

Origins and Evolution of the Global RNA Virome

*Yuri Wolf, Darius Kazlauskas,
Jaime Iranzo, Adriana Lucia-Sanz,
Jens Kuhn, Mart Krupovic,
Valerian Dolja, Eugene Koonin*

Virus metagenomics (viromics) revealed enormous diversity of +RNA, -RNA and dsRNA viruses and facilitated reconstruction of the evolutionary relationships in the global RNA virome

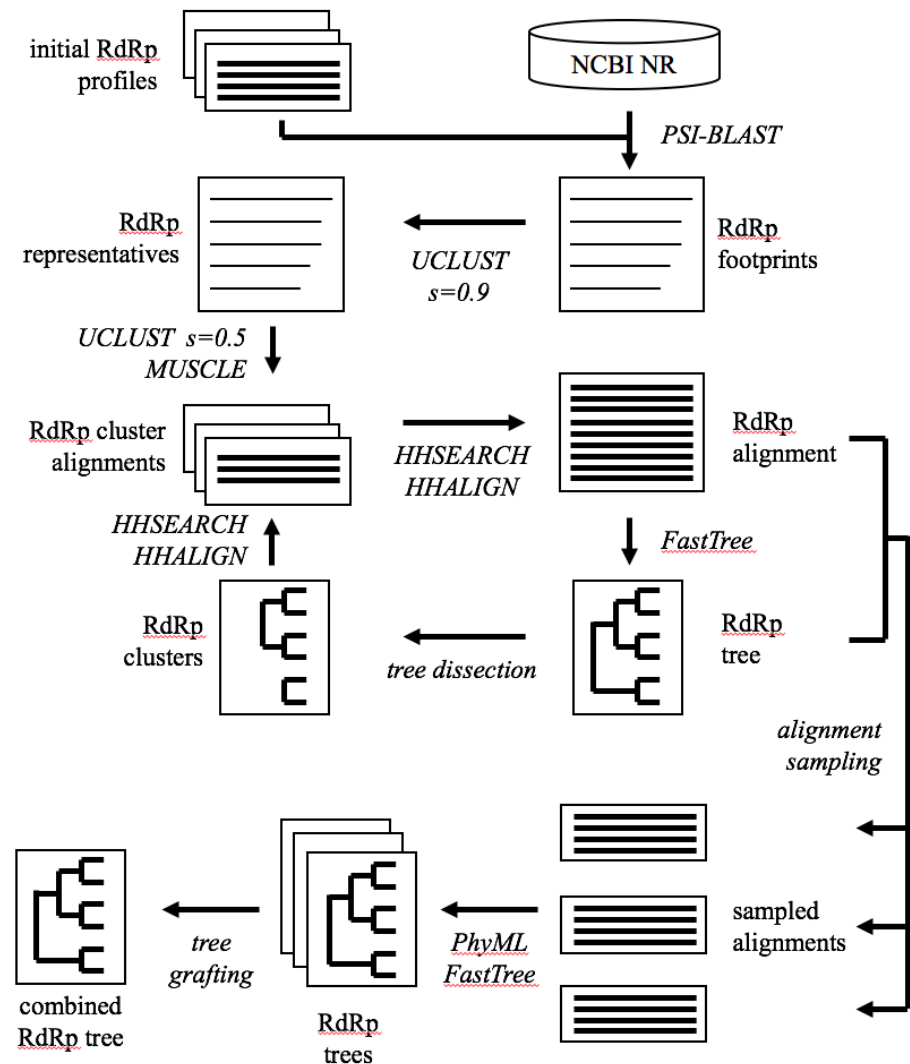


RdRp is the only protein universally conserved among all RNA viruses

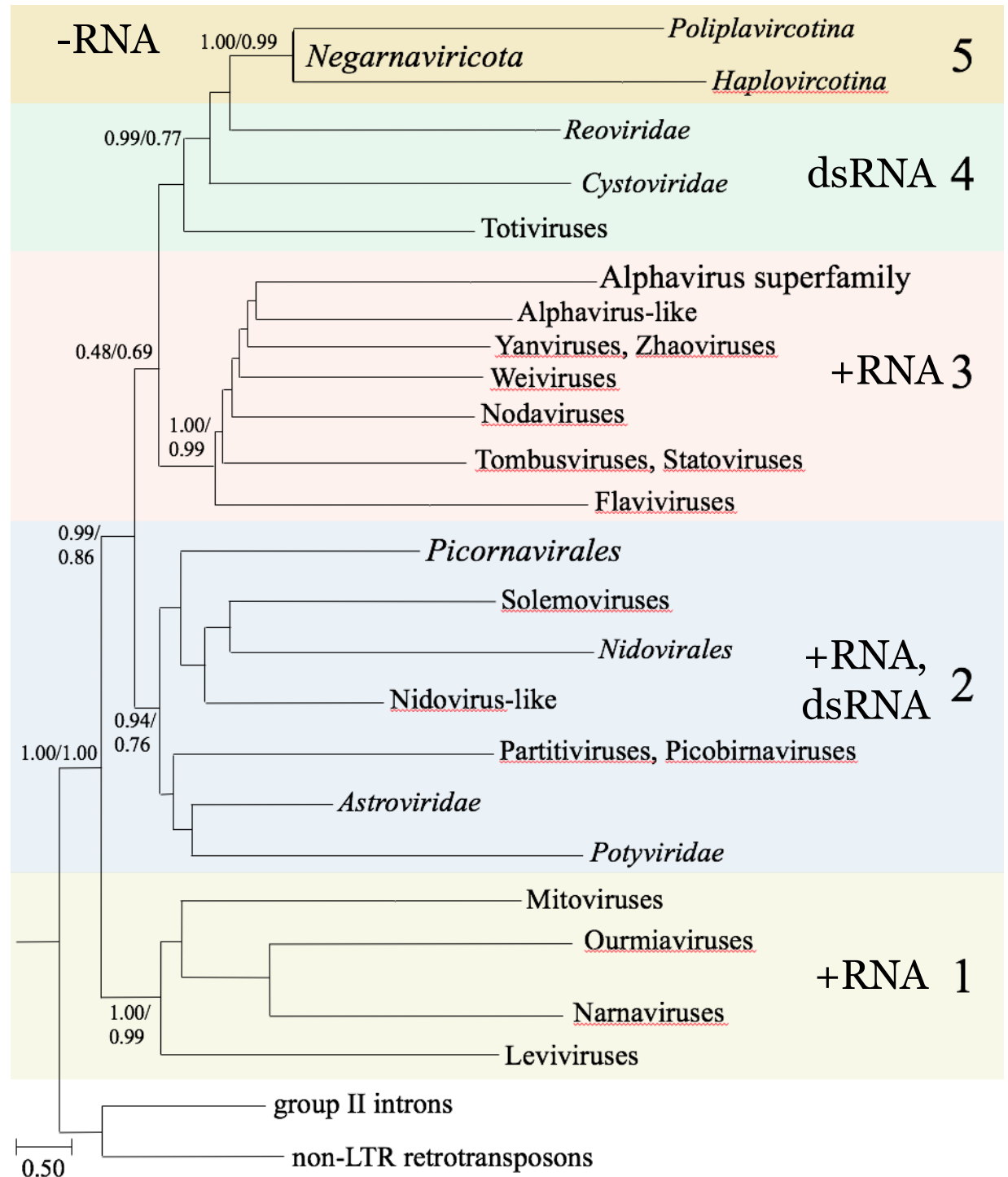
RdRp and RT are homologous palm domain proteins that shared common ancestor

