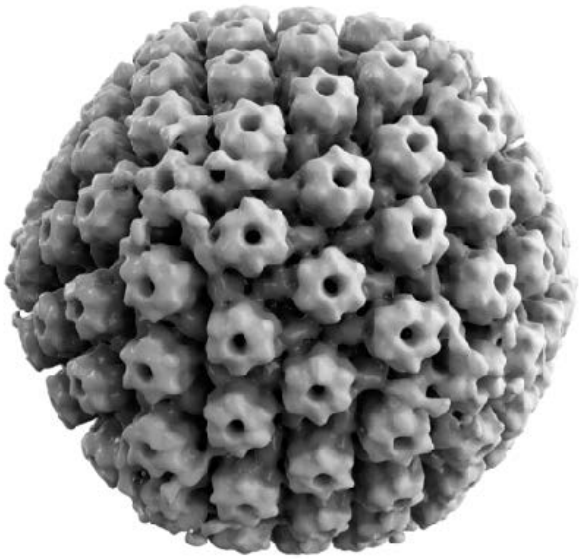
Building Standardized Dendograms for Viruses



David W. Ussery

Professor, Biomedical Informatics
Professor, Physiology & Biophysics
Director, ArC-GEM (Arkansas Center for Genomic Epidemiology & Medicine)
University of Arkansas for Medical Sciences

Wednesday, 18 September, 2019


SERY

AGENDA

Day 1 - Wednesday, September 18, 2019

Session 1 – Perspectives on Standardization

09:00	Welcome Remarks **Michael Tarlov**, NIST
09:10	*Current Practices for Regulatory Applications* **Arifa Khan,** FDA
09:40	*Overview of NIST Viral Standards* **Megan Cleveland**, NIST
10:10-10:30	Coffee break
10:30	*dPCR as a reference measurement procedure for viral detection and quantification* **Jim Huggett**, UK National Measurement Laboratory
11:00	*Standardization of Next Generation Sequencing Methods for the Quality Control of Live-attenuated Vaccines* **Javier Martin**, National Institute for Biological Standards and Control (NIBSC)
11:30	*Building Standardized Dendograms for Viruses* **David Ussery**, University of Arkansas for Medical Sciences
12:00 – 12:30	Q&A and Discussion
12:30 – 2:00	Lunch

11:30 Building Standardized Dendograms for Viruses

Outline:

1. Background: the explosion in sequences
2. More background - FAIR principles for data stewardship.
3. What is a Dendogram (and how is this different from a 'tree'?)
4. How can we standardize this?

frontiers in Microbiology

Direct RNA sequencing

TECHNOLOGY REPORT
published: 25 February 2019
doi: 10.3389/fmicb.2019.00260

Rapid Sequencing of Multiple RNA Viruses in Their Native Form

Thidathip Wongsurawat, Ph.D.

*Thidathip Wongsurawat[1†], Piroon Jenjaroenpun[1†], Mariah K. Taylor[2], Jasper Lee[2], Aline Lavado Tolardo[3], Jyothi Parvathareddy[4], Sangam Kandel[1,5], Taylor D. Wadley[1], Bualan Kaewnapan[6], Niracha Athipanyasilp[6], Andrew Skidmore[7], Donghoon Chung[7], Chutikarn Chaimayo[6], Michael Whitt[2], Wannee Kantakamalakul[6], Ruengpung Sutthent[6], Navin Horthongkham[6], David W. Ussery[1,8], Colleen B. Jonsson[2] and Intawat Nookaew[1,8]**

[1] Department of Biomedical Informatics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, United States, [2] Department of Microbiology, Immunology and Biochemistry, The University of Tennessee Health Science Center, Memphis, TN, United States, [3] Virology Research Center, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil, [4] Regional Biocontainment Laboratory, University of Tennessee Health Science Center, Memphis, TN, United States, [5] Department of Bioinformatics, University of Arkansas at Little Rock, Little Rock, AR, United States, [6] Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand, [7] Department of Microbiology and Immunology, University of Louisville, Louisville, KY, United States, [8] Department of Physiology and Biophysics, College of Medicine, University of Arkansas for Medical Sciences, Little Rock, AR, United States*

Piroon Jenjaroenpun, Ph.D.

Long-read nanopore sequencing by a MinION device offers the unique possibility to directly sequence native RNA. We combined an enzymatic poly-A tailing reaction with the native RNA sequencing to (i) sequence complex population of single-stranded (ss)RNA viruses in parallel, (ii) detect genome, subgenomic mRNA/mRNA simultaneously, (iii) detect a complex transcriptomic architecture without the need for assembly, (iv) enable real-time detection. Using this protocol, positive-ssRNA, negative-ssRNA, with/without a poly(A)-tail, segmented/non-segmented genomes were mixed and sequenced in parallel. Mapping of the generated sequences on the reference genomes showed 100% length recovery with up to 97% identity. This work provides a proof of principle and the validity of this strategy, opening up a wide range of applications to study RNA viruses.

Keywords: native RNA, genome, subgenomic mRNA, single-stranded RNA, virus, nanopore sequencing, rapid detection, MinION

Our solid protocol, ready for pooled viruses

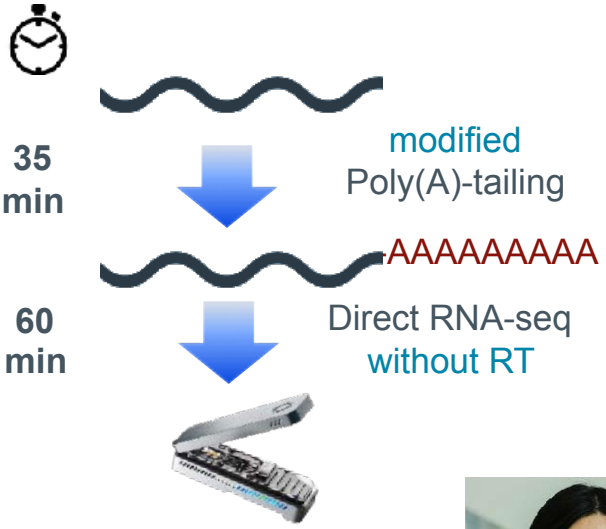
Intawat Nookaew, Ph.D.

