Mobile Genetic Elements 2:5, 247–252; September/October 2012; © 2012 Landes Bioscience

Viral evolution

Primordial cellular origins and late adaptation to parasitism

Arshan Nasir,[1,2] Kyung Mo Kim[3] and Gustavo Caetano-Anollés[1,2,*]

[1]Department of Crop Science; University of Illinois at Urbana-Champaign; Urbana, IL USA; [2]Illinois Informatics Institute; University of Illinois at Urbana-Champaign; Urbana, IL USA; [3]Biological Resource Center; Korea Research Institute of Bioscience and Biotechnology; Daejeon, Korea

Keywords: giant viruses, parasitism, phylogenomics, protein domains, reductive evolution

Abbreviations: ToL, tree of life; HGT, horizontal gene transfer; aaRSs, aminoacyl-tRNA synthetase; FF, fold family; FSF, fold superfamily; HMM, hidden Markov model; LUCA, last universal common ancestor; LUCELLA, last universal cellular ancestor; CEOs, capsid-encoding organisms; REOs, ribosome-encoding organisms

Submitted: 09/26/12

Revised: 11/05/12

Accepted: 11/06/12

http://dx.doi.org/10.4161/mge.22797

*Correspondence to: Gustavo Caetano-Anollés; Email: gca@illinois.edu

Commentary to: Nasir A, Kim KM, Caetano-Anolles G. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. BMC Evol Biol 2012; 12:156; PMID:22920653; http://dx.doi.org/10.1186/1471-2148-12-156

Explaining the origin of viruses remains an important challenge for evolutionary biology. Previous explanatory frameworks described viruses as founders of cellular life, as parasitic reductive products of ancient cellular organisms or as escapees of modern genomes. Each of these frameworks endow viruses with distinct molecular, cellular, dynamic and emergent properties that carry broad and important implications for many disciplines, including biology, ecology and epidemiology. In a recent genome-wide structural phylogenomic analysis, we have shown that large-to-medium-sized viruses coevolved with cellular ancestors and have chosen the evolutionary reductive route. Here we interpret these results and provide a parsimonious hypothesis for the origin of viruses that is supported by molecular data and objective evolutionary bioinformatic approaches. Results suggest two important phases in the evolution of viruses: (1) origin from primordial cells and coexistence with cellular ancestors and (2) prolonged pressure of genome reduction and relatively late adaptation to the parasitic lifestyle once virions and diversified cellular life took over the planet. Under this evolutionary model, new viral lineages can evolve from existing cellular parasites and enhance the diversity of the world's virosphere.

The Virus Problem

Viruses are intriguing biological entities that are borderline between inanimate and living matter. They have RNA- or DNA-based genomes with single- and double-stranded nucleic acids, but lack functional translation machinery responsible for protein synthesis, including ribosomes, and their own metabolism. Consequently, they require a host to replicate and spread as viral particles (virions) in large numbers populating the lands and the seas. They often integrate into cellular genomes and massively enrich the genetic repository of numerous organisms, including animals, plants and fungi.[1] They also cause important diseases and are economically relevant. Viruses are believed to have played important roles in the evolution of cellular organisms (hereinafter referred to as cells).[2-4] Despite their remarkable abundance in marine environments (~10^9 bacteriophages/L and > 50 genotypes/L)[5-8] and puzzling diversity (numerous morphological forms and replication strategies),[9] viruses, in general, have been excluded from phylogenetic and phylogenomic studies.[10-14] Many scientists support viral exclusion based on their minute size, parasitic nature, lack of metabolic activity and inability to self-replicate.[15] For these and other reasons (see ref. 15), viruses are considered unworthy of living status and their placement alongside cells in the "tree of life" (ToL) unwarranted. Unfortunately and unlike cells, viruses leave no fossil records. Their evolutionary trajectories must therefore be deduced from extant viral features, a proposition that is problematic. Historically, the question about the origin of viruses and life itself remains for the most part a philosophical debate and largely dealt with theoretical arguments rather than molecular data,