To obtain a scaffold for analyzing evolution of the virus RNA world, the phylogenetic tree representing entire diversity of RdRps and RTs needs to be built

An iterative, clustering-alignment-phylogeny pipeline was developed to obtain a single multiple alignment of 4,617 RdRp and 1,028 RT sequences



Building RdRp tree encompassing the known diversity of RNA viruses



No evolutionary trace of the common ancestor of retroelements and RNA viruses (retroelement seems more likely)

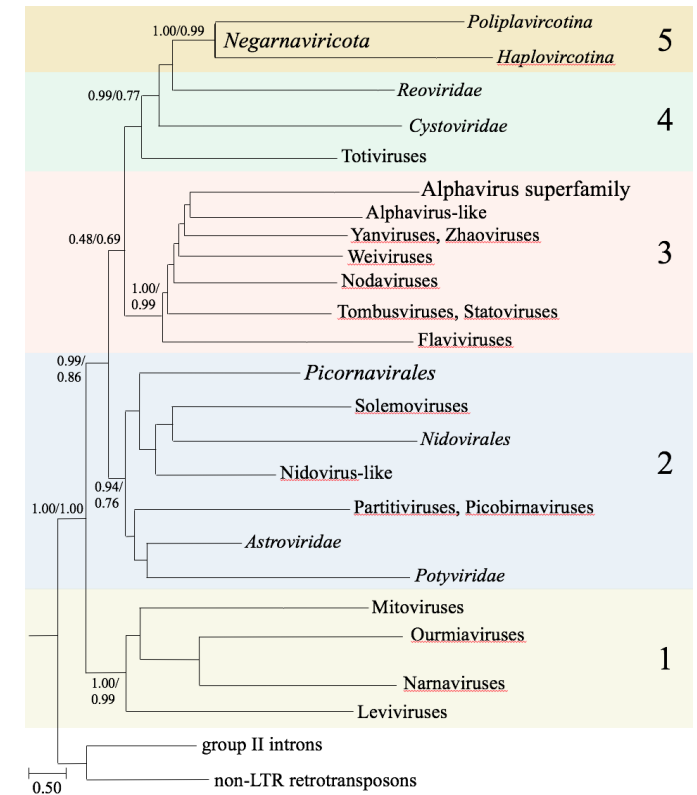
RNA viruses are monophyletic

