Selected discussion topics:

Evolutionary relationships between virus families

-

core genes and methods for homology detection

Yves Bigot – Ascoviridae study group

Bas E. Dutilh – Metagenomics study group

Level 1: Macroevolution of the viral world

- Koonin, Dolja, Krupovic. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology. 479-480:2-25.

  Used criterions and tools : phylogeny of RNA or DNA polymerases and comparative virus biology

- Nasir, Caetano-Anollés. 2015. A phylogenomic data-driven exploration of viral origins and evolution. Sci Adv. 1(8):e1500527.

  Used criterions and tools : proteic fold content (based on HMM modeling) and proteic fold phylogeny (based on distances)

The main remaining question : what is the impact of horizontal genetic transfers (HGT) and of evolutionary convergence?

Level 2: Evolution from the species to the superfamily levels: Impact of HGT, genomic mosaicism, and convergence

- Horizontal gene transfer (HGT), genomic mosaicism, and potential evolutionary convergence may be a problem whatever the genomic configuration (size, RNA or DNA, segmented or not, ss or ds)

➔ Solution : because viruses can acquire genes from their environment, the recommendation is to use core genes that may be less prone to HGT

➔ Core genes can encode structural proteins or enzymes whose function is conserved in all the studied virus clade

➔ We might expect fewer core genes at deeper taxonomic levels

Problems with this definition :

1 - HGT of core genes can occur (more or less frequently)

2 - Intragenic recombination events leading to protein domain exchanges can weaken phylogenetic studies

3 - Impact of co-evolution with virion proteins evolving in different context (virion shape)

Tools to calculate phylogeny for taxonomy purposes

- **Question : is there currently a need to fix a sequence alignment curated with Gblocks for phylogeny?**

➔ **Answer** : No

Morrison DA. 2009. Why would phylogeneticists ignore computerized sequence alignment? Syst Biol. 58:150-158.

Dessimoz C, Gil M. 2010. Phylogenetic assessment of alignments reveals neglected tree signal in gaps. Genome Biol. 11:R37.

Proposed solution 1 : Alignment with MUSCLE or tcoffee - No Gblocks step - Protest - NJ or ML or pars

