Viral metagenomics and the use of standards: from biology to clinical applications

Tasha M. Santiago-Rodriguez, Ph.D.

Bioinformatician R&D

Diversigen, Houston, Texas

Diversigen

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 largest, most diverse biological materials and information resource for cells and microbes – the "*gold standard*"
- Innovative R&D company featuring gene editing, microbiome, NGS, advanced models
- cGMP biorepository
- Partner with government, industry, and academia
- Leading global supplier of authenticated cell lines, viral and microbial standards
- Sales and distribution in 150 countries, 19 international distributors
- Talented team of 450+ employees, over one-third with advanced degrees

ATCC®

Agenda

- Diversigen overview
- What do we know so far about virome research?
- Biases in virome research
- Application of mock communities in virome research
- Potential applications of standards in the detection of ongoing and future pathogenic human viruses: considerations from SARS-CoV 2

Agenda

- Diversigen overview
- What do we know so far about virome research?
- Biases in virome research
- Application of mock communities in virome research
- Potential applications of standards in the detection of ongoing and future pathogenic human viruses: considerations from SARS-CoV 2

Diversigen as part of OraSure Technologies

DIVERSIGEN

MICROBIOME SERVICES

Metagenomic sequencing pipelines | Germ-free | qPCR | Project consulting | Bioinformatics

LOCATION

Houston, Texas - located in the heart of the Texas Medical Center

2014

YEAR FOUNDED

Operational since January 2015

PEOPLE

Highly talented and experienced staff, including 12 PhD scientists

ORIGIN

TMC TEXAS MEDICAL CENTER

21 Renowned Hospitals	8 Research Institutions	10 Academic Institutions	3 Medical Schools	1 Dental School	2 Pharmacy Schools

CMMr Alkek Center for Metagenomics and Microbiome Research

Centre for Metagenomics and Microbiome Research (CMMR)

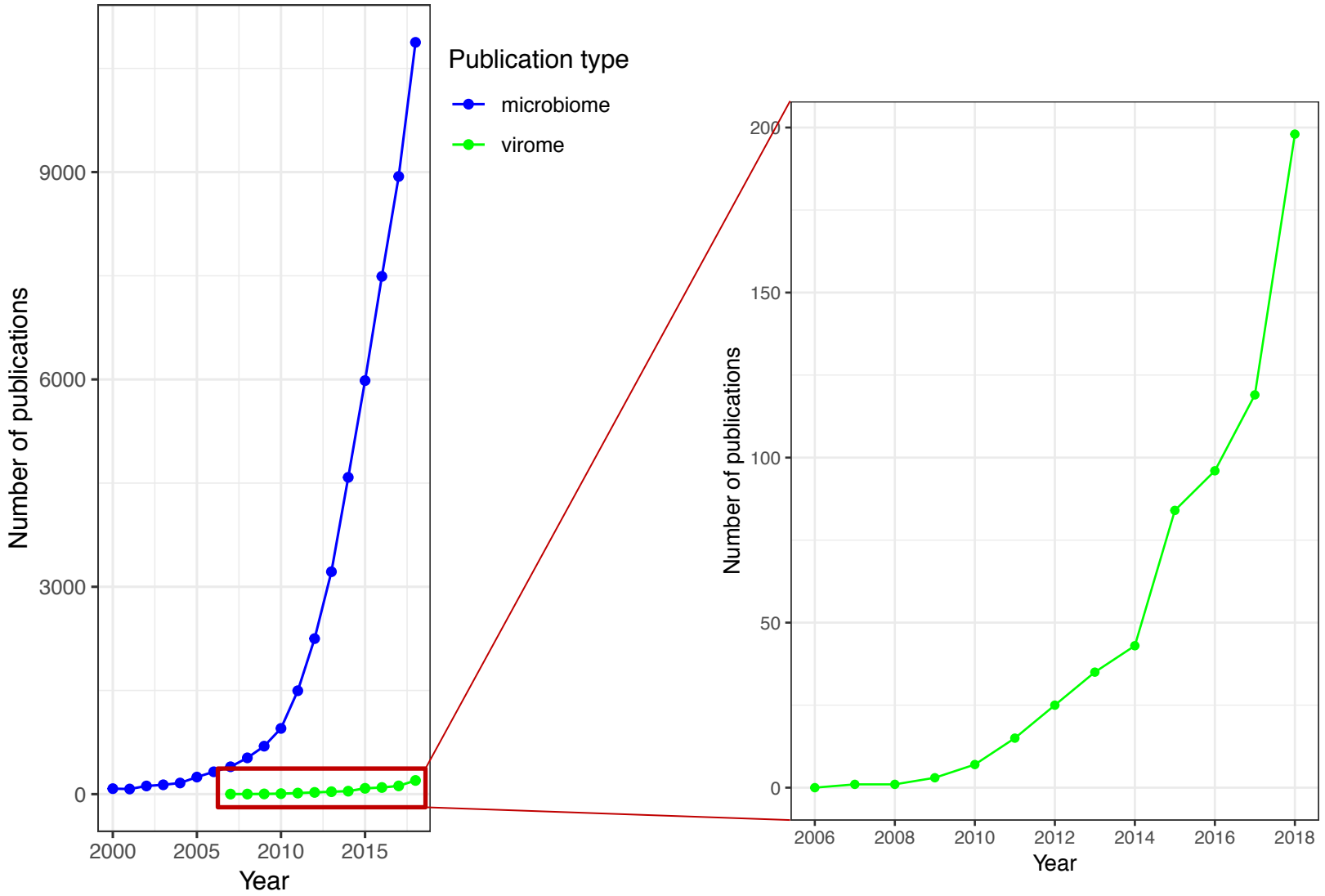
Microbiome Exploration | Microbial ecology, modeling, and dissection Therapeutic development | Policy and outreach | Education | Translation

OraSure Technologies | DNAgenotek™ | Diversigen® | novosanis

Agenda

- Diversigen overview
- What do we know so far about virome research?
- Biases in virome research
- Application of mock communities in virome research
- Potential applications of standards in the detection of ongoing and future pathogenic human viruses: considerations from SARS-CoV 2

Metagenomic studies focus on bacteria



- Viruses are more difficult to study
- No universal gene
- DNA or RNA
- Databases are biased towards pathogens and known phages

Diversigen

Why are viruses important?

Viruses can be found throughout the body (gut, oral cavity, skin, bladder, blood, respiratory tract)

Viruses, particularly phages, are estimated to outnumber bacteria by a factor of 10:1 in the human gut

Viral communities vary in the context of health and disease, even in the absence of known pathogens

Viral communities change in response to external factors (antibiotic-treatment, intimate contact and diet)

Viral profiles of stool can provide insight into diets (i.e., dairy product-associated phage, plant pathogens)

New studies are shedding light on the ability of the virome to influence immune response

Virus-mediated approaches show promise in treating antimicrobial resistant infections

Virome may provide answers to microbially-mediated phenomena where bacterial studies have come up short

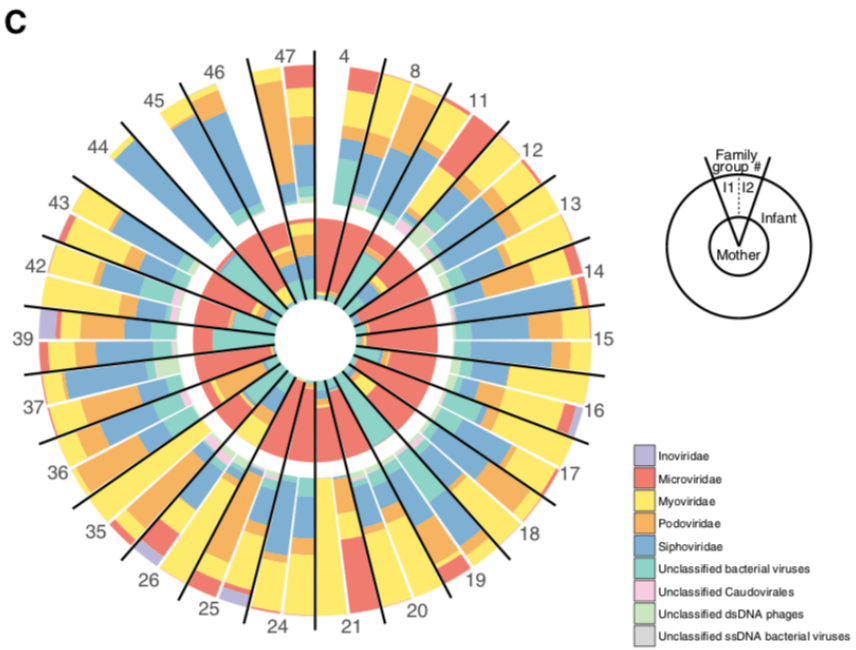
Diversigen®

Viruses are acquired from birth and modulated by diet

RESEARCH Open Access

Discordant transmission of bacteria and viruses from mothers to babies at birth

Rabia Maqsood[1,2†], Rachel Rodgers[3†], Cynthia Rodriguez[3], Scott A. Handley[4], I. Malick Ndao[3], Phillip I. Tarr[3,5], Barbara B. Warner[3], Efrem S. Lim[1,2*] and Lori R. Holtz[3*]

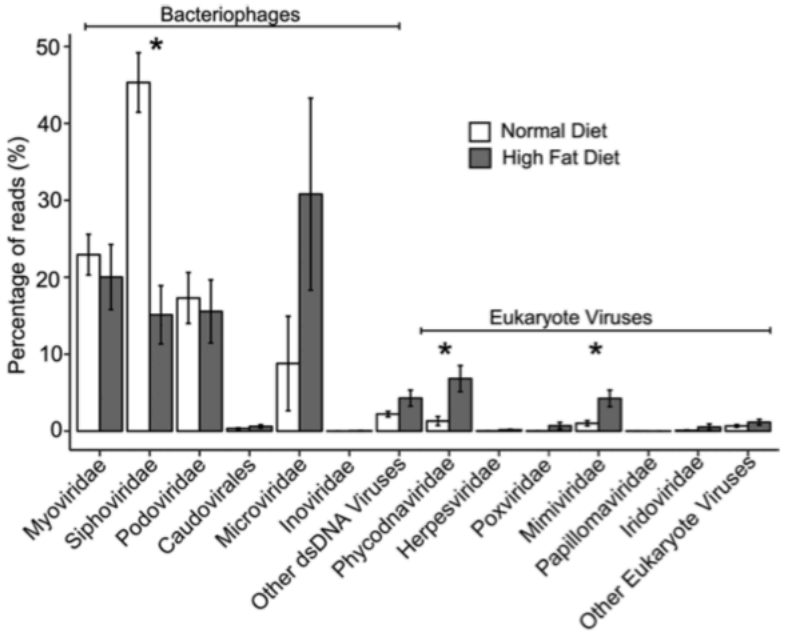


AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE
Host-Microbe Biology

Fecal Viral Community Responses to High-Fat Diet in Mice

Anjelique Schulfer,[a] Tasha M. Santiago-Rodriguez,[b] Melissa Ly,[b] Joshua M. Borin,[c] Jessica Chopyk,[b] Martin J. Blaser,[a,d] David T. Pride[b,e]



Diversigen®

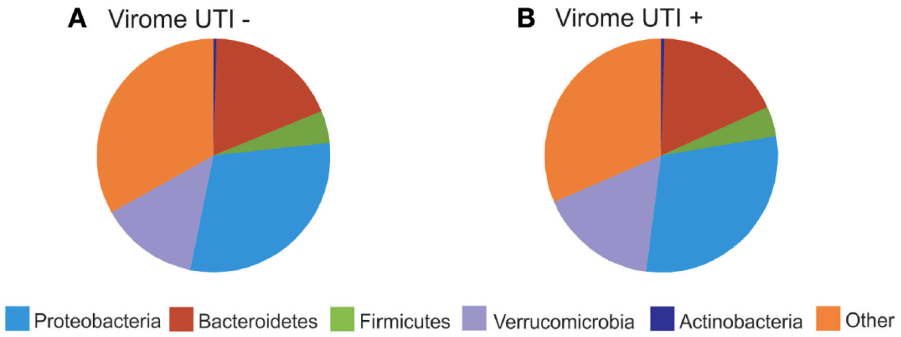
Viruses in samples previously thought to be sterile