+RNA viruses are paraphyletic relative to dsRNA and –RNA viruses

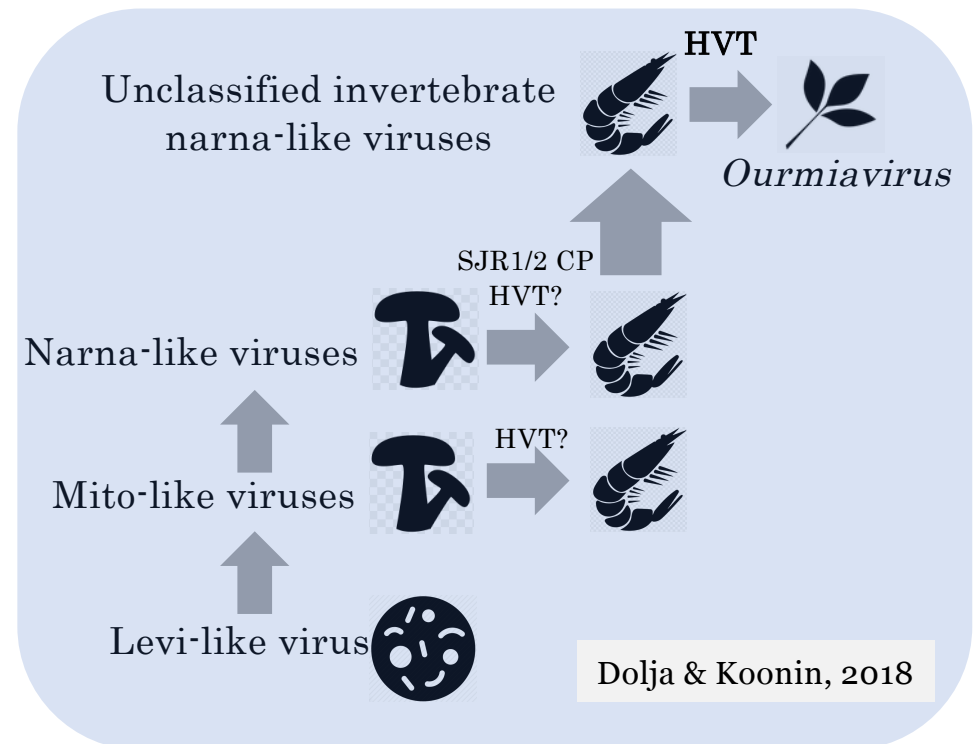
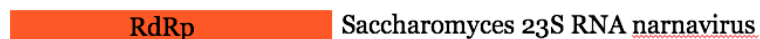
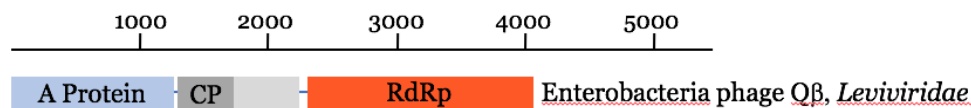
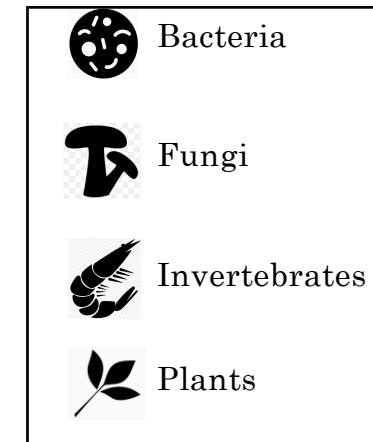
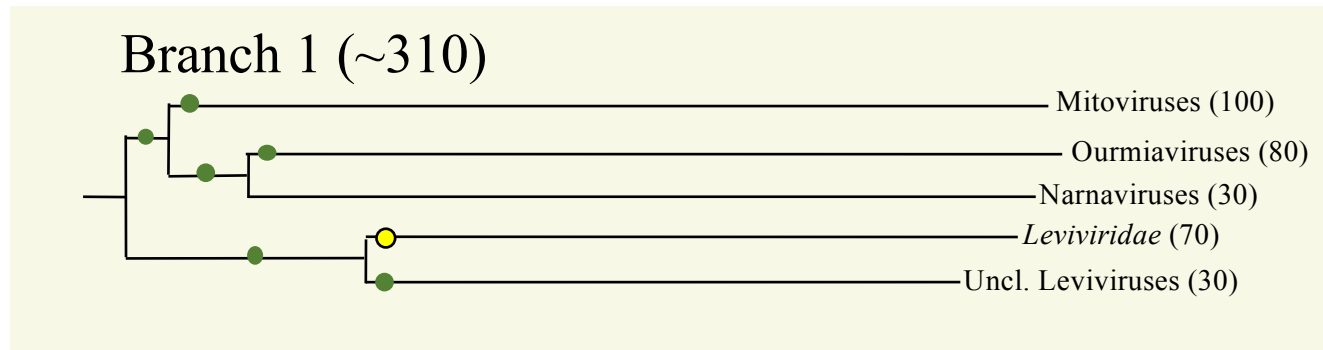
dsRNA viruses evolved from +RNA viruses on at least two occasions

-RNA viruses are monophyletic and might have evolved from dsRNA viruses

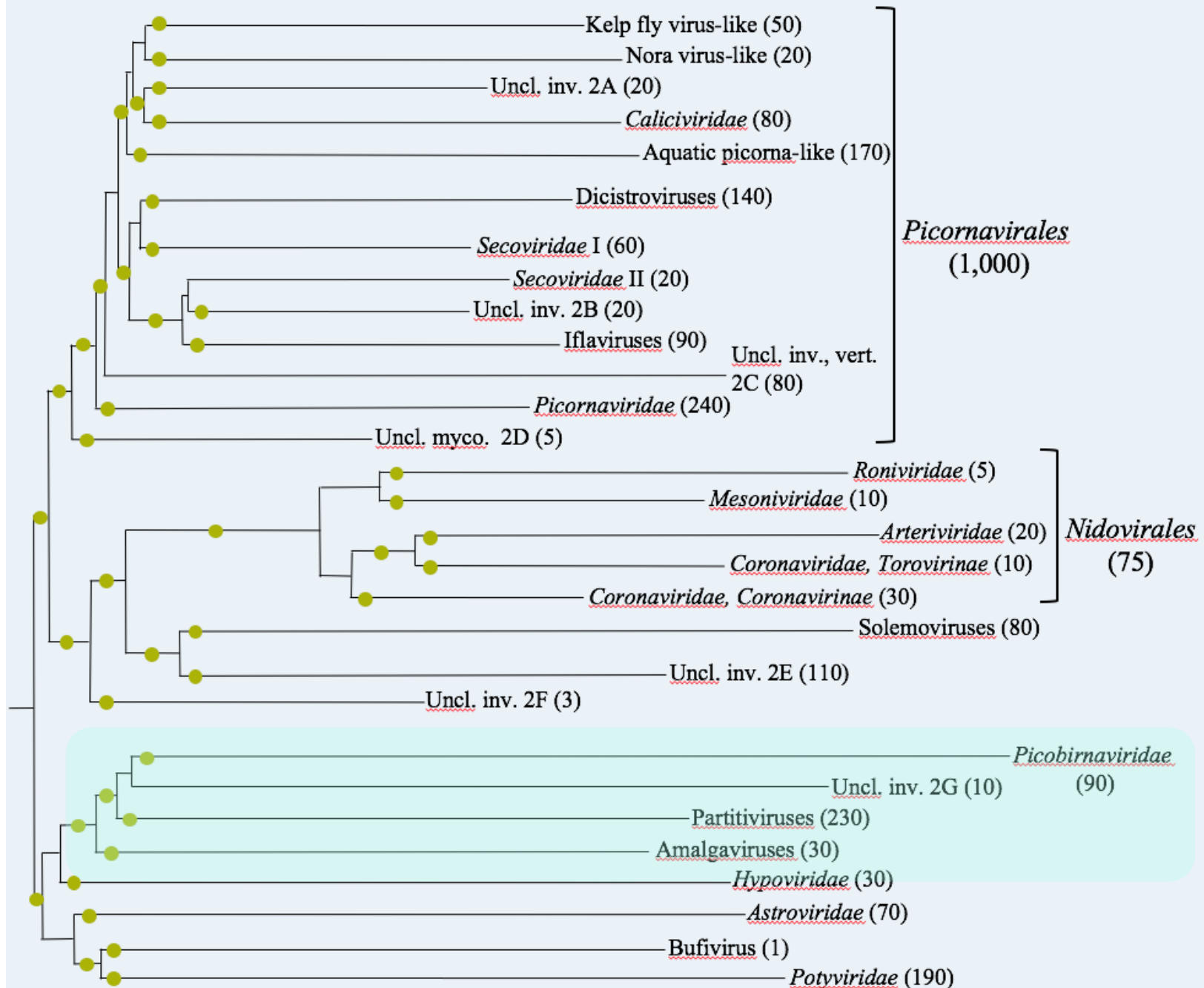
There are 5 major RdRp lineages of which Branch 1 is a sister group to the rest of RNA viruses and is a direct descendant of RNA phages



