REEXAMINING THE CLASSIFICATION OF VIRUSES AS NONLIVING BASED ON THEIR EVOLUTIONARY PATTERNS

HONORS THESIS

Presented to the Honors College of
Texas State University
in Partial Fulfillment
of the Requirements

for Graduation in the Honors College

by

Whitney Elaine Ortiz

San Marcos, Texas
December 2018

REEXAMINING THE CLASSIFICATION OF VIRUSES AS NONLIVING BASED ON THEIR EVOLUTIONARY PATTERNS

by

Whitney Elaine Ortiz

Thesis Supervisor:

_____

Kelly Woytek, Ph.D.
Department of Biology

Approved:

_____

Heather C. Galloway, Ph.D.
Dean, Honors College

ACKNOWLEDGMENTS

Writing this thesis has been a substantial undertaking that has required long hours and dedication on my part, however, I recognize that I could not have done it without a number of remarkable people whom I would like to acknowledge and thank.

First and foremost, I would like to thank Dr. Kelly Woytek who has taken on the role of my thesis supervisor. She made time in her busy schedule to work with me and help me sort through all of the papers and rabbit holes of information I had dug myself into. Without her I would have found myself lost. In addition, she found a fascination in my research that motivated me to push forward, if for no other reason than to answer the questions we both had. Thank you, Dr. Woytek for challenging my ideas and pushing me to find the answers.

I would also like to thank Dr. Ron Haas and Dr. Heather Galloway for all the work they put into helping all the thesis students with writing sessions and weekly meetups.

Finally, I would like to thank my friends and family. My parents and grandparents, who have been supportive throughout all the hardships that have come, not only with my thesis, but throughout my life. My little brother, William, who is the reason I push myself. I want to make you proud. And last but not least, thank you Alex. You have believed in me even when I didn't believe in myself.

Reexamining the Classification of Viruses as Nonliving Based on their Evolutionary Patterns

Abstract

Viruses are currently considered nonliving, acellular particles, and, as such, are not categorized under the three domains of life. This classification may be detrimental to viral research by limiting the resources invested and influencing the mentality of those involved. Classifying viruses as nonliving may make it difficult to see the viral influence evident in cellular evolution and regard viruses with the same evolutionary potential as cellular organisms. Existing data is inconclusive about the position of viruses in evolutionary history and the degree of relatedness between viral families. This is due to gaps in viral genome catalogues and the difficulties inherent in studying ancient evolution. Given the incomplete data set and the fact that viruses fall into a gray area when defining life, it may be necessary to examine the characteristics of living organisms and nonliving mechanisms and compare them to viruses. When examining current literature, the classification of viruses as nonliving seems incorrect when compared to their origins, evolutionary patterns, and characteristics. Viruses may then represent an evolved form of cellular life. There is little genome sequencing data, especially regarding ancient viral families, which makes constructing a phylogeny difficult. This data could be crucial to understanding viral origins and their connection to the cellular world. It is important that future research strives to collect a more comprehensive genome catalog for viruses and develop techniques to account for horizontal gene transfer and the rapid mutation found in viruses. Collecting accurate data may make it possible to examine viruses with a better perspective and open our minds when developing viral research.

Introduction

Viruses are considered nonliving organisms by most biologists. There are several reasons, with the most common being their lack of transcriptional machinery. Viruses depend on their hosts for all their machinery and to produce most of their proteins during infection. Viruses carry proteins in their virions, however these proteins are used mainly for gaining entry into a new cell and do not usually have transcriptional properties. Based on these factors and the gray area viruses reside in when looking at definitions of life, it is far simpler to label viruses as nonliving. There is some difficulty in constructing a definition of life in that the criteria can be too vague or can be interpreted multiple ways. The most common definitions of life are based off the seven characteristics of life including maintaining homeostasis, organization, the ability to reproduce, growth and change, a metabolism, environmental responsiveness, and the ability to evolve. Viruses are borderline on several of the criteria, and they especially fall short on their ability to reproduce on their own and their lack of metabolism. Their dependence on their hosts can make it easy to place viruses in a similar category to plasmids, being evolved by their hosts and not as an independent organism. Other ideas label viruses as living organisms only within their hosts. This is since viruses have all the characteristics of life only when they are infecting the host, the external virion is considered nonliving. This represents an interesting compromise, but one that is not logically sound. A living thing cannot become nonliving, only to regain its status as living once it infects a host.