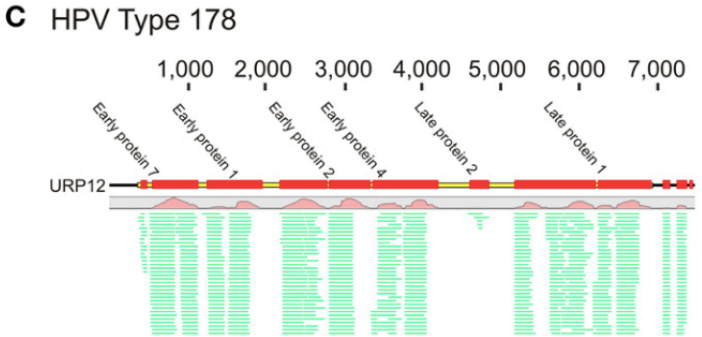
ORIGINAL RESEARCH ARTICLE
Front. Microbiol., 23 January 2015 | https://doi.org/10.3389/fmicb.2015.00014

The human urine virome in association with urinary tract infections

Tasha M. Santiago-Rodriguez[1], Melissa Ly[1], Natasha Bonilla[2] and David T. Pride[1,3*]

[1]Department of Pathology, University of California, San Diego, San Diego, CA, USA
[2]Department of Biology, San Diego State University, San Diego, CA, USA
[3]Department of Medicine, University of California, San Diego, San Diego, CA, USA



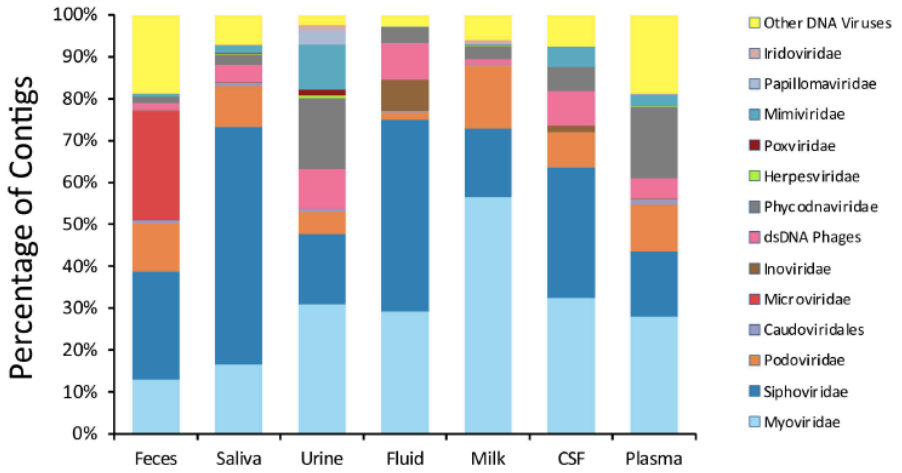


frontiers in Microbiology

ORIGINAL RESEARCH
published: 06 September 2019
doi: 10.3389/fmicb.2019.02061

The Virome of Cerebrospinal Fluid: Viruses Where We Once Thought There Were None

Chandrabali Ghose[1], Melissa Ly[2], Leila K. Schwanemann[2], Ji Hyun Shin[2], Katayoon Atab[2], Jeremy J. Barr[3], Mark Little[4], Robert T. Schooley[5], Jessica Chopyk[2] and David T. Pride[2,5]*



Diversigen

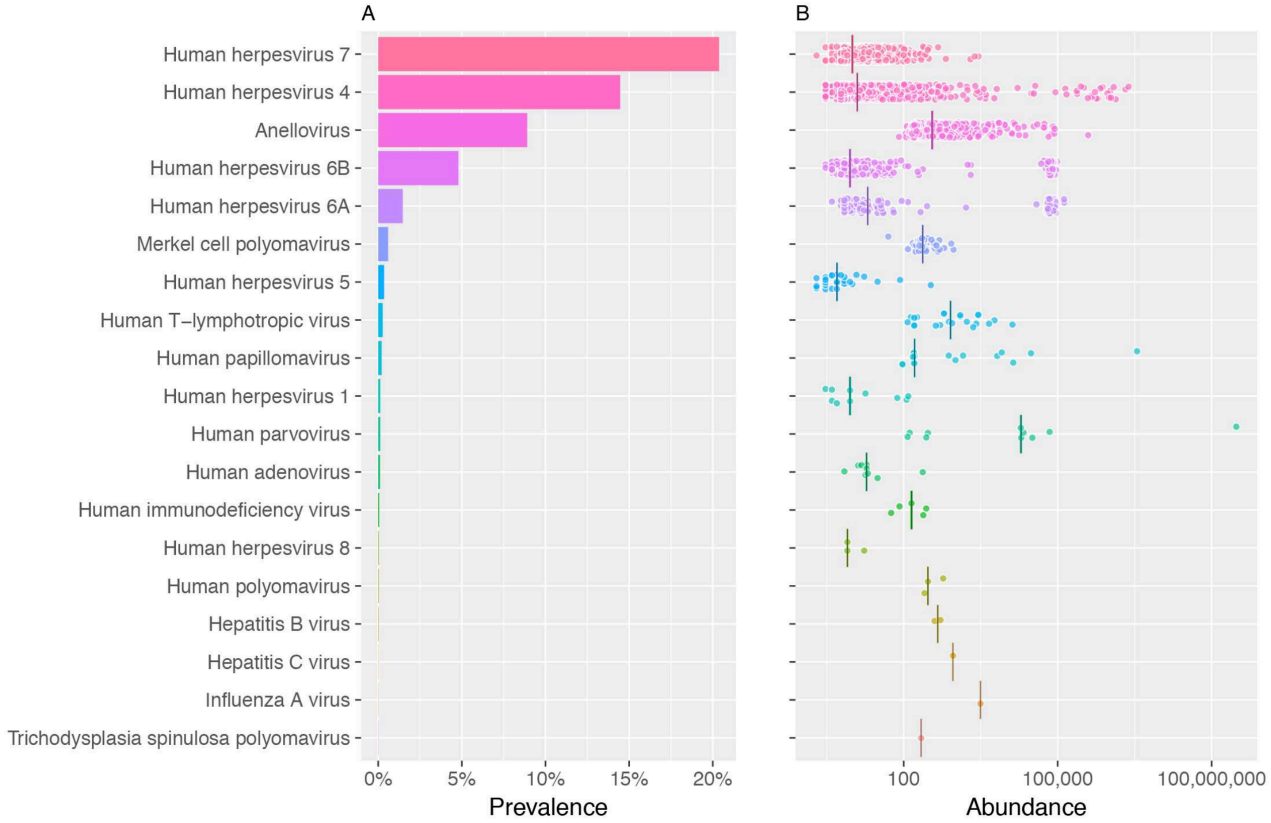
Viruses in samples previously thought to be sterile

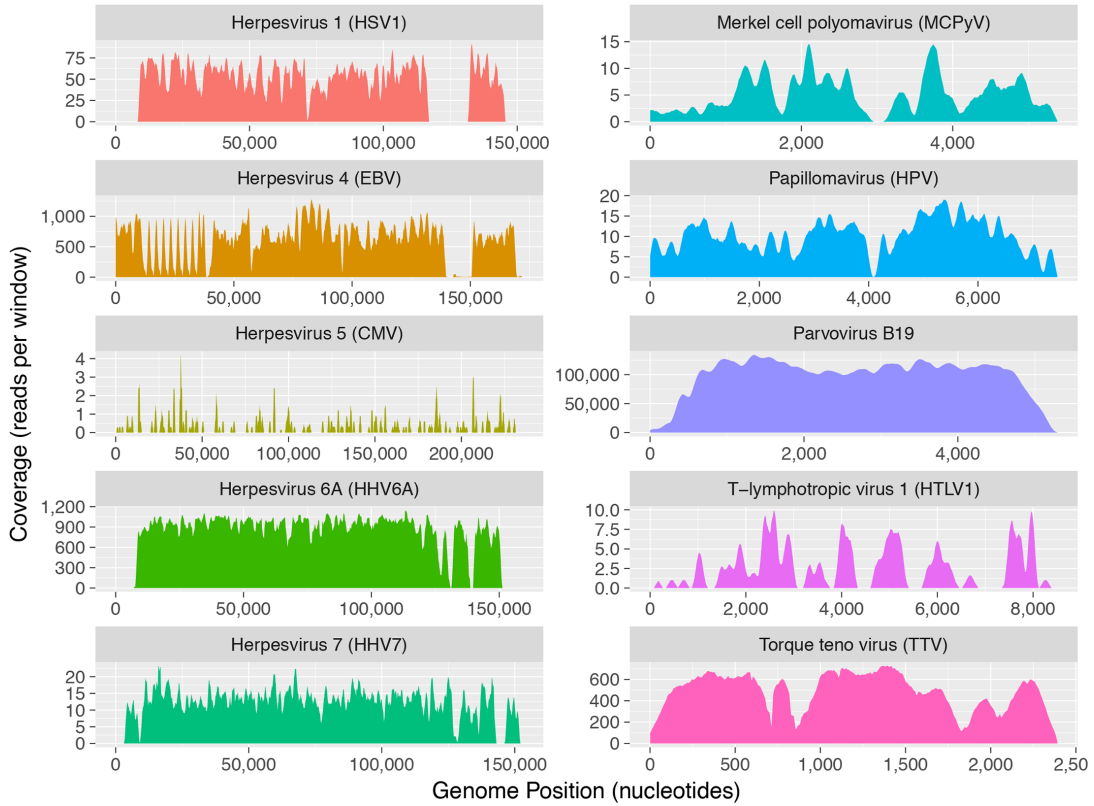
RESEARCH ARTICLE

The blood DNA virome in 8,000 humans

Ahmed Moustafa[1], Chao Xie[2], Ewen Kirkness[1], William Biggs[1], Emily Wong[1], Yaron Turpaz[2], Kenneth Bloom[1], Eric Delwart[3], Karen E. Nelson[4], J. Craig Venter[1,4]*, Amalio Telenti[1,4]*

1 Human Longevity Inc., San Diego, California, United States of America, **2** Human Longevity Singapore Pte. Ltd., Singapore, **3** Blood Systems Research Institute, Department of Laboratory Medicine, University of California San Francisco, San Francisco, California, United States of America, **4** J. Craig Venter Institute, La Jolla, California, United States of America





