

The viral World as a (Bipartite) Network of RNA structures

Roman Ochsenreiter

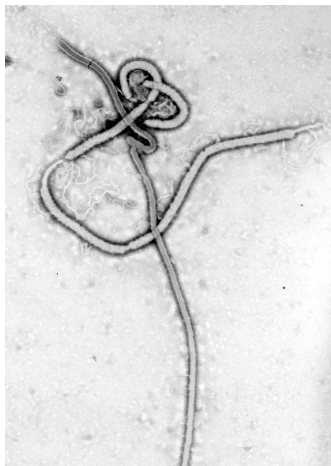
TBI Wien
University of Vienna

TBI Winterseminar 2018

Viruses I

Minimal Life Forms

- Obligate intracellular parasites
- No organelles or cell walls
- (Very) small size
- ss/ds DNA/RNA genomes



Electron micrograph of an Ebola Virion, size: 1000 nm
(image source: wikipedia)

Viruses II

Minimal Life Forms

Genome Size Comparison

Species	Type	Genome Size
Vernonia y. vein betasatellite	Virus (DNA)	1.5 kb
Zika Virus	Virus (RNA)	7 kb
<i>C. ruddii</i>	Prokaryote	150 kb (!)
Canarypox Virus	Virus (DNA)	350 kb
phage vB_KleM-RaK2	Virus (DNA)	380 kb
<i>B. aphidicola</i>	Prokaryote	650 kb
<i>M. pneumoniae</i>	Prokaryote	800 kb
Mimivirus	Virus (DNA)	1.1 Mb (!)
<i>E. coli</i>	Prokaryote	5 Mb
<i>S. cerevisiae</i>	Eukarya	11 Mb

Viruses and RNA Structure

- Viruses have tightly packed genomes
- Specific Functions are often provided by RNA Structures

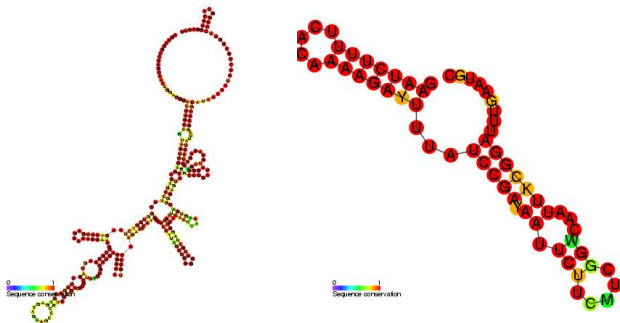


Figure: Pestivirus IRES, Poxvirus AX element¹

¹Rfam IDs: RF00209, RF00384

Viral Diversity

Viral Phylogeny

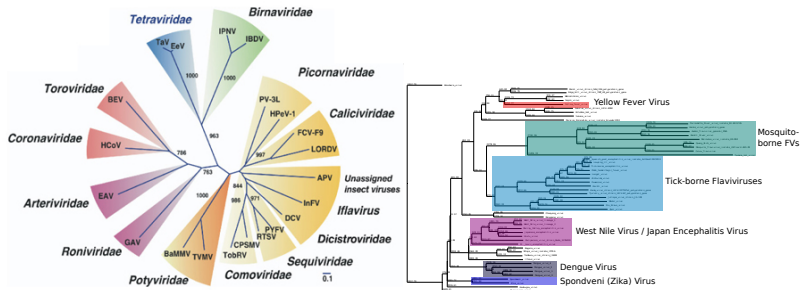
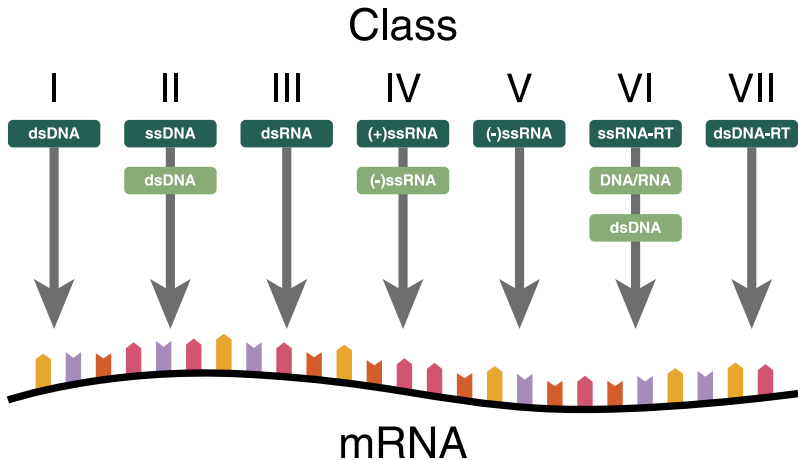