Many arguments attempt to consider viruses nonliving while granting them capabilities only observed in cellular organisms. Viruses are a special case and it may be necessary to examine them outside of the current definitions of life. Examining the

origins of viruses, along with their phylogeny, and comparing them to other nonliving mechanisms could provide insight and may indicate a need to rethink the criteria of life. If viruses evolve like cellular organisms and evolved from or with cellular organisms, it may be necessary to consider them living.

Origin of Viruses

There are difficulties inherent in studying the origins of organisms and the mechanisms within them. Currently, there is no definitive answer for how life originated and, given the nature of cells and molecules, it is likely impossible scientists will be able to investigate these early remnants of life. To develop hypotheses about the origin of life scientists create phylogenies using modern cells and analyze the capabilities of these cells to determine what may have been possible. These same techniques may be used to investigate the origin of viruses. There is no definitive answer for where or how viruses originated, and the two questions share many of the same difficulties. Investigating the origins of viruses has some additional problems, because viral genomes can be heavily integrated with their host genomes and their classification as nonliving means their origins can be investigated in conjunction with the origins of cells or under a separate mechanism. There are currently four predominating hypotheses for the origin of viruses. These hypotheses are the virus first hypothesis, the endogenous or escape hypothesis, the regressive hypothesis, and the protobiont hypothesis (7). These hypotheses differ in their proposed method of viral evolution and the timing of their appearance on an evolutionary timeline. The differences proposed by these hypotheses have ramifications for how scientists consider viruses. Hypotheses like the virus first and regression hypotheses are based on the idea that viruses are a part of the cellular family tree. If this is the case,

viruses would either need to be reclassified as living or it must be assumed living organisms can evolve into or from nonliving particles. The escape and protobiont hypotheses do not incorporate viruses into the cellular family tree, and these concessions would not need to be made. This makes determining the most likely hypothesis or combination of hypotheses the first step in reexamining the classification of viruses.

The virus first hypothesis states that viruses evolved before cells and function as precursors to cellular organisms (7). Some proponents of this hypothesis assert that the last universal common ancestor, LUCA, may not have been cellular in nature, and instead existed as competitive proteins and nucleic acids contained by inorganic compartments (7, 12). This hypothesis explains some questions concerning how competing molecules evolve into compartmentalized organisms by hypothesizing they become isolated in independent compartments. It does not explain the homologous proteins found amongst membrane proteins from the three domains of life, given this hypothesis assumes membranes were evolved after the division between bacteria and archaea (7). Assuming this hypothesis was found to be fully credible, these organisms would not be considered viruses by modern definitions. By today's definition, viruses can replicate only within a living cell. The nonliving, acellular organisms described by this hypothesis may be the ancestors to cellular organisms and viruses, but they cannot be called viruses themselves.

The escape hypothesis states that viruses are escaped cellular nucleic acid, usually in the form of plasmids or semi-autonomous chromosomes (7). This hypothesis gives viruses a polyphyletic origin, usually late on an evolutionary timeline. The polyphyletic origin described helps explain why no gene is shared amongst all viruses, and genes shared across diverse groups can be explained if the escape happened before the LUCA.

Plasmids as viral precursors also explains mechanisms found in both plasmids and in viruses like rolling cycle replication (13, 18). To explain homologous genes, this hypothesis relies heavily on the impact of horizontal gene transfer. HGT can explain genes found across viral groups when viruses take genes from their cellular host. This means viruses do not need to be related to share genes, they only need to infect hosts that carry the same genes. This explanation is not always represented by the patterns seen in viruses, and nucleocytoplasmic large DNA viruses, NCLDV, is a good example of this discrepancy (21). Members of this clade infect distantly related organisms, and yet, are described as members of different escape events (21). These viruses share homologous genes that cannot be explained by HGT and means they must be related before the divergence of their hosts. This hypothesis takes some steps to explain problems within viral evolution, however there are many gaps and this hypothesis is difficult to confirm given its dependence on HGT.