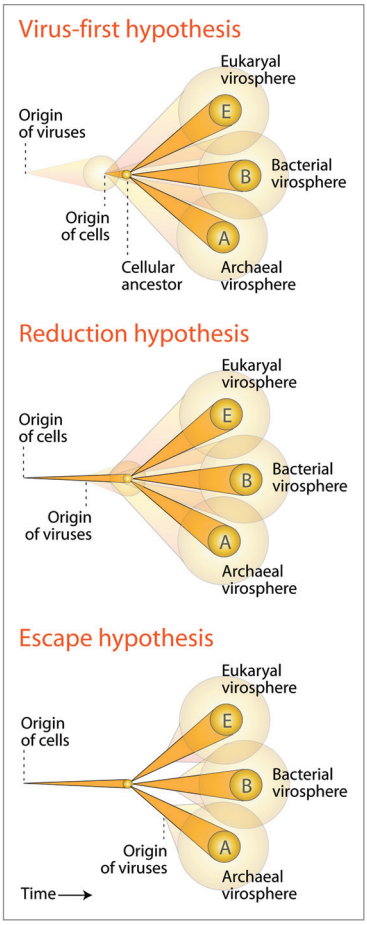

Figure 1. Three general frameworks to explain the origin of viruses. Many alternatives are possible within each hypothetical framework but are not made explicit in the diagrams. Virospheres are illustrated with clouds. We note that they can be physically linked but functionally disjoint. A, Archaea; B, Bacteria; E, Eukarya.

especially because viral genomic repertoires are limited and patchy.

Prevalent Views About the Origin of Viruses and Their Evolutionary Roles

Three general theories have been proposed to explain the origin of viruses[4] (**Fig. 1**). The "virus-first" hypothesis states that viruses predated cells and contributed to the rise of cellular life.[2,3] A significant proportion of all the viral genomes encode for genetic sequences that lack clear cellular homologs. Presence of such virus-specific sequences provides support to their unique origin.[2,3] Contrastingly, all known viruses need a cellular host to replicate, thus necessitating the existence of cells before virus survival.[4] Therefore, the virus-first hypothesis has been challenged and the existence of an ancient and independent viral world critiqued. An alternative general hypothesis associates the origin of viruses to cells and considers viruses to be the reduced forms of parasitic organisms.[16] This hypothesis, better known as the "reduction hypothesis," is supported by the recent discovery of giant viruses (e.g., mimiviruses and megaviruses)[17-19] with genomic and physical features that overlap those of numerous parasitic bacteria. A third prevalent hypothesis, the "escape hypothesis" suggests that viruses were once part of the genetic material of host cells but escaped cell control and later evolved by pickpocketing genes via horizontal gene transfer (HGT) (reviewed in refs. 2–4). HGT is believed by some scientists to be the predominant force shaping many viral genomes.[15,20] This hypothesis, however, fails to explain the presence of structures that are unique to viruses and are not present in cells.[3,4,21]

Despite disagreements, viruses are considered to be key contributors to the evolution of cells. Viruses, for example, could have mediated the evolutionary transition from RNA to DNA.[22] Adding to the already expanding roles of viruses, Patrick Forterre also proposed the "virocell" concept that links virions and cells.[23] While virions are protein-encapsidated infectious particles that contain the viral genome, virocells are bona fide cells that are under virus control have the potential to actively produce virions. In contrast, ribocells represent cells that require ribosomes to actively function and divide.[23] Virions and ribocells engage in dynamic life cycles specific to organismal groups while ribosomes and ribocells are part of a stable and tightly integrated universal system. These properties restrict evolutionary outcomes.

Structural Phylogenomics Reveals the Ancient Cellular Origin of Viruses

Hypotheses of viral origin have been hotly debated and contested.[2-4,15,24-28] Since none have full explanatory power, it is likely that a composite explanation may be more accurate. The discovery of mimiviruses and megaviruses, which mimic many parasitic cellular organisms and contain a partial translational apparatus, including several aminoacyl-tRNA synthetases (aaRSs) that are apparently functional,[17-19] now challenges the boundaries between cells and viruses. The discovery of giant viruses calls for the inclusion of viruses (at least those with larger genomes) into global phylogenetic studies.[3,29-31] In a recent breakthrough phylogenomic study, we used a census of protein domain structures in over a thousand genomes to study the origin of giant viruses.[32] Remarkably, viruses appear alongside with cells on a comparable evolutionary time scale and form a basal and distinct "supergroup" in a truly universal ToL. The phylogenomic analysis also produced network trees portraying universal ToLs very much alike those reconstructed in the past.[12-14] However, a distinct and unified viral supergroup was present at the base of the ToL before the emergence of superkingdoms, suggesting an ancient origin of giant viruses. To our knowledge, this is the first exercise that makes extensive and global use of molecular data to study viral evolution. In this study, we purposely sampled only dsDNA viruses that have large-to-medium genomes and are quite complex.[33,34] Their large proteomic makeup makes the sampling of viral domain structures comparable to cells.

