

The Phage Proteomic Tree: A Genome-Based Taxonomy for Phage

Student: Jade Cheng
Course: ICS 675
Assignment: Final Paper Review
Date: December 3, 2009



Introduction to Phages

Definition

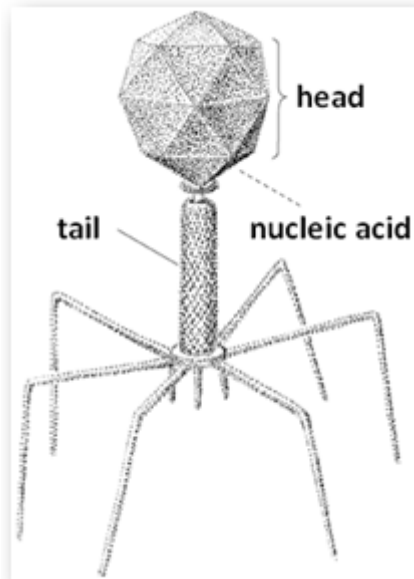
Phages are viruses that infect prokaryotes (bacteria).

Composition

All contain nucleic acid and protein.

Structure

Typical structure contains a Head and a Tail. Nucleic acid is in the head.



Phylogeny Study of Phages

Traditional Approach

Categorize phages by their physical characteristics, such as host range, capsid shape, genome size, and genome type.

Authority

Phage taxonomic system is approved and updated by ICTV.



Problems

Visualization of phage particles is not always possible. Information about phage lifestyle is not always available.

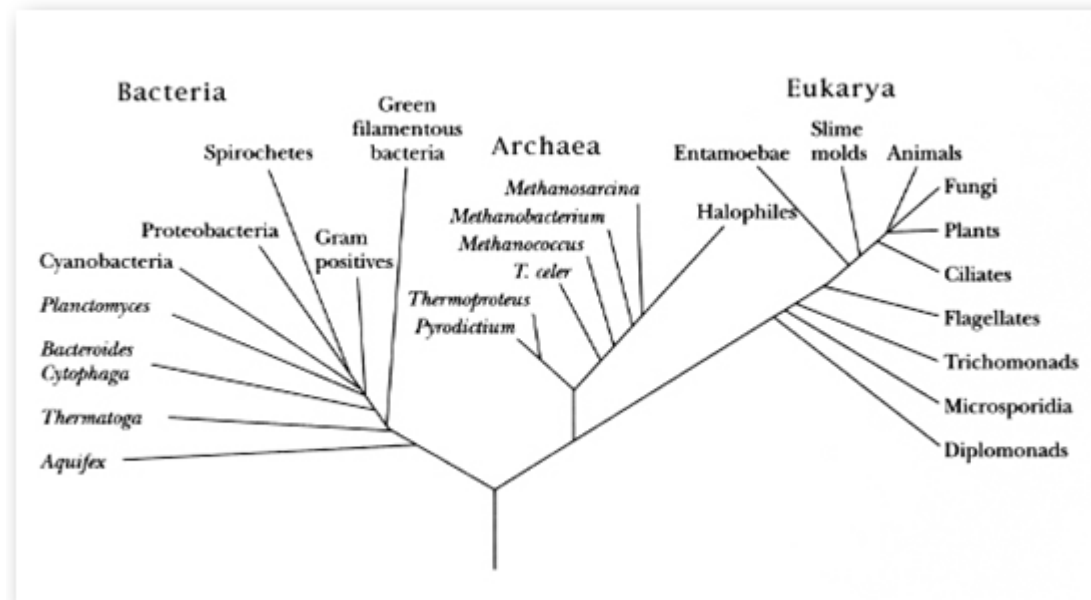
Resources/Opportunities

Large GenBank database of complete sequenced phages is accessible.

Obstacle for Phage Genome Based Taxonomical Study

Genomic Marker (rDNA, rRNA)

Ribosomal DNA, RNA are highly conservative and serve as evolutionary clocks.



A speculatively rooted tree for rRNA genes

Phages do not contain ribosomal sequences.

Phage biodiversity study based on genetic markers or structural proteins has achieved limited success.