6 ssRNA viruses

- 10.7-11.6 Kb
- OROV: 1, 4, 7kb

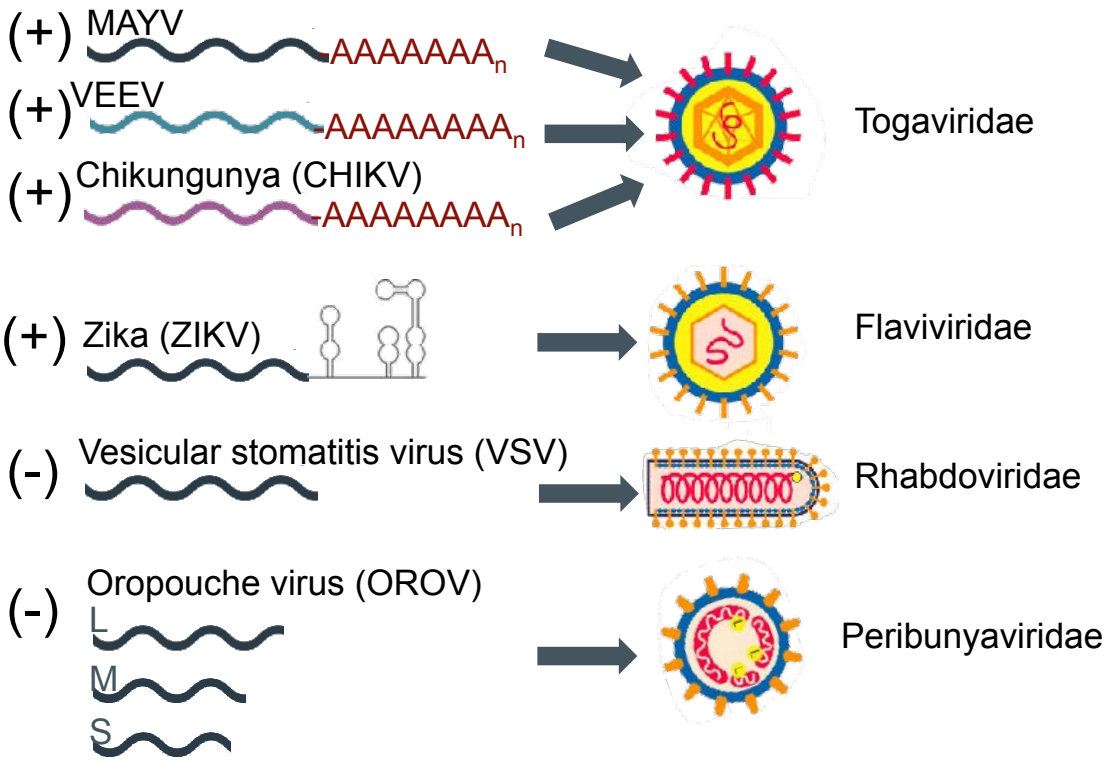
Colleen Jonsson, Ph.D.

UTHSC, Tennessee



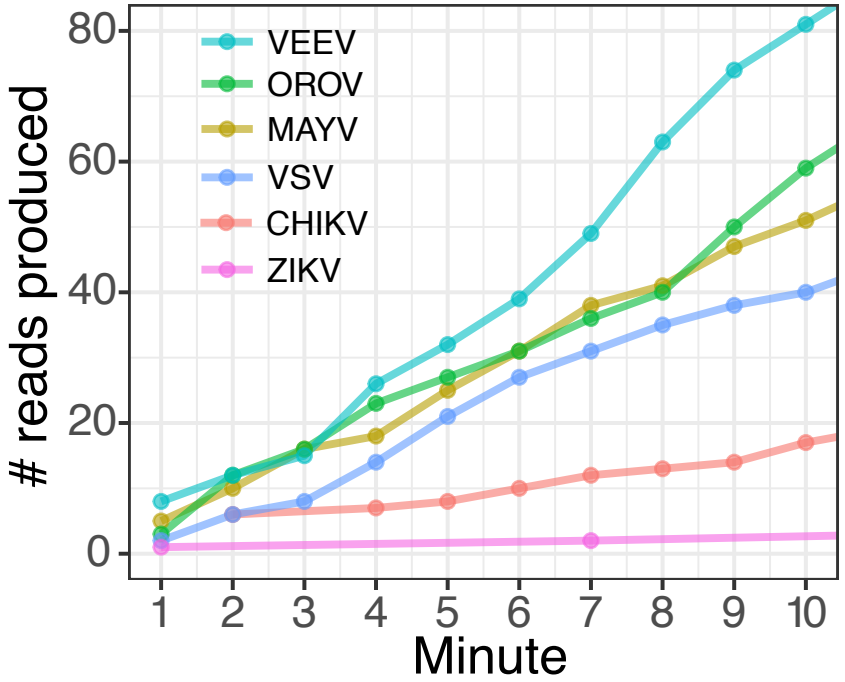
Piroon Jenjaroenpun, Ph.D.

Thidathip Wongsurawat, Ph.D.