What is the benefit of focusing on structure? Recent advancements in genomics and structural biology offer a wealth of molecular information that can be coupled with standard evolutionary bioinformatic tools to test alternative evolutionary models. However, it is crucial that the right molecular feature and approach be employed when studying deep evolutionary relationships. Inferring molecular phylogenies (statements of evolution) using protein domain structures has been shown in a number of studies to successfully recover reliable phylogenetic signatures.[12-14] Protein domains grouped into fold families (FFs, domains with high sequence conservation) and fold superfamilies (FSFs, domains with structural and functional evidence of common ancestry) are clearly useful study subjects for global phylogenomic analyses.[14,35,36] These protein fold structures are more

conserved than genetic sequences, which are highly variable and usually cannot hold deep historical evidence.[37] In addition, the structural phylogenomic methodology (see refs. 12–14) is robust against many artifacts resulting from sequence-based phylogenetic reconstruction[37] and provides an appropriate model for studying viral evolution. In the study of Nasir et al.,[32] viral FSF structures were assigned to genomic sequences using advanced hidden Markov models (HMMs) of structural recognition and cellular FSFs were downloaded directly from the SUPERFAMILY database.[38,39] The census of FSF abundance was then used to build phylogenies describing the evolution of protein domains and proteomes. **Figure 2** summarizes the main results of our study. Remarkably, the census in itself uncovers already important patterns. A total of 304 FSF domains were detected in the 56 viral proteomes, including 229 FSFs that were also present in all three cellular superkingdoms (Archaea, Bacteria and Eukarya). The majority (> 50%) of these "universal" FSFs were of ancient origin when they were traced on an evolutionary timeline obtained from phylogenies of protein domains (**Fig. 2**). The most ancient structures were important for metabolism and translation, some of which are part of membrane proteins, suggesting a cellular primordial origin of viruses. The axis of the timeline unfolds relative time in a 0–1 scale, from the origin of protein domains (nd = 0) to the present (nd = 1) (see refs. 12–14). These ancient and universal FSFs are a clear molecular testament of the very early coexistence of primordial viruses and cells before cellular diversification. The observation that FSFs shared with viruses represent a significant fraction of FSFs in each superkingdom is also remarkable (**Fig. 2**). These patterns underscore the central role of viruses in protein evolution. In addition, six virus-specific FSFs absent in cells make up capsids or are part of proteins necessary for cell attachment or inhibition of cellular apoptosis. These very few virus-specific FSFs appeared quite late in the timeline (at nd ~0.6) and almost concurrently with Archaea-specific and Eukarya-specific FSFs, confirming the cell-like nature of primitive viruses. Without virion structures and functions,