Advantage: results can be represented as trees or networks

Limits: the confidence of results is not probabilized

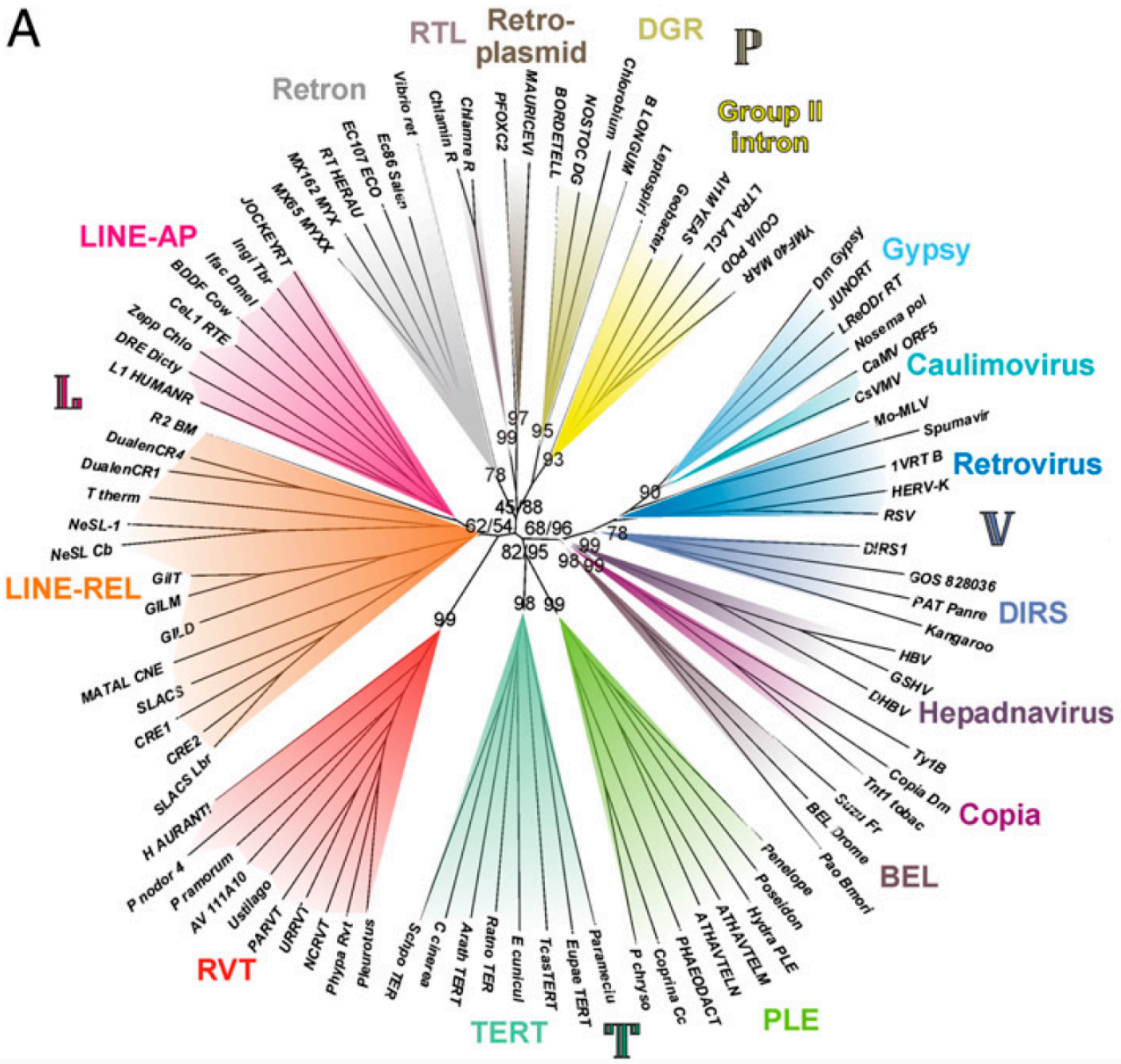
Proposed solution 2 : Baliphy (only the sequences are required ; software «learns» how to calculate