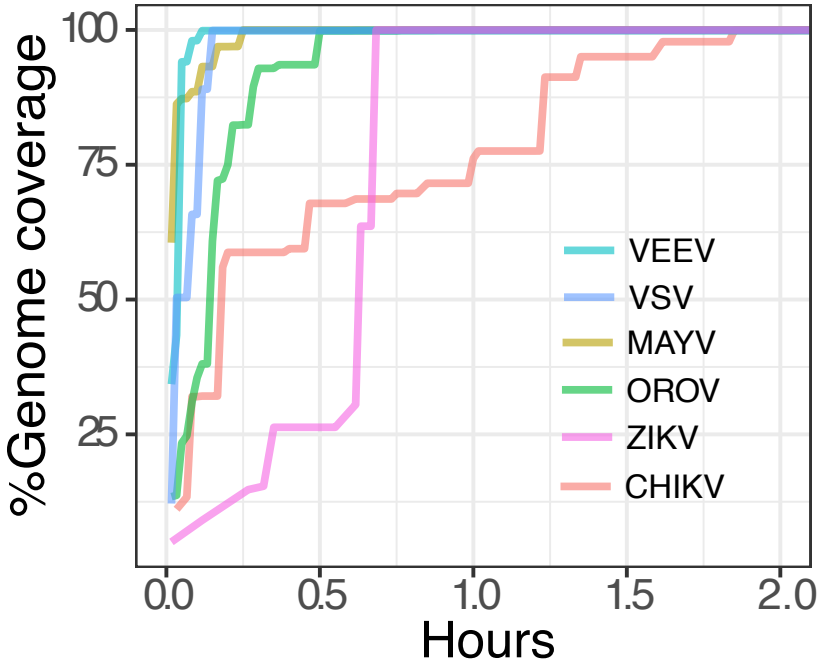


Rapid Third Generation Sequencing of mixtures of viruses

The reads of all viruses found within 2 min

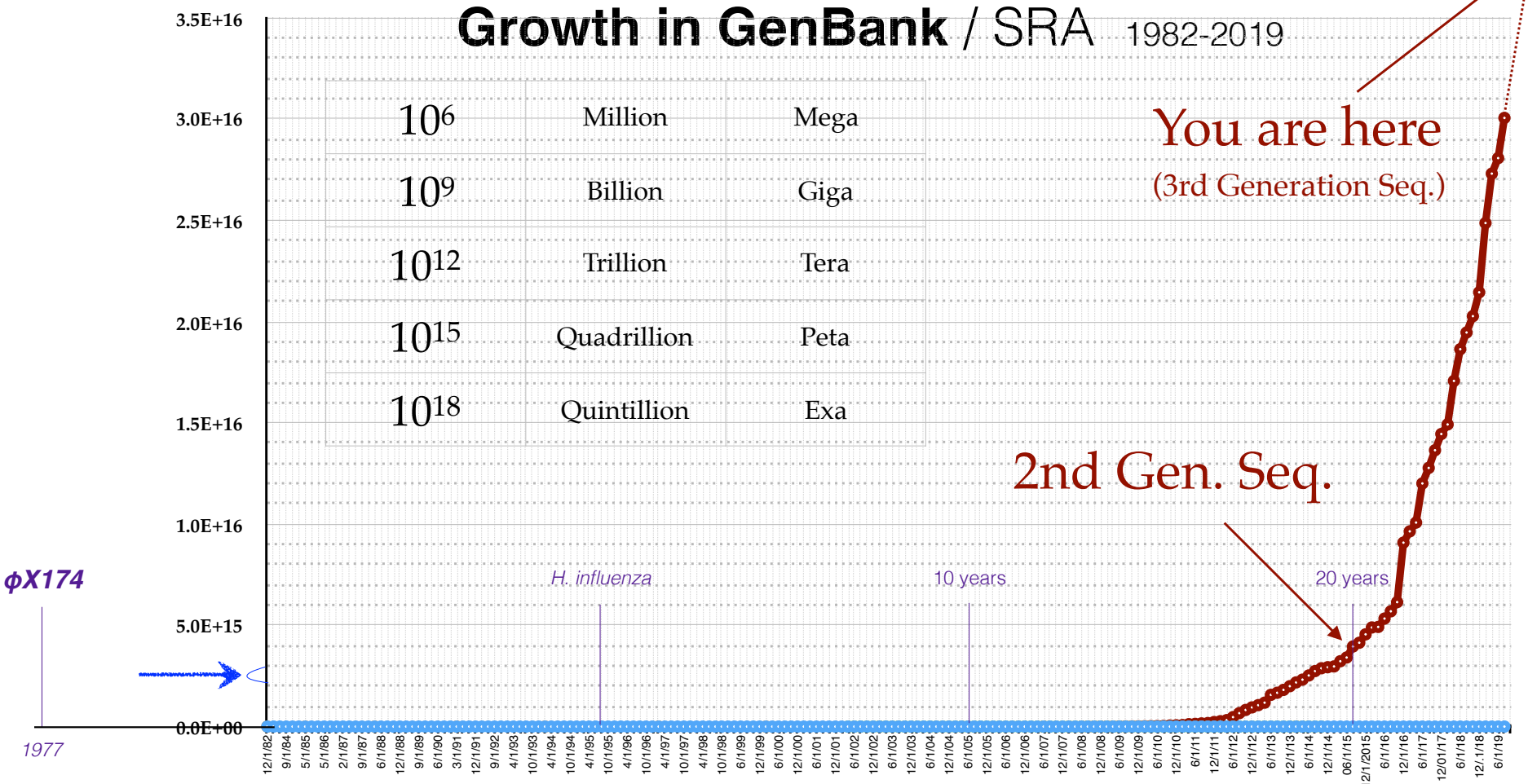


~2 hours for getting complete genome



1. Background: the explosion in sequences

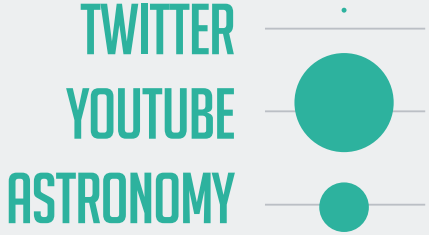
Biological Information = Sequences = "Big Data"



1,000,000,000,000,000 bp of DNA sequence

Modified from Figure 4 in Land *et al.*, Functional & Integrative Genomics, **15**:141-161 (2015).

To give some context, data scientists have drawn comparisons with other big data generators. So what are their predictions of data output by 2025? [7]



GENOMICS

=

~20 Zetabytes (10^{21})

FOR HUMAN GENOMES ALONE!

nature genetics

2. FAIR principles for data stewardship.

A Simple Question:

How many virus genomes are in GenBank?

FAIR principles for data stewardship

The FAIR data principles are simple guidelines for ensuring that machines can find and use data, supporting data reuse by individuals. More—and better—research can be generated by designing data and algorithms to be findable, accessible, interoperable and reusable, together with the tools and workflows that led to these data.

We are not lacking standards. Indeed, over 600 content standards for biological data types are listed by the BioSharing registry alone (https://biosharing.org). However, one recent attempt to set standards for all kinds of data generated by scholarly activity gets right to the point: much of our scholarship gets in the way of data reuse because it obscures the machine readability of the data. The consequence of this limitation is that the scale of data reuse by human researchers is restricted (*Sci. Data* **3**, 160018, 2016).

The authors, from diverse backgrounds (including representatives from *Scientific Data* and this journal), conclude that, rather than set yet more standards, we should deposit data and design tools for their formatting, distribution and storage according to the four basic principles of finding, accessing, integrating and reusing all scholarly data. This emphasis is designed to put a stop to the arms race between diversifying data types and metadata annotations on the one side and bespoke mining tools designed to parse those data and metadata on the other. The question to ask in order to be a data steward, to handle data or to simplify a set of standards is the same: "is it FAIR"?

Most types of reusable data that are expensive to produce now have purpose-built databases. FAIR principles dictate the publication of rich metadata to describe these data and to enable discovery of what is contained therein, even in the case of sensitive data that identify persons. The data fields and metadata schema should be accessible, together with the details of any access restrictions, whether or not the underlying data can actually be accessed. In contrast, many of the products of low-throughput bench science do not fit into these standard databases. The repositories so far created for such data are becoming increasingly diverse in purpose and form. The key to taming these is to realize that they will need to be searched by general-purpose open technologies because they contain unpredictable data types unsuited to specialized parsers.

Equally important to good scholarship is the publication of non-data research objects. Explicit analytical workflows, for example, are essential to most forms of knowledge generation. Publication of these according to FAIR principles is essential to ensure transparency of the work as well as maximal use to the community. The key to working with data is to realize that the human touch, the urge to annotate tables with footnotes and cram multiple elements and data types into every cell of a table, gets in the way of computation, automation and scaling up. And this impedes the usefulness of your work for other people. All research objects should be findable, accessible, interoperable and reusable (FAIR) both for machines and for people.

NIH U.S. National Library of Medicine | NCBI National Center for Biotechnology Information https://www.ncbi.nlm.nih.gov/genome/browse/#!/viruses/

Genome > **Genome Information by Organism**

32,539 Viruses

Overview (47210); Eukaryotes (9209); Prokaryotes (210413); **Viruses (32539)**; Plasmids (18279); Organelles (14660)

Filters

Choose Columns | Page 1 of 651 | 50

#	Organism Name	Organism Groups	BioSample	BioProjec	Assembly	Leve	Size(Mb)	GC%	Replicons
1	A-2 plaque virus	Viruses;Other;Picornaviridae			GCA_000861565.1		0.007374	42.90	NC_003988.1 /AF201894.1
2	ANMV-1 virus	Viruses;unclassified archaeal viruses;unclassified			GCA_003004805.1		0.038465	44.50	KP703175.1
3	Abaca bunchy top virus	Viruses;Other;Nanoviridae			GCA_000872625.1		0.006422	41.04	DNA N: NC_010314.1 /EF546808.1 DNA U3: NC_010315.1 /EF546809.1 DNA S: NC_010316.1 /EF546810.1 Show all 6 replicons
4	Abaca bunchy top virus	Viruses;Other;Nanoviridae			GCA_003985405.2		0.006409	40.99	DNA N: EF546802.1 DNA U3: EF546803.1 DNA S: EF546804.1 Show all 6 replicons
5	Abalone herpesvirus Victoria/AUS/2009	Viruses;Other;Malacoherpesviridae			GCA_000900375.1		0.211518	46.80	NC_018874.1 /JX453331.1
6	Abalone shriveling syndrome-associated virus	Viruses;Other;Other			GCA_000882555.1		0.034952	39.50	NC_011646.1 /EU350361.2
7	Abelson murine leukemia virus	Viruses;Other;Retroviridae			GCA_000848265.1		0.005894	55.20	NC_001499.1 /AF033812.1
8	Abisko virus	Viruses;Other;Other			GCA_002270725.1		0.010187	40.20	NC_035470.1 /KY662294.1
9	Abutilon Brazil virus	Viruses;Other;Geminiviridae			GCA_000889055.1		0.005271	46.46	DNA A: NC_014138.1 /FN434438.1 DNA B: NC_014139.1 /FN434439.1
10	Abutilon golden mosaic Yucatan virus	Viruses;Other;Geminiviridae			GCA_002821485.1		0.002629	47.20	DNA A: NC_038438.1 /KC430935.1
11	Abutilon mosaic Bolivia virus	Viruses;Other;Geminiviridae			GCA_000890575.1		0.005399	44.07	DNA A: NC_015045.1 /HM585445.1 DNA B: NC_015048.1 /HM585446.1
12	Abutilon mosaic Brazil virus	Viruses;Other;Geminiviridae			GCA_000895175.1		0.005282	45.01	DNA A: NC_016574.1 /JF694480.1 DNA B: NC_016577.1 /JF694479.1
13	Abutilon mosaic virus	Viruses;Other;Geminiviridae			GCA_000847225.1		0.005217	46.70	DNA A: NC_001928.2 /X15983.2 DNA B: NC_001929.2 /X15984.3
14	Abutilon yellows virus	Viruses;Other;Closteroviridae			GCA_002833665.1		0.001311	41.20	
15	Acanthamoeba castellanii mamavirus	Viruses;Other;Mimiviridae			GCA_002966335.1		1.19	28.00	JF801956.1
16	Acanthamoeba castellanii mimivirus	Viruses;Other;Mimiviridae			GCA_002966345.1		1.18	27.90	AP017644.1