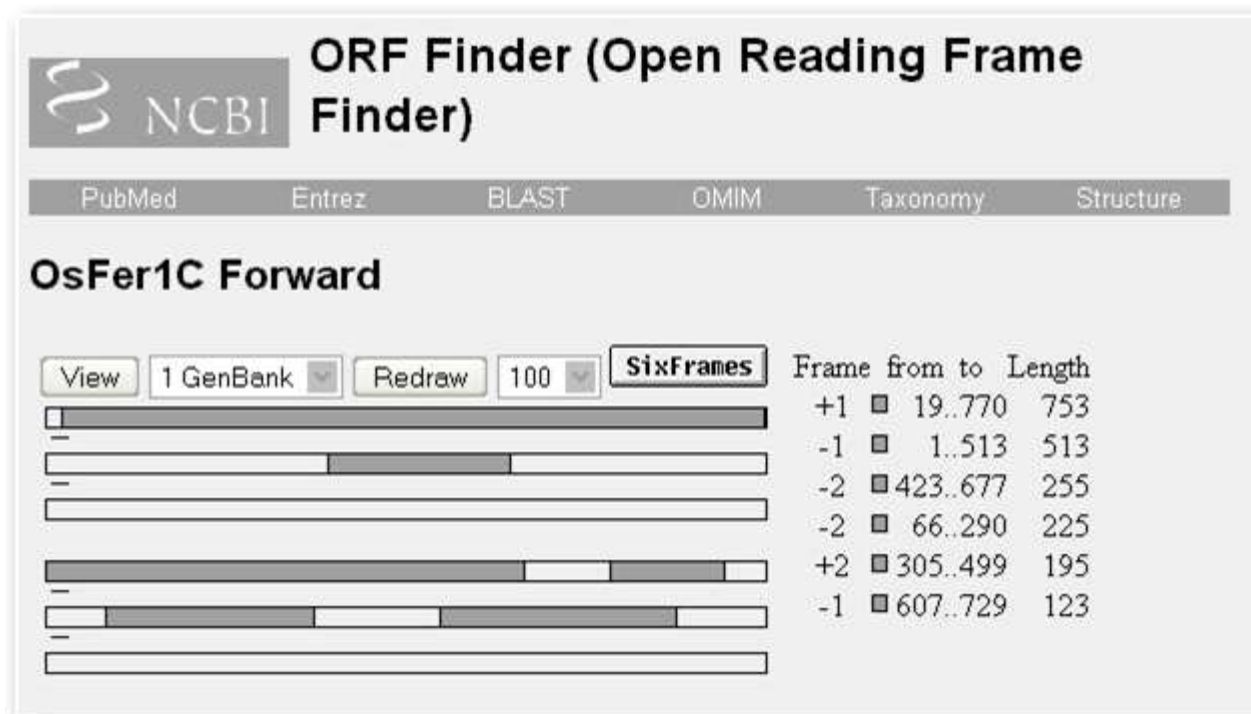
Approach Proposed: Predicted Proteome Analysis

“Proteome”

Proteome is the entire complement of proteins expressed by a genome.

“Predicted”

Open Reading Frames (ORF) were used, not the actual proteome of phages.



Tools

BLASTP Distances

BLASTP generates a Distance Matrix based on user-supplied protein sequences. The distance score (S) between pair-wise genomes is:

$$S_{pair-wise} = 1 - \frac{\# \text{ of significant hits}}{\text{total \# of genes in two genome}}$$