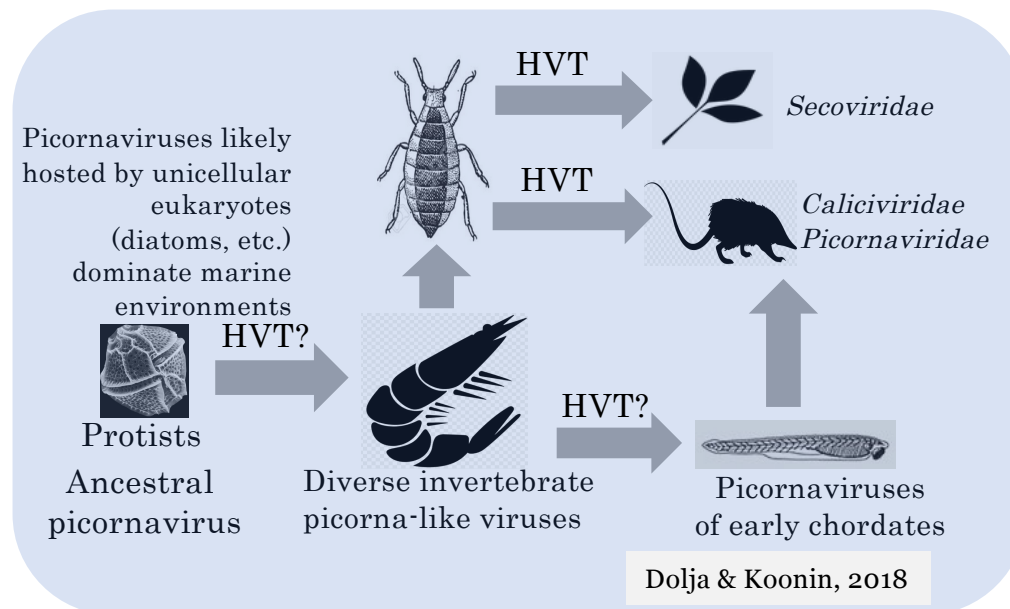
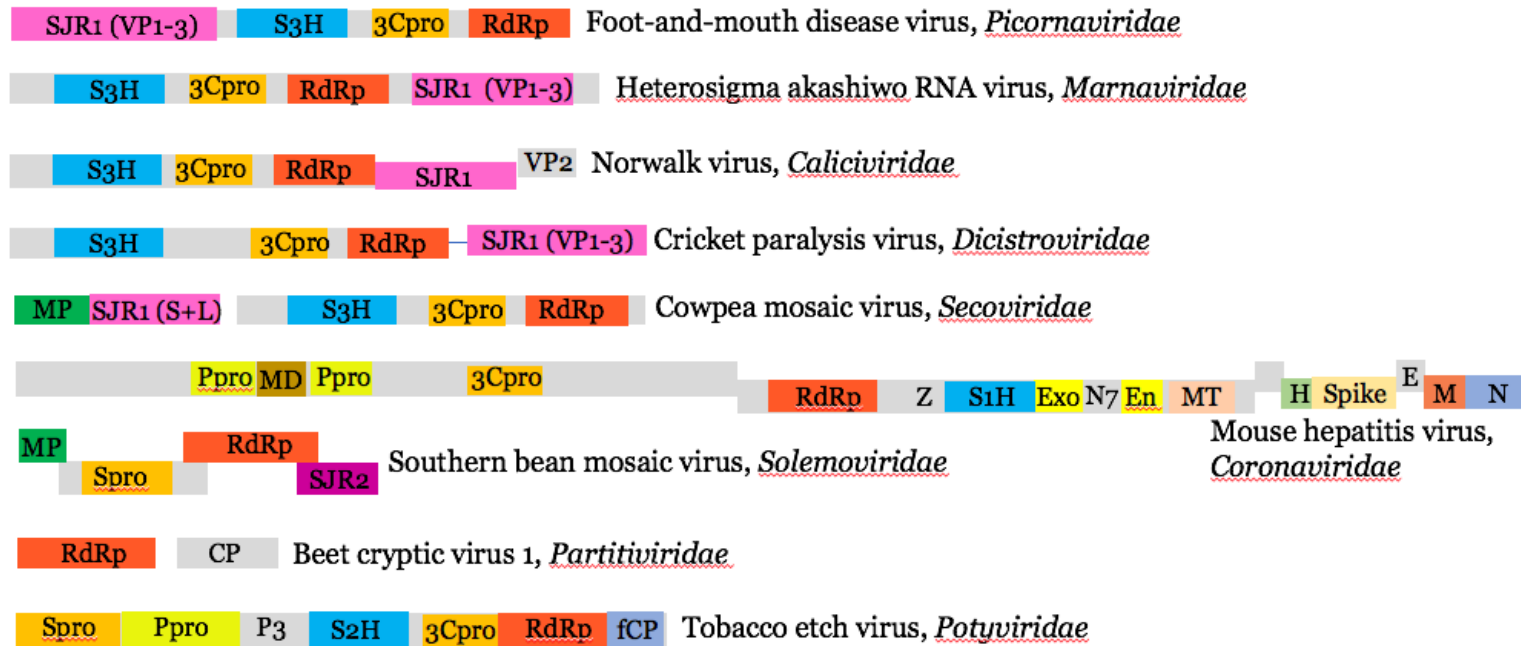
Evolution of the +RNA bacteriophage-derived viruses of eukaryotes