Diversigen

Human viruses can predict disease risk

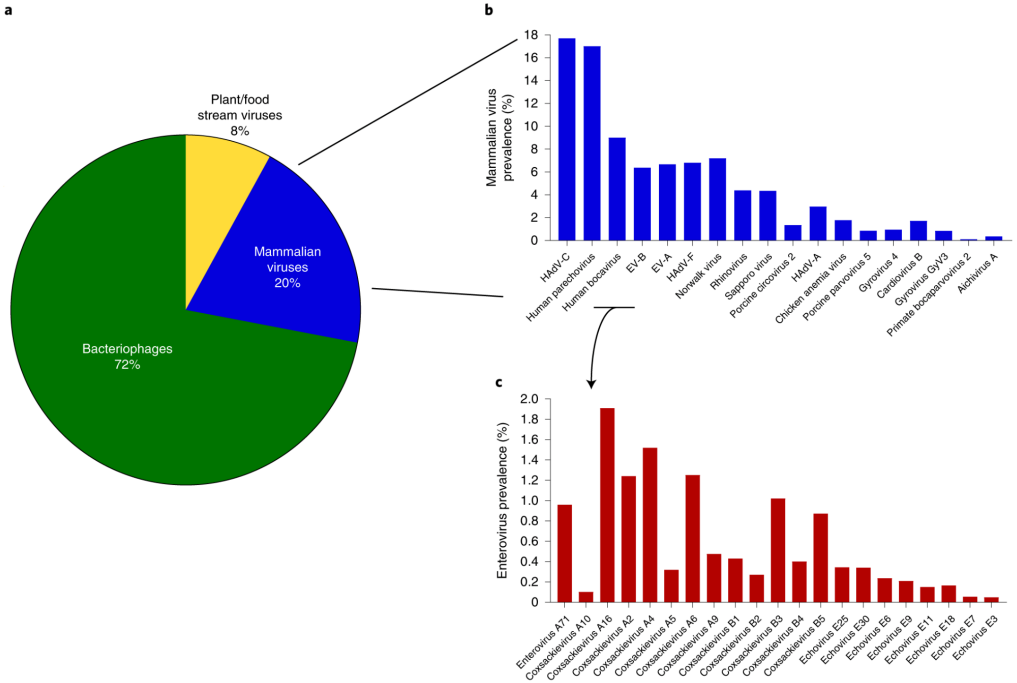
nature medicine

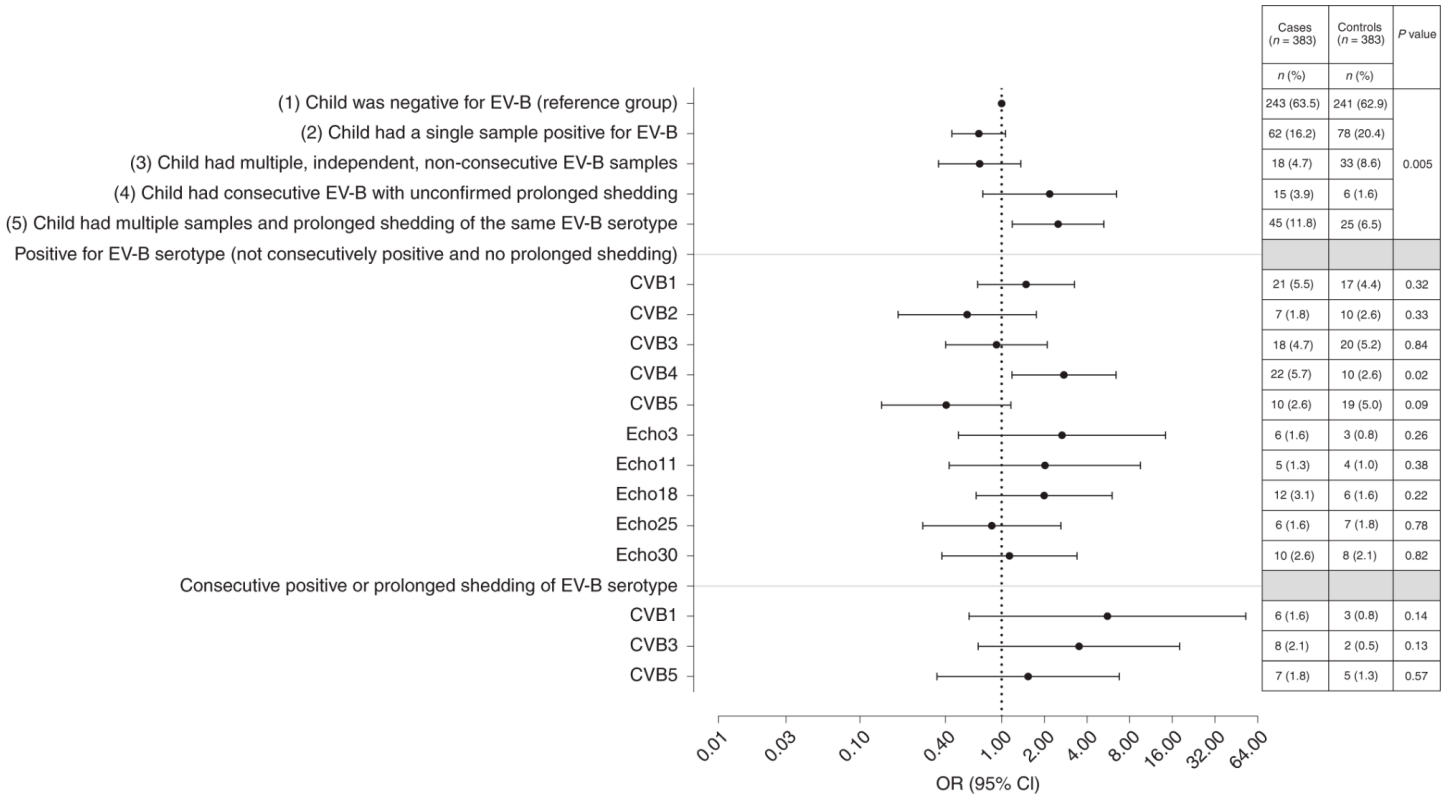
LETTERS

https://doi.org/10.1038/s41591-019-0667-0

Prospective virome analyses in young children at increased genetic risk for type 1 diabetes

Kendra Vehik[1*], Kristian F. Lynch[1], Matthew C. Wong[2], Xiangjun Tian[2], Matthew C. Ross[2], Richard A. Gibbs[3], Nadim J. Ajami[2], Joseph F. Petrosino[2], Marian Rewers[4], Jorma Toppari[5,6], Anette G. Ziegler[7,8,9], Jin-Xiong She[10], Ake Lernmark[11], Beena Akolkar[12], William A. Hagopian[13], Desmond A. Schatz[14], Jeffrey P. Krischer[1], Heikki Hyöty[15,16], Richard E. Lloyd[2] and the TEDDY Study Group[17]





	Cases (n = 383)	Controls (n = 383)	P value
	n (%)	n (%)	
(1) Child was negative for EV-B (reference group)	243 (63.5)	241 (62.9)	
(2) Child had a single sample positive for EV-B	62 (16.2)	78 (20.4)	
(3) Child had multiple, independent, non-consecutive EV-B samples	18 (4.7)	33 (8.6)	0.005
(4) Child had consecutive EV-B with unconfirmed prolonged shedding	15 (3.9)	6 (1.6)	
(5) Child had multiple samples and prolonged shedding of the same EV-B serotype	45 (11.8)	25 (6.5)	
Positive for EV-B serotype (not consecutively positive and no prolonged shedding)			
CVB1	21 (5.5)	17 (4.4)	0.32
CVB2	7 (1.8)	10 (2.6)	0.33
CVB3	18 (4.7)	20 (5.2)	0.84
CVB4	22 (5.7)	10 (2.6)	0.02
CVB5	10 (2.6)	19 (5.0)	0.09
Echo3	6 (1.6)	3 (0.8)	0.26
Echo11	5 (1.3)	4 (1.0)	0.38
Echo18	12 (3.1)	6 (1.6)	0.22
Echo25	6 (1.6)	7 (1.8)	0.78
Echo30	10 (2.6)	8 (2.1)	0.82
Consecutive positive or prolonged shedding of EV-B serotype			
CVB1	6 (1.6)	3 (0.8)	0.14
CVB3	8 (2.1)	2 (0.5)	0.13
CVB5	7 (1.8)	5 (1.3)	0.57

Diversigen

Viruses can be used to treat diseases

Bacteriophage targeting of gut bacterium attenuates alcoholic liver disease

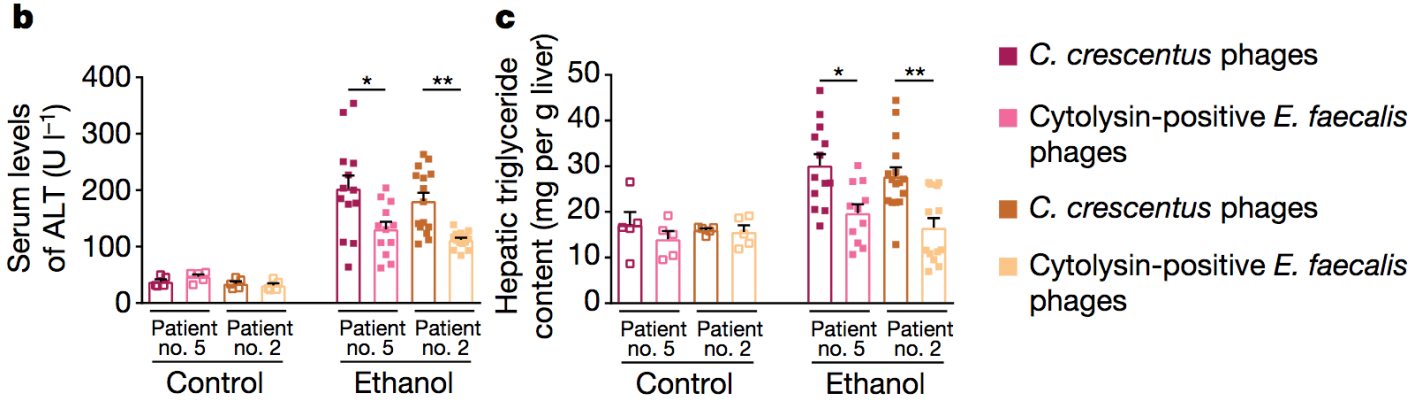
https://doi.org/10.1038/s41586-019-1742-x

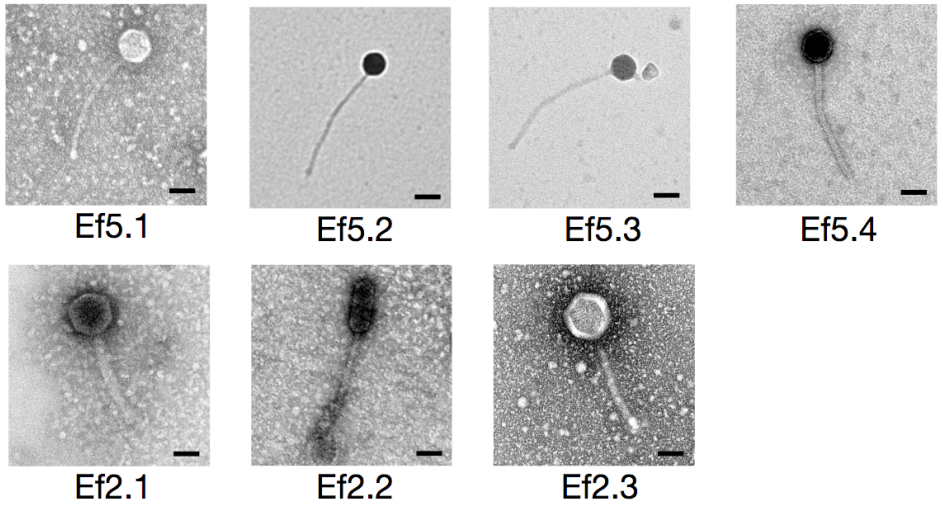
Received: 9 April 2019

Accepted: 2 October 2019

Published online: 13 November 2019

Yi Duan[1,2,29], Cristina Llorente[1,2,29], Sonja Lang[1], Katharina Brandl[3], Huikuan Chu[1], Lu Jiang[1,2], Richard C. White[4], Thomas H. Clarke[4], Kevin Nguyen[4], Manolito Torralba[5], Yan Shao[6], Jinyuan Liu[7], Adriana Hernandez-Morales[8], Lauren Lessor[9], Imran R. Rahman[10], Yukiko Miyamoto[1], Melissa Ly[11], Bei Gao[1], Weizhong Sun[1], Roman Kiesel[1], Felix Hutmacher[1], Suhan Lee[1], Meritxell Ventura-Cots[12], Francisco Bosques-Padilla[13], Elizabeth C. Verna[14], Juan G. Abraldes[15], Robert S. Brown Jr[16], Victor Vargas[17,18], Jose Altamirano[17], Juan Caballería[18,19], Debbie L. Shawcross[20], Samuel B. Ho[1,2], Alexandre Louvet[21], Michael R. Lucey[22], Philippe Mathurin[21], Guadalupe Garcia-Tsao[23,24], Ramon Bataller[12], Xin M. Tu[7], Lars Eckmann[1], Wilfred A. van der Donk[10,25,26], Ry Young[8,9], Trevor D. Lawley[6], Peter Stärkel[27], David Pride[1,11,28], Derrick E. Fouts[4] & Bernd Schnabl[1,2,28]*





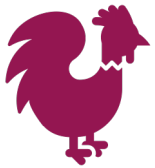
Diversigen

Ongoing and future applications of viral metagenomics

Natural history – what is present, where, and when?