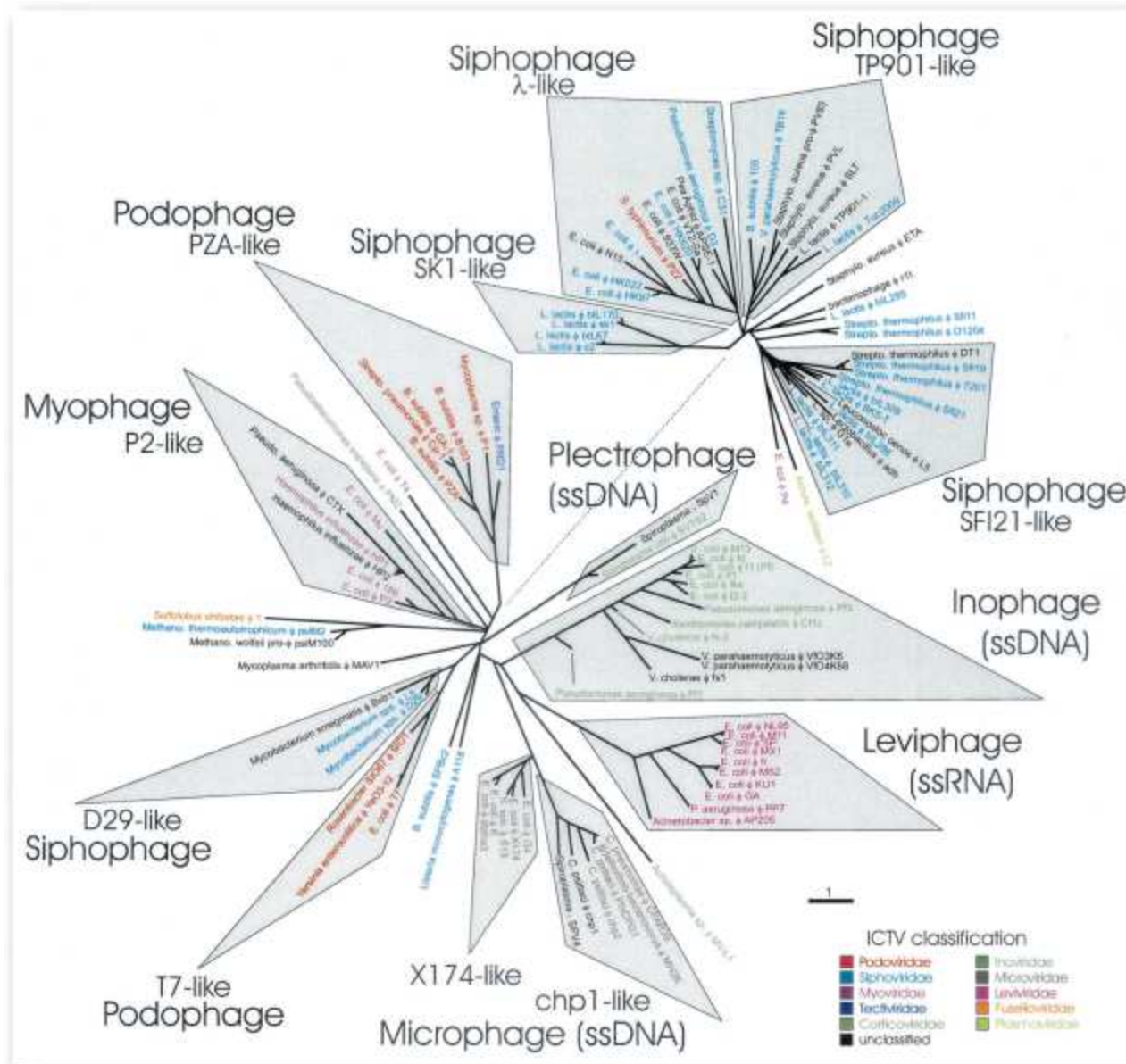
Proteomic Distances

CLUSTALW and PROTDIST generate a Distance Matrix based on user-supplied protein sequences.

Plotting the Tree

Tree generation methods, such as neighbor-joining, were used to draw the phylogeny tree based on the Distance Matrix.

Phage Proteomic Tree



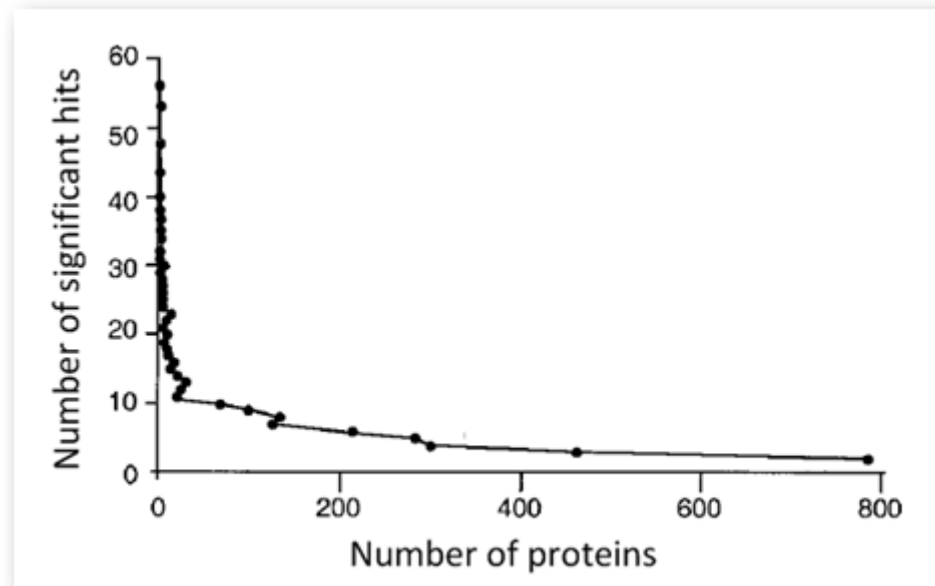
Conclusion 1: No Single Genomic Marker

Genomic Marker (rDNA, rRNA)

Ribosomal DNA, RNA are highly conservative and serve as evolutionary clocks.

