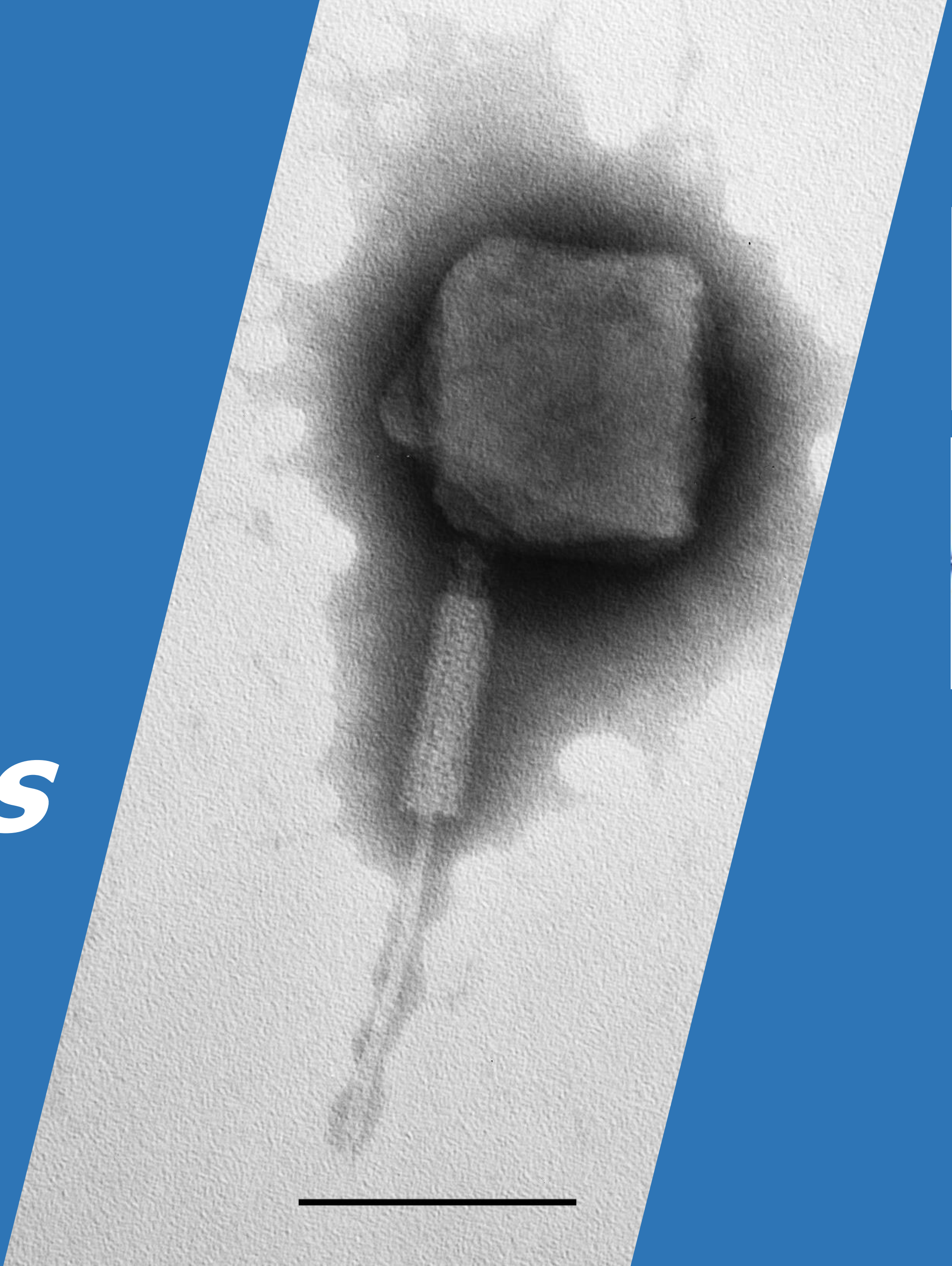
Classification of bacterial viruses in the class *Caudoviricetes*



Evelien Adriaenssens

Dann Turner

Introduction

- Classification of bacterial viruses is overseen of the ICTV Bacterial Viruses Subcommittee
- Major changes in 2021
  - Abolishment of the *Caudovirales* and the morphological families *Myoviridae*, *Siphoviridae* and *Podoviridae*
- Establishment of guidelines for genome-based classification
- New tools available to support classification, e.g.
  - INPHARED
  - GRAViTy
  - vConTACT2
  - virCLUST