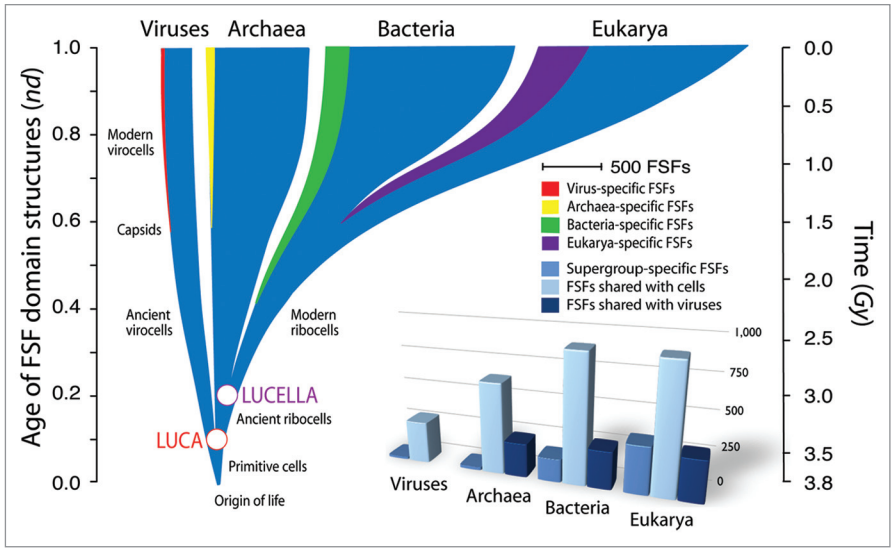

Figure 2. Evolution of the protein world. The diagram, drawn to approximate scale, shows a cartoon of a universal tree of life inferred from a phylogeny of protein domains. Time unfolds from bottom to top according to the age of FSF protein domains (nd) in a relative 0–1 scale and in geological time (billions of years, Gy) according to a molecular clock of folds.[44] The horizontal axis is proportional to the number of FSFs. Extant FSF repertoires are indicated for supergroups (superkingdoms and viruses). The FSFs that are unique to supergroups are highlighted with different color shades in the phylogeny. The common ancestor of the lineages of cells and large-to-medium-sized DNA viruses (LUCA) and the common ancestor of cellular organisms belonging to superkingdoms Archaea, Bacteria and Eukarya (LUCELLA) are indicated with circles at the base of the universal "tree of life." The bar plots show FSFs that are unique to supergroups or that are shared with viruses or cells. Note the significant number of structures shared by viruses and cells.

which are unique viral hallmarks,[40] primitive viruses had to multiply very much like cells. Consequently, they could not spread efficiently and in high numbers into the harsh environments of early Earth. Alternatively, they could have been also integral components of primitive cells.[26]

Origin of Modern Viruses from Primordial Cells

We envision a scenario in which the last universal common ancestor of life (LUCA) (from Latin: dative plural of lux, f, to become visible, shine) gave birth to (at least) two descendants: (1) the last universal cellular ancestor (LUCELLA) (from Latin: nominative plural of lucellum, i, dim. lucrum, small gain; a succession of small changes), and (2) the archaic virocell ancestor. LUCELLA was the ancestor of cells that evolved ribosomes and advanced protein biosynthesis, the ribocells. Its sibling, the archaic virocell was the ancestor of a lineage of cells that never unfolded ribosomal machinery and ultimately transformed into viral parasites and modern virocells (**Fig. 2**).

Because viruses infect all three superkingdoms of life, Forterre[23] proposed that virocells predated "modern cells" or the descendants of LUCA.[23] Our data advocates an expansion of this idea. Our timelines relate the origin of viruses to primitive ribosome-free cells committed to a reductive evolutionary path (**Fig. 2**). These cells were ancient virocells but without a capacity to produce virions (i.e., they lacked the reproductive feature of modern virocells). Our phylogenomic data suggest these virocell ancestors coexisted with evolving cellular lineages. Remarkably, there is accumulating microfossil evidence in 3–3.4 billion-year (Gy)-old black chert beds and shallow-marine siliclastic deposits of cells of spheroidal and spindle-like shapes.[41-43] These microfossils are biogenic microstructures of two broad size ranges, 5–25 μm in size and ~300 μm.[42,43] We contend that microfossil size variation could simply represent coexisting cellular lineages. Since a molecular clock of folds indicates that LUCELLA existed 2.9 Gy

ago[44] and cell size scales with genome complexity (Yafremava et al., manuscript submitted), microfossil size variation could result from early reductive evolutionary processes acting on genomic complements of the primitive virocell lineages.

Our phylognomic data also indicates that major capsid proteins and other proteins necessary for viral pathogenicity were acquired late (~1.6 Gy ago) and simultaneously with the appearance of superkingdom-specific FSFs. The late appearance of capsid proteins in our timelines disagrees with the prevalent views of capsid origin. Capsids are sheltering "envelopes" of viral genomes, which are necessary for viral spread and infection.[45] They are considered viral hallmarks[40] that are shared by many diverse viral groups and are used to unite viruses into capsid-encoding organisms (CEOs).[27] In contrast, ribosomes are the hallmarks of cells, defining cellular entities as ribosome-encoding organisms (REOs).[27] We propose that in the absence of capsids, the initial viral lineage was necessarily cellular and one of many "laboratories" exploring alternative biochemistries for life.[26] This ancient lineage failed to retain most of the translation machinery and never developed ribosomal protein biosynthesis, since ribosomal proteins and rRNA are absent in viruses. Its genome was probably an integral component that was compartmentalized. While many scenarios are possible, ancestral forms of volutin granules (acidocalcisomes) could have hosted primordial virocell nuclei and could have acted as evolutionary primordia for the stabilization of capsids and for the packaging of virocell genomes. Acidocalcisomes are ancient versatile organelles that store polyphosphates, calcium and metals, have regulatory roles during cell division and are present in all three superkingdoms.[46] Since membrane lipids are part of many viruses and are involved in the initial phases of viral infection,[47] we hypothesize that virocell membranes could have supported the appearance of first capsid proteins and could have facilitated the formation of "factories" in cellular hosts responsible for the first infectious viral cycles. This hypothesis implies that primitive viruses were in fact primordial cells with limited cytoplasmic structure but with a streamlined organelle-rich makeup. The constant reductive pressure on these cellular laboratories eventually led to secondary adaptations (i.e., parasitism) and the development of capsids and true virocells.[23]