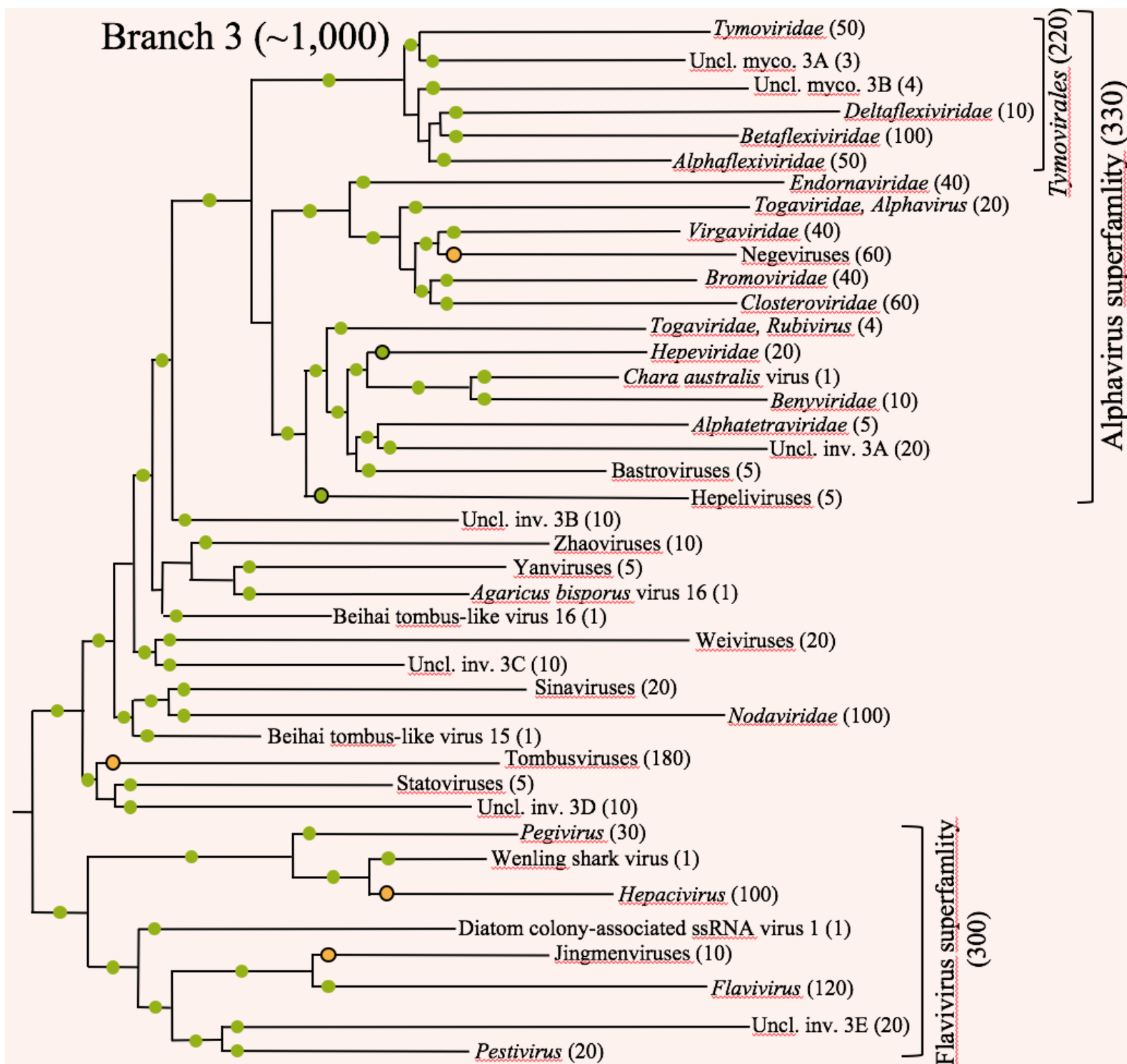


Branch 2 Picornavirus superfamily (~1,900)