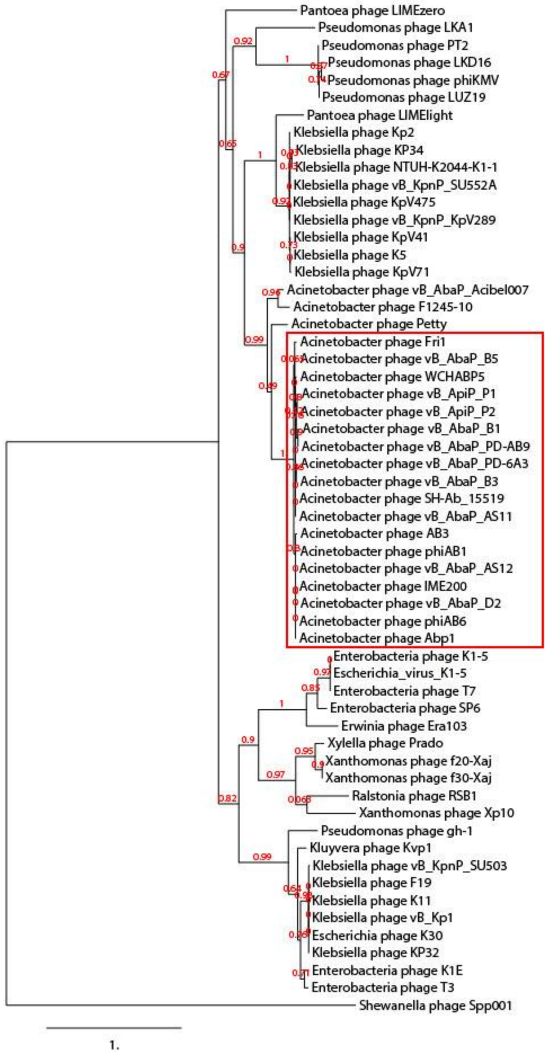
Classification Criteria

- **Species**
  - <95% sequence similarity
- **Genus**
  - >70% sequence similarity
  - Pangenome analysis for conserved ‘signature’ genes
- **Subfamily**
  - Relationships between discrete genera
  - Form clade in marker tree ML phylogeny

Nucleotide BLAST
nucleotide ▶ nucleotide

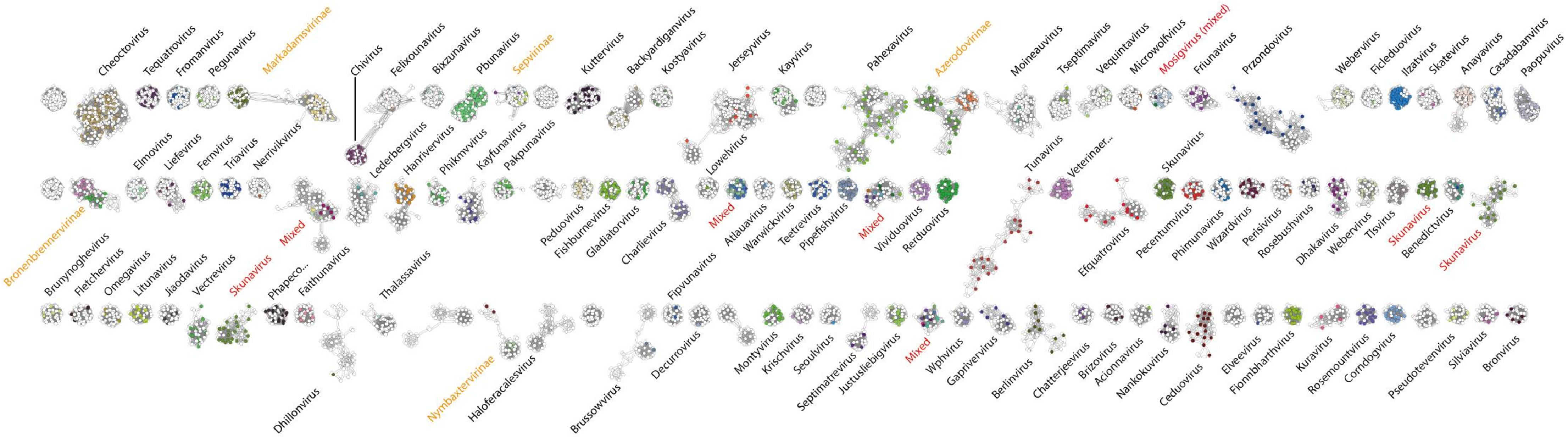
VIRIDIC **VICTOR** **ViPTree** **PHROGs**