Parasitism in Viruses: An Afterthought Triggered by Genome Reduction

One of the main properties used to define modern day viruses is their parasitic nature. Viruses are able to infect cells and take over cellular machinery of the host for their own replication. Our data suggest that viral parasitism was an afterthought likely triggered by both gradual loss of genes in ancient virocells and the opportunity to exploit the expanding ribocell molecular resources.[32] Thus, and in light of our results, current definitions of viruses must be revisited, as they are only applicable to extant viruses. Importantly, the structural makeup of the ancient viral lineage should be considered. Evolutionary timelines of domains uncovered a bimodal evolutionary pattern; the majority of the domain structures in viruses appeared either very early or very late in evolution. Timelines also revealed that while the global protein repertoire was in permanent expansion, genome reduction was the earliest primary force shaping both the viral and cellular proteomes. However, loss of ancient genes first started in viruses and was then followed by losses in superkingdoms, beginning with Archaea. These reductive trends are compatible with patterns of evolution of cells described previously[12,48] that are operating in microbial parasites and obligate parasites. Functional annotations also supported the view that very early in the timeline, viruses were functionally active and not much distinct from cells.[32] The ancient viral structures served metabolic, informational and gene regulation functions, very much like cells.[49] With the introduction of reductive evolutionary forces, most of the ancient structures were lost from viral proteomes and many were never adopted. This included the loss or lack of acquisition of advanced translational machinery. This explains why the largest viruses (mimiviruses and megaviruses) have retained only a partial encoding translation apparatus (up to 7 out of the 20 aaRSs).[18,19] This machinery is most likely the remnant of an advanced functional apparatus that was once present in the ancestor of these viruses.[19] Our contention is that genome reduction resulted in a transition to the parasitic lifestyle later in the evolutionary timeline. We have previously linked parasitism in cells to genome reduction and the appearance of domain structures unique to superkingdoms.[49] Viruses appeared to be no different.[32] The viral-specific structures acquired late in the timeline served supportive functions for viral pathogenicity. This included capsids, which appeared late and concurrently with mechanisms to suppress host defenses. Capsids crucially distinguish modern viruses from other mobile elements, such as plasmids, RNA satellites and transposons.[45] Because capsids are widespread among diverse groups of viruses, they are considered to be very ancient pre-LUCA discoveries. Our results, which are supported by molecular data and objective evolutionary bioinformatic approaches, indicate, however, that the appearance of capsid proteins postdated both LUCA and LUCELLA by at least 1.3 Gy. We therefore present an alternative view in which the appearance of capsid coincides with the appearance of modern cells and viral adaptations to parasitism. From this point onwards, archaic virocells started to acquire additional structures necessary for infecting the descendants of LUCELLA. Evolutionary forces that were predominant in this later stage included (but are not limited to) gene duplication, recombination and HGT. These forces were primarily responsible for enhancing the genetic repertoires of cells[50] once the adaptation to the parasitic mode in viruses was completed.

LUCA Predated LUCELLA

The HMM census, evolutionary timelines and the universal ToL support the ancient origin of viruses and point toward the existence of a new urancestral ~3.4-Gy-old entity that was already present before a redefined LUCELLA. This entity, the true "LUCA" of (all) life descended by gradual change into the cellular lineage that gave birth to LUCELLA, the three cellular

superkingdoms and modern ribocells[23] and the archaic virocell lineage that gave rise of virions and true virocells.[23] The original two main lines of descent preserved distinct features, which manifest today in genomic makeup. Since a modern virocell implies a transitive stage of the viral life cycle that includes both the cellular host and the viral pathogen, the primitive virocell must be considered a stable cell. This virocell lacked the complexities of a viral life cycle and had not yet developed its "virosphere" generating abilities.