Identifying potential viral associations in unexplained illnesses

Identifying novel viral relationships with health and disease



Veterinary medicine – identifying potential novel, emerging and re-emerging pathogens

Identifying viral associations as risk factors for disease

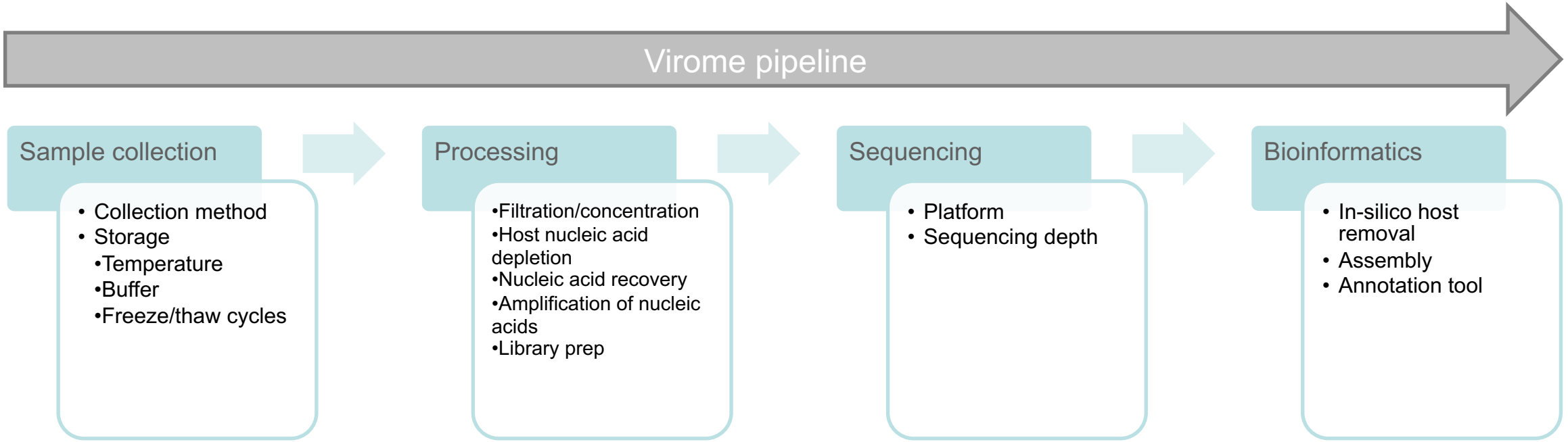
Surveillance for zoonotic diseases

Diversigen®

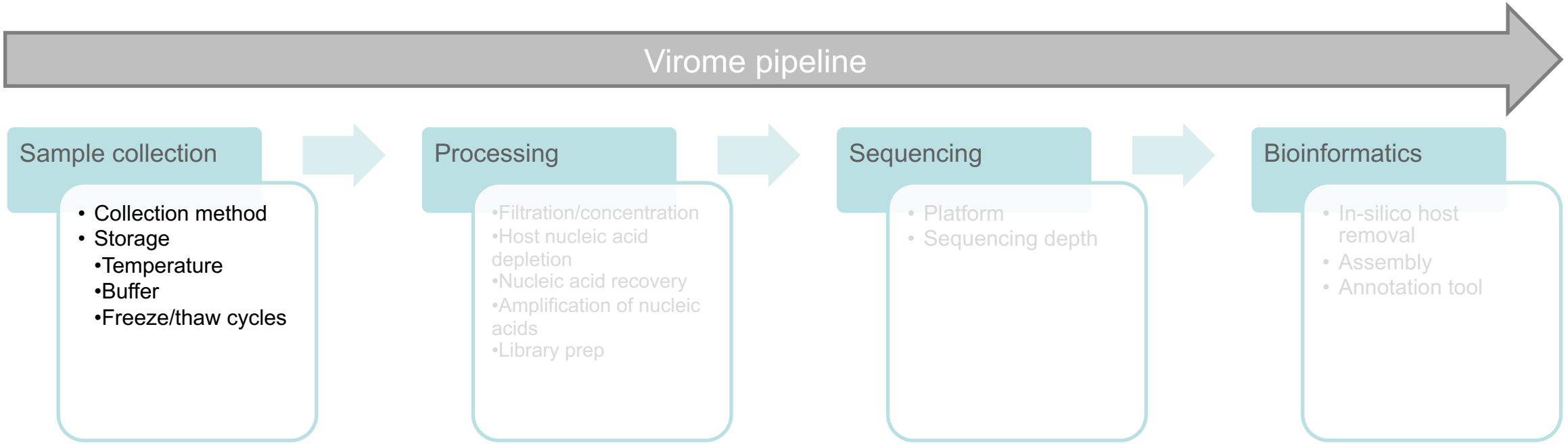
Agenda

- Diversigen overview
- What do we know so far about virome research?
- Biases in virome research
- Application of mock communities in virome research
- Potential applications of standards in the detection of ongoing and future pathogenic human viruses: considerations from SARS-CoV 2

Biases in virome research



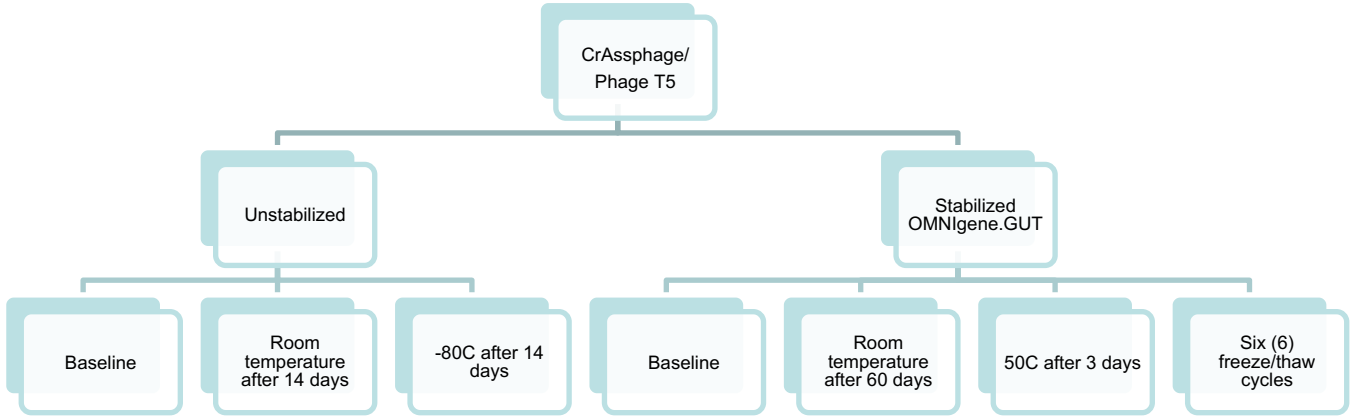
Biases in virome research - Sample collection

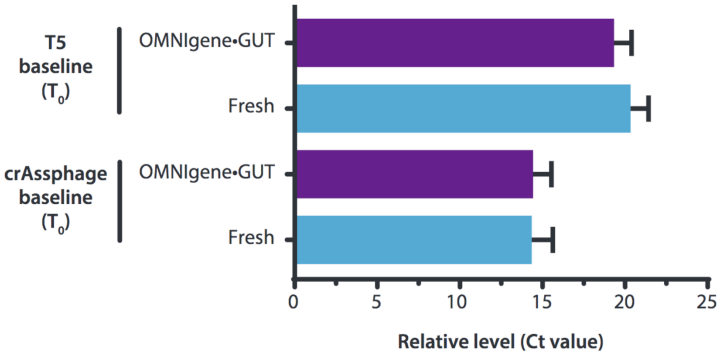


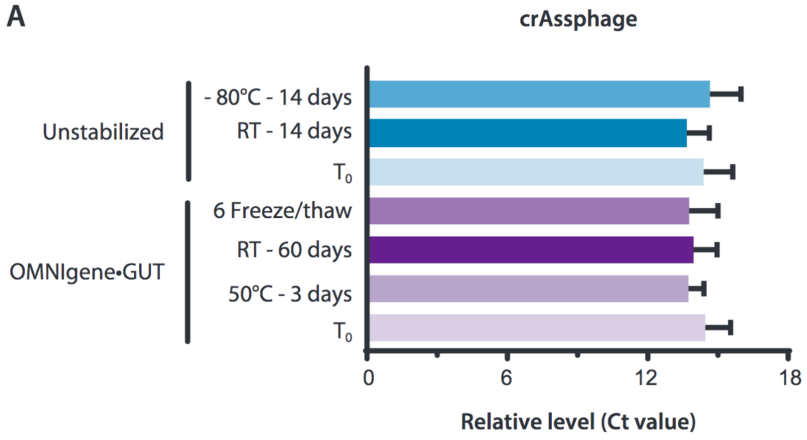
Biases in virome research - Sample collection

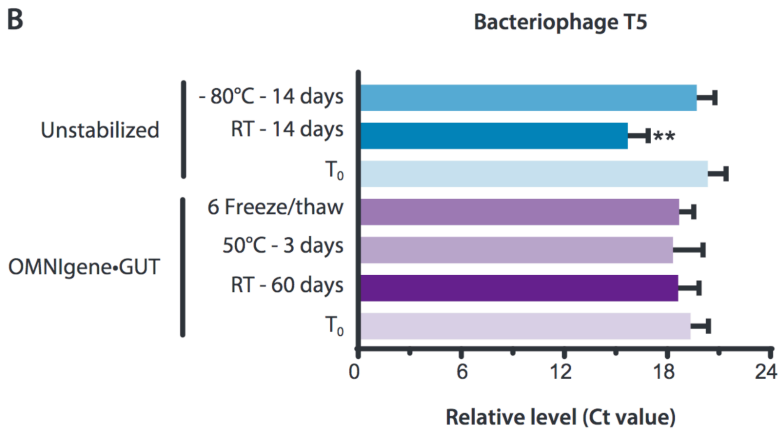
OMNIgene®·GUT

OMNIgene®·GUT accurately captures and stabilizes the human fecal dsDNA virome



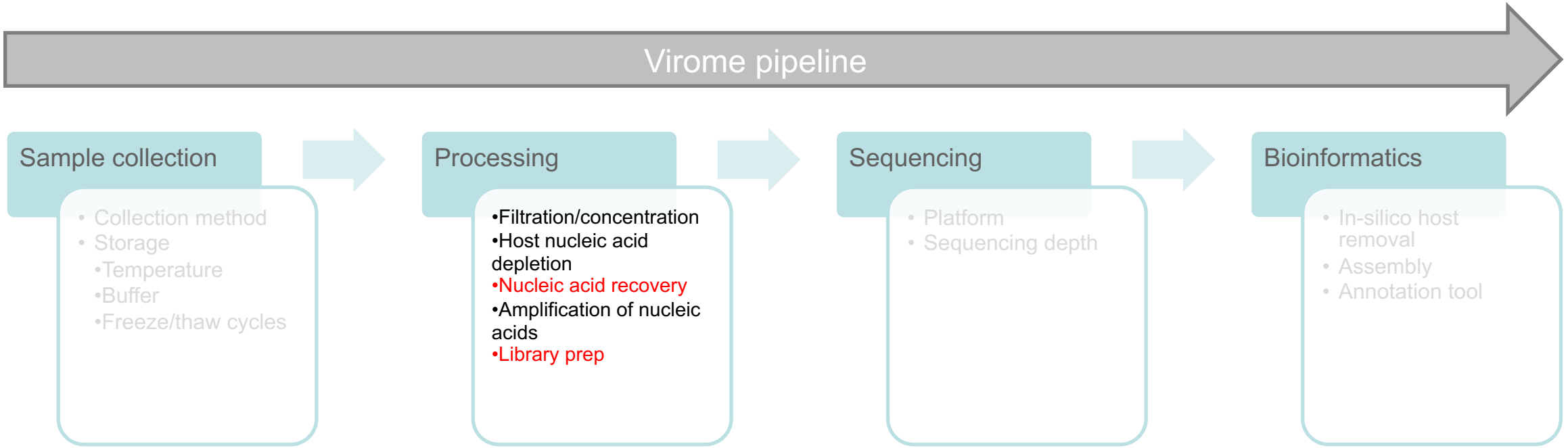






Diversigen®

Biases in virome research - Processing



Biases in virome research - Processing

Methodology | Open Access | Published: 28 June 2018

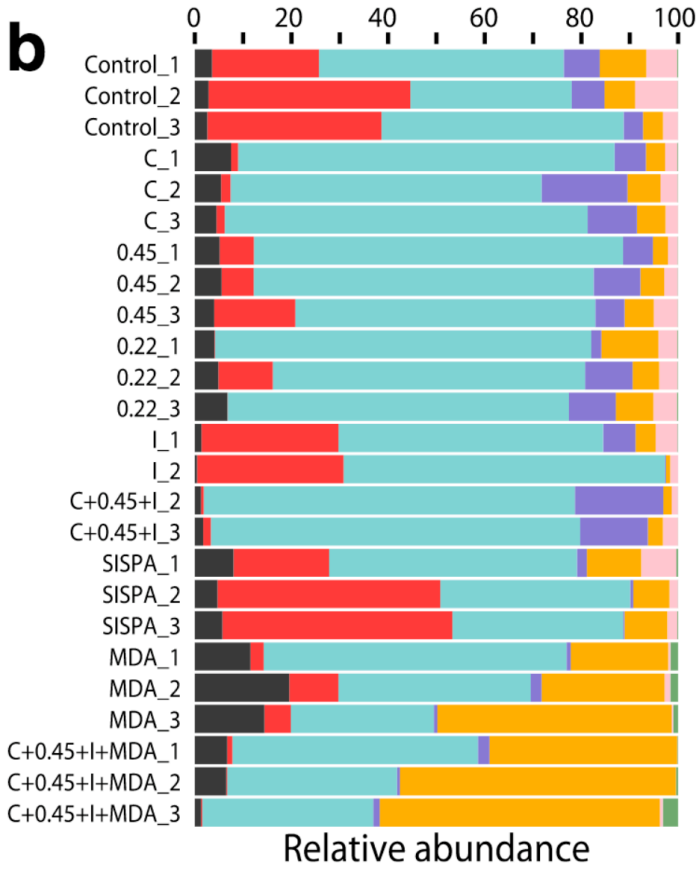
Evaluation of bias induced by viral enrichment and random amplification protocols in metagenomic surveys of saliva DNA viruses