the alignment and the tree in a same time)

Advantage: the confidence of results is probabilized

Limits: results can only be represented as trees

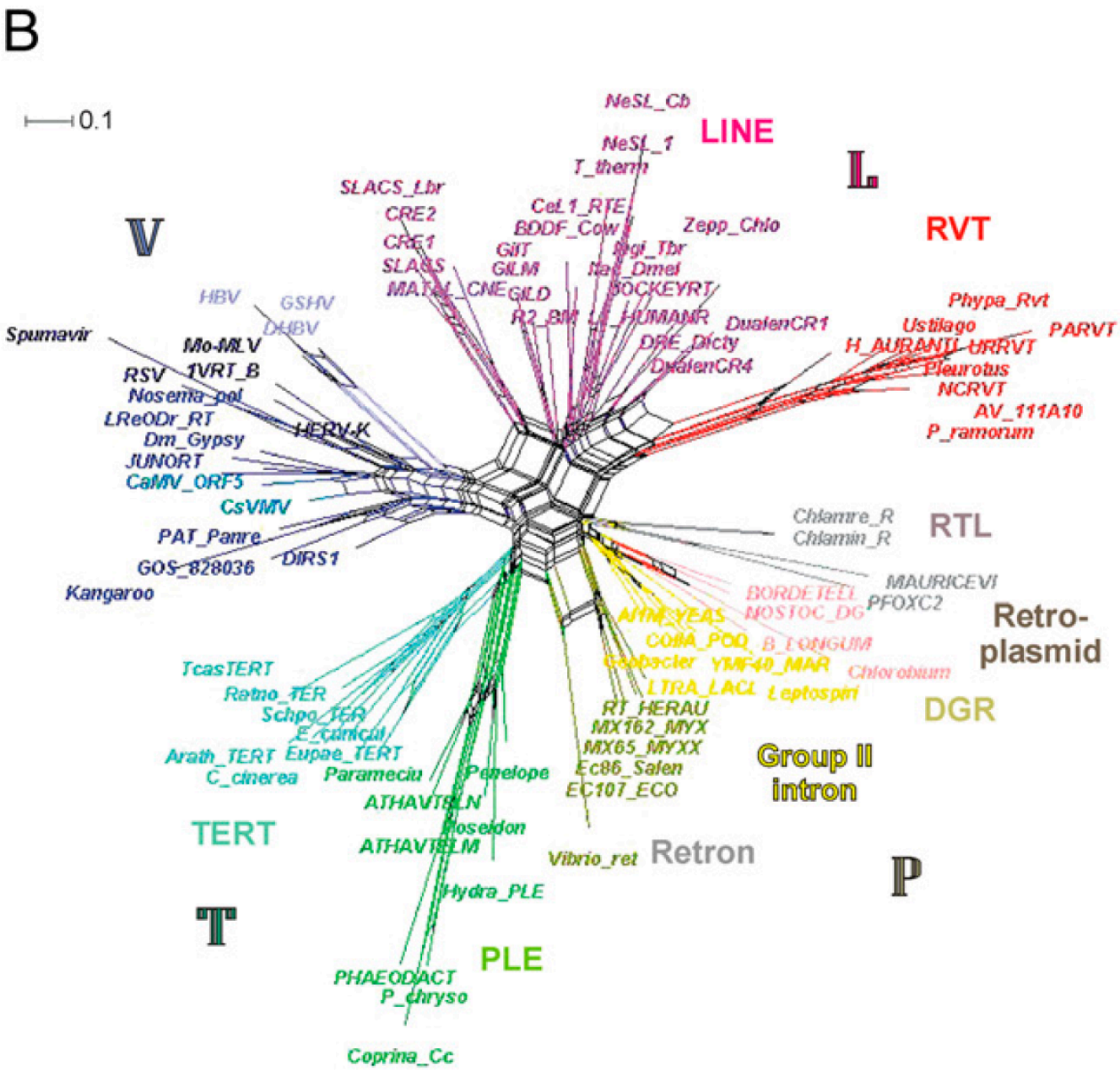
Representation of phylogenetic relationsips : network or tree ?

