Long-term evolution of viruses: A Janus-faced balance

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

The popular textbook image of viruses as noxious and selfish genetic parasites greatly underestimates the beneficial contributions of viruses to the biosphere. Given the crucial dependency of viruses to reproduce in an intracellular environment, viruses that engage in excessive killing (lysis) can drive their cellular hosts to extinction and will not survive. The lytic mode of virus propagation must, therefore, be tempered and balanced by non-lytic modes of virus latency and symbiosis. Here, we review recent bioinformatics and metagenomic studies to argue that viral endogenization and domestication may be more frequent mechanisms of virus persistence than lysis. We use a triangle diagram to explain the three major virus persistence strategies that explain the global scope of virus-cell interactions including lysis, latency and virus-cell symbiosis. This paradigm can help identify novel directions in virology research where scientists could artificially gain control over switching lytic and beneficial viral lifestyles.

Also see the Video Abstract: https://youtu.be/E1TOU1JDXo4

Keywords:
beneficial viruses; persistence triangle; viral domestication; viral endogenization; virus-host interactions

Introduction

Viruses engage in lytic interactions that can destroy the infected cells. Lysis underpins the name "virus" (*Latin*, venom, poisonous emanation) and supports the dominant view of textbooks and popular press that viruses are noxious parasites, selfish genetic agents and significant threats to crops, livestock, poultry, and human life. A focus on the "panspermic" lytic mode of virus propagation, however, greatly underestimates the global scope of virus-cell interactions and any possible beneficial roles that viruses may play in the biosphere [1, 2]. Viruses can also become dormant through episomal (plasmid-like) or proviral latency mechanisms of cellular and genomic integration, exemplified by lysogeny of temperate bacterial phages [3], and can lead to symbiosis (intimate associations), which can be taken to the extreme in symbiogenesis (fusion of partners) [2]. The three distinct forms of virus-cell interactions, *lysis*, *latency*, and *symbiosis*, can have contrasting long-term evolutionary consequences for both cells and viruses.

Lysis, if successful, often results in cell death and/or observable cytopathic effects under the microscope (e.g. syncytia formation, budding of enveloped viruses). Viral progeny in lytic interactions can be a source of evolutionary innovations and novelties [4] (e.g. evolution of antiviral defense systems [5] and viral mimicry of cellular proteins to escape host immune system [6]), as interactions drive "evolutionary arms races" in both cells and viruses [7]. In turn, *latency* often results in a state of viral inactivity or cellular integration that is covert, cannot be readily observed under the microscope, and can provide fitness advantages. For example, phage (virus) integration into bacterial chromosomes is known to enhance virulence of bacterial species and is also a mechanism of phage-mediated bacterial gene regulation [8]. Latency also involves domestication of full-length viral genomes or genes for functions beneficial to cells (e.g. [9]). Lytic and latent virus-cell interactions are generally restricted to specific hosts although some viruses can cross species borders [10, 11]. In contrast, *symbiosis* results in partnerships that can impact organisms separated by large evolutionary distances. Examples include viruses influencing archaeal and bacterial species of the eukaryotic microbiota [12] (similar to known examples of bacterial endosymbionts in plants and fungi [2]) and the use of lytic and/or latent properties of viruses by cells to gain a competitive edge against a "third" party (e.g. bacteriophages [13] providing immunity to metazoa against invading bacteria [14]).

DOI 10.1002/bies.201700026

[1)] Department of Biosciences, COMSATS Institute of Information Technology, Islamabad, Pakistan
[2)] Evolutionary Bioinformatics Laboratory, Department of Crop Sciences, University of Illinois, Urbana, IL, USA
[3)] Division of Polar Life Sciences, Korea Polar Research Institute, Incheon, Republic of Korea

***Corresponding authors:**
Arshan Nasir
E-mail: arshan.nasir@gmail.com
Gustavo Caetano-Anollés
E-mail: gca@illinois.edu

A struggle for persistence: A principle for long-term viral evolution?