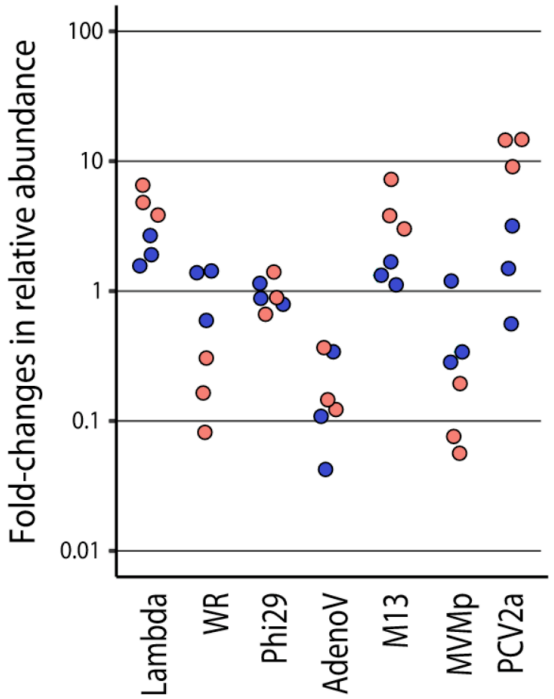
Marcos Parras-Moltó, Ana Rodríguez-Galet, Patricia Suárez-Rodríguez & Alberto López-Bueno

Microbiome **6**, Article number: 119 (2018) | Cite this article

Lambda WR Phi29 AdenoV
M13 MVMp PCV2a

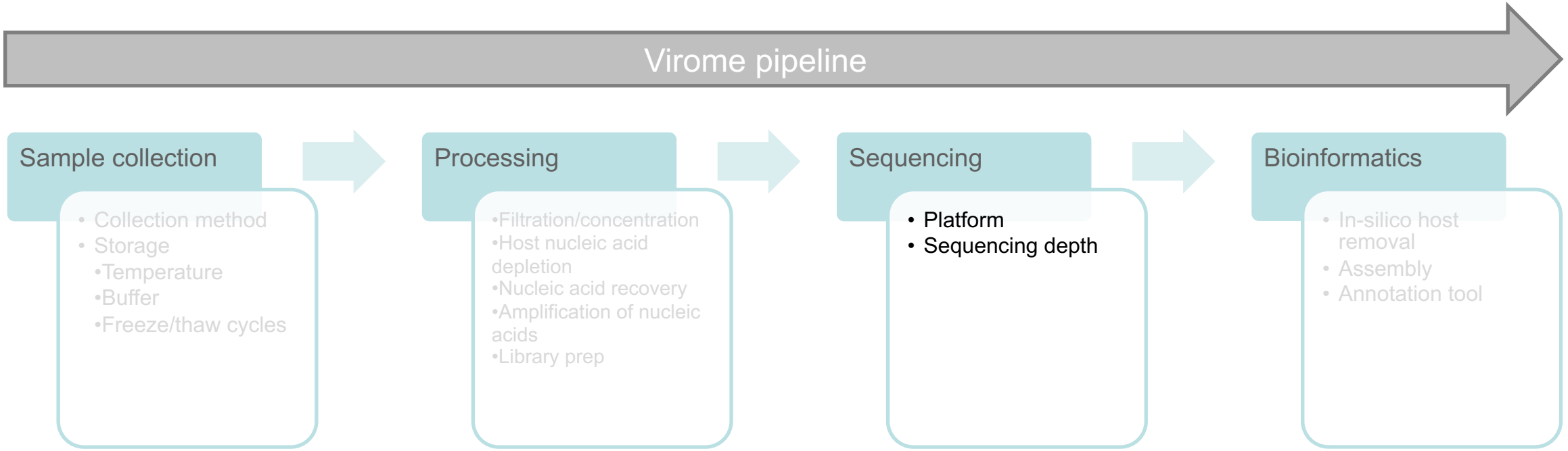
- Control=No treatment
- C= Two consecutive centrifugation rounds at 3000×*g* for 10 min
- 0.45=Filtered through 0.45um membranes
- 0.22=Filtered through 0.22um membranes
- I=Iodixanol cushion
- C+0.45+I=Centrifugation+Filtration+Iodixanol
- SISPA= Sequence-independent, single-primer amplification
- MDA=multiple displacement amplification
- C+0.45+I+MDA=Centrifugation+Filtration+Iodixanol+MDA





Diversigen

Biases in virome research - Sequencing



Biases in virome research - Sequencing

Journal Pre-proof

Comparison of Illumina MiSeq and the Ion Torrent PGM and S5 platforms for whole-genome sequencing of picornaviruses and caliciviruses

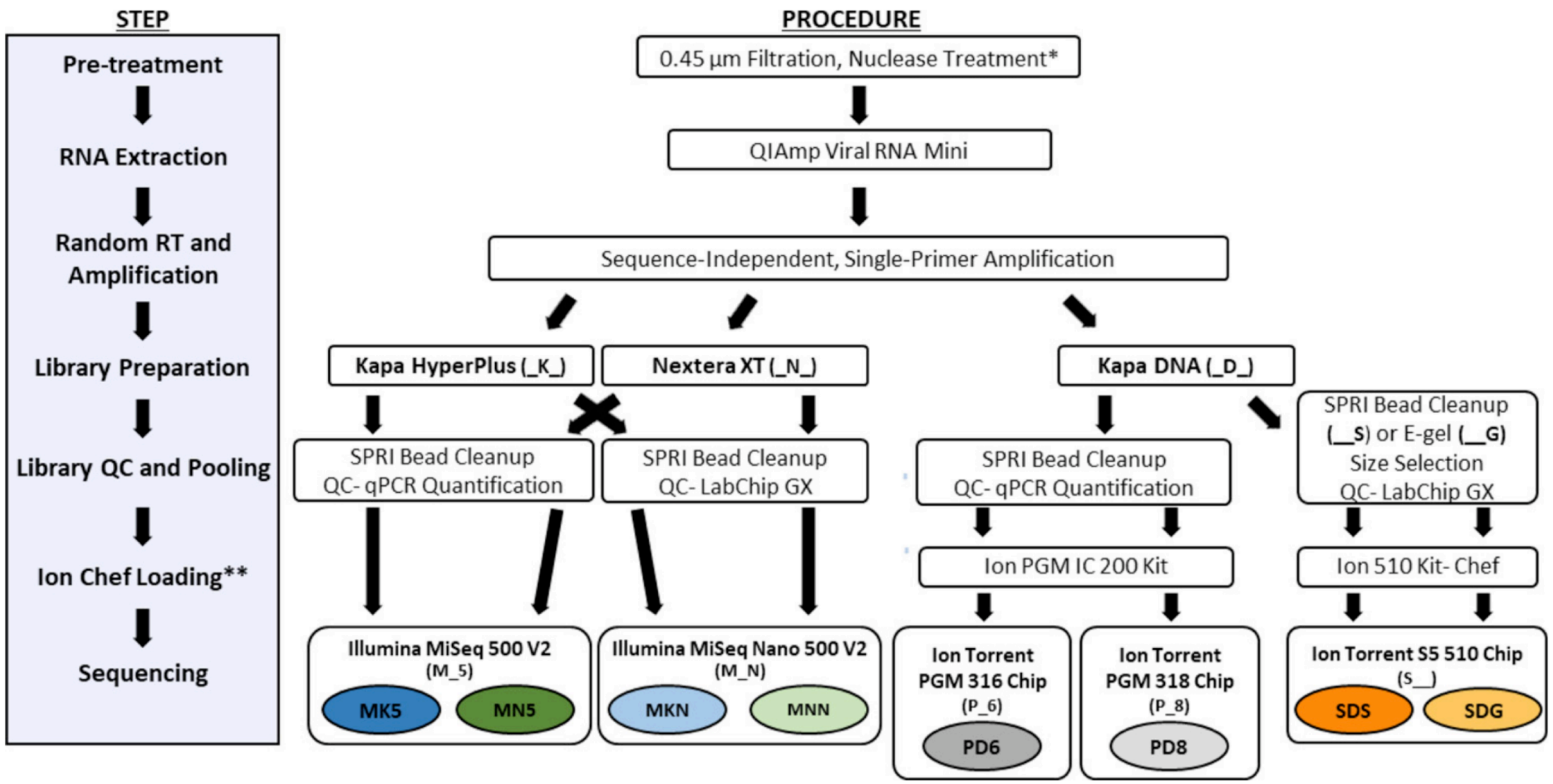
Rachel L. Marine (Conceptualization) (Investigation) (Writing - original draft) (Writing - review and editing), Laura C. Magaña (Conceptualization) (Investigation) (Writing - original draft) (Writing - review and editing), Christina J. Castro (Formal analysis) (Data curation), Kun Zhao (Formal analysis) (Visualization), Anna M. Montmayeur (Resources) (Writing - review and editing), Alexander Schmidt (Resources) (Writing - review and editing), Marta Diez-Valcarce (Resources) (Writing - review and editing), Terry Fei Fan Ng (Conceptualization) (Resources) (Writing - review and editing), Jan Vinjé (Supervision) (Writing - review and editing), Cara C. Burns (Supervision) (Writing - review and editing), W. Allan Nix (Supervision) (Writing - review and editing), Paul A. Rota (Supervision) (Writing - review and editing) (Funding acquisition), M. Steven Oberste (Supervision) (Writing - review and editing) (Funding acquisition)