Gladyshev EA and Arkhopova I 2011 PNAS 108:20311-20316.



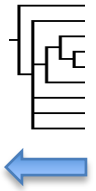
Representation of phylogenetic relationsips : network or tree ?

Gladyshev EA and Arkhopova I 2011 PNAS 108:20311-20316.



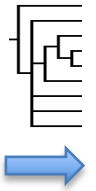
Challenges / Discussion points

1. **Identify homology between potentially rapidly evolving viral genomes**
   - Sensitive HMM profiles from alignments of good orthologs
   - HMM-HMM searches identify more distant homologs



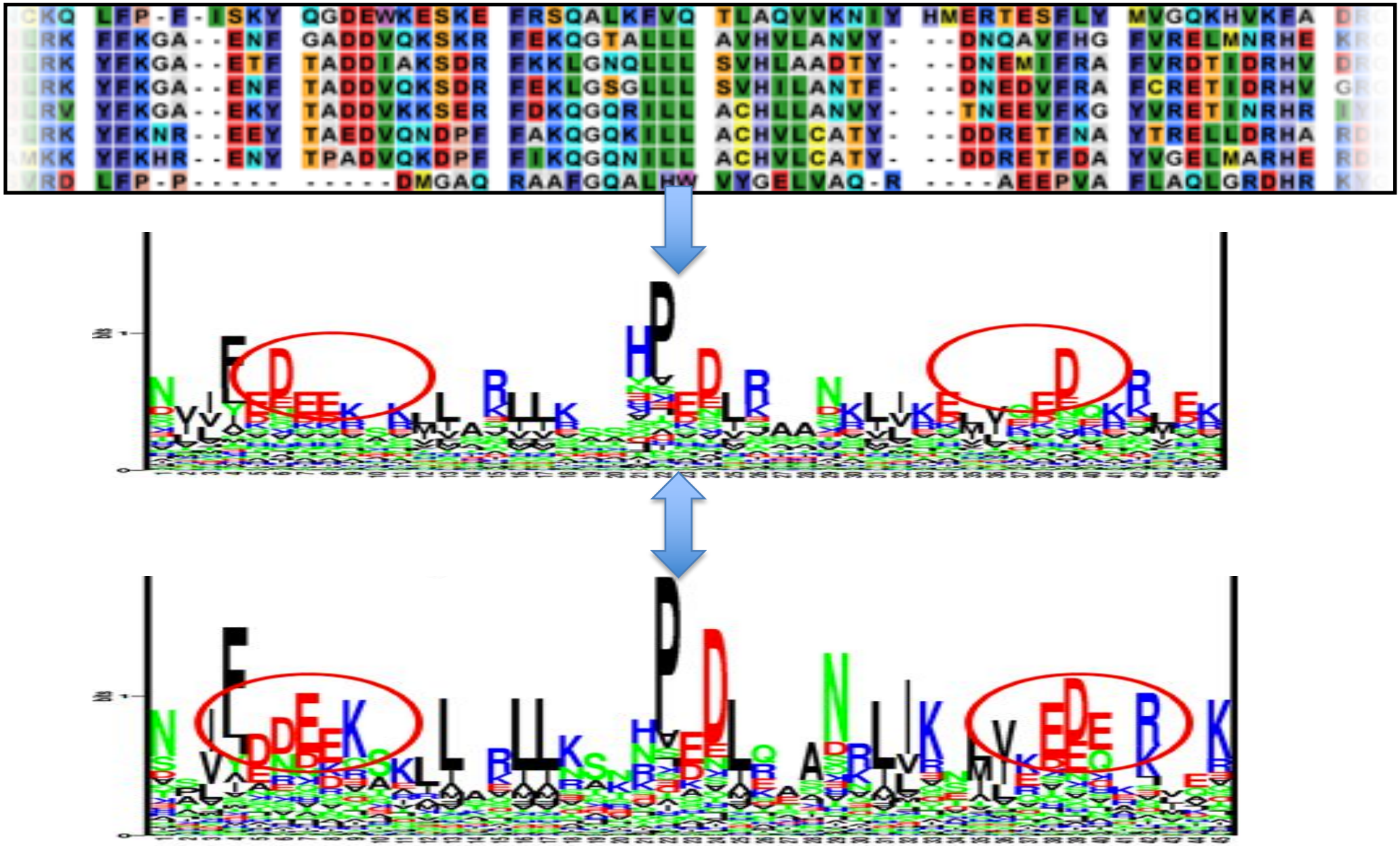
2. **Identify for which viral genes and clades HGT & mosaicism are important**
   - HGT adds "noise" to the "true" phylogenetic signal
   - Genes with discordant phylogenies relative to genome can be identified
   - When does noise become too strong? (viral genomes are small)

3. **Create reliable genome phylogenies from stable core genes**
   - Variation in gene content & synteny may form taxonomy baseline
   - How to combine individual gene trees into a genome tree?
   - Best phylogenies incorporate a relevant model of evolution
   - Models generally include rates of e.g. point mutations, but could also include indels and even HGT?