Conclusions

Our structural phylogenomic inferences enable the proposal of a composite theory for evolution of giant viruses and viruses in general. We propose that giant viruses (with their DNA genomes) are remnants of an ancient virocell lineage that once coexisted with cellular lineages either independently or compartmentalized within the primitive cells. This helps explain the presence of most ancient FSFs that were identified in both viruses and cells. This viral lineage suffered massive gene loss throughout evolutionary history. While we have not yet explored the ultimate cause of this reductive process, we conjecture that co-evolution of the ancient cells was instrumental for the development of nucleic acid repositories and modern genetics. Patterns of biochemical specialization could have initially favored small and compartmentalized genomic repertoires in the virocell lineage, putting in motion irreversible selective pressures for reductive evolution that were absent in the cellular lineage. This tendency required a focus on economy of resources and fast reproductive spread for persistence, which likely triggered increasingly smaller organismal entities, the need to adapt to a parasitic lifestyle and the development of the capsid container as strategy of ultimate persistence. This path to obligate parasitism mimics that of cellular parasites.[49] Since HGT appeared to play only marginal roles very late in evolution,[32] perhaps once the parasitic adaptation was completed, it is possible that other viral groups (such as those with RNA genomes) followed the same path. Under this model, evolution of parasitic cellular species into viruses may still be active.[51] Future analysis of the entire virosphere will yield significant insights into the evolution of all viruses and will test if indeed they have a single (monophyletic) or multiple (polyphyletic) origin.

Acknowledgments

Research is supported by the National Science Foundation (MCB-0749836 and OISE-1132791) of the United States and by a grant (PJ0090192012) from the Next-Generation BioGreen 21 Program, Rural Development Administration, Republic of Korea. Any opinions, findings, conclusions and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