NIH U.S. National Library of Medicine | NCBI National Center for Biotechnology Information

~~4,391,604 Viruses ?~~

Items: 1 to 20 of 4391604

<< First < Prev Page 1 of 219581 Next > Last >>

1. Acanthamoeba castellanii mimivirus DNA, nearly complete genome, strain: Mimivirus shirakomae
   1,182,849 bp linear DNA
   Accession: AP017645.1 GI: 1073515498
   Assembly Protein Taxonomy
   GenBank FASTA Graphics

2. Acanthamoeba castellanii mimivirus DNA, nearly complete genome, strain: Mimivirus kasaii
   1,182,801 bp linear DNA
   Accession: AP017644.1 GI: 1073514509
   Assembly Protein Taxonomy
   GenBank FASTA Graphics

3. Canarypox **virus**, complete genome
   359,853 bp linear DNA
   Accession: NC_005309.1 GI: 40555938
   Assembly BioProject Protein PubMed Taxonomy
   GenBank FASTA Graphics

4. Camelpox **virus** M-96 from Kazakhstan, complete genome
   205,719 bp linear DNA
   Accession: AF438165.1 GI: 18482913
   Assembly Protein PubMed Taxonomy
   GenBank FASTA Graphics

5. Canarypox **virus** strain ATCC VR-111, complete genome
   359,853 bp linear DNA
   Accession: AY318871.1 GI: 40233763
   Assembly Protein PubMed Taxonomy
   GenBank FASTA Graphics

6. Fowlpox **virus** isolate 282E4, complete genome
   308,827 bp linear DNA
   Accession: MG702259.1 GI: 1508731225

2,755,306 Viruses ?

Results by taxon

Taxonomic Groups [List]

- viruses *(2755306)*
- eukaryotes *(958029)*
  - animals *(708878)*
    - chordates *(414292)*
      - vertebrates *(413965)*
      - more... *(327)*
    - arthropods *(271220)*
    - more... *(23366)*
  - green plants *(174342)*
    - land plants *(173143)*
    - more... *(1199)*
  - fungi *(40625)*
  - trichomonads *(8470)*
  - kinetoplastids *(5689)*
  - more... *(20025)*
- other sequences *(297279)*
- bacteria *(188803)*
  - proteobacteria *(82902)*
    - g-proteobacteria *(53763)*
    - more... *(29139)*
  - firmicutes *(50605)*
  - actinobacteria *(26127)*
  - CFB group bacteria *(7433)*
  - more... *(21736)*
- unclassified *(41438)*
- archaea *(3682)*

1. Background: the explosion in sequences

A Simple Question:

How many virus genomes are in GenBank?

9,288 Viruses

12,145 Viruses in RefSeq

32,539 Viruses

346,695 Viruses ?

921,019 Viruses ?

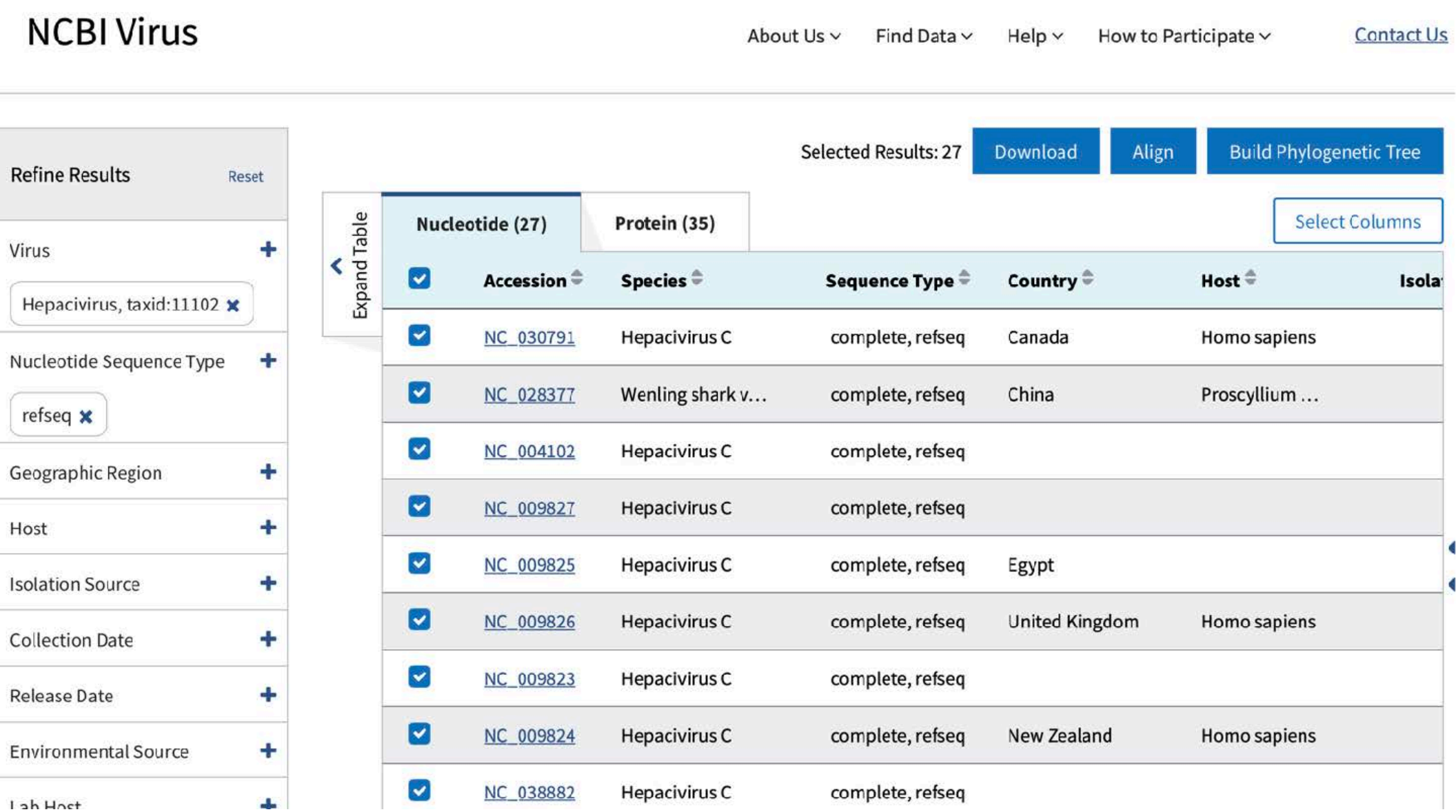
2,755,306 Viruses ?

