 million) are specific to DNA viruses. Overall, **65.1%** of reads (n= 17.2 million) were classified as host-associated reads.

Virus	GenBank ID	Genome size (kb)	# of reads (million)	% expected reads	% mapped reads (normalized)	Ave. Genome Coverage X
Human herpesvirus 5	X17403	229.4	7.9	50%	46.5%	7874
Human mastadenovirus F	NC_001454	34.2	1.3	50%	53.5%	9027

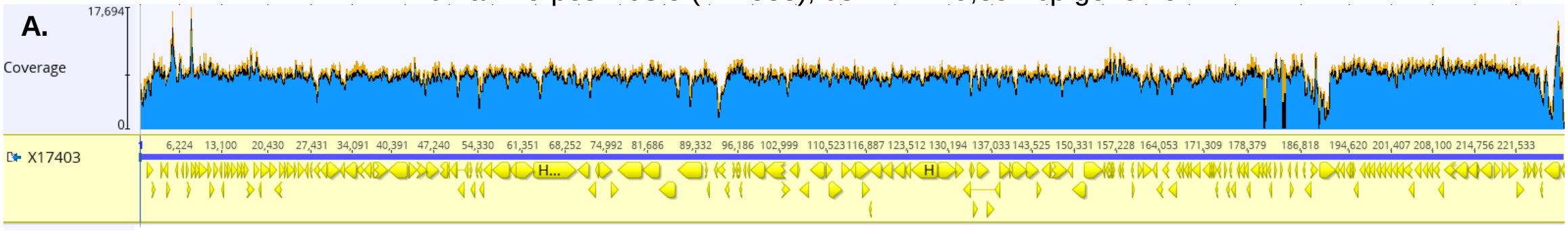
RNA Sequencing Analysis: 5.7% of reads (n= 2.2 million) are specific to RNA viruses. Overall, **94.3%** of reads (n= 37.0 million) were classified as host-associated reads.

Virus	GenBank ID	Genome size (kb)	# of reads (million)	% expected reads	% mapped reads (normalized)	Ave. Genome Coverage X
Influenza B virus B/Florida/4/2006	CY018365.1-CY018372.1	14.2	0.439	25%	20.6%	6239
Zika virus	KX830960	10.8	0.627	25%	38.9%	11814
Human respiratory syncytial virus	KT992094	15.2	0.443	25%	19.5%	5884
Reovirus 3	HM159613.1-HM159622.1	23.6	0.739	25%	21.0%	6204

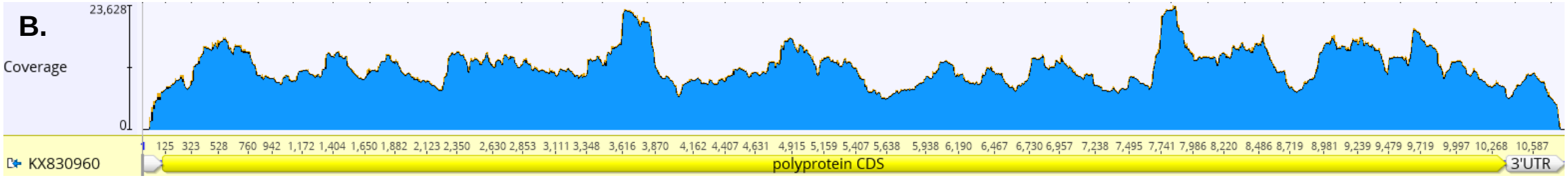
Next-generation sequencing analysis

DNA and RNA shotgun sequences were aligned to (A) Human herpesvirus 5 and (B) Zika virus reference genomes to evaluate the depth and distribution of the NGS coverage.

Human herpesvirus 5 (VR-538), dsDNA 229,354 bp genome



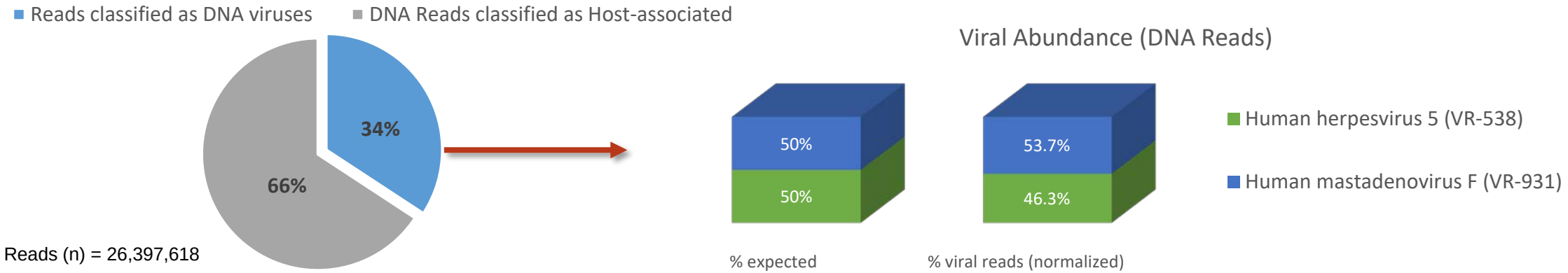
Zika virus (VR-1838), ss(+)RNA 10,807 bp genome