PII: S0166-0934(20)30117-8

DOI: https://doi.org/10.1016/j.jviromet.2020.113865

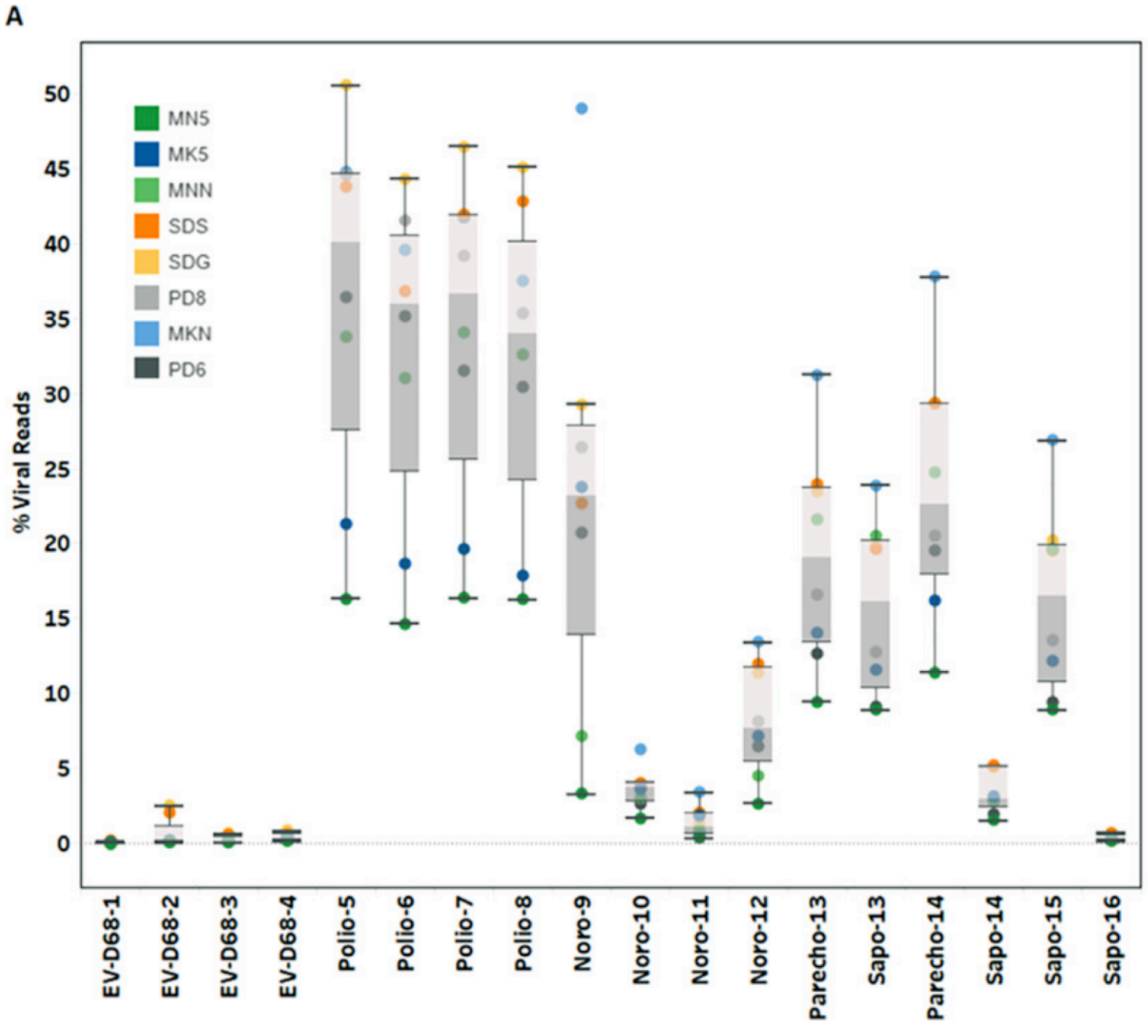
Reference: VIRMET 113865

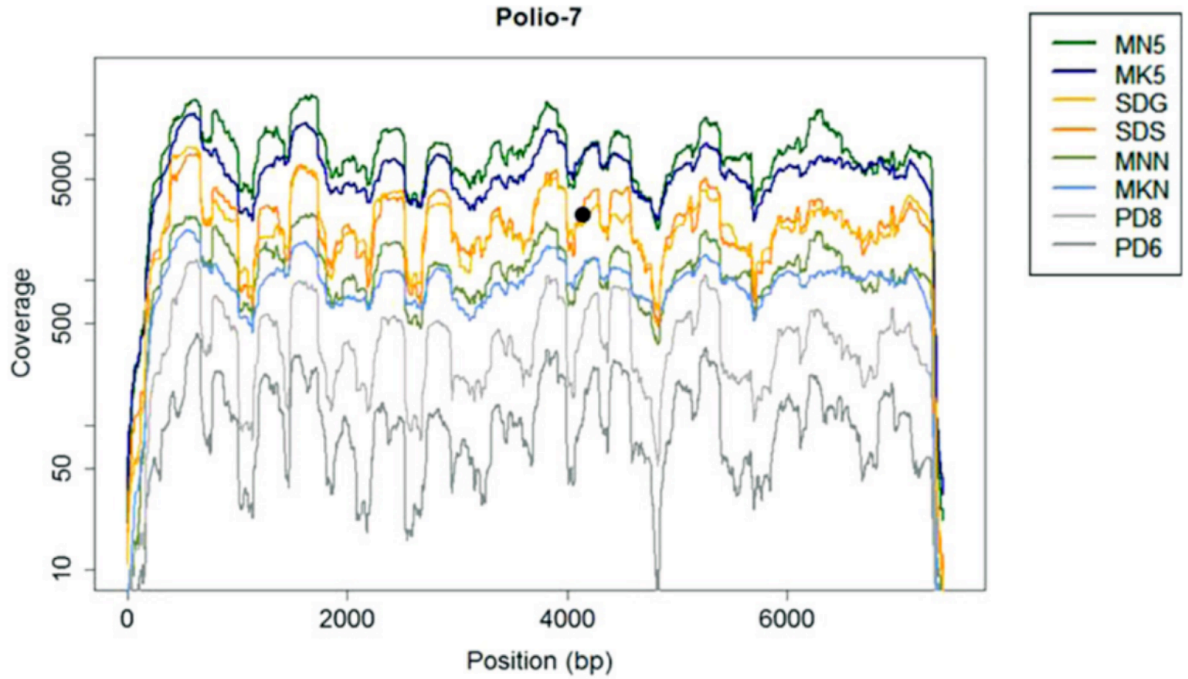
To appear in: *Journal of Virological Methods*



Diversigen

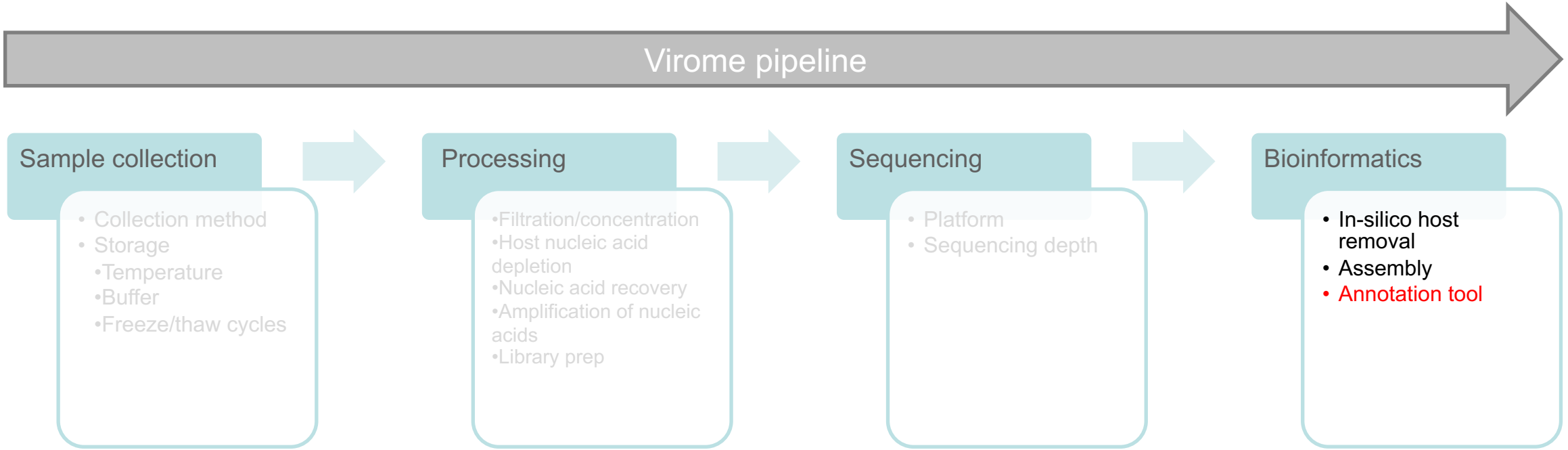
Biases in virome research - Sequencing





Diversigen®

Biases in virome research - Bioinformatics



Agenda

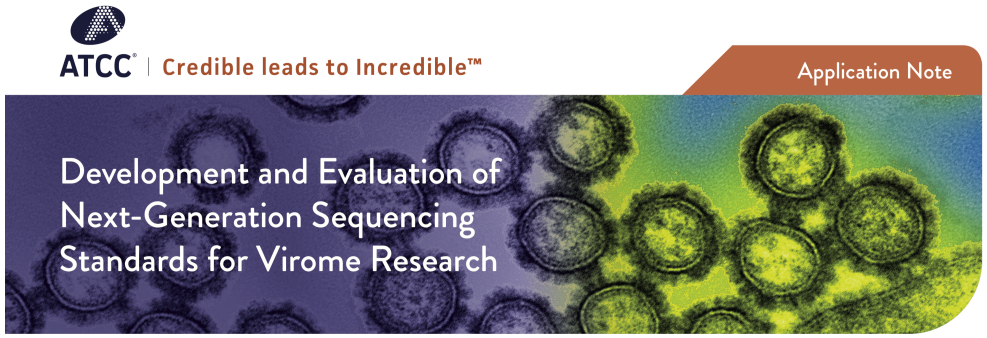
- Diversigen overview
- What do we know so far about virome research?
- Biases in virome research
- Application of mock communities in virome research
- Potential applications of standards in the detection of ongoing and future pathogenic human viruses: considerations from SARS-CoV 2

Mock communities in microbiome studies

- The realization of including reference materials, such as mock communities, arose very recently
- In articles published in two important microbiome and microbiology journals in 2018, only 30% reported a negative control, and 10% reported the use of a positive control
- Mock communities can be used as positive controls
  - Are still not implemented on a regular basis
  - Available through institutions, laboratories and commercial facilities
  - Most are intended for microbiome analysis
  - Very few have been developed for virome analysis

Diversigen

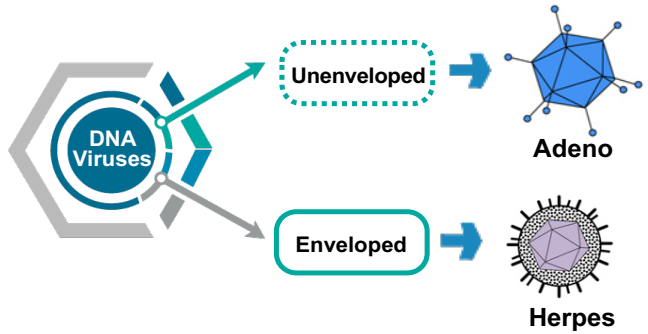
Viral mock communities (ATCC)

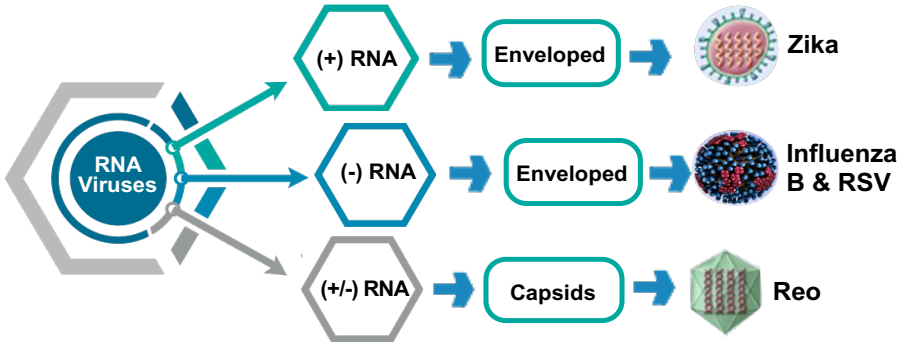


Juan Lopera, PhD,[1] Briana Benton, BS,[1] Jung-Woo Sohn, PhD,[1] Stephen King, MS,[1] Tasha M. Santiago-Rodriguez, PhD,[2] Emily B. Holister, PhD,[2] Matthew C. Wong, BS,[3] Nadim Ajami, PhD,[3] Cara Wilder, PhD[1]

Table 1: Selection attributes for strains included in the ATCC® Virome Standards.

Virus Name	ATCC® No.	Genome Type	Host (ATCC® No.)*	Virion Structure	Reference GenBank ID	Genome Size (Kbp)	Relevance
Human herpesvirus 5	VR-538™	ds DNA	MRC-5 (CCL-171™)	Enveloped	X17403.1	229.4	Ubiquitous infection in adult humans, and significant pathogen within immunocompromised populations[10]
Human mastadenovirus F	VR-931™	ds DNA	HEK-293 (CRL-1573™)	Unenveloped	NC_001454.1	34.2	Human gastrointestinal infection and severe infection in children and immunocompromised patients[11]
Influenza B virus B/Florida/4/2006	VR-1804™	ss (-) RNA (8 segments)	SPF embryo-nated chicken eggs	Enveloped	CY018365.1-CY018372.1	14.2	Causes worldwide human epidemics of influenza with high rates of illness and death[12]
Zika virus	VR-1838™	ss (+) RNA	Vero (CCL-81™)	Enveloped	KX830960.1	10.8	Mosquito-borne viral infection that can cause congenital microcephaly in fetuses and infants[13]
Human respiratory syncytial virus	VR-1540™	ss (-) RNA	HEp-2 (CCL-23™)	Enveloped	KT992094.1	15.2	Causes severe respiratory tract infections in humans[14]
Reovirus 3	VR-824™	ds RNA (10 segments)	LLC-MK2 Derivative (CCL-7.1™)	Capsids	HM159613.1-HM159622.1	23.6	Human respiratory and gastrointestinal infection; oncolytic virus[15,16]





Diversigen

Beta-testing ATCC viral mock community (nucleic acids)

Sample collection

- Nucleic acids (ATCC-MSA 1008)