The protobiont hypothesis asserts that viruses co-evolved with the first cellular organisms or their precellular ancestors (2). Another key aspect of this hypothesis is that viruses do not have a cellular stage in their phylogeny; they neither evolved from or evolved into cellular organisms (2). Viruses and the first cellular organisms would have most likely evolved together in an RNA based world (3). Competition between the two groups would have led to high levels of mutation and the creation of new genes and cellular mechanisms. This hypothesis helps to explain the hallmark genes that exist in viruses (13). With viruses existing before the LUCA, we would expect to see hallmark viral genes, genes shared across diverse viral groups with only distant cellular homologs, in viruses across the three domains of life. These hallmark genes are found in viruses

across all domains of life and are typically seen in genome replication, virion formation, and packaging. The best-known example of this is the jelly-roll capsid protein. This structure is found in viral groups spanning the three domains of life and its dispersal is echoed in Superfamily 3 helicase, UL9-like superfamily 2 helicase, Archaeo-eukaryotic DNA primase, etc. (13). These dispersal patterns are not easily explained by HGT and are much more likely to result from a common ancestor. Viruses existing before the LUCA is also important in explaining homologous genes found between diverse viral groups across the domains. Homologous genes are dispersed in much the same was as hallmark genes across viruses in the three domains of life.

The regressive hypothesis argues that viruses evolved from cellular organisms through reductive evolution (7). This hypothesis offers flexibility in the origin of viruses on an evolutionary timeline and allows for the polyphyletic origin of viruses. There is debate about whether hallmark functions and structures found in viruses are indicative of monophyletic origins or are the result of horizontal gene transfer, and this hypothesis can account for both of those scenarios. If viruses are polyphyletic, then regression may have happened many times resulting in diverse virus families without hallmark genes in common. Any genes found across diverse virus families can then be explained by horizontal gene transfer. When looking at viruses as a monophyletic group, it is assumed the viruses are reductive evolutions of the first cellular organisms. In this case, similarities across diverse virus families are derived from a common ancestor. The absence of these genes in some viruses can be explained as further reductive evolution whereas giant viruses can be explained by horizontal gene transfer from cellular organisms into viruses.

Each of these hypotheses can be changed to better suit personal biases and in no way have conclusive evidence or arguments. Determining the most likely time frame for the evolution of viruses may make it easier to determine which is most likely. It has become widely accepted that the first cells lived in an RNA world around four billion years ago, where RNA was used as the carrier of genetic information as opposed to DNA (9). Assuming a RNA world, the question then becomes, how did DNA become the primary form of genetic storage? DNA is more stable allowing organisms to achieve greater genome lengths and DNA has greater ability for genome repair. Although DNA is a more favorable form of information storage, the intermediary transitions stages of DNA do not appear to have any evolutionary benefit for cellular organisms. If viruses co-evolved with early viral cells, then they could provide an explanation for the evolutionary shift of RNA to DNA (7). Modern DNA viruses have been shown to alter their genome to become resistant to host nucleases, increasing their fitness and ability to avoid the host (25). Similar modifications may have occurred within early parasitic organisms. Ribonucleotide reductases convert the ribonucleotides found in RNA to deoxyribonucleotides (24). This transition would result in U-DNA and supported by the fact that dUMP is used to synthesize dTMP in modern cells (7, 24). This transition in viruses has an immediate fitness benefit whereas it would not show such extreme selection preferences in RNA cells. The next transition would be to develop thymidylate synthase activity which would produce modern DNA containing thymidine. It is assumed that the development of thymidylate synthase activity happened independently multiple times explaining the non-homologous synthase of ThyA and ThyX (17). Many DNA viruses encode their own ribonuclease reductase and thymidylate synthase which

supports this transition within viruses. Retroviruses also represent the intermediary steps of RNA to DNA evolution. They show how a virus with an RNA genome could use DNA intermediaries to circumvent the host. Cells could have evolved DNA through interaction with viral genes or by capturing DNA viruses (7).

The advantage for parasitic organisms to evolve DNA and the remnants of this transition existing in viruses indicate that these initial parasitic organisms were likely the ancestors to viruses. Based on this, the most likely path for viral origins and evolution is a combination between the protobiont and regressive hypotheses. In this hypothesis, viral ancestors co-evolved with the first cellular organisms and may have evolved machinery typically found in cellular organisms. A possible explanation for their transition may be that they were less competitive with these cellular organisms and became parasitic as a result. This parasitism gave them an advantage in evolving DNA and over time they lost the machinery associated with cellular organisms, like ribosomes, and became modern obligate parasites. While there is currently no definitive way to determine viruses are ancient and evolved along with the first cellular organisms, analyzing phylogeny, the genetic connections between virus families and cellular organisms, and major evolutionary steps in cellular evolution can indicate whether this is a likely scenario.

