

Agriculture & Agri-Food Canada Phage Genomics Workshop

Andrew M. Kropinski

Department of Pathobiology
University of Guelph, Canada

Email:
Phage.Canada@gmail.com



Agriculture and
Agri-Food Canada
Agriculture et
Agroalimentaire Canada

UNIVERSITY
of GUELPH

PART 3

Taxonomy

Outline

- ☐ History of phage taxonomy
- ☐ Who is responsible for the taxonomy of phages?
- ☐ How do I classify my phage?
- ☐ How do I get my classification officially recognized?

Nobody likes change but it is inevitable

- ❑ "There is nothing permanent except change"
(Heraclitus of Ephesus, c. 535 – c. 475 BCE)
- ❑ "Progress is impossible without change"
(George Bernard Shaw, 1856 – 1950)
- ❑ "Intelligence is the ability to adapt to change"
(Stephen Hawking, 1942 – 2018).



Taxonomy



- ❑ Taxonomy is a human invention
- ❑ It is not stable, with each generation comes a new viewpoint; plus, the plethora of new phages in GenBank give us a new perspectives on diversity & diversification
- ❑ Bacterial viral taxonomy is the purview of ICTV – Bacterial Viruses Subcommittee
- ❑ Taxonomy is controversial



Franz Joseph Haydn (1732–1809) – The Creation/Die Schöpfung



- ❑ Adam Lau, bass, singing the part of Adam about the newly created animals: lion, tiger, stag, horse, cattle, sheep, insects, and ...
- ❑ <https://www.youtube.com/watch?v=fROA7ZQ8Htk>

Sense and Nonsense in Bacterial Taxonomy (Samuel Tertius Cowan 1971)



- ❑ Taxonomy is written by taxonomists for taxonomists; in this form the subject is so dull that few, if any, non-taxonomists are tempted to read it, and presumably even fewer try their hand at it.
- ❑ ...It is the most subjective branch of any biological discipline
- ❑ ...full-time taxonomists like to think of ourselves as artists able to perceive form, shape, colour and relationships that are hidden from the gaze of the more mundane scientists.
- ❑ ...like artists, we are argumentative and aggressive, we feel self-important and are scathing of the views of others.

- ❑ "I see all these characteristics in myself, thereby confirming that I am a successful taxonomist, and since the subject is armchair biology I expect to continue at it to a ripe old age." (1905-1976)

Take-home messages

- ❑ Taxonomies change – e.g. *Pseudomonas acidovorans*, became *Comamonas acidovorans*, and now *Delftia acidovorans*
- ❑ A taxon (a taxonomic group of any rank) **is not** an individual. *Homo sapiens* in a **taxon**, each one of you is an **exemplar** – an individual human being.



The Treachery of Images aka *This is not a Pipe* - Belgian surrealist painter René Magritte (Created: 1928–1929)

History of tailed phage taxonomy

1998

- ☐ 1 Order (*Caudovirales*), 3 Families, 15 Genera & 265 Species
- ☐ Classification largely based upon morphology,
- ☐ Life style (lytic, **temperate**) and host



2011

- ☐ 1 Order, 3 Families, 5 Subfamilies, 37 Genera & 135 Species
- ☐ Classification using comparative proteomics
- ☐ Classical Families were not monophyletic



History of phage taxonomy 2

2017

- ☐ First new Family introduced – *Ackermannviridae*
- ☐ Classification using comparative genomics

2019

- ☐ ICTV introduced 15 ranks (realm, subrealm, kingdom, subkingdom, phylum, subphylum, class, subclass, order, suborder, family, subfamily, genus, subgenus, and species). Where realm (viruses) = domain (bacteria)

2022

- ☐ *Caudovirales* and its three Families abolished

History of phage taxonomy 3

2023

- ☐ Latinate binomial taxonomy mandated

Exemplar

Taxon

Salmonella phage P22

Lederbergvirus P22


Pseudomonas phage Jeff

Pbunavirus jeff

2025


- ☐ 1 Class (Caudoviricetes), 11 Orders, 105 Families, 132 Subfamilies, 1680 Genera & 5799 Species
- ☐ None of this would be possible without new bioinformatic/taxonomic tools and insights

Phage T7 - 2008



☐ In 2008 – genus “T7-like phages” with three species: T7, Kvp1 and gh-1

Phage T7 - 2025



☐ In 2025 – one Order (Autographivirales), four Families, 13 Subfamilies, 227 genera and 723 species