Sequencing

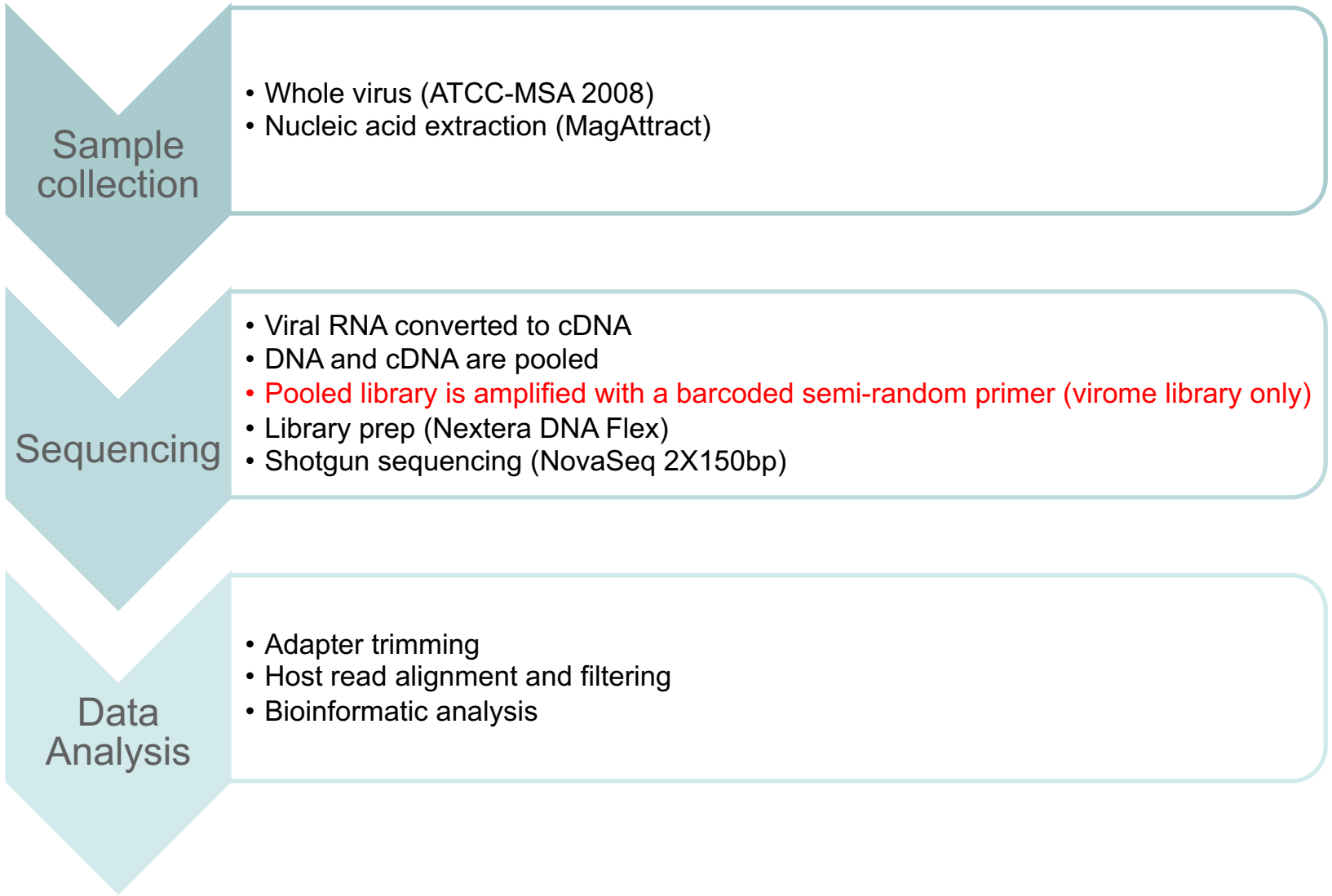
- Viral RNA converted to cDNA
- DNA and cDNA are pooled
- Pooled library is amplified with a barcoded semi-random primer (virome library only)
- Library prep (Nextera DNA Flex)
- Shotgun sequencing (NovaSeq 2X150bp)

Data Analysis

- Adapter trimming
- Host read alignment and filtering
- Bioinformatic analysis

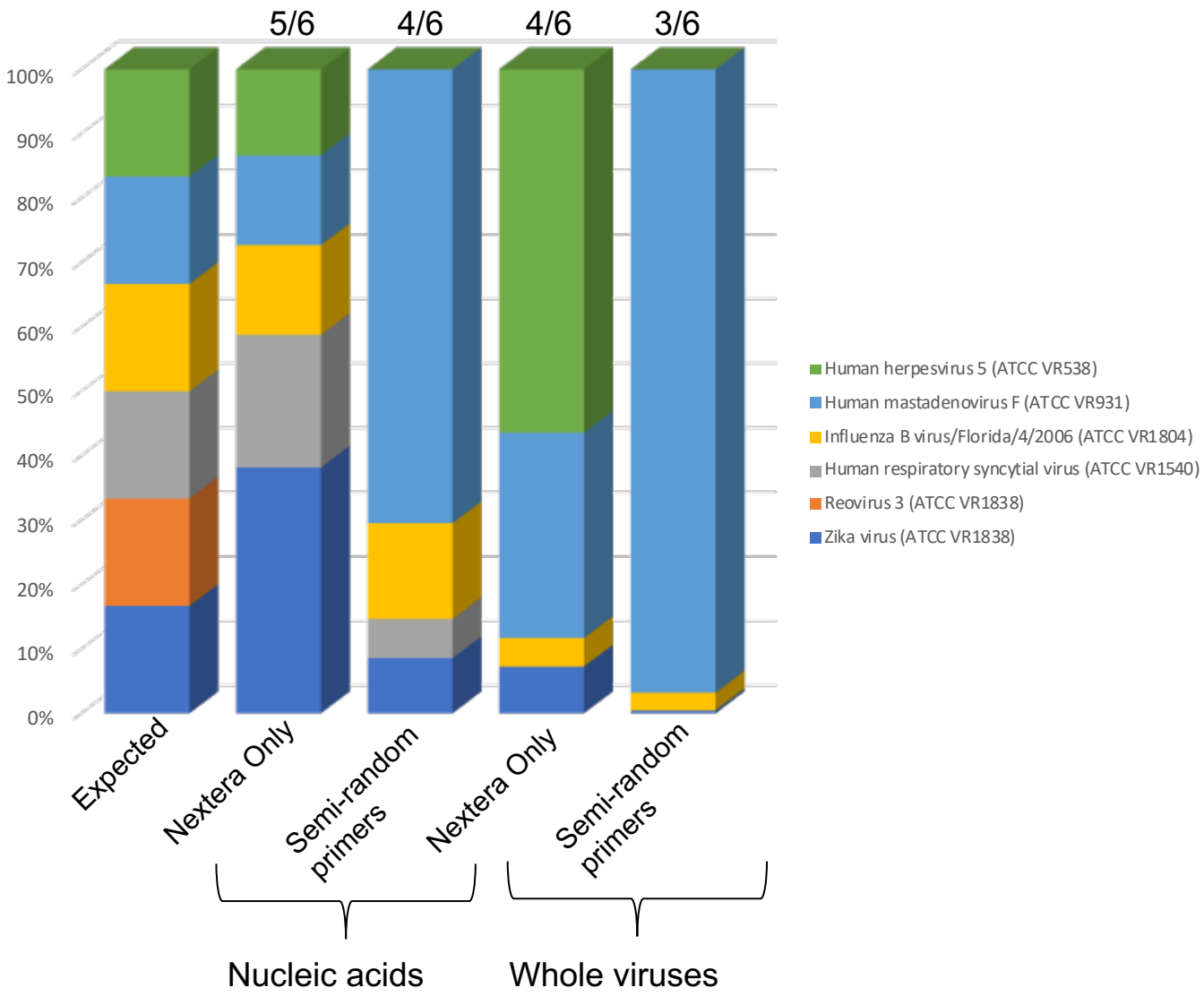
Diversigen®

Beta-testing ATCC viral mock community (whole virus)



Diversigen®

ATCC viral mock communities - MetaPhlAn2



MetaPhlAn 2.0

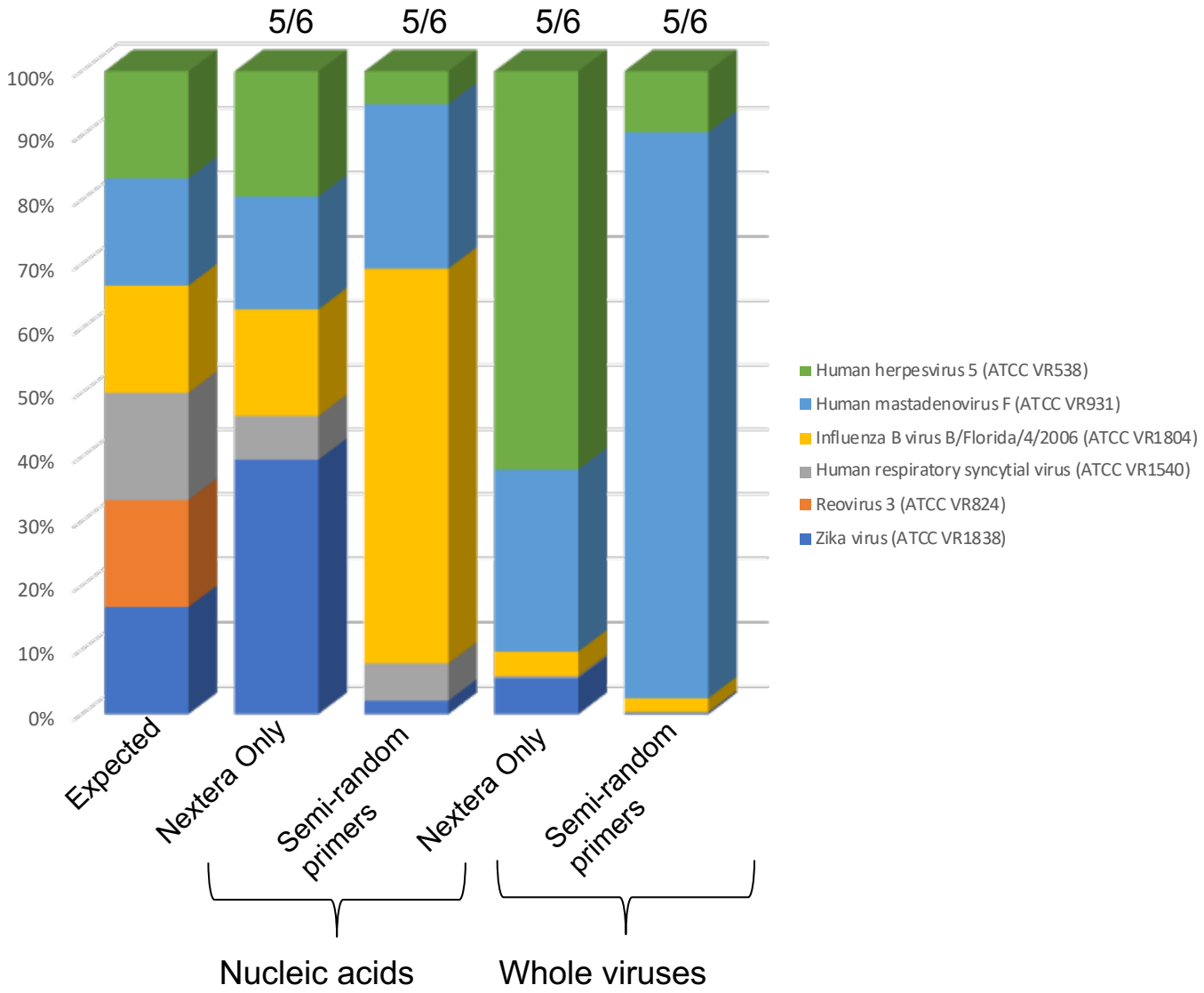
MetaPhlAn (Metagenomic Phylogenetic Analysis) is a computational tool for profiling the composition of microbial communities from metagenomic shotgun sequencing data. MetaPhlAn relies on unique clade-specific marker genes identified from ~17,000 reference genomes (~13,500 bacterial and archaeal, ~3,500 viral, and ~110 eukaryotic), allowing:

- up to 25,000 reads-per-second (on one CPU) analysis speed (orders of magnitude faster compared to existing methods);
- unambiguous taxonomic assignments as the MetaPhlAn markers are clade-specific;
- accurate estimation of organismal relative abundance (in terms of number of cells rather than fraction of reads);
- species-level resolution for bacteria, archaea, eukaryotes and viruses;
- extensive validation of the profiling accuracy on several synthetic datasets and on thousands of real metagenomes.