1. Griffiths DJ. Endogenous retroviruses in the human genome sequence. Genome Biol 2001; 2:S1017; PMID:11423012; http://dx.doi.org/10.1186/gb-2001-2-6-reviews1017.
2. Koonin EV, Senkevich TG, Dolja VV. The ancient Virus World and evolution of cells. Biol Direct 2006; 1:29; PMID:16984643; http://dx.doi.org/10.1186/1745-6150-1-29.
3. Koonin EV, Senkevich TG, Dolja VV. Compelling reasons why viruses are relevant for the origin of cells. Nat Rev Microbiol 2009; 7:615, author reply 615; PMID:19561624; http://dx.doi.org/10.1038/nrmicro2108-c5.
4. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. Virus Res 2006; 117:5-16; PMID:16476498; http://dx.doi.org/10.1016/j.virusres.2006.01.010.
5. Anderson NG, Cline GB, Harris WW, Green JG. Isolation of viral particles from large fluid volumes. In: Berg G, ed. Transmission of Viruses by the Water Route. New York: Interscience, 1967:75.
6. Breitbart M, Rohwer F. Here a virus, there a virus, everywhere the same virus? Trends Microbiol 2005; 13:278-84; PMID:15936660; http://dx.doi.org/10.1016/j.tim.2005.04.003.
7. Bergh O, Børsheim KY, Bratbak G, Heldal M. High abundance of viruses found in aquatic environments. Nature 1989; 340:467-8; PMID:2755508; http://dx.doi.org/10.1038/340467a0.
8. Rohwer F, Thurber RV. Viruses manipulate the marine environment. Nature 2009; 459:207-12; PMID:19444207; http://dx.doi.org/10.1038/nature08060.
9. Wessner DR. Discovery of the giant mimivirus. Nature Education 2010; 3:61.
10. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci USA 1977; 74:5088-90; PMID:270744; http://dx.doi.org/10.1073/pnas.74.11.5088.
11. Woese C. The universal ancestor. Proc Natl Acad Sci USA 1998; 95:6854-9; PMID:9618502; http://dx.doi.org/10.1073/pnas.95.12.6854.
12. Wang M, Yafremava LS, Caetano-Anollés D, Mittenthal JE, Caetano-Anollés G. Reductive evolution of architectural repertoires in proteomes and the birth of the tripartite world. Genome Res 2007; 17:1572-85; PMID:17908824; http://dx.doi.org/10.1101/gr.6454307.
13. Wang M, Caetano-Anollés G. Global phylogeny determined by the combination of protein domains in proteomes. Mol Biol Evol 2006; 23:2444-54; PMID:16971695; http://dx.doi.org/10.1093/molbev/msl117.
14. Kim KM, Caetano-Anollés G. The evolutionary history of protein fold families and proteomes confirms that the archaeal ancestor is more ancient than the ancestors of other superkingdoms. BMC Evol Biol 2012; 12:13; PMID:22284070; http://dx.doi.org/10.1186/1471-2148-12-13.
15. Moreira D, López-García P. Ten reasons to exclude viruses from the tree of life. Nat Rev Microbiol 2009; 7:306-11; PMID:19270719; http://dx.doi.org/10.1038/nrmicro2108.
16. Bândea CI. A new theory on the origin and the nature of viruses. J Theor Biol 1983; 105:591-602; PMID:6672474; http://dx.doi.org/10.1016/0022-5193(83)90221-7.
17. La Scola B, Audic S, Robert C, Jungang L, de Lamballerie X, Drancourt M, et al. A giant virus in amoebae. Science 2003; 299:2033; PMID:12663918; http://dx.doi.org/10.1126/science.1081867.
18. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, et al. The 1.2-megabase genome sequence of Mimivirus. Science 2004; 306:1344-50; PMID:15486256; http://dx.doi.org/10.1126/science.1101485.
19. Arslan D, Legendre M, Seltzer V, Abergel C, Claverie JM. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc Natl Acad Sci USA 2011; 108:17486-91; PMID:21987820; http://dx.doi.org/10.1073/pnas.1110889108.
20. Moreira D, Brochier-Armanet C. Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. BMC Evol Biol 2008; 8:12; PMID:18205905; http://dx.doi.org/10.1186/1471-2148-8-12.
21. Abroi A, Gough J. Are viruses a source of new protein folds for organisms? - Virosphere structure space and evolution. Bioessays 2011; 33:626-35; PMID:21633962; http://dx.doi.org/10.1002/bies.201000126.
22. Forterre P. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie 2005; 87:793-803; PMID:16164990; http://dx.doi.org/10.1016/j.biochi.2005.03.015.
23. Forterre P. Virocell concept, The. eLS 2012.
24. Moreira D, López-García P. Comment on "The 1.2-megabase genome sequence of Mimivirus". Science 2005; 308:1114, author reply 1114; PMID:15905382; http://dx.doi.org/10.1126/science.1110820.
25. Koonin EV, Dolja VV. Evolution of complexity in the viral world: the dawn of a new vision. Virus Res 2006; 117:1-4; PMID:16497406; http://dx.doi.org/10.1016/j.virusres.2006.01.018.
26. Koonin EV. On the origin of cells and viruses: primordial virus world scenario. Ann N Y Acad Sci 2009; 1178:47-64; PMID:19845627; http://dx.doi.org/10.1111/j.1749-6632.2009.04992.x.
27. Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. Nat Rev Microbiol 2008; 6:315-9; PMID:18311164; http://dx.doi.org/10.1038/nrmicro1858.
28. Claverie JM, Ogata H, Audic S, Abergel C, Suhre K, Fournier PE. Mimivirus and the emerging concept of "giant" virus. Virus Res 2006; 117:133-44; PMID:16469402; http://dx.doi.org/10.1016/j.virusres.2006.01.008.
29. Claverie JM, Ogata H. Ten good reasons not to exclude giruses from the evolutionary picture. Nat Rev Microbiol 2009; 7:615, author reply 615; PMID:19561626; http://dx.doi.org/10.1038/nrmicro2108-c3.