Determining Viral Phylogeny

Creating phylogenetic trees for viral families proves difficult for several reasons. Some viruses incorporate their genomes directly into their hosts while others may indirectly introduce genes during their replication process and through HGT. This can make distinguishing between viral origin and cellular origin of genes difficult. The lack of clarity surrounding the origin of these genes can make it difficult to determine if

homologous genes reflect common ancestry in the case of a viral gene or horizontal gene transfer in the case of a gene with cellular origins.

Genome sequencing is the standard method used in creating phylogenetic trees and determining ancestry and relatedness. It can be used to investigate relatedness between viral groups, however, it's reliability can diminish when studying older relationships. Genome sequencing has been used to investigate NCLDV, and their relationships to each other and a common ancestor. NCLDVs all share five core genes and tend to share about fifty other ancestral proteins and genes (11). NCLDVs also tend to show higher levels of genetic overlap with cellular organisms, seen especially in giant viruses which contain genes involved in DNA repair, translation, protein folding, and polysaccharide synthesis (5). The relatedness between NCLDVs and the connection of giant mimivirus to cellular organisms makes them a good starting point for creating a viral phylogeny in conjunction with cellular organisms. Cells and NCLDVs use ribonucleotide reductase, discussed above as biosynthesis of DNA precursors, and these enzymes can be used for phylogenies. Phylogenies created using genes associated with ribonucleotide reductase support placing a viral clade branching from Eukarya and Archaea (4). The presence of ribonucleotide reductase would indicate that this clade contained ancestral DNA replication machinery (4).

Unfortunately, genome sequencing is not necessarily a good mechanism for studying viral phylogenies. In addition to the issues of rapid genome mutation and HGT which can disrupt the clarity of the phylogeny, viral genome catalogs are biased and fragmentary with an emphasis on disease-causing viruses (22). When a conscientious effort is placed in diversifying a genome catalog of viral groups, it reveals new

evolutionary connections and can expand the understanding of the evolution of that group (22). The lack of comprehensive data has led to the investigation into different methods of creating and analyzing viral phylogenies.

Structure-based phylogenies examine the changes between protein structures and use them to create a phylogeny. Protein sites evolve at different rates depending on the function of the protein. Proteins involved in crucial and highly specific mechanisms tend to evolve very slowly and can be used to examine more distant ancestral relationships (6). Before examining phylogenies constructed using structure-based techniques, it is important to note that using this technique in viruses is not widespread. This is a fairly common technique in analyzing the relationships between bacterial families, and the technique appears to be applicable to viruses (1, 16, 27).

Structure-based phylogenies are best constructed from highly conserved viral proteins such as capsid proteins which have highly specific interactions. Comparative analysis between the sequence of homologous capsid proteins and their structure reveal that structure-based phylogenies are similar to those created by sequences (23). These homologous capsid proteins indicate a possible common ancestry for Retroviruses, Caulimoviruses, Pseudoviruses, and Metaviruses (14, 15). Although some viral families have homologous capsid proteins, they are diverse viruses with unrelated varieties (14, 15). These types of relationships can be seen when looking at virus fold superfamilies. Examining these fold superfamily relationships indicate that viruses descended from ancient cells through reductive evolution (19). The evolutionary patterns shown by viruses resemble those seen in cellular life. These interactions indicate that viruses evolve

the same ways cellular life does, and that they are not evolved by life as some hypotheses state. Viruses evolve independently and of their own volition (19).

There are several issues with investigating viral phylogenies. Looking at sequence and structure-based phylogenies we can make connections between vast viral groups and create families, but it is difficult to determine when viruses evolved and whether their origin is mono- or polyphyletic. The main reason for these difficulties is the gapes in sequenced viral genomes. There are clear biases in the genomes that have been sequenced in favor of eukaryotic disease-causing viruses. There is a large lack of genomes from Archaea viruses and these gaps make it difficult to put together a complete picture.