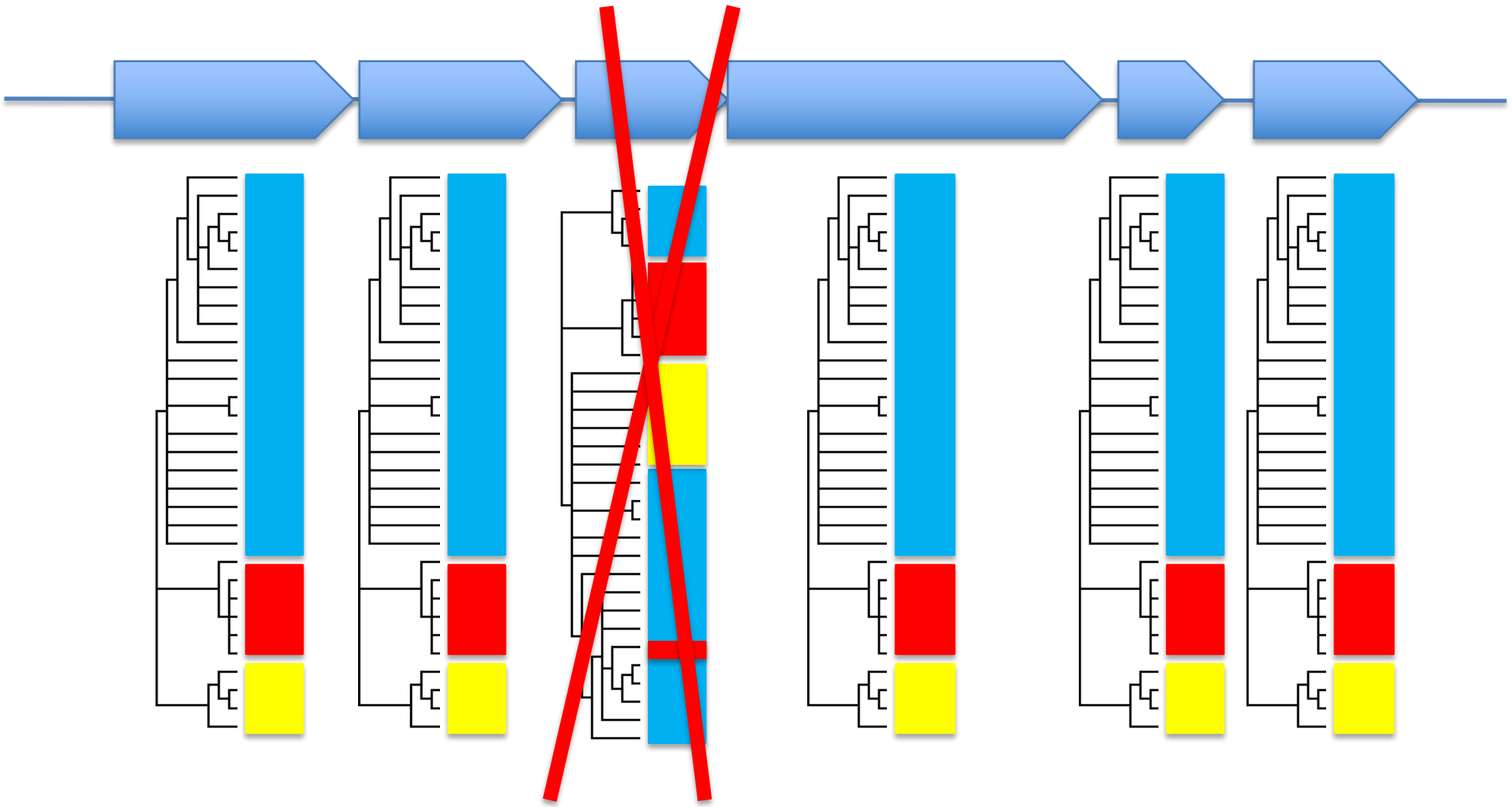
4. **Decide: what do we want from the phylogeny at the end of the day?**
   - Is a tree or network (= quantitative) going to be enough to place genomes into classes (= qualitative) ?
   - Are taxa true things or concepts of the mind?