30. Claverie JM, Abergel C. Mimivirus and its virophage. Annu Rev Genet 2009; 43:49-66; PMID:19653859; http://dx.doi.org/10.1146/annurev-genet-102108-134255.
31. Ludmir EB, Enquist LW. Viral genomes are part of the phylogenetic tree of life. Nat Rev Microbiol 2009; 7:615, author reply 615; PMID:19561625; http://dx.doi.org/10.1038/nrmicro2108-c4.
32. Nasir A, Kim KM, Caetano-Anolles G. Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. BMC Evol Biol 2012; 12:156; PMID:22920653; http://dx.doi.org/10.1186/1471-2148-12-156.
33. Iyer LM, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. Virus Res 2006; 117:156-84; PMID:16494962; http://dx.doi.org/10.1016/j.virusres.2006.01.009.
34. Iyer LM, Aravind L, Koonin EV. Common origin of four diverse families of large eukaryotic DNA viruses. J Virol 2001; 75:11720-34; PMID:11689653; http://dx.doi.org/10.1128/JVI.75.23.11720-11734.2001.
35. Andreeva A, Howorth D, Chandonia JM, Brenner SE, Hubbard TJ, Chothia C, et al. Data growth and its impact on the SCOP database: new developments. Nucleic Acids Res 2008; 36(Database issue):D419-25; PMID:18000004; http://dx.doi.org/10.1093/nar/gkm993.
36. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995; 247:536-40; PMID:7723011; http://dx.doi.org/10.1016/S0022-2836(05)80134-2.
37. Caetano-Anollés G, Nasir A. Benefits of using molecular structure and abundance in phylogenomic analysis. Front Genet 2012; 3:172; PMID:22973296; http://dx.doi.org/10.3389/fgene.2012.00172.
38. Gough J, Karplus K, Hughey R, Chothia C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol Biol 2001; 313:903-19; PMID:11697912; http://dx.doi.org/10.1006/jmbi.2001.5080.
39. Gough J, Chothia C. SUPERFAMILY: HMMs representing all proteins of known structure. SCOP sequence searches, alignments and genome assignments. Nucleic Acids Res 2002; 30:268-72; PMID:11752312; http://dx.doi.org/10.1093/nar/30.1.268.
40. Krupovic M, Bamford DH. Order to the viral universe. J Virol 2010; 84:12476-9; PMID:20926569; http://dx.doi.org/10.1128/JVI.01489-10.
41. Sugitani K, Grey K, Nagaoka T, Mimura K. Three-dimensional morphological and textural complexity of Archean putative microfossils from the Northeastern Pilbara Craton: indications of biogenicity of large (>15 microm) spheroidal and spindle-like structures. Astrobiology 2009; 9:603-15; PMID:19778272; http://dx.doi.org/10.1089/ast.2008.0268.
42. Javaux EJ, Marshall CP, Bekker A. Organic-walled microfossils in 3.2-billion-year-old shallow-marine siliciclastic deposits. Nature 2010; 463:934-8; PMID:20139963; http://dx.doi.org/10.1038/nature08793.
43. Wacey D, Kilburn M, Saunders M, Cliff J, Brasier M. Microfossils of sulphur-metabolizing cells in 3.40 billion-year-old rocks of western australia. Nat Geosci 2011; 4:698-702; http://dx.doi.org/10.1038/ngeo1238.
44. Wang M, Jiang YY, Kim KM, Qu G, Ji HF, Mittenthal JE, et al. A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. Mol Biol Evol 2011; 28:567-82; PMID:20805191; http://dx.doi.org/10.1093/molbev/msq232.
45. Abrescia NG, Bamford DH, Grimes JM, Stuart DI. Structure unifies the viral universe. Annu Rev Biochem 2012; 81:795-822; PMID:22482909; http://dx.doi.org/10.1146/annurev-biochem-060910-095130.
46. Seufferheld MJ, Kim KM, Whitfield J, Valerio A, Caetano-Anollés G. Evolution of vacuolar proton pyrophosphatase domains and volutin granules: clues into the early evolutionary origin of the acidocalcisome. Biol Direct 2011; 6:50; PMID:21974828; http://dx.doi.org/10.1186/1745-6150-6-50.
47. Roine E, Bamford DH. Lipids of archaeal viruses. Archaea 2012; 2012:384919; PMID:23049284; http://dx.doi.org/10.1155/2012/384919.
48. Wang M, Kurland CG, Caetano-Anollés G. Reductive evolution of proteomes and protein structures. Proc Natl Acad Sci USA 2011; 108:11954-8; PMID:21730144; http://dx.doi.org/10.1073/pnas.1017361108.
49. Nasir A, Naeem A, Khan MJ, Lopez-Nicora HD, Caetano-Anolles G. Annotation of protein domains reveals remarkable conservation in the functional makeup of proteomes across superkingdoms. Genes 2011; 2:869; http://dx.doi.org/10.3390/genes2040869.
50. Wang M, Caetano-Anollés G. The evolutionary mechanics of domain organization in proteomes and the rise of modularity in the protein world. Structure 2009; 17:66-78; PMID:19141283; http://dx.doi.org/10.1016/j.str.2008.11.008.
51. Bandea CI. The origin and evolution of viruses as molecular organisms. Nature proceedings 2009; http://hdl.handle.net/10101/npre.2009.3886.1