Viral Characteristics

The gaps in the viral genome literature make it difficult to make conclusions using phylogenetic data. Without using data, it is necessary to critically examine viruses compared with both living organisms and nonliving mechanisms similar to viruses. Viruses are often compared to prions and prion disease. Prions, being composed only of a single protein, are acellular and do not have a metabolism or replicative mechanisms of their own. The argument is that viruses and prion are both dependent on their host for their survival to a degree that other obligate parasites are not. While this may be accurate, comparing the way prions evolve compared to viruses may give insight into whether assumptions made about one is applicable to the other. Prions involved in prion disease, PrPres, face selection pressures which select for the most stable form of the misfolded protein. The wildtype prion protein, PrPC, also faces selection pressures and anti-prion systems which work against the negative effects of prion disease (26). The selection pressures faced by PrPres have to do with their ability to retain their misfolded character,

misfold PrPCs, and evade degradation by the immune system. PrPres evolve when variants of the misfolded protein are present in an organism and face outside environmental pressures, allowing some forms to spread quicker than others (17). The variants are created when environmental pressures which are problematic to the exact duplication of PrPres and cause changes to the prions (17). The idea of proteins evolving without the need of nucleic acid is interesting and has some possible implications in prion research, however, this method is far different than the evolution observed in viruses. Viruses evolve through Darwinian evolution in the same way cellular organisms do. Some research looking into their evolutional patterns suggest that not only is the mechanism the same, but the relationships observed between viruses and each of the domains is the same as the patterns seen between any two domains (19). This would seem to indicate that, on an evolutionary gauge, viruses behave like cellular organisms.

Plasmids are also compared to viruses frequently, and the escape hypothesis earlier discussed argues that viruses may have been escaped plasmids. One reason for this comparison is that viruses and plasmids share several proteins and features like rolling circle replication (10). Unlike prions, plasmids evolve using the same mechanisms as cellular organisms as they are a part of their host's DNA storage, so this does not distinguish them from viruses. The main difference between viruses and plasmids are their independence from their host. Plasmids evolve with their host in the same way chromosomal DNA does, however, plasmid DNA is not a requirement for host survival. Plasmids carry extra genes that can confer an advantage, like antibiotic resistance, and alternative forms of genetic exchange. These functions are advantageous, but superfluous. If a cell is dividing under stress or lacks the energy and nutrients required to

replicate plasmids, then the resulting daughter cell may not contain a plasmid. Plasmids contain genes for the benefit of their cell and evolve within the cell in the same way all other components do. Viruses evolve for their own benefit and replicate despite any stresses their host may be under. Viruses are not evolved by their host any more than any cellular obligate parasite. The argument that viruses and plasmids can be viewed as nearly the same ignores which one, host or parasite, benefits from its presence.

Prions and plasmids can offer insight into the functioning of viruses and cellular organisms, but they are functionally different from either group. Viruses function in the same way as cellular obligate intracellular parasites, however, they depend on their host machinery instead of bringing their own. While this is an important difference, it does not change the fact that viruses function the same as cellular organisms. Reaching a conclusion based on the information currently available is impossible. When analyzing the information we have, viruses represent a gray area between life and nonlife. Assuming viruses are distinct from cellular organisms may be limiting our perspective on their influence and limiting the data deemed important to take. Without opening a discussion on the possibility that viruses should be classified as living, it is impossible to consider their full potential and impact.

Discussion

Viruses are difficult to classify as they fall in a gray area when using our current definition of life. When examining the various hypotheses regarding viral evolution, the most likely scenario appears to be a combination of the protobiont and regression hypotheses with viral ancestors being cellular organisms. These conclusions would indicate that viruses are likely ancient organisms which would be evolved from and

evolve with cellular organisms. If this was true, it may be necessary to reexamine how life is classified, because it is counterintuitive to label something that evolved from life as nonlife. To make concrete determinations regarding this hypothesis, data is needed. Phylogenetic data is very limited with regards to viruses, and the data currently accessible is biased against bacterial and especially archaea viruses. Phylogenetic data also points to viruses being ancient. An important point in the data are hallmark genes that are only found in viruses. These tend to be responsible for genomes replication, virion formation, and packaging (13). These genes are not adequately explained by the regression hypothesis and the virus first hypothesis does not work under our current definitions. The protobiont hypothesis in combination with the regression hypothesis best explains our current data on viruses and helps explain some major evolutionary steps. Viral ancestor's possible role in these evolutionary stops, like the transition of RNA to DNA, also may indicate that viral ancestors are ancient and are likely before the split between bacteria, archaea, and eukaryotes. The regression aspect also helps explain why there are hallmark genes, but no genes common to all viruses.