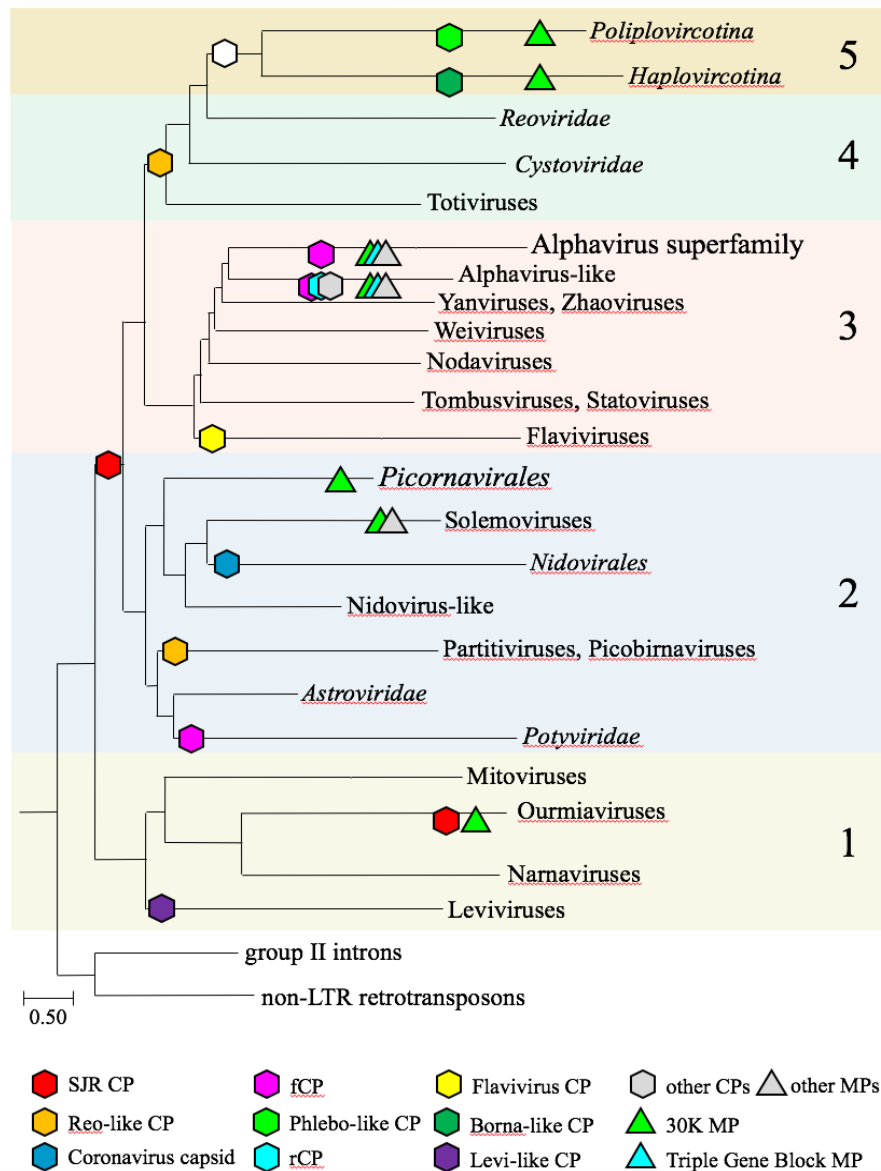


Origin and evolution of the *Picornavirales*

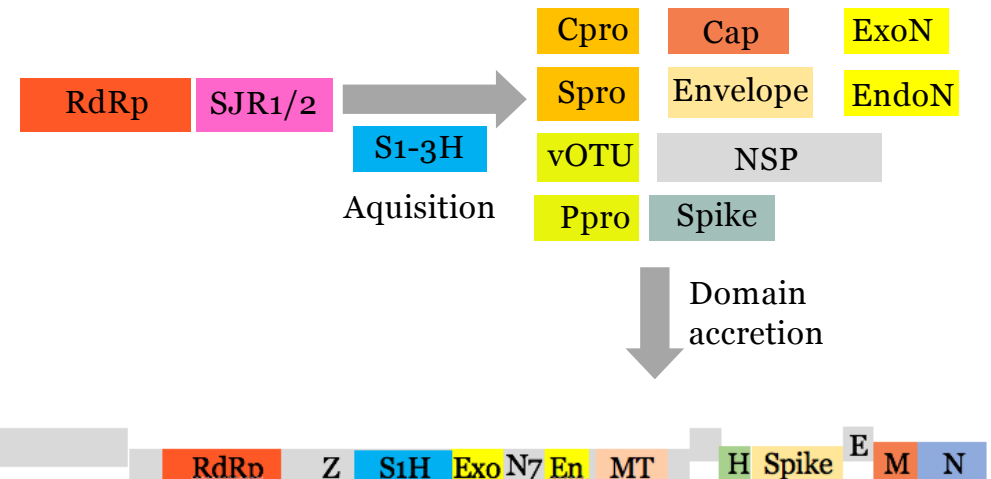




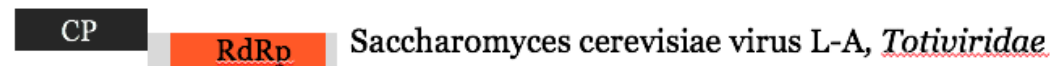
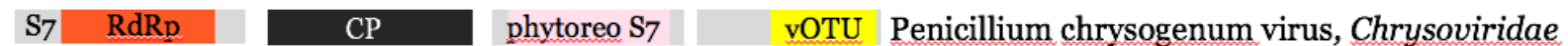
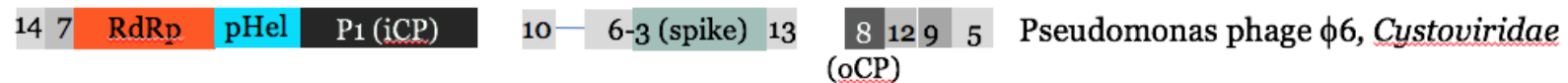
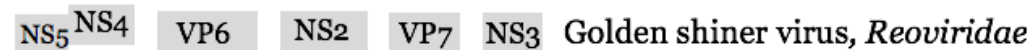