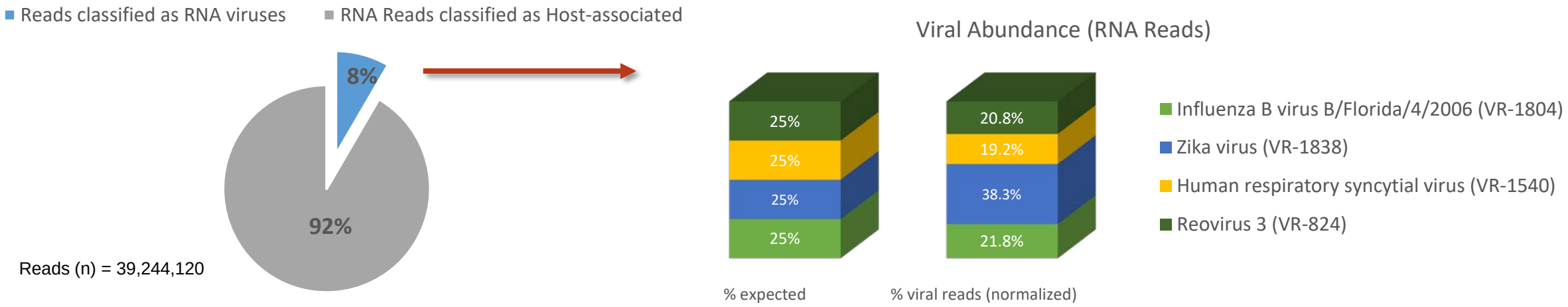


Alignments to reference genomes were generated with Geneious v11.1.4.

Bioinformatics analysis

Individual abundances in the virome nucleic acid mixture (MSA-1008™) were determined using the VirMAP analysis tool.