Figure: Picorna-like Viruses (left)¹, Flavivirus (right)

¹Fauquet, Schrock; 2006

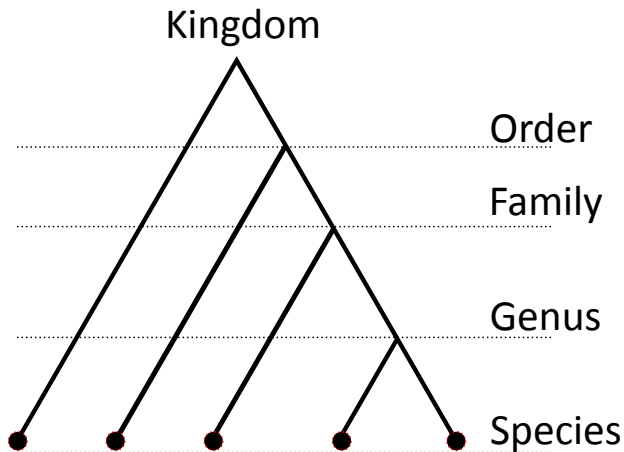
Viral Diversity

Baltimore Classification



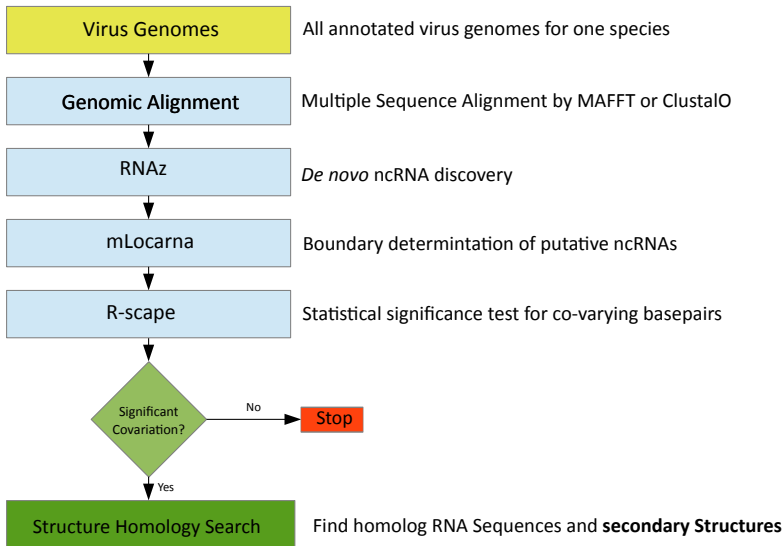
Viral Diversity

ICTV Classification



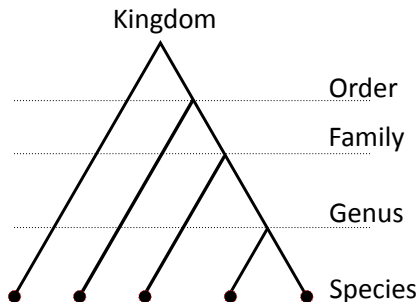
Based on Virus Biology and Sequence Criteria

Finding interesting Structures



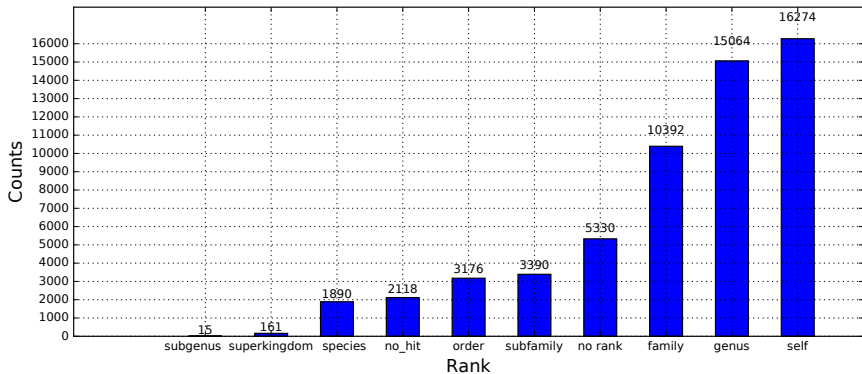
Searching for similar structures

- Use RNAz Windows/Loci as template for Covariance models
- Run with CMsearch against Database of viral Genomes
- Analyze taxonomic relation of CMsearch hits