Roary **ProteinOrtho** **vConTACT2** **GRAVITY**



Genus level classification

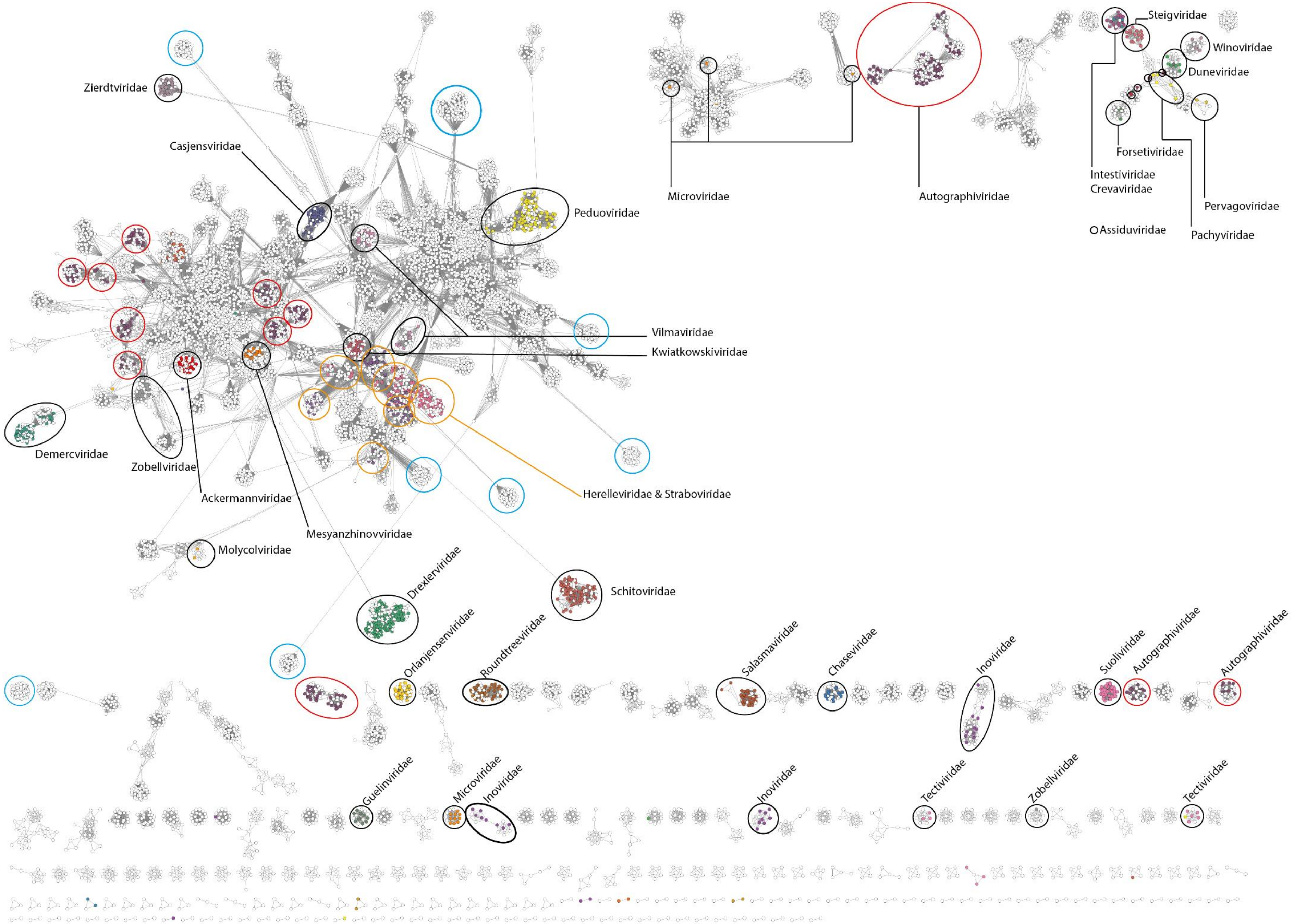
- **MASH** network analysis may allow for rapid delineation of genera, species and strains
- Ideally requires verification using a second method (e.g. **VIRIDIC**)