Because of the crucial dependency of viruses to reproduce in an intracellular environment, the three forms of virus-cell interactions, lysis, latency, and symbiosis are in conflict. Inspired by a previous explanatory framework of trade-offs of engineering strategies [15], here we propose that these interactions can lead to *propagation*, *dormancy*, and *dependency* trade-off solutions fostering flexibility, robustness, and economy (defined in [15]), respectively, that are beneficial to the long-term evolution of viruses. This triangle of viral persistence (Fig. 1) depicts a "Janus-Faced" balance of power between the lytic pathogenic and the cooperative and more altruistic non-lytic transects. Janus is the Roman God of beginnings, transitions, and time, usually portrayed with two faces, one looking into the future and the other into the past. A global viral quasispecies locates in the trade-off triangle according to its physiology, ecology, and history. Such balance of trade-offs seeks explanation of the "evolutionary dilemma that too much success is a potential disaster: an organism that drives its prey or hosts to extinction does not survive" [16]. That is, a long-term persistent evolutionary push towards the *propagation* vertex, as generally believed, could lead to extinction of virus hosts. Thus, the evolutionary opportunities for mutational innovation (flexibility) provided by viral propagation must be offset by counteracting pushes toward the *dormancy* and *dependency* vertices (Fig. 1), which foster robustness through cellular latency and economy through sharing of resources between interacting partners, respectively. We speculate that these non-lytic modes are preferred or more frequent outcomes in virus evolution than anticipated but have been greatly underestimated because: (i) they do not yield the phenotypic (cytopathic) effects of viral infection; (ii) sequencing databases hold information for only a tiny fraction of extant viruses (e.g. the human virome is far from complete [17]); and (iii) the current bias is to study viruses of clinical, economical, and agricultural importance (i.e. lytic viruses).

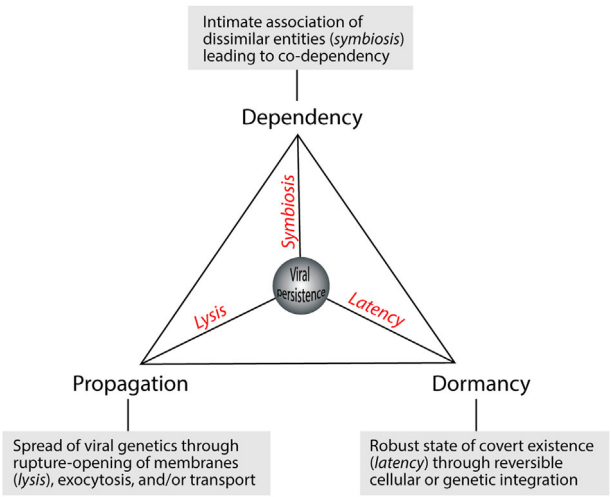

Figure 1. A triangle of viral persistence. The model explains three trade-off solutions for increasing viral performance throughout evolutionary history (persistence): *propagation*, *dormancy* and *dependency*. The sphere represents viral quasispecies collectives that locate in Pareto fronts within a 3-dimensional fitness landscape (a 2-polytope) according to physiological, ecological, and historical factors (see [67] for a mathematical elaboration of Pareto geometries of performance spaces for best-fitness solutions). Lysis, latency, and symbiosis are approaches we showcase to achieve solution goals. Lysis describes a tendency toward spread of virus genetic material through destruction of infected cells, though milder mechanisms (e.g. budding via exocytosis, cell-to-cell transport) are possible [68, 69]. It offers opportunities for mutational innovation (evolutionary flexibility). Latency describes a tendency toward virus dormancy inside the cell that favors robustness either in the form of episome or endogenized genetic material. Note that latency-to-lysis conversion is possible and can push viral quasispecies clouds toward corresponding vertices. Symbiosis describes tendencies toward mutualism, commensalism, amensalism, and parasitism in which viruses remain active and provide intimate association through altruistic, cooperative, or antagonistic behaviors of symbiotic partners. The approach offers sharing of resources (economy). Alternatives to the proposed model include persistence spaces that are simpler (line segment describing lytic and non-lytic strategies) or more complex (*n*-polytopes describing a multiplicity of strategies).

Is there a preference for latency and symbiosis in the long-term evolution of viruses?