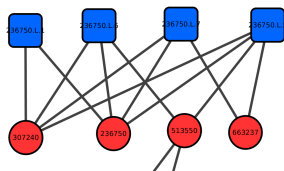
Results

Structure Homology



Bipartite Network

- Two types of nodes
- No edges between nodes of the same type
- Preferable to projections to unipartite network
- Here: Red nodes (Virus Genomes) and blue nodes (Covariance Models)



Building the Bipartite Network

Ingredients

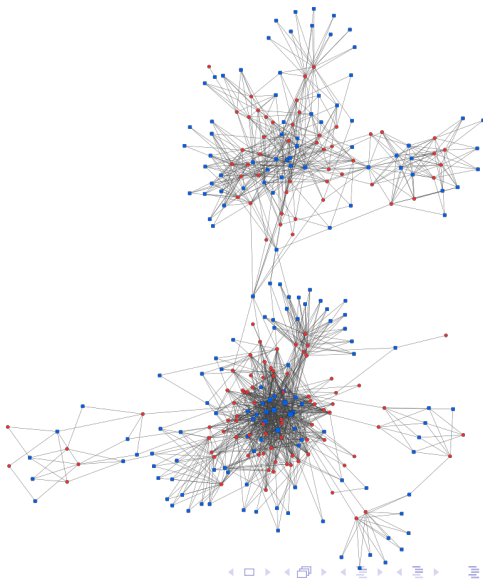
- Node Type 1: Virus Genomes
- Node Type 2: Covariance Models
- Edge if relevant CMsearch hits (aka. structural similarities)
- Optional: CMsearch E-values as edge weights

Key Questions

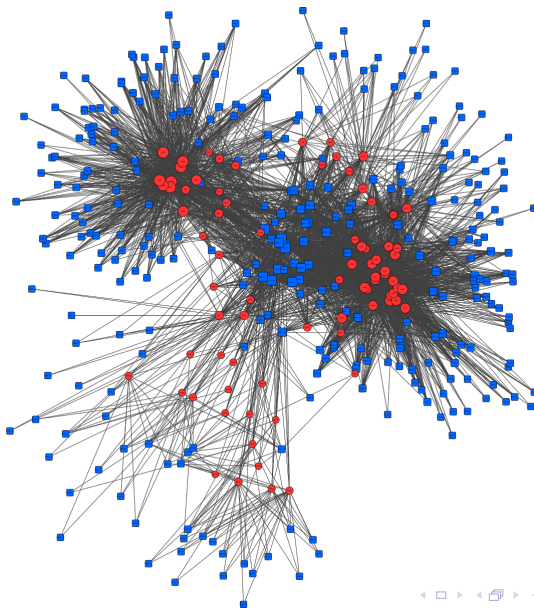
- Can we identify viral communities?
- Can we judge in any meaningful way how viruses are related to each other?
- Can we bypass the need for a phylogenetic tree?

Plant Virus Subnetwork

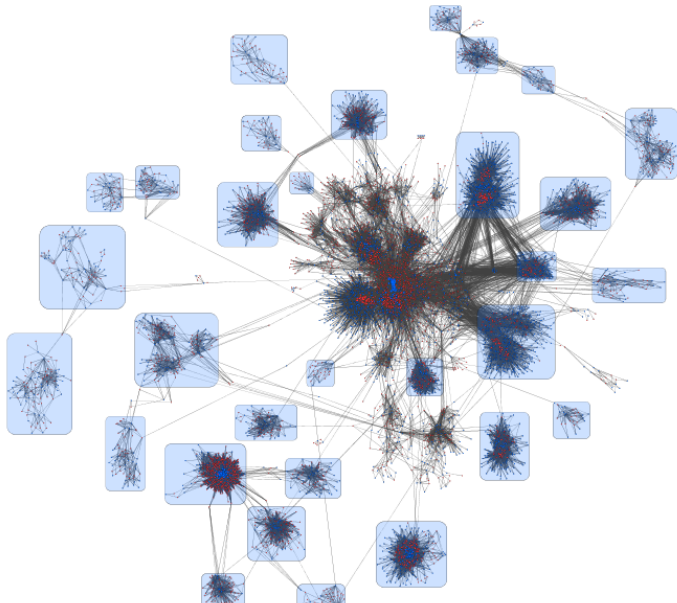
- Plant Viruses
- Upper Community:
Flower Viruses
- Lower Community:
Crop Viruses