There is limited data regarding viral phylogenies which makes it difficult to put together a full phylogenetic viral tree or place them on the tree of life. The gaps in the data are exasperated by horizontal gene transfer and the rapid mutation of viral genomes which makes definitive conclusions impossible. Many articles examining phylogenetic data can interpret the same data in multiple contradictory ways, and that is not conducive to the scientific process. The gaps in data is likely a symptom of the flippancy with which this topic is regarded given some believe defining life is unnecessary in this context. While defining viruses may not suddenly change virology, it is evident that labeling

viruses as nonliving has an effect, conscious or not, on the way viruses are discussed and studied. This label may be limiting the scientific perspective on viruses unnecessarily.

The current lack of data means that we must compare viruses to other nonliving mechanisms to determine whether viruses behave more like them or more like cellular organisms. Two mechanisms often associated with viruses are prions and plasmids. When examining the way prions evolve and function, the association between them is clearly superficial. Although they both cause disease and are considered acellular, prions manifest as a malfunction within the body where viruses are more obviously parasitic. The comparison between viruses and plasmids show more insight. They share genes and mechanisms which made hypotheses like the escape hypothesis seem probable and lessens the possibility of viruses as cellular life. A large distinction between viruses and plasmids is their purpose within the cell or host. Plasmids confer evolutionary benefit onto their hosts and evolve within their host in the same manner as all other host nucleic acid. They also sport the benefit of being nonessential and give the host flexibility. Viruses function as obligate intracellular parasites which evolve in response to their host much like any other parasite. The similarities between plasmids and viruses is likely a result of HGT and mixing of viral and host genomes. While these similarities may offer insight, it is shortsighted to take these relations as evidence without conducting further investigation.

Viruses cause some of the deadliest diseases in human history and, in recent studies, have demonstrated the capability to function as cures for other diseases. They likely influenced the evolution of all cellular life and are an integral part of life despite being classified as nonlife themselves. This classification makes them less autonomous

and lesser in our minds, even if it is unconsciously. This may hinder our ability to research them as it is easier to discount their influence. Even if this was not the case, there are clear biases in the viral genomes sampled which are not as evident in cellular life. In addition, some studies have found reasonable placements for some viral families on the tree of life, however, they do not place them since they are nonliving. These arbitrary restrictions place real restriction on research. Scientists are open minded by nature and avoiding investigating this topic because of old definitions in contrary to that character.

Literature Cited

1. Agarwal, G., Rajavel, M., Gopal, B., & Srinivasan, N. (2009). Structure-Based Phylogeny as a Diagnostic for Functional Characterization of Proteins with a Cupin Fold. *PLoS ONE, 4*(5). doi:10.1371/journal.pone.0005736
2. Bândea, C. I. (1983). A new theory on the origin and the nature of viruses. *Journal of Theoretical Biology, 105*(4), 591-602. doi:10.1016/0022-5193(83)90221-7
3. Bandea, C. I. (2009, October 21). A Unifying Scenario on the Origin and Evolution of Cellular and Viral Domains. Retrieved from http://precedings.nature.com/documents/3888/version/1
4. Boyer, M., Madoui, M., Gimenez, G., Scola, B. L., & Raoult, D. (2010). Phylogenetic and Phyletic Studies of Informational Genes in Genomes Highlight Existence of a 4th Domain of Life Including Giant Viruses. *PLoS ONE, 5*(12). doi:10.1371/journal.pone.0015530
5. D., Raoult, S., Audic, C., Robert, C., Abergel, P., Renesto, H., Ogata, . . . JM., Claverie (2004). The 1.2-Megabase Genome Sequence of Mimivirus. *Science, 306*(5700), 1344-1350. doi:10.1126/science.1101485
6. Echave, J., Spielman, S. J., & Wilke, C. O. (2016). Causes of evolutionary rate variation among protein sites. *Nature Reviews Genetics, 17*(2), 109-121. doi:10.1038/nrg.2015.18
7. Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research, 117*(1), 5-16. doi:10.1016/j.virusres.2006.01.010