4,391,604 Viruses ?

This article describes four foundational principles—**F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

NIH U.S. National Library of Medicine | NCBI National Center for Biotechnology Information

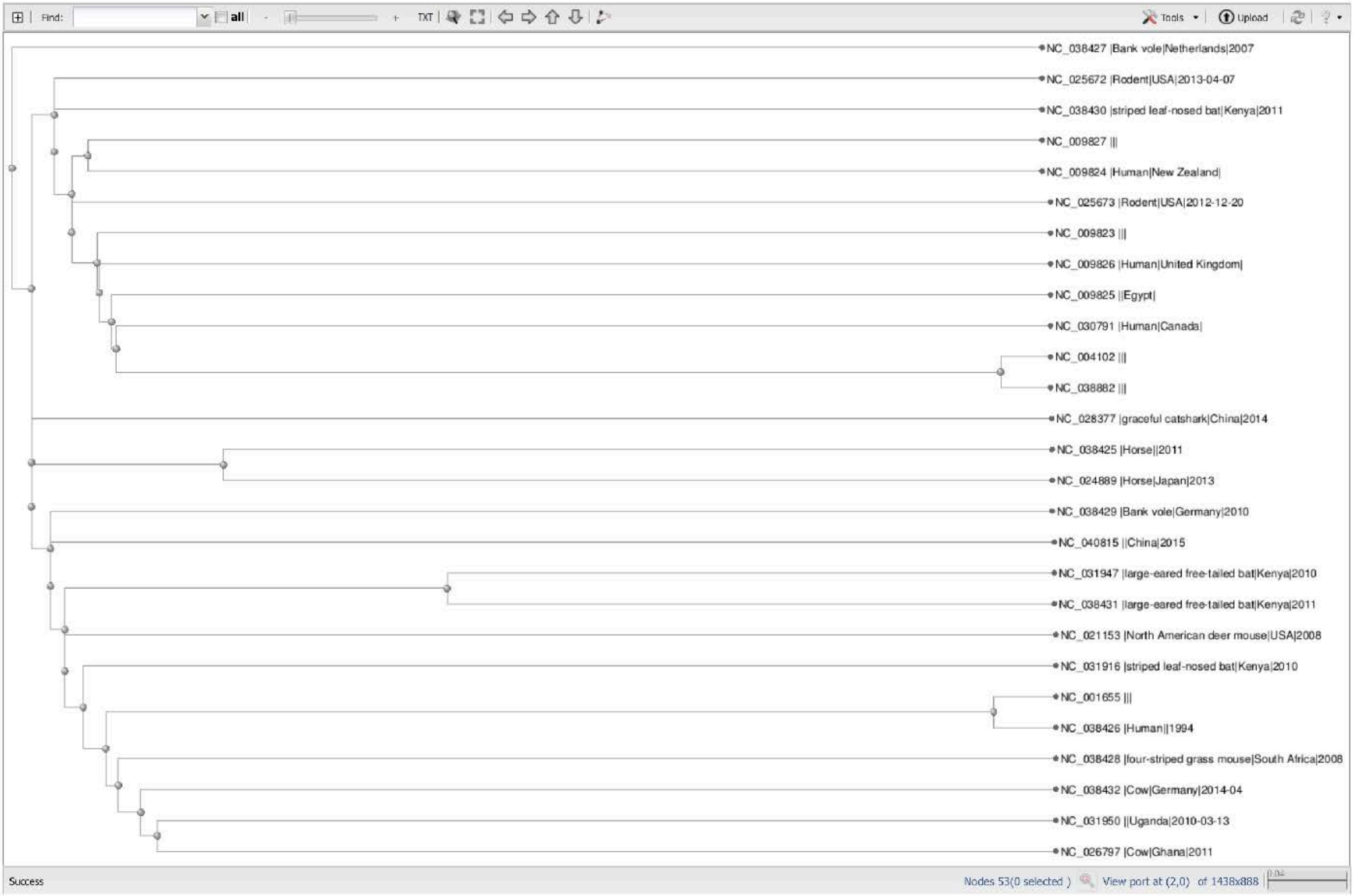
3. What is a Dendogram (and how is this different from a 'tree'?)



NIH U.S. National Library of Medicine | NCBI National Center for Biotechnology Information

27 *Hepatitis* RefSeq genomes

Phylogenetic Tree



Liver Disease

- Alagille Syndrome +
- Autoimmune Hepatitis +
- Biliary Atresia +
- Cirrhosis +
- Hemochromatosis
- Primary Sclerosing Cholangitis +
- Hepatitis (Viral) −
  - What Is Viral Hepatitis?
  - Hepatitis A
  - Hepatitis B
  - Hepatitis C
  - Hepatitis D
  - Hepatitis E
- Liver Transplant +
- Nonalcoholic Fatty Liver Disease & NASH +
- Nonalcoholic Fatty Liver Disease & NASH in Children +
- Primary Biliary Cholangitis +
- Wilson Disease +

Hepatitis (Viral)

View or Print All Sections

What is Viral Hepatitis?

Viral hepatitis is an infection that causes liver inflammation and damage. Several different viruses cause hepatitis, including hepatitis A, B, C, D, and E. The hepatitis A and E viruses typically cause acute infections. The hepatitis B, C, and D viruses can cause acute and chronic infections.

Hepatitis A

Hepatitis A causes only acute infection and typically gets better without treatment after a few weeks. The hepatitis A virus spreads through contact with an infected person's stool. You can protect yourself by getting the hepatitis A vaccine.

Hepatitis B

Hepatitis B can cause acute or chronic infection. Your doctor may recommend screening you for hepatitis B if you are pregnant or have a high chance of being infected. You can protect yourself from hepatitis B by getting the hepatitis B vaccine.

- Hepatitis B: What Asian and Pacific Islander Americans Need to Know

Hepatitis C

Hepatitis C can cause acute or chronic infection. Your doctor may recommend screening you for hepatitis C if you have a high chance of being infected or were born between 1945 and 1965. Early diagnosis and treatment can prevent liver damage.

Hepatitis D

The hepatitis D virus is unusual because it can only infect you when you also have a hepatitis B virus infection. A coinfection occurs when you get both hepatitis D and hepatitis B infections at the same time. A superinfection occurs if you already have chronic hepatitis B and then become infected with hepatitis D.

Hepatitis E

Hepatitis E is typically an acute infection that gets better without treatment after several weeks. Some types of hepatitis E virus are spread by drinking water contaminated by an infected person's stool. Other types are spread by eating undercooked pork or wild game.