Partial network of bacterial viruses generated using an all-versus-all MASH analysis and visualised using Cytoscape. The clusters are *generally* concordant with genus and subfamily level assignments. (Unpublished)

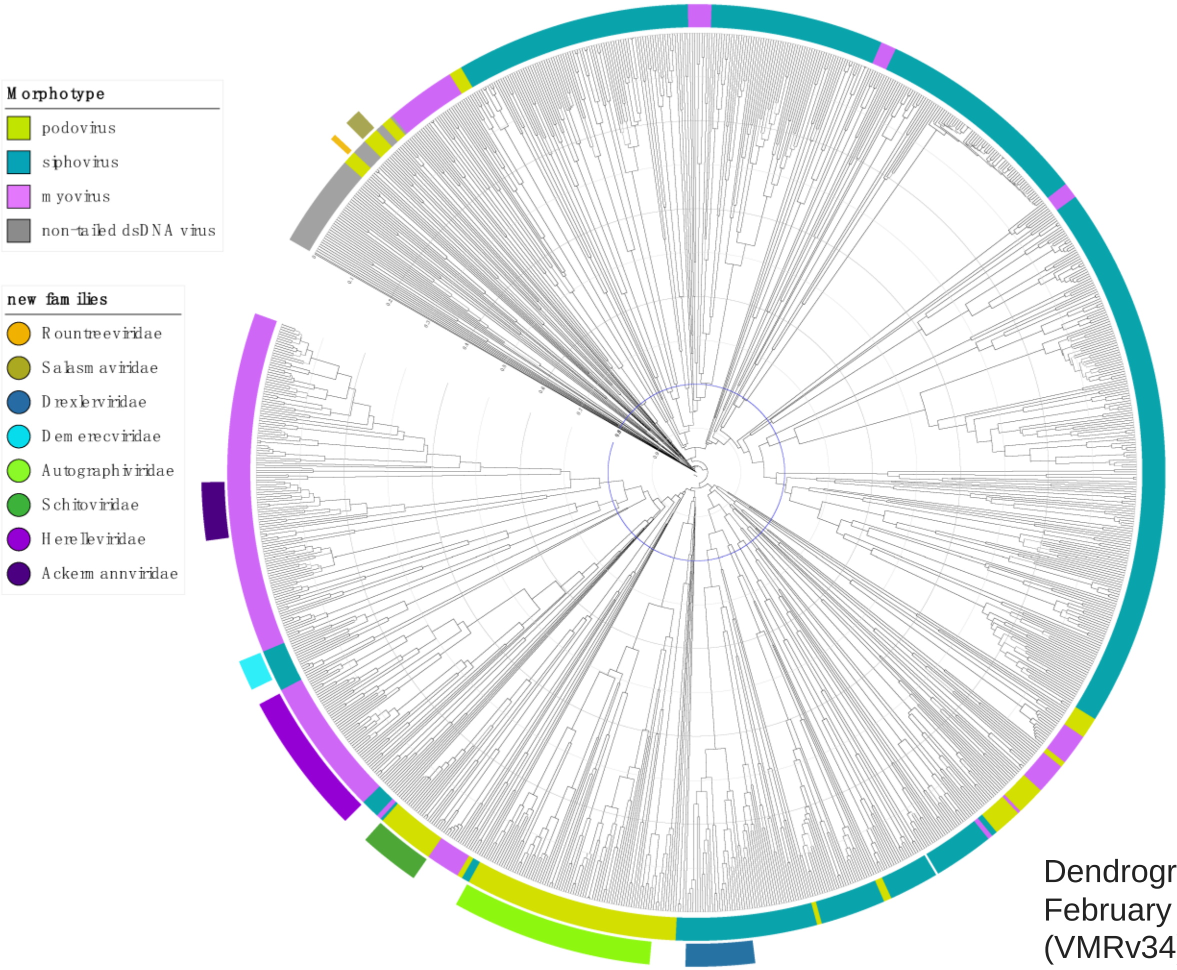
Family Level Classifcation: Global Views

