Review of: "A five-fold expansion of the global RNA virome reveals multiple new clades of RNA bacteriophages"

Ergin Sahin[1]

1 Ankara University

Potential competing interests: The author(s) declared that no potential competing interests exist.

Viruses ubiquitously infect members from all domains of life. They have a vital role in the evolution and diversity of every life form and global biogeochemical cycling. Since viruses rely on and evolve with their hosts for propagation, they are crucial in gaining insight into the origins of life on earth.

Today, it is beyond doubt that one of the most significant contributions, the next generation sequencing (NGS) made to life sciences, is the vast expansion of the sequence data related to viruses, the most diverse biological entities on earth. Accumulation of these huge amounts of viral sequence data has mostly been achieved by high throughput sequence analyses (e.g. metagenomics and metatranscriptomics) of environmental samples. These analyses pawed the way for making deep evolutionary inferences regarding the phylogenies of viruses which otherwise may not be possible. Owing to the technical reasons related to the sequencing of RNA molecules, studying the diversity and evolution of RNA viruses with a comprehensive approach was belated in contrast to those implemented for DNA viruses.

Phylogenetic inferences of RNA viruses rely on sequence analyses of the RNA-dependent RNA polymerases (RdRp) as these enzymes are ubiquitous components and hallmarks of RNA viruses. RNA viruses are currently classified by The International Committee on Taxonomy of Viruses (ICTV) within the realm *Riboviria* which comprises two kingdoms namely *Orthornavirae* which includes the RdRp encoding “true” RNA viruses and *Pararnavirae* which includes reverse transcribing viruses. Within the last couple of years, deep evolutionary phylogenetic inferences made based on the high throughput analyses of RdRp sequences retrieved mostly from metatranscriptome data of environmental samples revealed a total of five monophyletic phyla within the kingdom *Orthornavirae*[1]. This classification has been approved by ICTV and suggests to group positive-sense single-stranded RNA (+ssRNA) viruses within three phyla namely *Lenarviricota*, *Kitrinoviricota,* and *Pisuviricota*, negative-sense single-stranded RNA (-ssRNA) viruses in a single phylum called *Negarnaviricota*, and double-stranded RNA (dsRNA) viruses in one another phylum called *Duplornaviricota*.

One of the most significant contributions to our understanding of the diversity and evolution of RNA viruses has been made by the RNA Virus Discovery Consortium and their most recent findings, currently available as a preprint (Ribovira.org), deserve special attention. By mining over 5000 metatranscriptome data obtained from diverse terrestrial and aquatic biomes, as well as plant and animal microbiomes, the research team was able to identify roughly 330k novel RdRps which increased the current RNA virus diversity by fivefold. These newly identified RdRp sequences increased the

total number of virus species by more than ninefold (based on average nucleotide identity <90% analyses) and the total number of virus families, orders, and classes by roughly fivefold respectively. Detailed phylogenetic analyses of these newly identified RdRp sequences supported the monophyletic state of the previously established five phyla within the kingdom *Orthornavirae*. The lowest common ancestor analyses revealed the deepest branching order as *Pisu-* and *Kitrinoviricota* forming the crown group, and *Lenar-* and *Negarnaviricota* occupying the basal position. Furthermore, the authors proposed to establish two novel phyla (namely p.0002 and RvANI90_0011770) one of which (p.0002) was positioned below the base of *Kitrinoviricota*. Conserved Shine Dalgarno (SD) sequence analyses revealed that the newly proposed phylum p.0002 includes viruses that infect prokaryotic hosts as opposed to its closest relative *Kitrinoviricota* whose members are known to infect eukaryotic hosts.

The efforts put forward by the RNA Virus Discovery Consortium on elucidating a more accurate phylogeny of RNA viruses are praiseworthy. However, regarding the methodological approach of their study, there appear to be some limitations that should be mentioned. Firstly, as the authors acknowledge, analyses of a vast number of RdRp sequences that exhibit a great degree of diversity are problematic in terms of the difficulties encountered during the alignment process for homology assignment. Individual sequences (singletons) that do not become part of a large cluster decrease the phylogenetic resolution and this situation can only be overcome by populating these poorly sampled groups which require expansion of sampling. Therefore, the deep phylogenetic inferences of RNA viruses proposed by The RNA Virus Discovery Consortium will most likely require further revisions in the light of future studies. Secondly, deep phylogenetic inferences are prone to long-branch attraction artefacts and therefore it would be useful to employ a secondary approach such as protein 3D analyses to support the alignment-based phylogenetic methods. Also, a phylogenetic analysis that operates based on an artificial intelligence-assisted algorithm would be beneficial for increasing the accuracy of the evolutionary inferences. As such, these approaches were successfully implemented in another comprehensive study focusing on deep evolutionary relationships of RNA viruses found in marine ecosystems[2]. Through a survey of roughly 28 terabases of the global Ocean RNA sequences revealed by The Tara Ocean Foundation, Zayed et.al proposed five novel RNA virus phyla in addition to the other five previously established phyla by combining a machine-learning-based approach and traditional phylogenetic analyses. One of the most striking findings of this recent study is the polyphyletic state of the dsRNA virus phylum *Duplornaviricota* which is suggested to be monophyletic by The RNA Virus Discovery Consortium. This finding was reported to be supported further by 3D structure analyses of the RdRp sequences related to dsRNA viruses. Another important finding of this research team is the identification of a novel (-) ssRNA virus phylum (namely "*Arctiviricota*"), the members of which are reported to be prevalent in the arctic region and phylogenetically distinct from the established *Negarnaviricota*. Collectively, the findings of this research team infer that (+) ssRNA, (-) ssRNA and dsRNA viruses are polyphyletic and emerged independently on multiple occasions. Lastly, based on their findings, the authors proposed yet another RNA virus phylum named "*Taraviricota*" which was speculated to be the missing link between the kingdoms *Orthornavirae* and *Pararnavirae*.