Evidence for the evolutionary push towards the *dormancy* and *dependency* vertices of the triangle (Fig. 1) comes from the historical record. An expected long-term outcome of viral latency is cellular integration of viral genetic material that can be domesticated/coopted by cells [18, 19], suggesting this persistence mode could be a widely employed cellular mechanism. Indeed, the prevalence and abundance of endogenous retroviruses (ERVs) in mammalian genomes [20], evolution of "integration hotspots" in bacterial chromosomes (e.g. prophages) [18], plasmids co-existing harmoniously in diverse prokaryotes, and virus-derived genes in a number of cellular genomes [21–23] provide support to the idea that historically both full-length viral genomes and viral genes have either established permanent residence in hosts or were domesticated/coopted by cells. Additional support comes from the recent metagenomic analysis of diverse microbial communities revealing that increases in microbial abundance were linked to a decline in virus-to-microbe ratio and increases in abundance of hallmark genes involved in viral lysogeny [24]. Although being the first evidence of this kind, the study indicated viral preference for dormancy in special circumstances and offered unique insights into the ecological dynamics of viral lifestyles. In fact, temperate phages tend to dominate viromes extracted from fecal microbiota samples [25, 26] and ~60–70% of sequenced bacterial genomes are estimated to contain prophages [27].

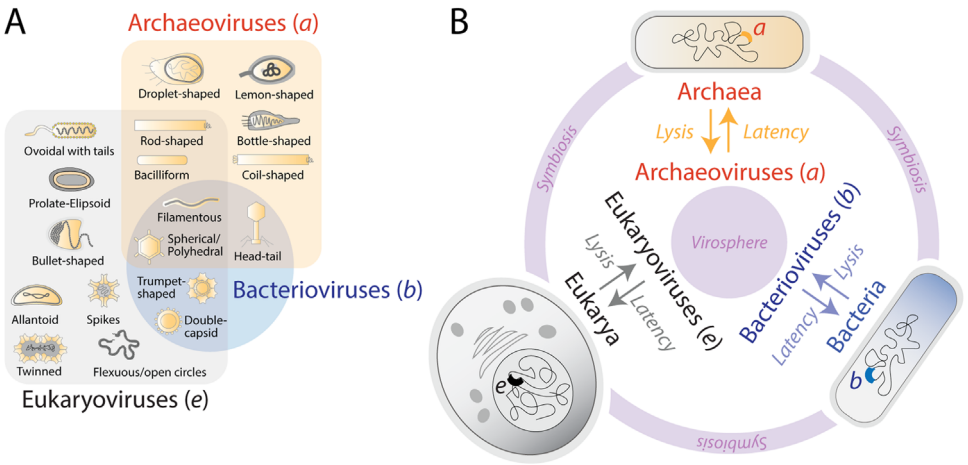

Figure 2. The world of virus-host interactions. **A:** The Venn diagram shows the diversity of viruses infecting the three superkingdoms of life, archaeoviruses (*a*), bacterioviruses (*b*), and eukaryoviruses (*e*), grouped on the basis of common or unique virion morphotypes (modified from [41]). Note that there are only two morphotypes common to all three viral groups and that there are no bacterial-specific morphotypes. **B:** Diagram depicting lysis, latency, and symbiosis interactions of the virosphere. Viruses establish interactions with their hosts in each superkingdom, sometimes becoming latent inside cells (*a*, *b*, and *e* inside cell diagrams). A ring of symbiosis unifies organisms and microbiomes across all superkingdoms providing opportunities of sharing and exchange. Note that archaea and bacteria have a mobilome that is more similar to each other and different from the eukaryotic mobilome.