- **Families** need to be cohesive and monophyletic
- Should share a significant number (how many?) of orthologous genes
- **Two** main tools available: **GRAViTy** and **vConTACT2**
- **vConTACT2** combined with the **INPHARED** pipeline
- Scripts available to map ICTV VMR data to resultant network
- A wish list:
  - Integration of nucleotide sequence similarity
  - Hierarchical clustering output of nearest neighbours based on PC content to a query genome
  - A measure of synteny



Network analysis produced using vCONTact2 with input genomes from the INPHARED pipeline (March 2022 release) and mapped against family level taxonomy of bacterial viruses. (Unpublished)

GRAViTy



- Measures of shared protein content and genome organisation
- Another wish list…
  - More frequent updates to reference database
  - Address barriers to local implementation
  - Scaling for large datasets
  - HMM recalculations

Dendrogram generated by GRAViTy (**http://gravity.cvr.gla.ac.uk**, accessed on 5 February 2021) for DB-B: Baltimore Group Ib—Prokaryotic and archaeal dsDNA viruses (VMRv34) and annotated using iTOL

Current Issues

- Classification lags significantly behind current INSDC releases
  - c. 7,000 bacterial viruses classified by BVS
  - c. 20,000 genomes in GenBank (March 2021)
- Recombination/HGT?
  - Not an evident problem for strictly lytic bacterial viruses
  - More apparent as a confounding factor for temperate viruses
- Metagenomic studies
  - How to cope with the increases in metagenome-derived information?
- A global view
  - How to assess congruity between different methods
  - Maintaining and updating source databases for bioinformatics tools

Acknowledgements

Universidad de los Andes Colombia
Alejandro Reyes Muñoz

DSMZ
Johannes Wittmann

University of Leicester
Andrew Millard

USC Australia
Ipek Kurtboke

University of Helsinki
Hanna Oksanen
Minna Poranen

KU Leuven
Rob Lavigne
Cédric Lood

Utrecht University
Bas Dutilh

UCLouvain
Annika Gillis

University of the Western Cape
Leonardo van Zyl

Pontificia Universidad Católica de Chile
Andrea Moreno-Switt

University of Zagreb
Petar Knezevic

Makerere University
Jesca Nakavuma

Flinders University
Rob Edwards

LV BMC Latvian Biomedical Research and Study Centre
research and education in biomedicine from genes to human
Jānis Rūmnieks

Universidade Federal de Viçosa
Poliane Alfenas-Zerbini

apc Microbiome Ireland
Interfacing Food & Medicine
Andrey Shkoporov

北京化工大学
Beijing University of Chemical Technology
Yigang Tong

AmpliPhi Biosciences Corporation
Susan Lehman

icmr Indian Council of Medical Research | NICED National Institute of Cholera and Enteric Diseases
BL Sarkar

Université Blaise Pascal Clermont-Ferrand UBP
François Enault

The Ohio State University
Matthew Sullivan

References

- Aiewsakun P, Adriaenssens EM, Lavigne R, Kropinski AM, Simmonds P. (2018) Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. *Journal of General Virology* 99(9):1331-1343
- Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. (2019) Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology* 37(6):632-639
- Cook R, Brown B, Redgwell T, Rihtman B, Barnes M, Clokie M, Stekel DJ, Hobman , Jones MA, Millard A. (2021). INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *Phage* 2(4):214-223
- Minh BQ , Schmidt HA, Chernomor, O Schrempf D, Woodhams MD, von Haeseler A, Lanfear A. (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37:1530-1534.
- Moraru C, Varsani A, Kropinski AM. (2020) VIRIDIC-A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses* 12(11):1268
- Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. (2017) ViPTree: the viral proteomic tree server. *Bioinformatics* 33(15):2379-2380
- Simmonds P, Aiewsakun P. (2018) Virus classification - where do you draw the line? *Archives of Virology* 163(8):2037-2046
- Turner D, Kropinski AM, Adriaenssens EM. (2021) A Roadmap for Genome-Based Phage Taxonomy. *Viruses* 18;13(3):506

Questions