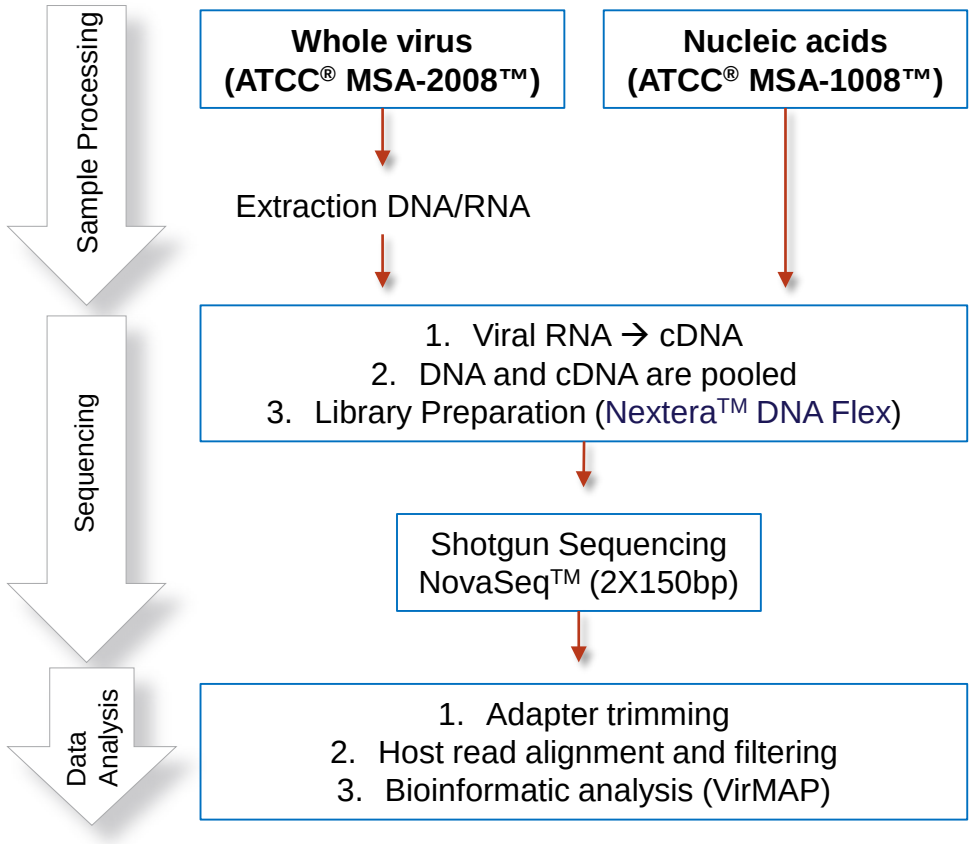


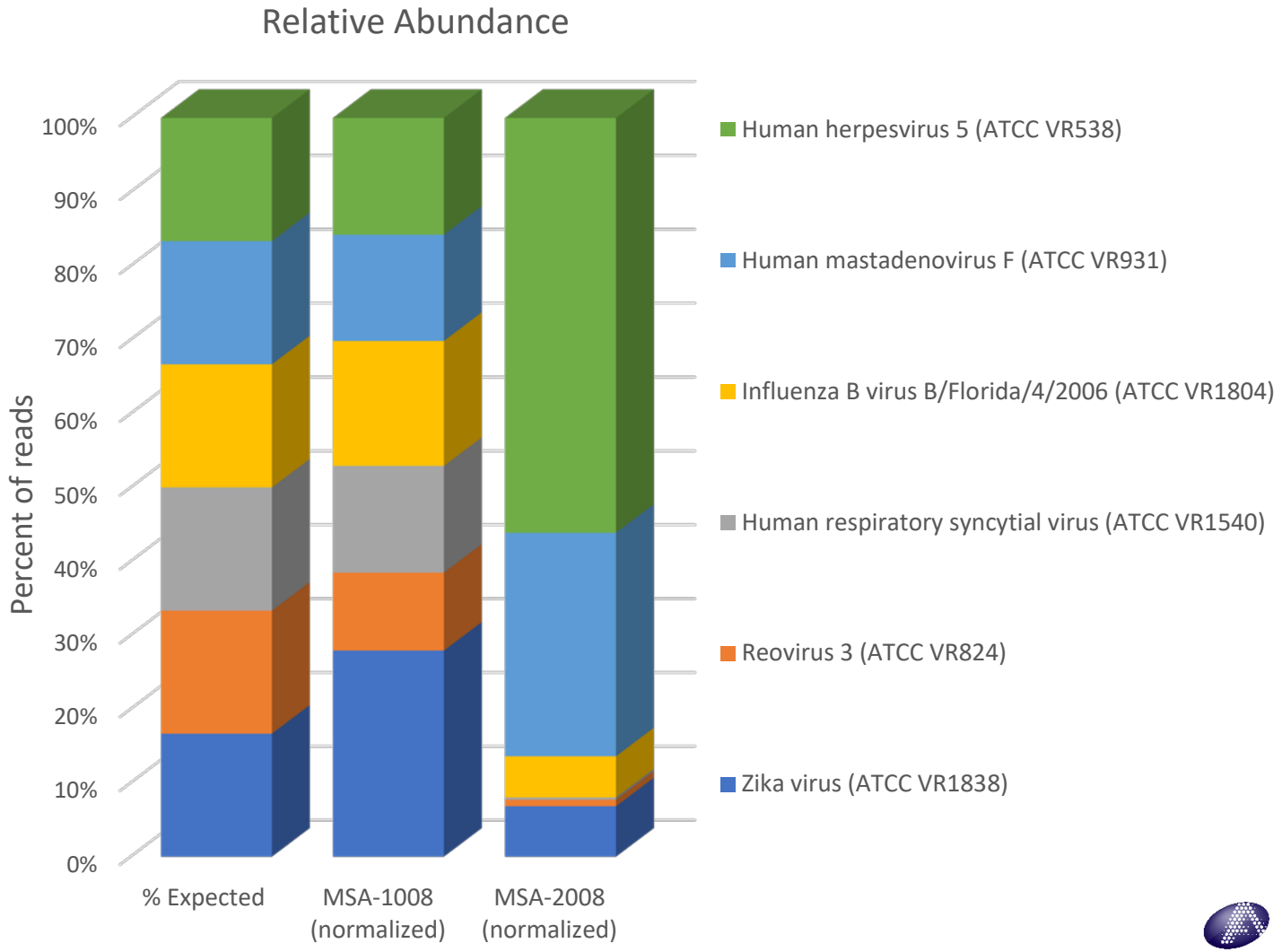
Analysis of the DNA and RNA shotgun data were generated with the VirMAP tool (Ajami, *et al.,* 2018. Maximal viral information recovery from sequence data using VirMAP. Nature Communications. https://doi.org/10.1038/s41467-018-05658-8. Data analysis was provided by Nadim Ajami, PhD.