☐ *Escherichia* phage T7 – *Teseptimavirus T7*

☐ *Kluyvera* phage Kvp1 – *Berlinvirus Kvp1*

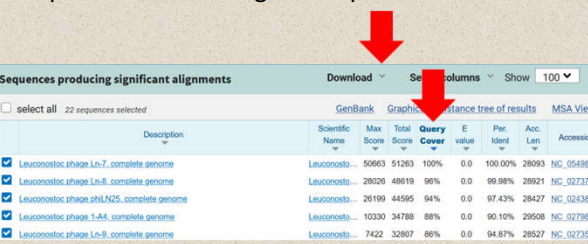
☐ *Pseudomonas* phage gh-1 – *Ghunavirus gh1*

➤ The name of your phage is: (Host genus) + phage + name **DO NOT** name your phage with the taxon name

Practical viral taxonomy

☐ BLASTn versus Caudoviricetes (taxid:2731619)

☐ you may have to increase the “Max target sequences” under “algorithm parameters”



☐ Download Fasta (complete sequence) as “seqdump.txt”

Accession number versus RefSeq number

- ☐ GenBank is part of the International Nucleotide Sequence Database Collaboration, which comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI.
- ☐ exchange information
- ☐ RefSeq records are distinguished from INSDC records by: Accession format: The most distinguishing feature of a RefSeq record is the distinct accession number format that begins with two characters followed by an underscore (e.g., NP_). INSDC accession numbers never include an underscore
- ☐ The RefSeq database - NCBI

Massaging fasta file

- ☐ Change fasta line **from**:

```
>NC_054988.1 Leuconostoc phage Ln-7, complete genome
GGTTAATAGTAGTCTTTTGAATCTATGCCCAATTTACGAAAGTCCGAATAATACGTAATACGAATTAAATTTATCTC
TCATTATTAGAAATATATGGCAAAATAAAAAGACCTAACATAAGTTACATCAAATGTCAAGCCTTTATTGTTATTAAATT
```

to:

```
>NC_054988.1 Leuc phg Ln-7
GGTTAATAGTAGTCTTTTGAATCTATGCCCAATTTACGAAAGTCCGAATAATACGTAATACGAATTAAATTTATCTC
TCATTATTAGAAATATATGGCAAAATAAAAAGACCTAACATAAGTTACATCAAATGTCAAGCCTTTATTGTTATTAAATT
```

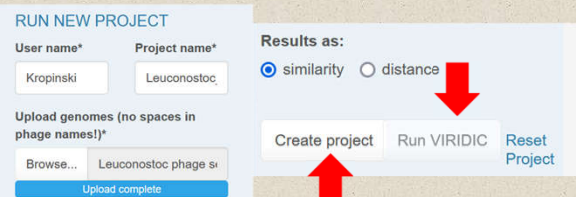
- ☐ N.B. If you are satisfied with just the accession or RefSeq number don't do anything

Submit to -

- ☐ VIRIDIC

➤ <https://rhea.icbm.uni-oldenburg.de/viridic/>

➤ complete and upload



- ☐ Save project number: **Current project ID: P4253**
 ___Kropinski___Leuconostoc_phages___
20250207183018



Shared proteins - CoreGenes

CoreGenes 5.0

(<https://coregenes.ngrok.io/>) CoreGenes 3.5
(<http://binf.gmu.edu:8080/CoreGenes3.5/>)

The number of homologs in each column is : 26
Total number of genes in genome 1 : 40
% in common, with respect to genome 1: 65
Total number of genes in genome 2 : 41
% in common, with respect to genome 2: 63.41



Leuconostoc phage Lmd1	Leuconostoc phage LNS4
NC_018273	NC_024388
PE-YP_006560458.1	PE-YP_009044872.1
PRODUCT:HNH endonuclease	PRODUCT:putative phage HNH endonuclease
PE-YP_006560460.1	PE-YP_009044875.1
PRODUCT:RusA-like Holliday junction resolvase	PRODUCT:RusA-like Holliday junction resolvase
PE-YP_006560461.1	PE-YP_009044876.1
PRODUCT:DNA helicase	PRODUCT:DNA helicase