Phage Genomic Marker

Researchers continuously search for a single phage genomic marker.



Conclusion from the image

No single protein was found in all 105 genomes; in fact none were even close.

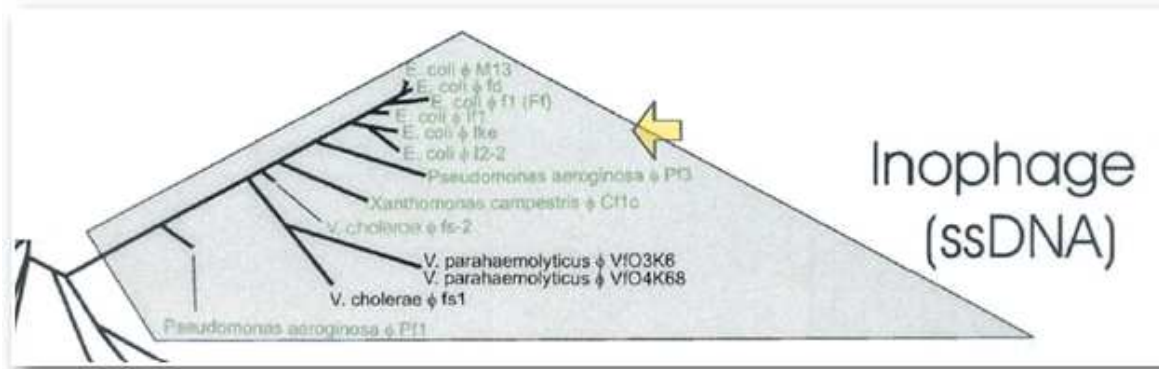
Conclusion 2: Compatible with ICTV standards

Criteria extracted from the literature that were used to evaluate the Phage Proteomic Trees^a

Criteria ^b	BLASTP result with E-value cutoff of:			Protein distance result with penalty value of:			
	0.001	0.01	0.1	0	5	10	100
dsDNA, ssDNA, and ssRNA phage are fundamentally different	N ^c	N ^c	N ^c	N ^d	Y	Y	Y
ssRNA phage fall into two groups: (i) leviviruses (fr, MS2, and GA) and (ii) alloleviviruses (SP, NL95, M11, and MX1); PP7 is an out group (7)	N	N	N	Y	Y	Y	Y
Leviviruses consist of two groups: group I (fr and MS2) and group II (GA and KU1) (28)	N	N	Y	Y	Y	Y	Y
φCPG1, φAR39, and Chp2 are more closely related to each other than to the avian <i>Chlamydia psittaci</i> φChp1 (53)	Y	N	N	Y	Y	Y	Y
α3 and φK ssDNA phage are more closely related to each other than to φX174 and G4 (36)	N	Y	N	Y	Y	Y	Y
ssDNA phages I2-2 and Ike are more closely related to each other than to F1 (58)	N	N	N	Y	Y	Y	Y
ssDNA phage Pf3 is only distantly related to M13, f1, and fd (42)	Y	Y	Y	Y	Y	Y	Y
Fs-2 is similar to f1, fd, M13, Ike, and Pf3 (33)	Y	Y	Y	Y	Y	Y	Y
ssDNA phage SVTS2 is related to <i>Spiroplasma citri</i> φSpVI (56)	Y	Y	Y	Y	Y	Y	Y
Among the podoviruses, PZA is more closely related to B103 than to GA1 (51)	Y	Y	Y	Y	Y	Y	N
<i>Mycoplasma</i> φP1 is related to other terminal protein-containing phage (e.g., φ29) (62)	Y	Y	Y	N	N	Y	Y
Podoviruses SIO1, T7, YeO3-12 are related (48, 49, 54)	Y	Y	Y	Y	Y	Y	Y
Lambdoid phage include lambda, 933W, N15, HK022, HK97, VT2-Sa, P22, D3, APSE-1, and HK620 (8, 13, 15, 37, 59, 64)	Y	Y	Y	N	Y	Y	N ^e
D29, L5, and Bxb are closely related (24, 44)	Y	Y	Y	Y	Y	Y	Y
Relative relationships of Sfi21 > adh > PVL ≥ φ105 (18)	Y	Y	N	N	Y	Y	Y
<i>Methanobacterium</i> φM2 is closely related to <i>Methanobacterium wolfeii</i> prophage ψM100 (52)	Y	Y	Y	Y	Y	Y	Y
HP1, P2, and 186 are related (21)	Y	Y	Y	Y	Y	Y	Y