Virus-cell interactions that push viral persistence towards the *dependency* vertex can alter the definition of virus "host" and involve virus-cell symbiosis (Fig. 2). Viral host jumps are common (e.g. influenza viruses evolving to infect new species) but are mostly restricted to organisms related by taxonomy. Indeed, no virus is currently known to produce progeny (virions) in organisms belonging to more than one superkingdom. However, this should not mean that viruses do not influence cells they do not lyse. For example, a recent study reported direct virus-metazoan symbiosis limiting pathogenic bacterial growth on mucosal surfaces, a phenomenon apparently conserved from cnidarians to humans [14]. Such symbiotic relationships represent virus-host dependencies involving simultaneous interactions of viruses with organisms from more than one superkingdom (bacteria and eukarya), which in this case provide a novel non-host derived virus-based immunity to metazoa while lytic effects are observed only in bacteria. Such interactions are relatively well documented for bacterial endosymbionts and their host pathogens of other groups of organisms (e.g. [28]). For example, the plant pathogenic fungus *Rhizopus microsporus* harbors the proteobacterial endosymbiont *Burkholderia rhizoxinica*. The endosymbiont produces the virulence factor (rhizoxin) that is antimitotic in nature and arrests cell cycle in the plant host of fungi leading to rice seedling blight disease [29]. It will be intriguing to extend this tripartite interaction between the bacterial endosymbiont, pathogenic fungi, and plants to prophages inserted into the bacterial genomes. Prophages are known to enhance the virulence of their hosts [30] and in doing so can modulate animal and plant microbiomes (e.g. a 3-way virus-fungus-plant symbiosis [31]). More recently, a eukaryotic association module was detected in prophages (bacteriophage WO) inserted in the genomes of the bacterial parasite *Wolbachia* that infects arthropods reporting the first documented example of lateral gene transfers between eukaryotes and bacterioviruses [32]. Another interesting example are polydnaviruses integrated into the genomes of parasitoid wasps [33, 34]. Parasitic wasps have undoubtedly domesticated polydnaviruses [35] using them to coat wasp genes to produce customized viral particles. These "genetically modified" viral particles [33] deactivate the caterpillar's immune system when wasps lay eggs and thus help wasps to reproduce. In this example, polydnaviruses and parasitic wasps function as a single unit, blurring the definition of organisms [34] and behaving as holobionts [36]. In another example, ASPE phage genes appear to protect aphids (hosts of endosymbiotic bacterium *Hamiltonella defensa*) against parasitoid wasps [37].

In summary, virus-host interactions do not always yield hallmark phenotypic symptoms of viral infections and can influence hosts they do not lyse. Moreover, integration or domestication of viral genetic elements often benefits the cells, well illustrated by the examples of parasitoid wasps [33, 34] and prophages in bacteria [38]. These observations confront our textbook perception of viruses as selfish genetic parasites and call for a wider recognition of a multiplicity of viral roles including their utility as symbionts and beneficial drivers of host evolution [2, 39]. Their study may also reveal a preference for virus-cell dormancy and dependency [24] that is worthy of exploration.

Lytic interactions hold deep historical accounts of how superkingdoms customized virospheres by gain-and-loss of viral lineages

Lytic interactions that drive ongoing evolutionary arm races between cells and viruses (*sensu* [4]) and push viral persistence towards the *propagation* vertex of the triangle (Fig. 1) could have

triggered major evolutionary innovations. For example, cellular organisms could have evolved strategies that permanently block some viral infections. If true, there would be strong detectable biases in the distribution of viral replicons in major groups of cellular organisms. Indeed, RNA viruses are either absent in archaea or rare in bacteria, retrotranscribing and RNA viruses are abundant in animal and plant hosts, dsRNA viruses are abundant in Fungi, and DNA viruses are rare in plants (Table 1, see also [40]). These biases hint that virus-cell conflicts have historically led to gain/loss of viral lineages, customizing the virospheres of superkingdoms of cellular life [41]. For example, the ancestors of archaea were likely thermophilic organisms [42, 43] (see [44] for an example phylogeny). Perhaps migration to warmer habitats provided a fitness advantage to ancestral archaeal cells to get rid of the primordial RNA viruses, especially because RNA is quite unstable at extreme temperatures [45]. Similarly, the evolutionary development of a thick peptidoglycan layer of bacterial ancestors that is seemingly impenetrable to many viruses could have blocked many viral interactions [46]. Viral persistence could have also driven cellular complexity. The significant abundance and diversity of RNA viruses and retroviruses in eukaryotes (i.e. 55 distinct dsRNA, ssRNA, and retrotranscribing viral families out of total 77) (Table 1) suggest they triggered arms races responsible for eukaryotic organismal complexity [4]. This is especially relevant since RNA and retroviruses are known to mediate genetic rearrangements and induce epigenetic changes [19, 45, 47].