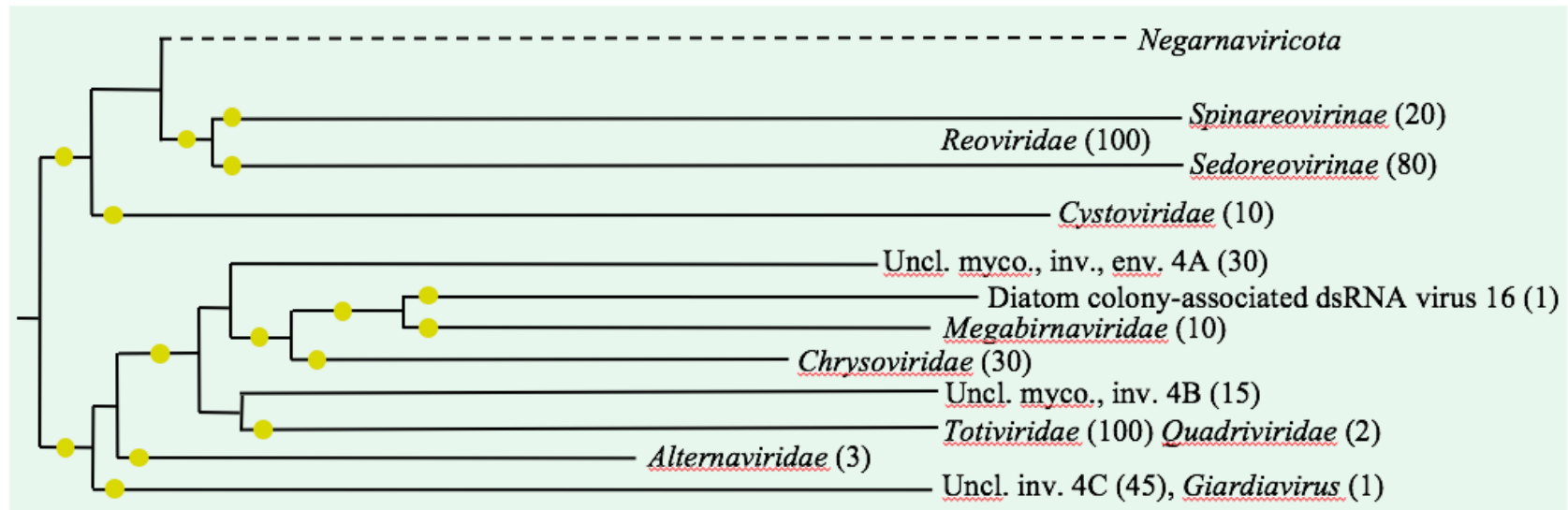
Principal evolution pathway of eukaryotic +RNA viruses



Domain gain/loss analysis (GLOOME) on a phylogenetic tree scaffold suggests that the ancestral +RNA viruses of eukaryotes possessed simple genomes containing only RdRp and SJR-CP