Conclusion 2 Examples

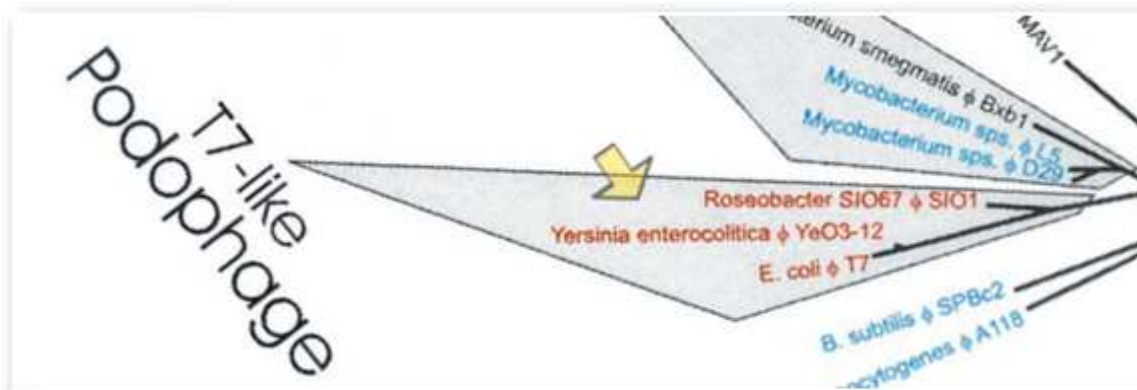
Example One



ssDNA phage Pf3 is only distantly related to M13, f1, and fd

Y Y Y Y Y Y Y

Example Two



Podoviruses SIO1, T7, YeO3-12 are related

Y Y Y Y Y Y Y

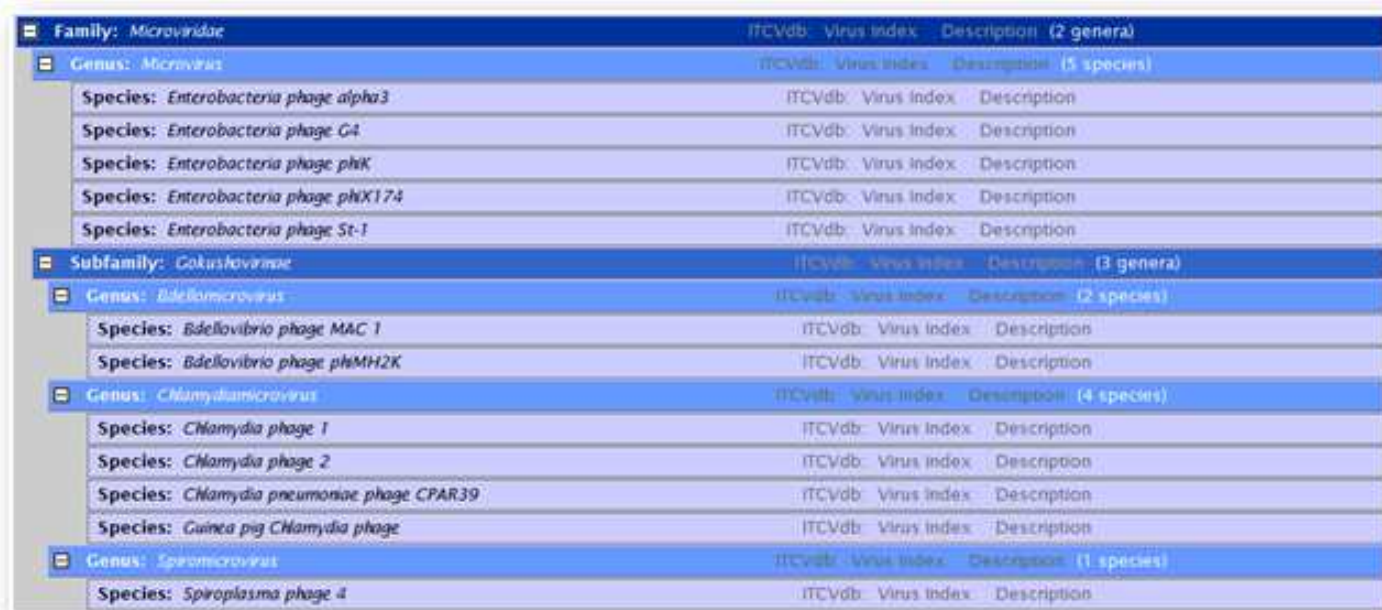
Proposals

Phage Taxonomy Scheme

The proteomic tree should be used as the genomic-based taxonomy system of phages.

Phage Naming Scheme

Phage genotypes should be named using the first host, then ϕ (indicating it is a phage), and then the first reference; e.g. *E. coli* ϕ X174.



The image shows a screenshot of the International Committee on Taxonomy of Viruses (ICTV) phage taxonomy scheme. It displays a hierarchical tree structure of phage classification. The root is the Family Microviridae, which is divided into several genera and subfamilies. Each level is expanded to show its members, including species names and their corresponding ICTVdb Virus Index and Description.

Family: Microviridae	ICTVdb: Virus Index	Description (2 genera)
Genus: Microvirus	ICTVdb: Virus Index	Description (5 species)
Species: Enterobacteria phage alpha3	ICTVdb: Virus Index	Description
Species: Enterobacteria phage G4	ICTVdb: Virus Index	Description
Species: Enterobacteria phage phiK	ICTVdb: Virus Index	Description
Species: Enterobacteria phage phiX174	ICTVdb: Virus Index	Description
Species: Enterobacteria phage St-1	ICTVdb: Virus Index	Description
Subfamily: Gokushovirinae	ICTVdb: Virus Index	Description (3 genera)
Genus: Bdellovirovirus	ICTVdb: Virus Index	Description (2 species)
Species: Bdellovibrio phage MAC 1	ICTVdb: Virus Index	Description
Species: Bdellovibrio phage phiMH2K	ICTVdb: Virus Index	Description
Genus: Chlamydiovirus	ICTVdb: Virus Index	Description (4 species)
Species: Chlamydia phage 1	ICTVdb: Virus Index	Description
Species: Chlamydia phage 2	ICTVdb: Virus Index	Description
Species: Chlamydia pneumoniae phage CPAR39	ICTVdb: Virus Index	Description
Species: Guinea pig Chlamydia phage	ICTVdb: Virus Index	Description
Genus: Spirovirus	ICTVdb: Virus Index	Description (1 species)
Species: Spiroplasma phage 4	ICTVdb: Virus Index	Description