Diversigen: Standardization of methodologies for analysis of human virome using ATCC® Virome Standards

Analysis of MSA-1008™ and MSA-2008™ identified all of the viruses in each mix.

Diversigen Virome Analysis Workflow



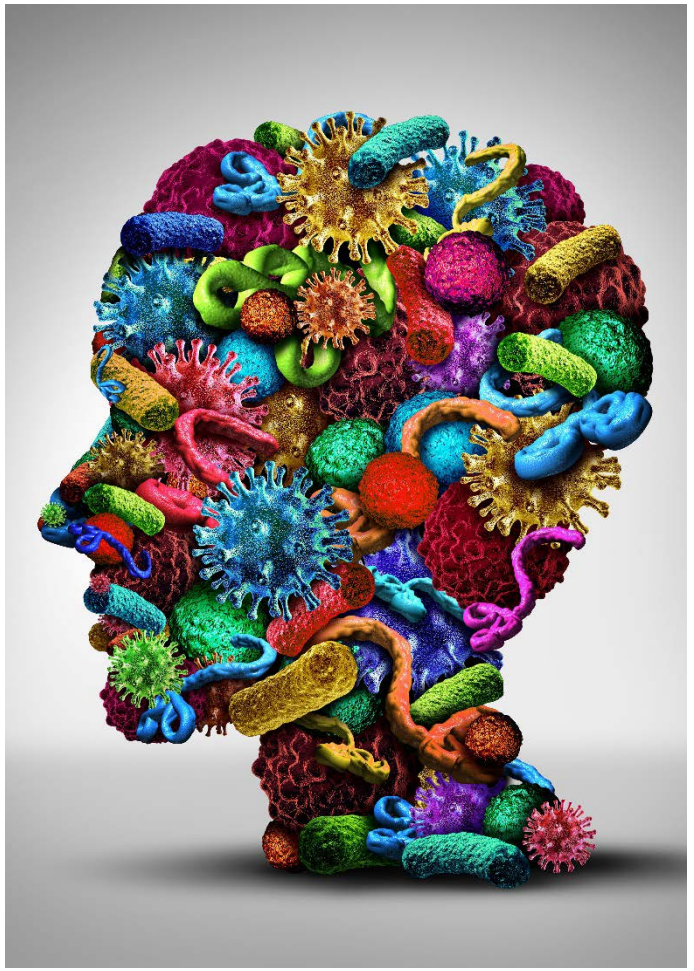


Conclusions

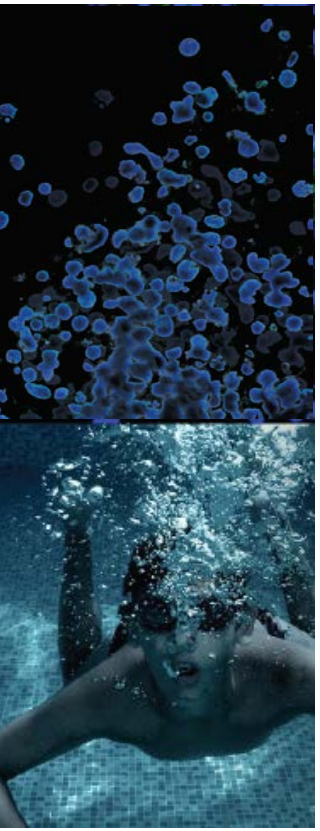
- The standardized concentrations 2 x 10^4 genome copies/µL per virus in MSA-1008™ is sufficient for NGS library preparation and data analysis.
- All DNA and RNA viruses in both MSA-1008™ and MSA-2008 ™ were identified.
- Using the VirMAP analysis platform, we:
  - removed host DNA
  - identified all individual viruses
  - assessed the quality and sequencing depth of each virus
- ATCC® virome standards are suitable for assay validation, assessing run-to-run reproducibility, and assessing different bioinformatics analysis tools.
- The use of controls is essential.

Acknowledgements

- Briana Benton, BS
- Juan Lopera, PhD
- Sujatha Rashid, PhD
- Jessica Shifflett, MS
- Teresa Towle, BS
- Dev Mittar, PhD
- Reed Shabman, PhD
- Tasha Santiago-Rodriguez, PhD (Diversigen)
- Emily Hollister, PhD (Diversigen)
- Nadim Ajami, PhD (BCM - Alkek Center for Metagenomics and Microbiome Research)
- Matthew Wong, BS (BCM - Alkek Center for Metagenomics and Microbiome Research)



ATCC®



Thank you!

Questions?

Credible Leads to Incredible™