8. Herman, J. L., Challis, C. J., Novák, Á, Hein, J., & Schmidler, S. C. (2014). Simultaneous Bayesian Estimation of Alignment and Phylogeny under a Joint Model of Protein Sequence and Structure. *Molecular Biology and Evolution, 31*(9), 2251-2266. doi:10.1093/molbev/msu184
9. Higgs, P. G., & Lehman, N. (2014). The RNA World: Molecular cooperation at the origins of life. *Nature Reviews Genetics, 16*(1), 7-17. doi:10.1038/nrg3841
10. Kado, C. I. (1998). Origin and evolution of plasmids. *Antonie Van Leeuwenhoek, 73*(1), 117-126. doi:10.1023/a:1000652513822
11. Katzourakis, A., & Aswad, A. (2014). The origins of giant viruses, virophages and their relatives in host genomes. *BMC Biology, 12*(51). doi:10.1186/s12915-014-0051-y
12. Koonin, E. V., & Martin, W. (2005). On the origin of genomes and cells within inorganic compartments. *Trends in Genetics, 21*(12), 647-654. doi:10.1016/j.tig.2005.09.006
13. Koonin, E. V., Senkevich, T. G., & Dolja, V. V. (2006). The ancient Virus World and evolution of cells. *Biology direct*, *1*, 29. doi:10.1186/1745-6150-1-29
14. Krupovic, M., & Koonin, E. V. (2017). Homologous Capsid Proteins Testify to the Common Ancestry of Retroviruses, Caulimoviruses, Pseudoviruses, and Metaviruses. *Journal of Virology, 91*(12). doi:10.1128/jvi.00210-17
15. Krupovic, M., & Koonin, E. V. (2017). Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences, 114*(12), 2401-2410. doi:10.1073/pnas.1621061114

16. Lakshmi, B., Mishra, M., Srinivasan, N., & Archunan, G. (2015). Structure-Based Phylogenetic Analysis of the Lipocalin Superfamily. *Plos One, 10*(8). doi:10.1371/journal.pone.0135507

17. Makarava, N., & Baskakov, I. V. (2013). The Evolution of Transmissible Prions: The Role of Deformed Templating. *PLoS Pathogens, 9*(12). doi:10.1371/journal.ppat.1003759

18. Moreira, D., & López-García, P. (2009). Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology, 7*(4), 306-311. doi:10.1038/nrmicro2108

19. Nasir, A., & Caetano-Anolles, G. (2015). A phylogenomic data-driven exploration of viral origins and evolution. *Science Advances, 1*(8). doi:10.1126/sciadv.1500527

20. Nasir, A., Kim, K., & Caetano-Anolles, G. (2012). Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC Evolutionary Biology, 12*(1), 156-174. doi:10.1186/1471-2148-12-156

21. Scola, B. L., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., . . . Raoult, D. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature,455*(7209), 100-104. doi:10.1038/nature07218

22. Shi, M., Lin, X., Tian, J., Chen, L., Chen, X., Li, C., . . . Zhang, Y. (2016). Redefining the invertebrate RNA virosphere. *Nature, 540*(7634), 539-543. doi:10.1038/nature20167

23. Sinclair, R. M., Ravantti, J. J., & Bamford, D. H. (2017). Nucleic and Amino Acid Sequences Support Structure-Based Viral Classification. *Journal of Virology, 91*(8). doi:10.1128/jvi.02275-16

24. Torrents, E. (2014). Ribonucleotide reductases: Essential enzymes for bacterial life. *Frontiers in Cellular and Infection Microbiology, 4*. doi:10.3389/fcimb.2014.00052

25. Warren, R. A. (1980). Modified Bases in Bacteriophage DNAs. *Annual Review of Microbiology, 34*(1), 137-158. doi:10.1146/annurev.mi.34.100180.001033

26. Wickner, R. B., & Kelly, A. C. (2015). Prions are affected by evolution at two levels. *Cellular and Molecular Life Sciences, 73*(6), 1131-1144. doi:10.1007/s00018-015-2109-6

27. Xu, L., Kuo, J., Liu, J., & Wong, T. (2012). Bacterial phylogenetic tree construction based on genomic translation stop signals. *Microbial Informatics and Experimentation, 2*(1), 6-20. doi:10.1186/2042-5783-2-6