The current ICTV phage naming scheme

Discussion

Does this phage proteomic tree reflect an evolutionary history?

- According to the paper, the answer is yes (of course).
- Its reasoning is weak.
- It states subtle improvements in the algorithm should provide a better tree.

Opposing Voices

Some people suggest it might be impossible to have a strictly hierarchical taxonomic system of phages.

Imbroglios of Viral Taxonomy: Genetic Exchange and Failings of Phenetic Approaches

Jeffrey G. Lawrence,* Graham F. Hatfull, and Roger W. Hendrix

Pittsburgh Bacteriophage Institute and Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Received 13 March 2002/Accepted 23 April 2002

The practice of classifying organisms into hierarchical groups originated with Aristotle and was codified into nearly immutable biological law by Linnaeus. The heart of taxonomy is the biological species, which forms the foundation for higher levels of classification. Whereas species have long been established among sexual eukaryotes, achieving a meaningful species concept for prokaryotes has been an onerous task and has proven exceedingly difficult for describing viruses and bacteriophages. Moreover, the assembly of viral “species” into higher-order taxonomic groupings has been even more tenuous, since these groupings were based initially on limited numbers of morphological features and more recently on overall genomic similarities. The wealth of nucleotide sequence information that catalyzed a revolution in the taxonomy of free-living organisms necessitates a reevaluation of the concept of viral species, genera, families, and higher levels of classification. Just as microbiologists discarded dubious morphological traits in favor of more accurate molecular yardsticks of evolutionary change, virologists can gain new insight into viral evolution through the rigorous analyses afforded by the molecular phylogenetics of viral genes. For bacteriophages, such dissections of genomic sequences reveal fundamental flaws in the Linnaean paradigm that necessitate a new view of viral evolution, classification, and taxonomy.