➤ High % of homologous proteins - Subfamily

Phylogeny

"One Click" at <https://ngphylogeny.fr/>

Lemoine F et al. 2019. Nucleic Acids Res.
47(W1): W260-W265. doi: 10.1093/nar/gkz303.
PMID: 31028399

latter program interfaces with iTOL



Phylogeny 2

in some cases a phylogenetic tree based upon a single protein doesn't support the proposed taxonomy; repeat with a polyprotein

Take home message 2

- ☐ Name your isolated after the primary isolation host i.e. *Escherichia* phage xxx
- ☐ Check Adriaenssens & Brister (2017). How to Name and Classify Your Phage: An Informal Guide (<https://www.mdpi.com/1999-4915/9/4/70>)

Take home message 3

- ☐ Consult the ICTV website (<https://ictv.global/>) for the latest taxonomy
- ☐ Consider sending an email to the Bacterial Viruses Subcommittee
- ☐ Be willing to accept that you isolate may be currently a genomic orphan
- ☐ Consider submitting an International Community on Taxonomy of Viruses (ICTV) Taxonomy Proposal (TaxoProp)

ICTV Documentation

- ☐ ICTV Taxonomy Proposal: two parts (Word & Excel)
<https://ictv.global/taxonomy/templates>

ICTV Documentation 2

- ❑ Species demarcation criteria: Two phages are assigned to the same species if their genomes are more than 95% identical over their genome length for isolates. These values can be calculated by a number of tools, such as BLASTn [1,2] – usually calculated using intergenomic distance calculator VIRIDIC [3].
- ❑ Genus demarcation criteria: In search for criteria that create cohesive and distinct genera that are reproducible and monophyletic, the Bacterial Viruses Subcommittee has established 70% nucleotide identity of the genome length as the cut-off for genera. Genus-level groupings should always be monophyletic in the signature genes, as tested with a phylogenetic tree. [10]
- ❑ Subfamily demarcation criteria: Subfamilies are to be created when two or more genera are related below the family level. In practical terms, this usually means that they share a low degree of sequence similarity (usually about 40-50%) and that the genera form a clade in a marker tree phylogeny. [10]
- ❑ Family demarcation criteria: The family is represented by a cohesive and monophyletic group in the main predicted proteome-based clustering tools (VirClust, VIPTree, GRAVITY dendrogram, vConTACT2 network). Members of the family share a significant number of orthologous genes (the number will depend on the genome sizes and number of coding sequences of members of the family). [10]

Comparative genomics and proteomics

- ❑ Classical
 - progressiveMauve - Darling AE et al. 2010. PLoS One **5(6)**: e11147. doi: 10.1371/journal.pone.0011147. PMID: 20593022
<https://usegalaxy.eu/>
 - Easyfig - Sullivan MJ et al. 2011. Bioinformatics **27(7)**: 1009-10. doi: 10.1093/bioinformatics/btr039. PMID: 21278367
<http://mjsull.github.io/Easyfig/files.html>

Comparative genomics and proteomics

- ❑ Modern
 - clinker at CAGECAT - CompAraTive GEne Cluster Analysis Toolbox
 - Gilchrist CLM & Chooi YH. 2021. Bioinformatics. **37(16)**: 2473-2475. doi: 10.1093/bioinformatics/btab007. PMID: 33459763
<https://cagecat.bioinformatics.nl/>
 - input: gbk files

clinker 2

Job W03054245325K24 is currently running.


- 1. Sanitizing input files
- 2. Parse genome files
- 3. Align clusters
- 4. Summarizing results
- 5. Save session file
- 6. Writing results

[Bookmark this page](#)

Pseudomonas phage SN

Pseudomonas phage tabernarius

DiGAlign - Dynamic Genomic Alignment



- ☐ requires: poly fasta-formatted genomes
- ☐ Nishimura Y et al. 2024. Microbes Environ. **39(1)**: ME23061. doi: 10.1264/jsme2.ME23061. PMID: 38508742
- ☐ <https://www.genome.jp/digalign>

Type of BLAST

☒ BLASTn ☐ tBLASTx


- ☐ Valid email address required

DiGAlign 2

A notification email will be sent when the computation finishes.

Check input sequence	✓ validated
BLAST	✓ finished
Tree	✓ generated
Gene finding and annotation	✓ finished

All computation is finished. ---> [start to browse results](#)

 **Browse alignment**

[browse an alignment](#)