Viral persistence could have driven cellular diversification

The strong bias in the distribution of viral replicon types in prokaryotes (mostly DNA) and eukaryotes (mostly RNA) (Table 1) can test scenarios of origin of superkingdoms and viruses (Fig. 3). For example, is the (near)-absence of RNA viruses in prokaryotes due to loss of viral lineages [40] or late de novo gain of viral families in eukaryotes [48]? How to reconcile loss of RNA viruses from prokaryotes under the 3-domain "Woeseian" canonical tree [49], the 2-domain archaeal-ancestor scenario (AAS) [50], or the ring of life models of evolution [51]? Or alternatively, is data compatible with a root of the ToL in the branch leading to archaea [52]?

We speculate that the late origin of a large number of eukaryotic RNA and retroviruses from mixing of prokaryotic viruses [48] seems unlikely because: (i) RNA and retroviruses are likely very ancient and mediated the transition to the DNA world via retrotranscription [53]; (ii) a total of 68 protein fold superfamilies (FSFs) [54] encoded by all seven viral replicon types are present in archaeoviruses, bacterioviruses, and eukaryoviruses (the *abe* group, Fig. 3) suggesting an origin of viral lineages before the origin of modern cells [55]; and (iii) under the 2-domain AAS or canonical 3-domain trees, eukaryoviruses should exhibit (at least) some overlap with archaeoviruses but only two FSFs (involved in DNA replication/repair and metabolism) and two virion morphotypes (rod-shaped and bacilliform) are shared by archaeoviruses and eukaryoviruses (*ae* group, Fig. 3). The *ae* FSFs are coded by dsDNA (and not RNA) viruses while the common morphotypes likely evolved via convergence [41]. Notably, while the archaeal and eukaryotic virospheres appear starkly different, the mobilomes of archaea and bacteria show remarkable resemblances (e.g. common viral families, abundances of plasmids, and 23 common FSFs encoded by dsDNA and ssDNA viruses) in addition to employing common CRISPR-*Cas* antiviral

Table 1. Counts of viral replicon types (RC) and families (FC) in major host groups[a]

	dsDNA		ssDNA		dsRNA		ssRNA		Retro-transcribing	
Host[b]	RC	FC	RC	FC	RC	FC	RC	FC	RC	FC
Archaea	40	10	0	0	0	0	0	0	0	0
Bacteria	1731	6	82	2	5	1	12	1	0	0
Algae	23	1	0	0	1	1	2	2	0	0
Fungi	0	0	0	0	50	7	30	5	0	0
Invertebrates	103	8	72	2	17	3	99	12	0	0
Invertebrates and plants	0	0	0	0	10	1	47	3	0	0
Invertebrates and vertebrates	0	0	0	0	0	0	5	2	0	0
Plants	1	1	402	2	62	4	540	14	63	1
Protozoa	12	2	0	0	29	1	0	0	0	0
Vertebrates	378	7	149	3	23	2	289	12	75	2
Vertebrates and humans	44	5	59	2	9	3	191	14	4	2
Vertebrates and invertebrates	1	1	0	0	13	1	25	5	0	0
Vertebrates, invertebrates, and humans	0	0	0	0	6	1	80	4	0	0

[aa]Counts as reported in the file transfer protocol repository of NCBI Genome Reports on March 28, 2016 (ftp://ftp.ncbi.nlm.nih.gov/genomes/GENOME_REPORTS/).
[b]Unclassified and unassigned viruses, satellites, viroids, and environmental isolations were excluded from counts. Some viral families and replicons may repeat in more than one group (e.g. Retroviridae and Hepadnaviridae that infect both vertebrates and vertebrates and humans).