	Nucleic acids (MSA1008)		Whole viruses (MSA2008)	
	Nextera Only	Semi-random primers	Nextera Only	Semi-random primers
False positives	1	1	1	1
False negatives	1	2	2	3

Diversigen®

ATCC viral mock communities - Kraken2



Improved metagenomic analysis with Kraken 2

Derrick E. Wood, Jennifer Lu & Ben Langmead

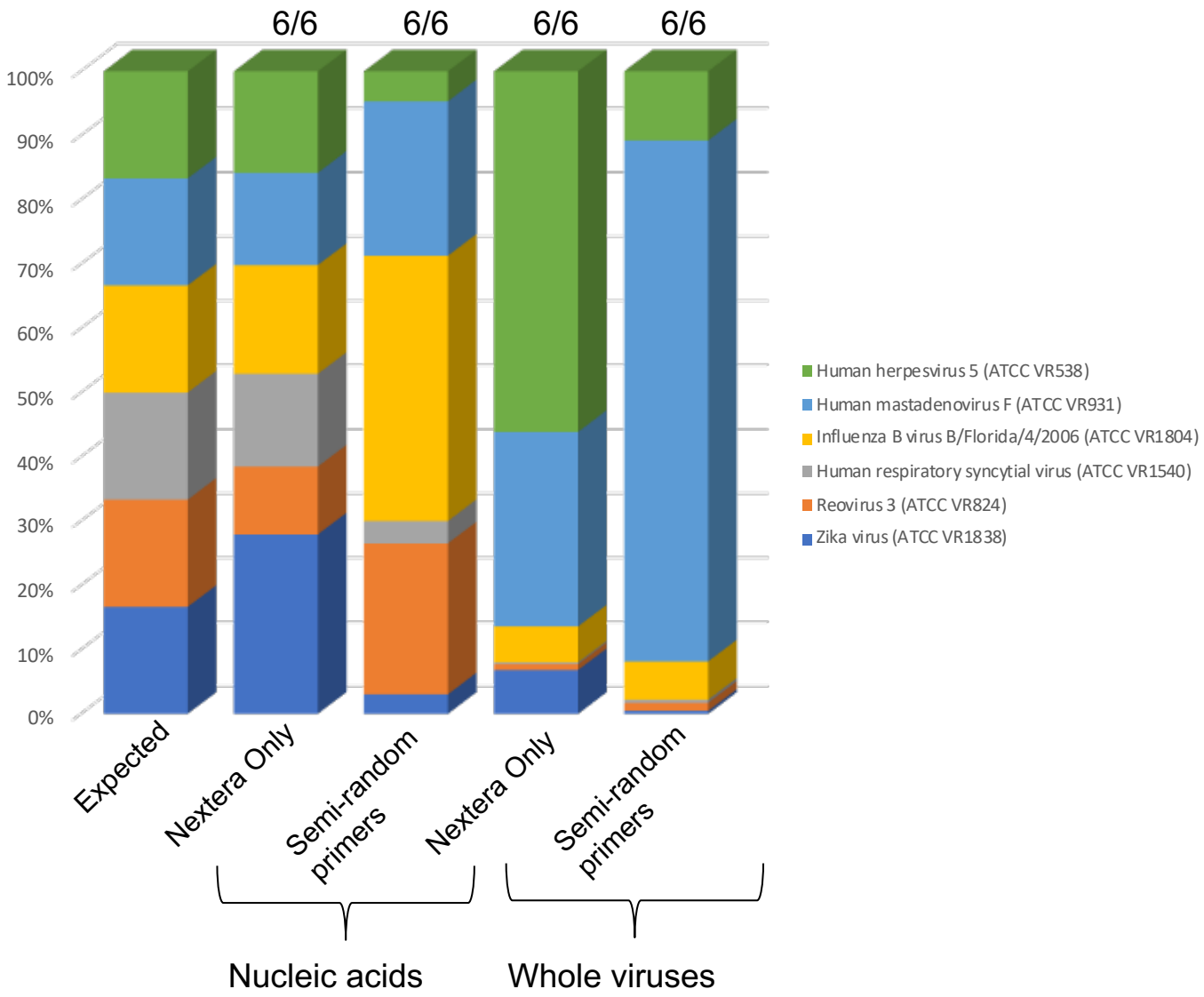
Genome Biology **20**, Article number: 257 (2019) | Cite this article

6610 Accesses | **12** Citations | **75** Altmetric | Metrics

	Nucleic acids (MSA1008)		Whole viruses (MSA2008)	
	Nextera Only	Semi-random primers	Nextera Only	Semi-random primers
False positives	356	258	278	352
False negatives	1	1	1	1

Diversigen

ATCC viral mock communities - VirMap



Maximal viral information recovery from sequence data using VirMAP

Nadim J Ajami, Matthew C. Wong, Matthew C. Ross, Richard E. Lloyd & Joseph F. Petrosino

Nature Communications **9**, Article number: 3205 (2018) | Cite this article

	Nucleic acids (MSA1008)		Whole viruses (MSA2008)	
	Nextera Only	Semi-random primers	Nextera Only	Semi-random primers
False positives	1	0	0	1
False negatives	0	0	0	0

Diversigen

Agenda

- Diversigen overview
- What do we know so far about virome research?
- Biases in virome research
- Application of mock communities in virome research
- Potential applications of standards in the detection of ongoing and future pathogenic human viruses: considerations from SARS-CoV 2

Viral metagenomics as a surveillance tool of SARS-CoV 2

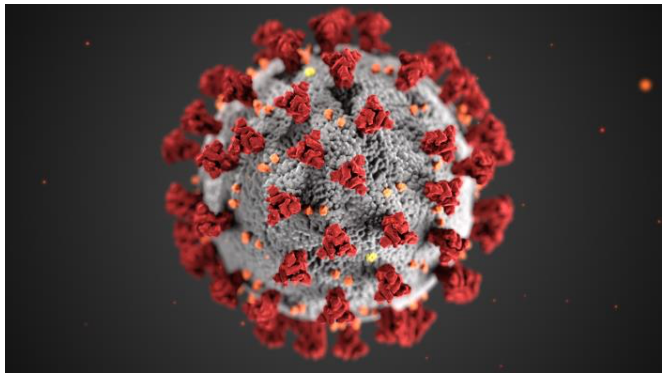
Capabilities in SARS-CoV 2 detection reside in building a custom database with the virus sequence, as well as in the annotation tool used

To confirm the ability to detect SARS-COV-2, we conducted multiple tests:

-Profiling known positive samples (NCBI SRA)*

-Profiling suspected negative samples (NCBI SRA)

The availability of sequence libraries of varying depth allowed us to evaluate the effects of sequencing depth on virus detection and genome recovery



https://www.cdc.gov/media/subtopic/images.htm

Article | Open Access | Published: 10 August 2018

Maximal viral information recovery from sequence data using VirMAP

Nadim J Ajami ✉, Matthew C. Wong, Matthew C. Ross, Richard E. Lloyd & Joseph F. Petrosino

Nature Communications **9**, Article number: 3205 (2018) | Cite this article

Diversigen

Profiling SARS-CoV2 positive and negative samples

Capabilities in SARS-CoV-2 detection reside in building a custom database with the virus sequence and the annotation tool used

To confirm the ability to detect SARS-COV-2, we conducted multiple tests:

-Profiling known positive samples (NCBI SRA)*

-Profiling suspected negative samples (NCBI SRA)

The availability of sequence libraries of varying depth allowed us to evaluate the effects of sequencing depth on virus detection and genome recovery

SAMPLE ID	SRA RUN IDENTIFIER(S)	EXPECTED SARS-COV-2 STATUS	SARS-COV-2 DETECTED	DNA VIRUSES DETECTED	OTHER RNA VIRUSES DETECTED
WIV02	SRR11092058, SRR11092063	+	✅	✅	✅
WIV04	SRR11092057, SRR11092062	+	✅	✅	✅
WIV05	SRR11092061	+	✅	✅	✅
WIV06	SRR11092056, SRR11092060	+	✅	✅	✅
WIV07	SRR11092064, SRR11092059	+	✅	✅	✅
COPD18	SRR5677628	−	−	✅	✅
COPD25	SRR5677642	−	−	✅	−

Diversigen

Sequencing depth affects genome coverage of SARS-CoV-2

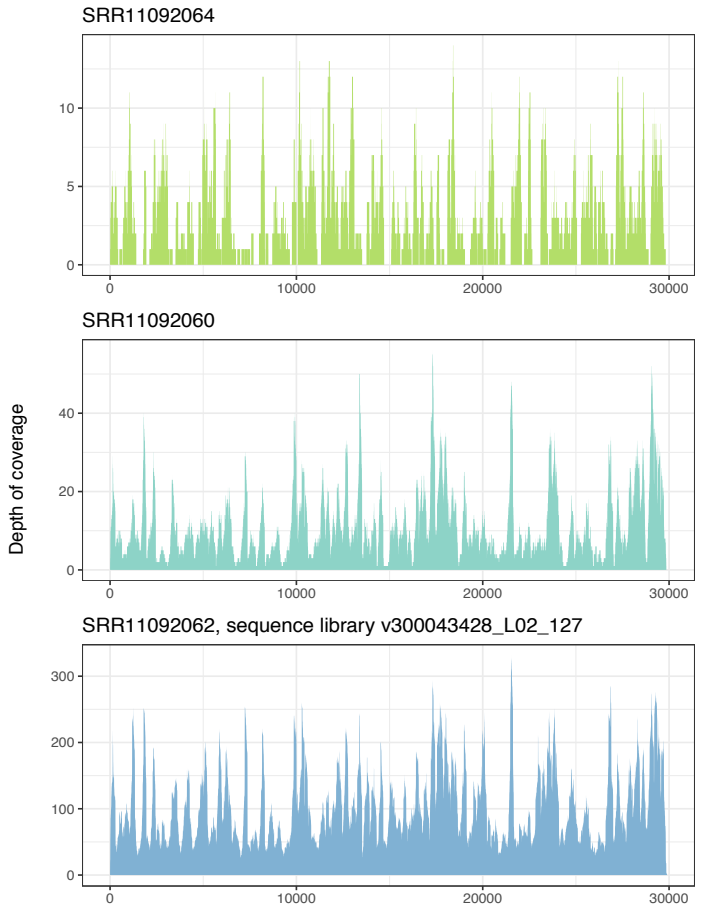
Capabilities in SARS-CoV-2 detection reside in building a custom database with the virus sequence and the annotation tool used

To confirm the ability to detect SARS-COV-2, we conducted multiple tests:

-Profiling known positive samples (NCBI SRA)*

-Profiling suspected negative samples (NCBI SRA)

The availability of sequence libraries of varying depth allowed us to evaluate the effects of sequencing depth on virus detection and genome recovery



Sequencing depth: 1.04 Gb

% of SARS-COV-2 genome recovered: 83%

% Hominid reads in library (NCBI): 56%

Average depth of coverage: 3.35x

Sequencing depth: 8.90 Gb

% of SARS-COV-2 genome recovered: 99.89%

% Hominid reads in library (NCBI): 60%

Average depth of coverage: 11.79x

Sequencing depth: 18.4 Gb

% of SARS-COV-2 genome recovered: 100%

% Hominid reads in library (NCBI): 67%

Average depth of coverage: 105x

Diversigen

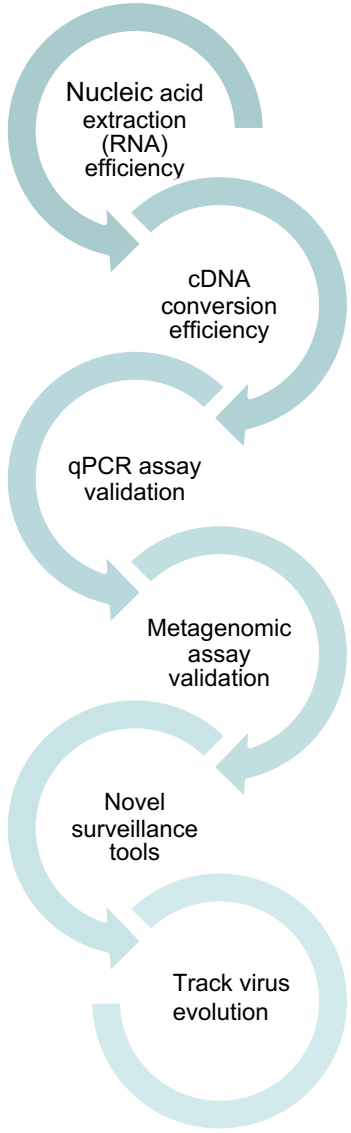
Application of standards in SARS-CoV-2 research

ATCC® | Credible leads to Incredible™

Credible solutions for critical public health emergencies

The outbreak of severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) in December 2019 has put the health and safety of the global community at risk. The virus has already spread beyond the borders of China, with numerous countries reporting confirmed cases of coronavirus disease (COVID-19). With the virus continuing to spread and the number of confirmed cases and deaths rising, the World Health Organization has declared the outbreak a public health emergency of international concern.

As in previous public health emergencies such as Zika, SARS, MERS, and the H1N1 2009 pandemic, ATCC stands ready to partner with the dedicated scientists working toward preventing and containing this devastating outbreak. Only through the combined efforts of the global scientific community can we discover the tools and treatments needed to keep humankind healthy and safe. View our resources below to discover how we can support your work toward the development of novel diagnostics and effective therapeutics.



Diversigen®

Application of viral standards moving forward

Expansion of viral databases and surveillance activities

Continued improvements related to collection, stabilization, extraction

New reagents and approaches

Layering virome information onto metagenomic studies to provide new biological and clinical insights

Viral standards for viral metagenomics and clinical applications will continue to evolve

Diversigen®

QUESTIONS?

ATCC Microbiome Research Solutions

Human Microbiome Research

- Virome Standards
- Site-specific Standards
- Mycobiome Standards
- Pathogen Detection Standards

Assay Standardization

- Mock Microbial Communities
- Spike-in Standards

Environmental Microbiome Research

- ABRF-MGRG Standards

www.atcc.org/microbiome