Sequence profiles give more weight to conserved residues



Gene order conservation provides extra confidence

Identify phylogenetically noisy genes



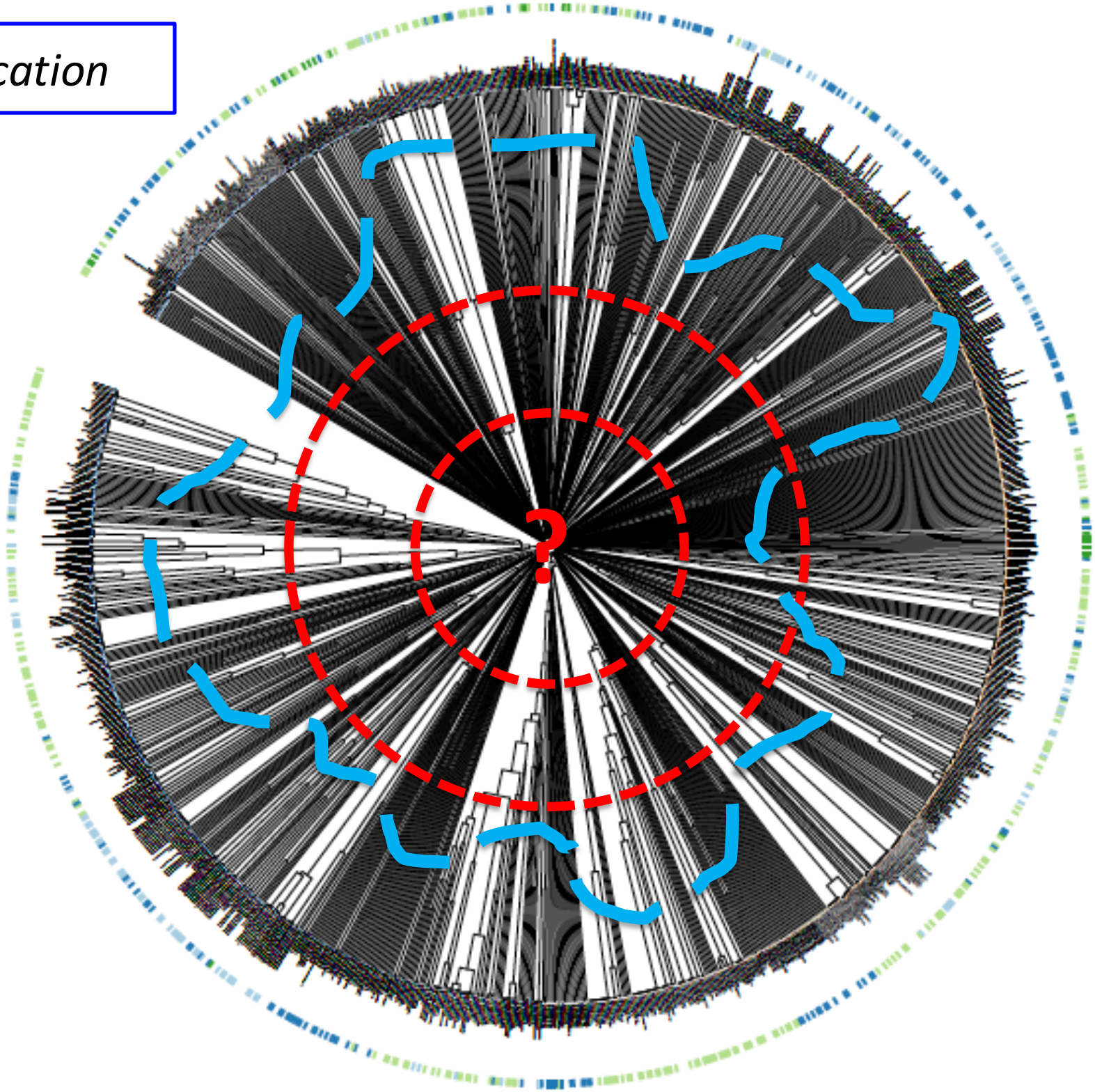
What if most genes are discordant / is the consistent signal strong enough?
How many genes can be removed before none are left?