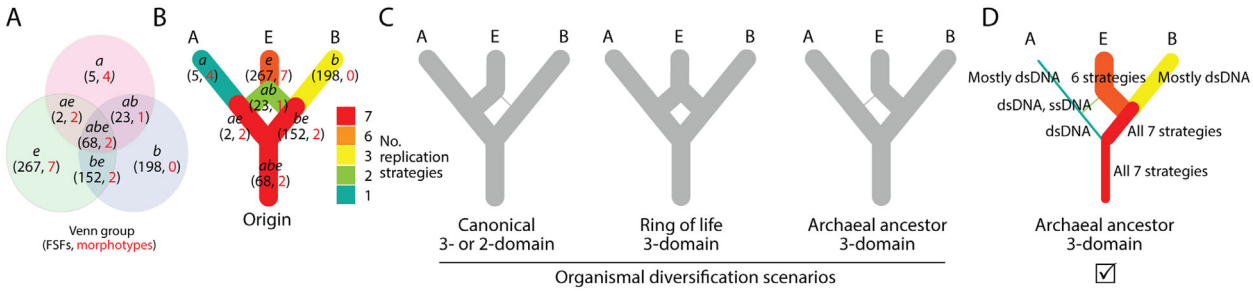

Figure 3. Evaluation of competing scenarios of cellular evolution under the gain-and-loss model of viral lineages. **A:** A Venn diagram shows FSFs and virion morphotypes of archaeoviruses (*a*), bacterioviruses (*b*), and eukaryoviruses (*e*), many of which are shared between them, traced onto corresponding Venn groups. **B:** FSFs and virion morphotype tracings are mapped onto the branches of a ToL with possible reticulation (accounting for all possible Venn groups), which describes the evolution of archaea (A), bacteria (B), and eukarya (E). This mapping is a comparative (not evolutionary) exercise since change is not reconstructed back in time in the branches of the trees (failing to account for the interplay of gains and losses along branches). **C:** The ToL with reticulations (described in B) embodies three alternative superkingdom diversification scenarios. **D:** The tracing exercise reveals that the best match of tracings to evolutionary scenarios is the archaeal ancestor 3-domain organismal diversification [52]. Remarkably, the mapping of viral replicon type strategies on this tree is also the most parsimonious. Data shown in the Figure is taken from [41, 55].

defense [56]. These data suggest that prokaryotes selected virospheres similarly in evolution, experiencing an early loss of viral lineages (mostly RNA). We note, however, that evolutionary, ecological, and physiological processes likely started earlier in archaea than bacteria given the relatively larger size of the *be* group (152 FSFs) and the fact that only the *abe* and *be* groups are coded by all seven replicon type strategies. These RNA virus lineages were, therefore, retained in Eukarya leading us to pick the archaeal-ancestor 3-domain tree as the best parsimonious explanation supported by data.

Taken together, current data on lytic virus-cell interactions indicate that superkingdoms have likely customized their virospheres by gain-and-loss of viral lineages. These interactions have thus tailored the long-term evolutionary history of modern cells dating back to the earliest stages of cellular diversification. We note, however, that virospheres remain vastly unexplored for many host organisms and hence zeroes for any group (e.g. RNA viruses in archaea, Table 1) should be considered underestimates. Future discovery of novel viruses from metagenomic samples will directly test the proposed gain-and-loss model of viral lineages and its impact on cellular diversification.

Can the persistence triangle help artificial construction of beneficial virus-cell interactions?