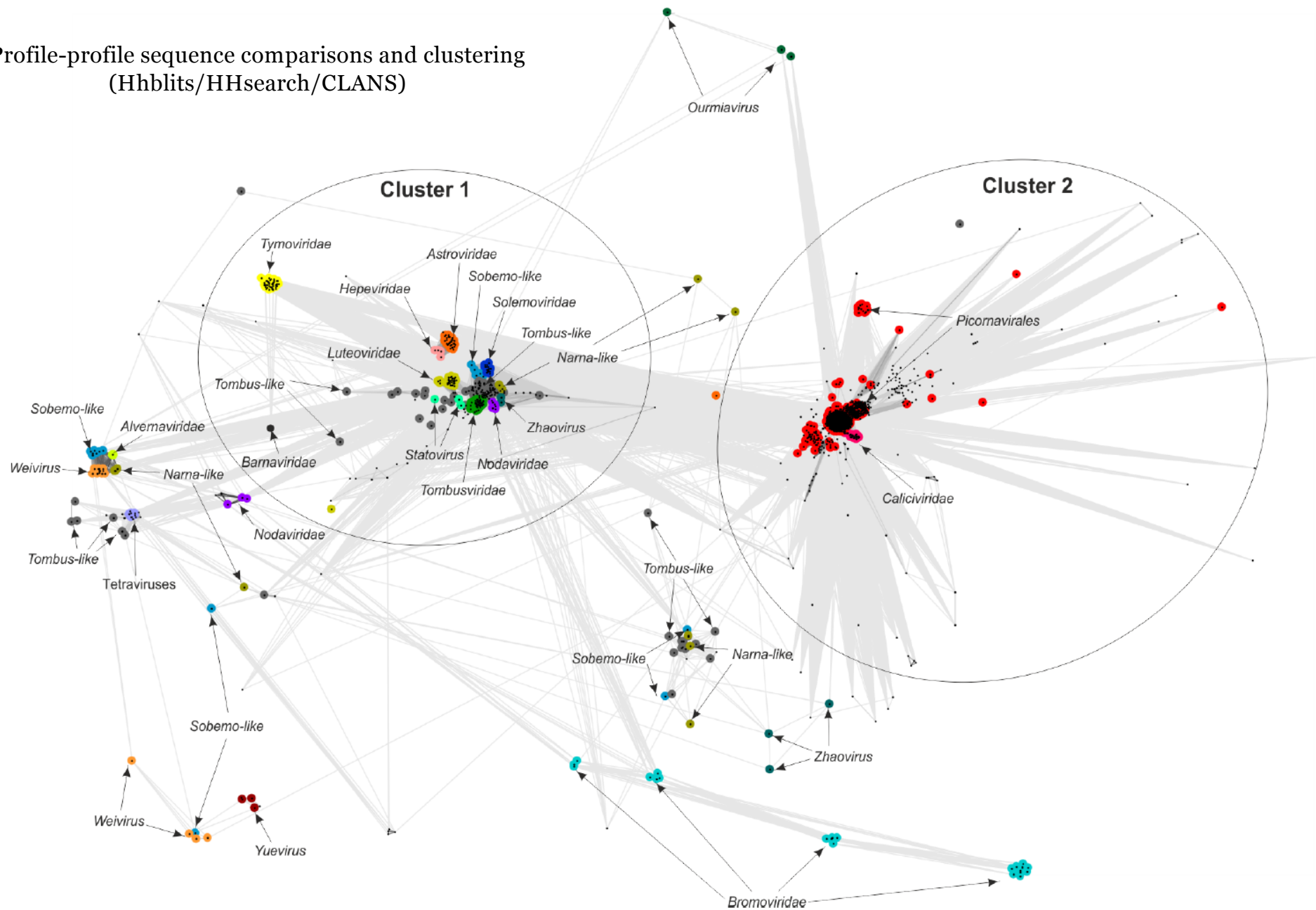


Branch 4 (~350)



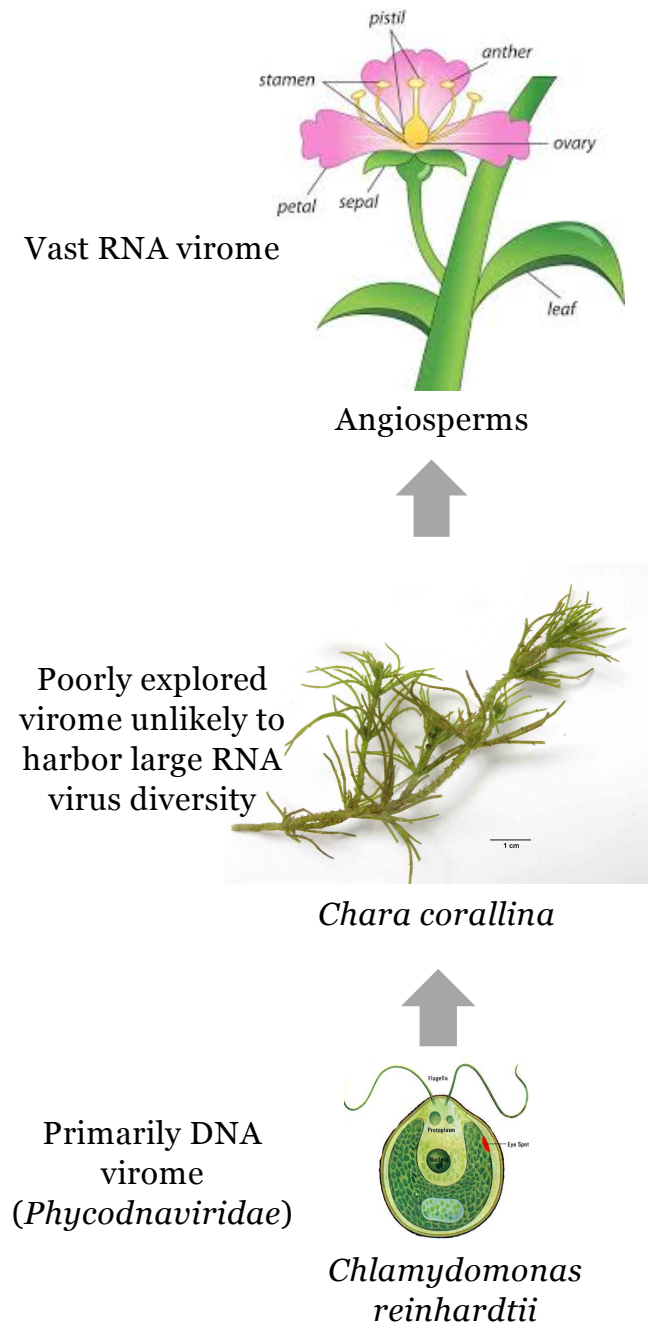
Evolution of SJR Capsid Proteins

Profile-profile sequence comparisons and clustering
(Hhblits/HHsearch/CLANS)






HGT and HVT

- The gene module shuffling (HGT) that has already been recognized as a major evolutionary trend, appears to be virtually limitless in light of metaviromics
- In all 5 tree branches, mammalian and plant viruses are embedded within radiation of invertebrate viruses implying HVT from invertebrate virome to plants and vertebrates
- Vectoring invertebrates (arthropods, nematodes, etc.) likely played a central role in formation of RNA viromes of the terrestrial plants and vertebrates



RNA virus evolution coming into focus

-  The three apparent routes through which the RNA viruses of eukaryotes have evolved:
 - i) continuous line of descent from +RNA bacteriophages (Narna/Ourmiaviruses);
 - ii) assembly of the ancestral virus genome from genes derived from bacteria and their selfish genetic elements (Picornaviruses);
 - iii) more recent origin of –RNA viruses and Alphavirus superfamily (perhaps, at the root of metazoa, given their apparent absence from protists)
-  These three major routes seem to have converged into superhighway (via HGT and HVT) during diversification of invertebrates yielding the vast invertebrate pan-virome
-  There are five major phyla of RNA viruses of which +RNA viruses with simplest genome replication/expression strategy appear to be the most ancient viruses that begot dsRNA viruses twice, with one lineage being ancestral to –RNA viruses

HVT as a driver of virus evolution