OXFORD

FEMS JOURNALS investing in science

FEMS Microbiology Reviews, fuv031

doi: 10.1093/femsre/fuv031
Review Article

FEMS Microbiology Reviews, Volume 39, Issue 5, pages 764 - 778 [September,2015]

Se-Ran Jun
Assistant Professor
UAMS

★ EDITOR'S CHOICE ★

REVIEW ARTICLE

Ebolavirus comparative genomics

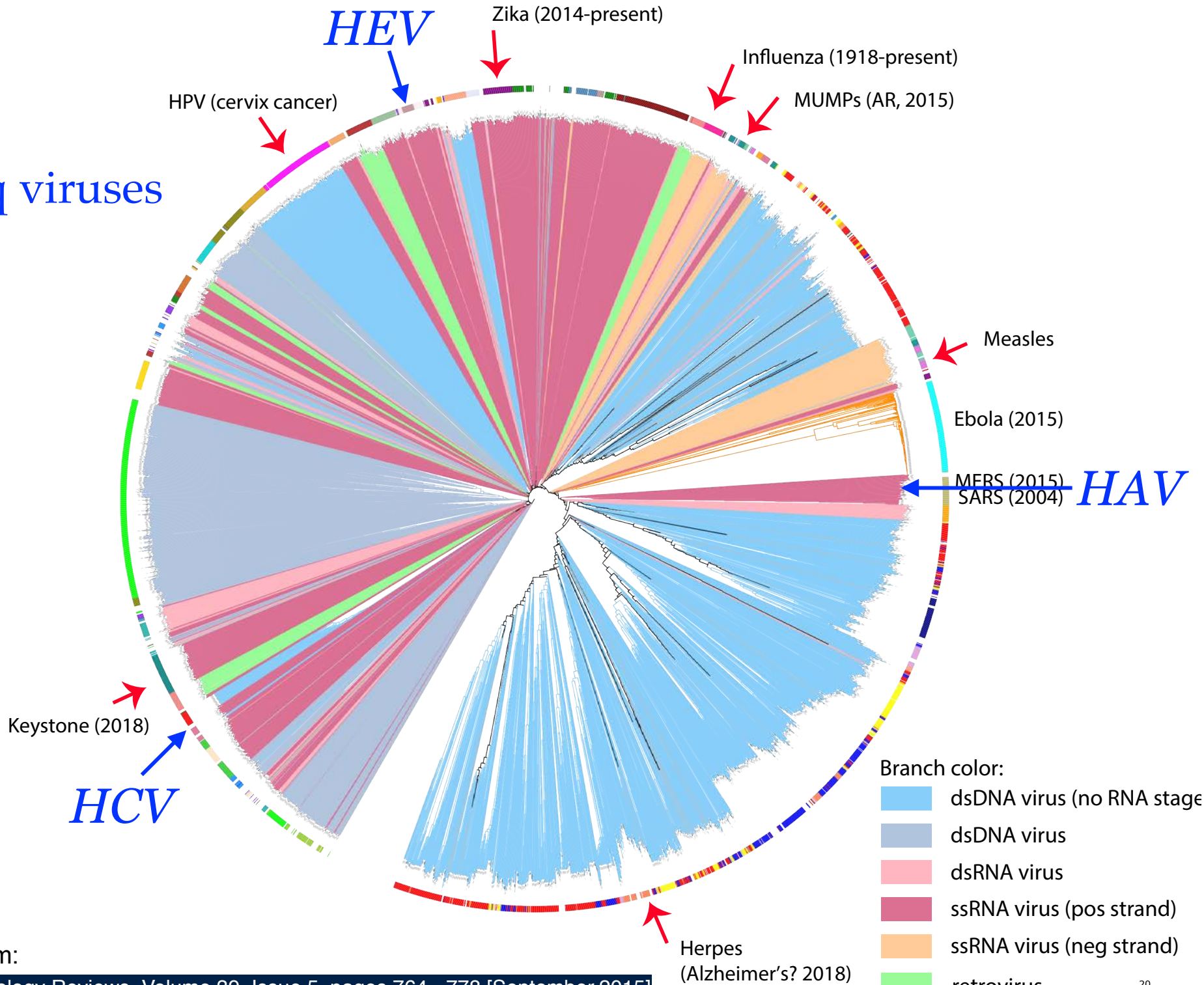
Se-Ran Jun[1,2,†], Michael R. Leuze[3,†], Intawat Nookaew[1], Edward C. Uberbacher[1], Miriam Land[1], Qian Zhang[1,4], Visanu Wanchai[1], Juanjuan Chai[3], Morten Nielsen[5,6], Thomas Trolle[5], Ole Lund[5], Greg Buzard[7], Thomas D. Pedersen[5,8], Trudy M. Wassenaar[9] and David W. Ussery[1,4,5,*]

ABSTRACT

The 2014 Ebola outbreak in West Africa is the largest documented for this virus. To examine the dynamics of this genome, we compare more than 100 currently available ebolavirus genomes to each other and to other viral genomes. Based on oligomer frequency analysis, the family *Filoviridae* forms a distinct group from all other sequenced viral genomes. All filovirus genomes sequenced to date encode proteins with similar functions and gene order, although there is considerable divergence in sequences between the three genera *Ebolavirus, Cuevavirus* and *Marburgvirus* within the family *Filoviridae*. Whereas all ebolavirus genomes are quite similar (multiple sequences of the same strain are often identical), variation is most common in the intergenic regions and within specific areas of the genes encoding the glycoprotein (GP), nucleoprotein (NP) and polymerase (L). We predict regions that could contain epitope-binding sites, which might be good vaccine targets. This information, combined with glycosylation sites and experimentally determined epitopes, can identify the most promising regions for the development of therapeutic strategies.

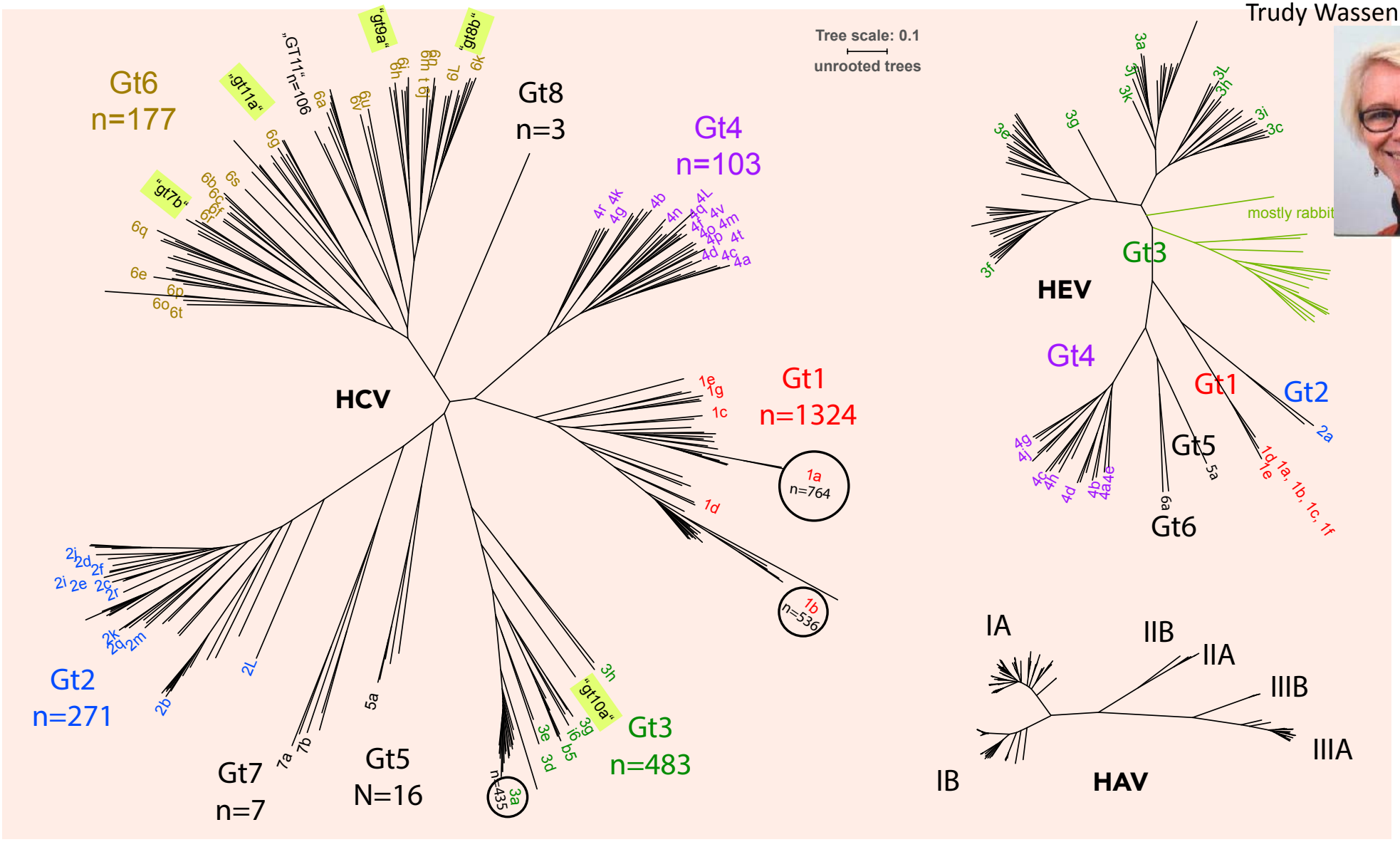
Keywords: Ebola; comparative genomics; viral genomes; epitope prediction; Ebola virus disease (EVD); Filovirus