Gene trees and genome trees

- **How do we think viral genomes evolve?**

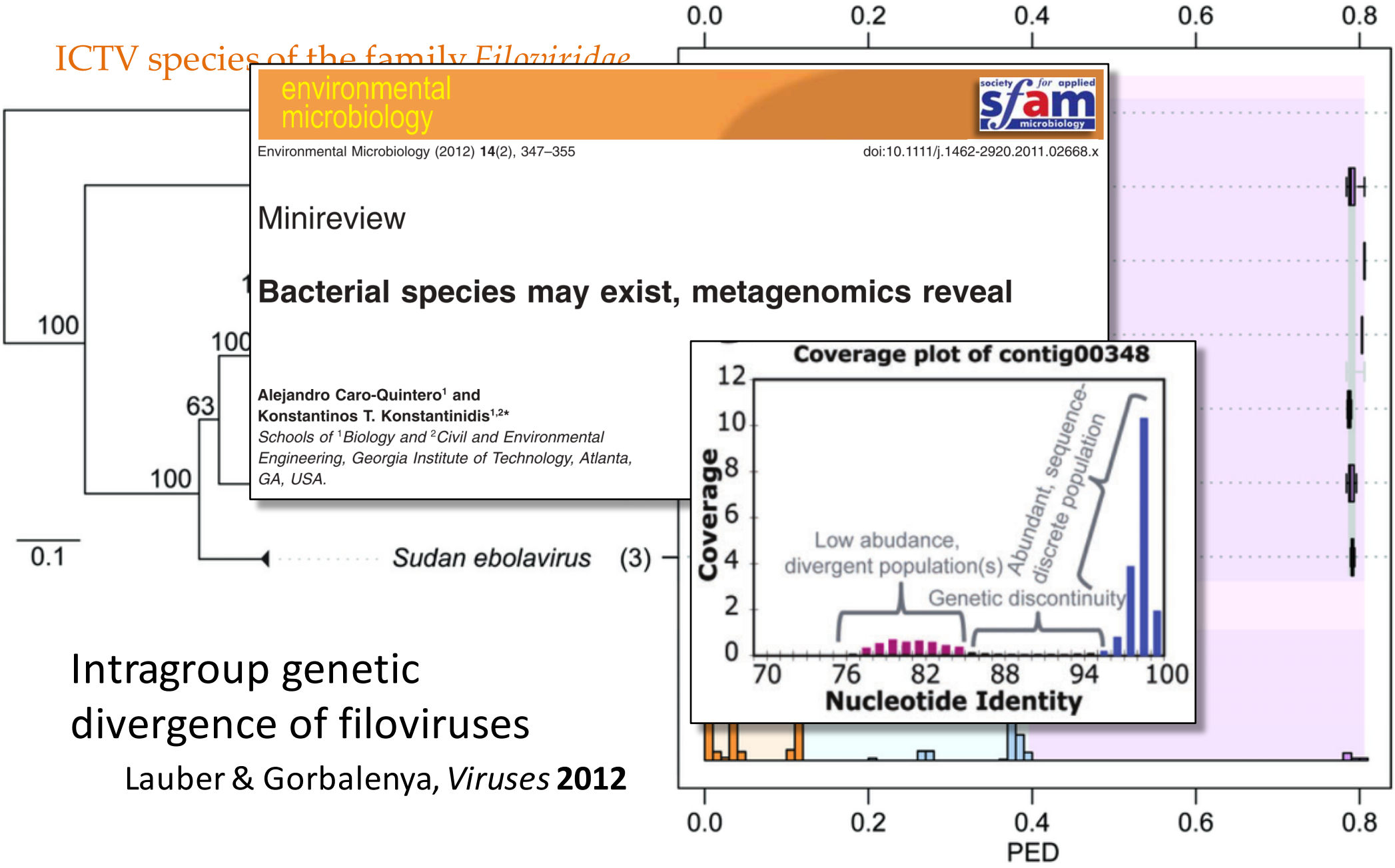
- **If we have a good evolutionary model we can use it to see how likely a given phylogeny is**
  - Maximum likelihood tree: the one most consistent with model
  - Model could include rates of point mutations, but conceivably also indels and HGT (to my knowledge such a model is currently not available)
  - Current consensus in phylogenomics is to weigh every mutation (in a core gene equally) but this decision process can be much more advanced incorporating study group knowledge

Clade demarcation



Can phylogenetic trees place genomes into classes?

ICTV species of the family *Filoviridae*



environmental microbiology

Environmental Microbiology (2012) **14**(2), 347–355 doi:10.1111/j.1462-2920.2011.02668.x

Minireview

Bacterial species may exist, metagenomics reveal

Alejandro Caro-Quintero[1] and Konstantinos T. Konstantinidis[1,2]*

Schools of [1]Biology and [2]Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

Intragroup genetic divergence of filoviruses

Lauber & Gorbalenya, *Viruses* **2012**

Challenges / Discussion points

1. **Identify homology between potentially rapidly evolving viral genomes**
   - Sensitive HMM profiles from alignments of good orthologs
   - HMM-HMM searches identify more distant homologs

2. **Identify for which viral genes and clades HGT & mosaicism are important**
   - HGT adds "noise" to the "true" phylogenetic signal
   - Genes with discordant phylogenies relative to genome can be identified
   - When does noise become too strong? (viral genomes are small)

3. **Create reliable genome phylogenies from stable core genes**
   - Variation in gene content & synteny may form taxonomy baseline
   - How to combine individual gene trees into a genome tree?
   - Best phylogenies incorporate a relevant model of evolution
   - Models generally include rates of e.g. point mutations, but could also include indels and even HGT?

4. **Decide: what do we want from the phylogeny at the end of the day?**
   - Is a tree or network (= quantitative) going to be enough to place genomes into classes (= qualitative) ?
   - Are taxa true things or concepts of the mind?