One of the most remarkable findings of the current study conducted by The RNA Virus Discovery Consortium is the identification of the new lineage of partitiviruses (clade genPartiti.0019: identified in the metatranscriptomes of a hot

spring biome) whose most probable hosts are thought to be prokaryotic organisms (Particularly *Roseiflexus* sp., hot spring species). Members of the family *Partitiviridae* are known to infect eukaryotic hosts including plant, fungus, and invertebrate species. Two lines of bioinformatic evidence including the presence of conserved Shine Dalgarno sequences in the 5' UTRs of the analyzed partitiviral genome segments and the existence of CRISPR spacer sequences (Type III-RT CRISPR Array) that match with the analyzed partitiviral genome sequences suggest that members of these partitivirus lineages infect prokaryotic hosts.

Some other valuable findings of the current study conducted by The RNA Virus Discovery Consortium include the identification of various new viral protein domains. In this regard, the demonstration of the presence of a serine-threonine kinase (STK) domain in some members of a partitivirus clade (genPartiti.0029) is quite intriguing. Considering the diverse sets of protein targets of serine-threonine kinases, it is difficult to make the right inference on the function(s) of partitivirus encoded STKs without experimentally addressing it; although their possible roles in promoting cell division or regulating cell stress responses during the infectious cycle would be hypothesized.

One correction that is to be taken into account for this preprint is that virus-related NUDIX hydrolase domains were previously reported in other studies. The presence of virus-related NUDIX domains was firstly reported in poxviruses[3] and later in certain mycovirus taxa that are closely related to megabirnaviruses[4-7].

In conclusion, although the current state of the virus classification is much better than it was two decades ago, evolutionary virologists are still away from seeing an accurate picture of the evolutionary history of RNA viruses. As long as new viral sequences are surveyed using advanced sequencing and improved phylogenetic analysis approaches, virologists will get a more comprehensive insight into the evolution of RNA viruses.

REFERENCES:

1- Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucía-Sanz, A., Kuhn, J. H., Krupovic, M., ... & Koonin, E. V. (2018). Origins and evolution of the global RNA virome. MBio, 9(6), e02329-18.

2- Zayed, A. A., Wainaina, J. M., Dominguez-Huerta, G., Pelletier, E., Guo, J., Mohssen, M., ... & Wincker, P. (2022). Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. Science, 376(6589), 156-162.

3- Parrish, S., Resch, W., & Moss, B. (2007). Vaccinia virus D10 protein has mRNA decapping activity, providing a mechanism for control of host and viral gene expression. Proceedings of the National Academy of Sciences, 104(7), 2139-2144.

4- Lim, J. M., Jamal, A., Phoon, X., Korhonen, K., & Coutts, R. H. (2011). Incidence of Phlebiopsis gigantea large virus-1

in a collection of Phlebiopsis gigantea isolates. Archives of virology, 156(11), 2091-2094.

5- Magae, Y. (2012). Molecular characterization of a novel mycovirus in the cultivated mushroom, Lentinula edodes. Virology Journal, 9(1), 1-6.

6- Li, Y., Xu, P., Zhang, L., Xia, Z., Qin, X., Yang, G., & Mo, X. (2015). Molecular characterization of a novel mycovirus from Rhizoctonia fumigata AG-Ba isolate C-314 Baishi. Archives of virology, 160(9), 2371-2374.

7- Petrzik, K., Sarkisova, T., Starý, J., Koloniuk, I., Hrabáková, L., & Kubešová, O. (2016). Molecular characterization of a new monopartite dsRNA mycovirus from mycorrhizal Thelephora terrestris (Ehrh.) and its detection in soil oribatid mites (Acari: Oribatida). Virology, 489, 12-19.