RefSeq viruses



modified from:

FEMS Microbiology Reviews, Volume 39, Issue 5, pages 764 - 778 [September,2015]



The genetic diversity of HCV is the highest of the three species. The genotypes 1 to 7 are well resolved, except for a few genomes whose incorrect genotype were all described in the same ref.[4] Gts 4 and 1 have been described as younger than the other gts,[1] which is reflected by their position on the tree. The abundance of genotypes (indicated for some large clusters) does not reflect frequency of infection, as genome sequence datasets are biased.

TM Wassenaar, SR Jun, M Robeson, D W Ussery, "Comparative genomics of Hepatitis A virus, Hepatitis C virus and Hepatitis E virus provides insights into the evolutionary history of Hepatovirus species", submitted to Microbiology Open, August, 2019

Trudy Wassenaar, Ph.D.

Se-Ran Jun, Ph.D.

Mike Robeson, Ph.D.

Phylogenetic analysis

The genomes were aligned by Mafft (Yamada et al. 2016) and FastTree was used to build phylogenetic Maximum-Likelihood (ML) trees (Price et al. 2009). This infers approximately maximum likelihood phylogenetic trees and is much faster than other algorithms; we used the generalized time-reversible (GTR) model of nucleotide evolution and the Shimodaira-Hasegawa test for statistical confidence of internal nodes. Information on genotypes and subtypes that were included in Genbank annotations were used to map these on the trees. For visual representation, the HCV and HEV trees are shown after collapsing branches at 90% identity.

11:30 Building Standardized Dendograms for Viruses

TM Wassenaar, SR Jun, M Robeson, D W Ussery, "Comparative genomics of Hepatitis A virus, Hepatitis C virus and Hepatitis E virus provides insights into the evolutionary history of Hepatovirus species", submitted to Microbiology Open, August, 2019

ACS central science

THE HUB

In collaboration with C&EN

DOI: 10.1021/acscentsci.9b00557
ACS Cent. Sci. 2019, 5, 929−932

Standards Seekers Put the Human Microbiome in Their Sights

Published: June 17, 2019

Alla Katsnelson

NIST and others work to develop a reference standard for the human gut microbiome.

One day in the future, doctors may add a step to your routine checkup. In addition to measuring your blood pressure, putting a stethoscope to your chest, and running some blood tests, they may examine your poop. The stool sample you provide, loaded with the bacteria and other microbial matter that populate your gut, could divulge whether you have a particular disease or provide clues about your diet, stress levels, or other health markers. On the basis of that stool sample, doctors might prescribe medicines like the ones we already have—drugs that target specific proteins in our guts, for instance. But they might also give you medicines that rebalance the types and amounts of bacteria in your gut microbiome.

Dozens of companies are racing to develop drugs (or microbes prescribed as drugs) that target the microbiome and that change its composition, aiming to treat a wide range of conditions, including recurring bacterial infections, autism, depression, Parkinson's disease, asthma, cancer, and more. So far, no such drugs have been approved by regulatory agencies, although a few are close. And academic laboratories are still trying to work out some basic questions, including the full lineup of microbial species inhabiting the gut and how they interact with each other to

The US National Institute of Standards and Technology has standards for peanut butter, limestone, and more. Will poop be next? Credit: J. Stoughton/NIST.

an actual human gut—that could be used as a yardstick for aligning results across the field. The US National Institute of Standards and Technology (NIST) has set out to create this and other tools that could serve as standards for microbiome measurements. Such tools will be crucial for gaining a full picture of the microbiome's basic features, as well as getting a handle on how to modulate it for therapies.

A complex ecosystem

"The human gut microbiome has been described—and I believe rightfully so—as the most complex ecosystem on Earth," says Scott Jackson, who leads the Complex Microbial Systems Group at NIST. "It raises measurement