Flavivirus Subnetwork



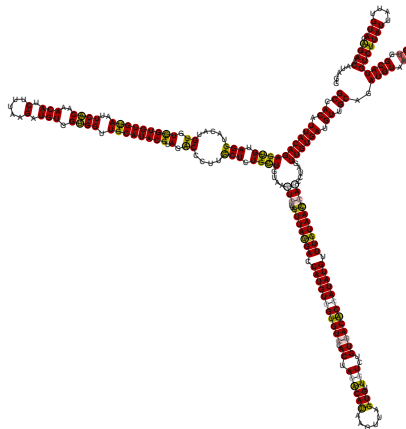
The Bipartite Virus World



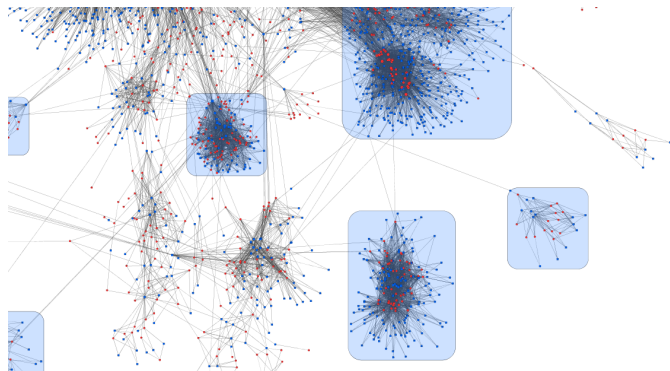
Analog structure?

Structure 694006.L.29

- Rousettus bat coronavirus HKU9 (Source organism)
- Rabbit Coronavirus (E-value: 10^{-18})
- Human herpesvirus 4 type 2 (E-value: 10^{-8})
- Structure seems strongly conserved

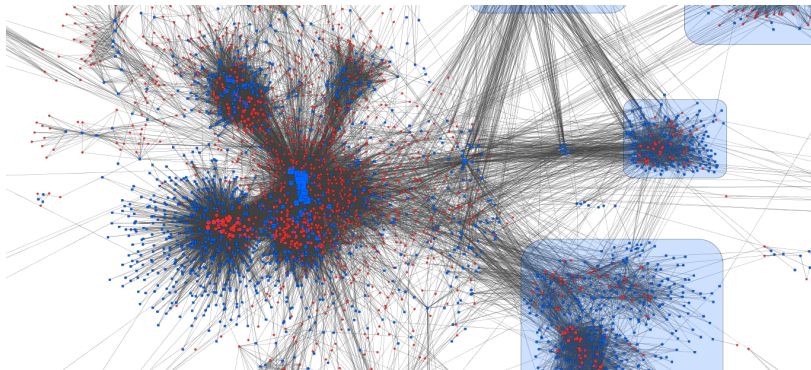


Showcase analog structure



Structure 694006.L.29 appears as edge between compact and almost not interconnected communities

Problems in Community detection



- Dense and highly interconnected Areas
- Meaningful Parameters for community detection?

Summary

- RNA structure can be used to reconstruct viral communities
- The viral world seems to be rather modular with few heavily interconnected communities
- Can the network substitute a tree? Maybe. . .

Where to go from here

- Automated Community Detection
- Automated Community Analysis
- Gene-based Network?
- Protein-based Network?
- Comparison RNA-structure vs. Protein based?

Thanks to...

- Hans-Jörg Hellinger (CUBE, Uni Wien)
- Thomas Rattei (CUBE, Uni Wien)
- Ivo Hofacker

...and you!



universität
wien

FWF

Der Wissenschaftsfonds.