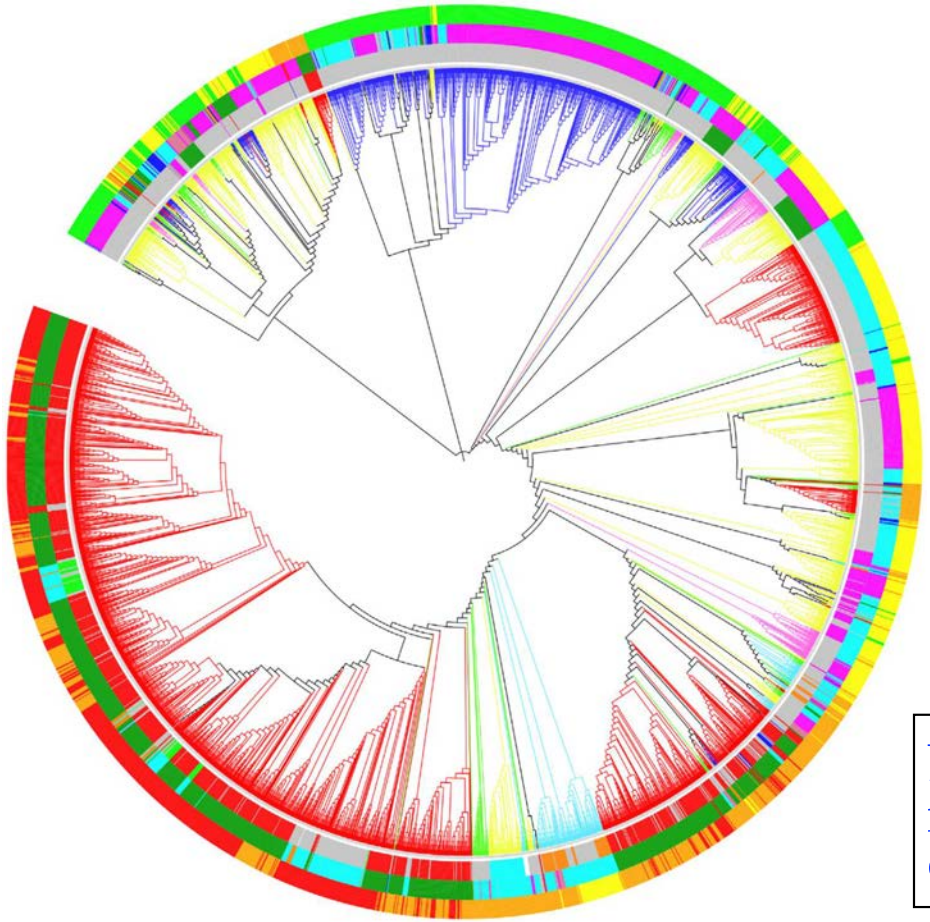
Published: June 17, 2019

But even before researchers extract DNA, they must consider how a sample is collected. Is it preserved at the point of collection or carried fresh to the lab? Is it an actual stool sample or the toilet tissue used to wipe? Different DNA profiles also arise from different procedures for sequencing samples and from differences in the bioinformatic tools for identifying which microbes are present. "A colleague and I tracked it—there are currently 97 different ways to analyze the same raw data, and they will give you 97 different answers," NIST's Jackson says.

That variation adds up in unpredictable ways, and it has undoubtedly muddled the literature. "That's why you don't get reproducible results," says Rashmi Sinha, an epidemiologist at

DOI: 10.1021/acscentsci.9b00557
ACS Cent. Sci. 2019, 5, 929−932

4. How can we standardize this?



Some suggestions:

1. Anchor to a reference

RefSeq viruses

2. Check genome quality

3. Check alignment length

4. Use FAIR guidelines

Benchmark with standardized, reproducible methods, including 'containers' for computational flow, with reference data

Figure 9. Optimal dendrogram of the 3,905 RefSeq viral genomes ($k=9$). The braches are colored by Baltimore classifications. The circles, from inside to outside, are colored by different orders, hosts, and genome sizes as follows: The ranches by Baltimore classification: dsDNA (no RNA stage), red; dsRNA, green; RT viruses, pink; ssDNA, blue; ssRNA negative-strand, bright blue; ssRNA positive-strand, yellow. The First circle, inside to outside, by order: *Caudovirales*, red; *Herpesvirales*, green; *Ligamenvirales*, blue; *Mononegavirales*, orange; *Nidovirales*, cyan: *Picornavirales*, pink; *Tymovirales*, dark green; unclassified, silver. The second circle, inside to outside, by host: protist, orange; archaea, red; bacteria, dark green; fungi, blue; animal, cyan; animal and plants, pale violet red; plant, pink; environment or NA, silver.The third circle, inside to outside, by genome size: Q1, green; Q2, yellow; Q3, orange; Q4, red.

Q Zhang, SR Jun, M Leuze, D Ussery, I Nookaew, "Viral Phylogenomics Using an Alignment-Free Method: A Three-Step Approach to Determine Optimal Length of *k-mer*." *Sci. Rep.* **7**, 40712; doi: 10.1038/srep40712 (2017).

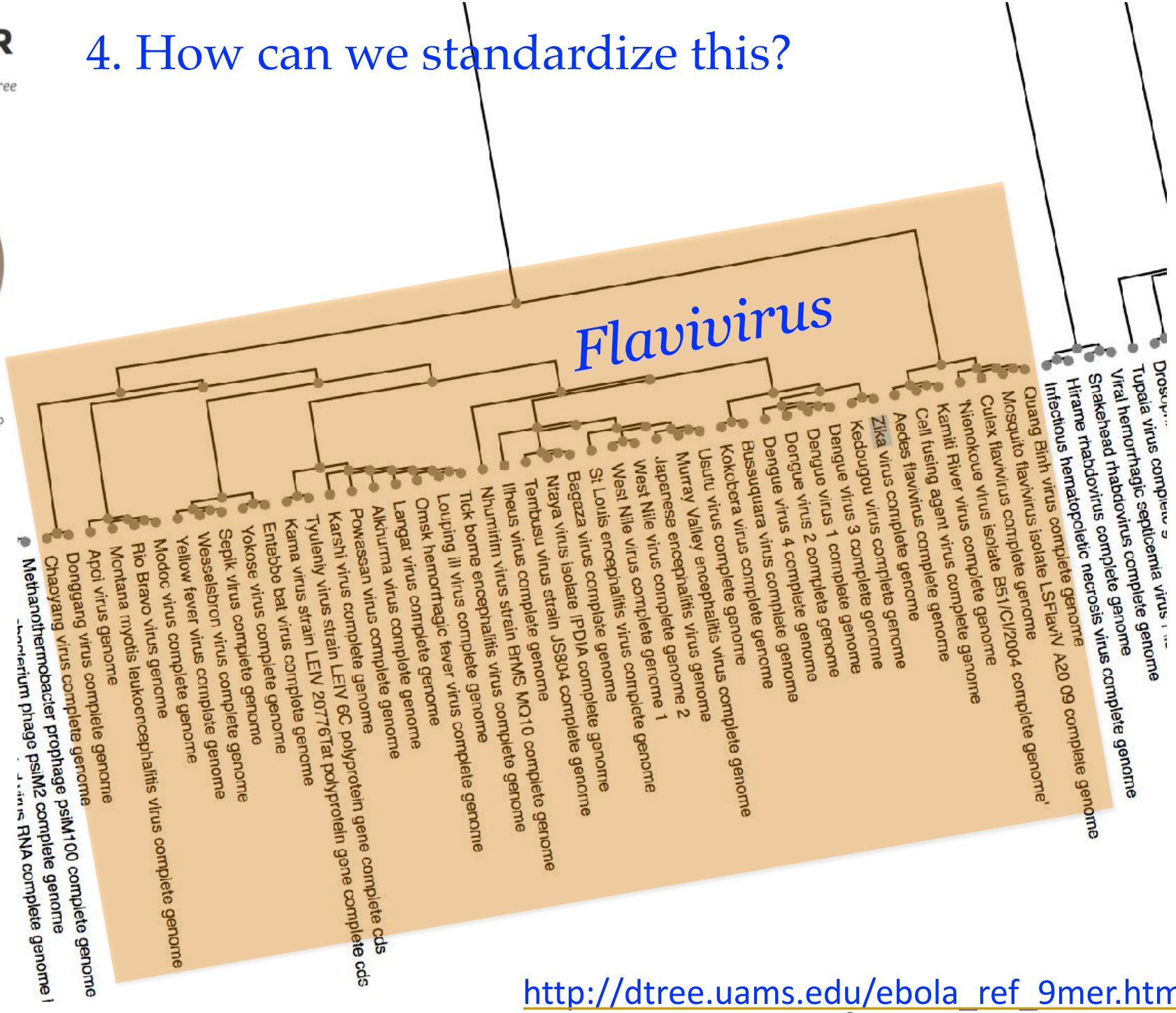
DEVELOPER

Who is currently in charge of Dtree

Visanu Wanchai

The Comparative Genomics group

4. How can we standardize this?



http://dtree.uams.edu/ebola_ref_9mer.html