Modifying viral persistence by changing the triangle's trade-offs can have important medical applications. Because the differences in the cellular membrane and molecular biology of the three superkingdoms apparently block viruses from lysing organisms in more than one superkingdom, the artificial construction of virus-host alliances against a "third party" could benefit antimicrobial research. For example, the idea of virus-mediated cleansing of microbiota to treat bacterial infections has gained popularity (reviewed in [25]; see [57] for practical challenges and concerns). The example discussed above where bacterioviruses residing in the mucosal membranes of metazoa kill invading bacteria lends additional support to virus-host mutualism [14]. While the idea may seem a distant engineering possibility, the social, molecular, and genetic processes behind such transitions are increasingly becoming better understood. For example, a recent study demonstrated that lytic-to-lysogenic viral switching in microbe-rich seawater samples increased significantly with increases in host density [24]. These results support a "piggyback-the-winner" model and challenge the long-held "killing the winner" model of viral switching. In another recent study, viruses utilized small peptides to communicate and coordinate decisions about entering into lysis or lysogeny [58]. Perhaps the best-studied lytic-to-lysogenic "switch" is that of the bacteriovirus λ, which is dependent on host environment and number of infecting viral particles ([59]; recently reviewed in [3]). Moreover, the virion is the crucial distinction between viruses and plasmids, which is sometimes rooted in the presence or absence of a single capsid-encoding gene [60]. Knockout of capsid genes (a new switch?), along with genes that trigger lysis, could theoretically transform viruses into plasmids and vice versa (e.g. [61]). Given the ongoing metagenomics trends towards the discovery of novel viruses in environmental samples, these distant possibilities may become practical sooner than later but will need to be evaluated on a case-by-case basis to avoid possible viral health side-effects.

Conclusions and outlook

Viruses interact with cells directly and indirectly sometimes involving multiple host layers. These interactions include virus-host symbiosis and virus domestication to pursue common objectives (similar to documented examples of bacteria-eukarya partnerships) and lead to interesting evolutionary, ecological, and social consequences for interacting partners. While much has been written about the lytic virus-cell interactions, a better understanding of the beneficial virus-host partnerships holds enormous clinical and medical value as it opens new doors for therapeutic research in microbiology. An extended survey of the virosphere will help populate the triangle of persistence. Technical demands include accurate viral detection in

metagenomic samples surveyed broadly from geographically diverse habitats. This can be problematic because viruses do not encode a universal gene marker such as ribosomal RNA. The solution could be to focus instead on in silico detection of protein folds present in the viral capsid/coat proteins because capsids have been termed the virus "self" [62] and protein folds involved in capsid assembly tend to be remarkably conserved throughout the virosphere [60]. A shift in strategy may therefore improve viral detection and discovery in environmental samples [63]. It is also important to pursue landmark-sampling efforts to complete the human virome especially from infants, frequent travelers, individuals in contact with livestock and poultry, immunocompromised individuals, and from geographically diverse regions. A long-term objective is to understand how and when viruses switch to the endogenous or endosymbiotic mode. Keeping in mind the crucial dependency of viruses on host cells, we hypothesize that this switching could perhaps be the long-term preferred evolutionary route outcome for viruses. Unlocking novel mechanisms of viral endogenization will require clever integration of bioinformatics and wet lab experiments. It will be necessary to map molecular data (e.g. capsid genes) to virus-host partnerships [55], to identify viral genome integration sites in high-throughput sequencing data (e.g. VirusSeq [64] and Virus-Finder [65]), to simulate behavior of viral particle conformational dynamics under varying conditions using atomic scale molecular dynamics (MD) simulations [66], and to trace the evolutionary spread of viral folds in cellular life [55] to better understand viral lifestyles. We hope that our arguments will encourage an updated thinking about the virosphere, increase interest and focus in discovering non-lytic and beneficial virus-cell interactions (see also [1, 2]), and inspire novel microbiological approaches to study viruses and manage viral diseases.

Acknowledgments

Research is supported by the Higher Education Commission Start-up Research Grant Program to AN (Project No. 21-519/SRGP/R&D/HEC/2014), a grant from KOPRI research program (PE17020) to KMK, and grants from the National Science Foundation (OISE-1132791) and United States Department of Agriculture (ILLU-802-909 and ILLU-483-625) and Blue Waters allocation to GCA.

The authors have declared no conflict of interest.

References

1. **Roossinck MJ.** 2011. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol* **9**: 99–108.
2. **Roossinck MJ.** 2015. Move over, Bacteria! Viruses make their mark as mutualistic microbial symbionts. *J Virol* **89**: 6532–5.
3. **Gandon S.** 2016. Why be temperate: lessons from bacteriophage λ. *Trends Microbiol* **24**: 356–65.
4. **Forterre P**, **Prangishvili D.** 2009. The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. *Ann NY Acad Sci* **1178**: 65–77.
5. **tenOever BR.** 2016. The evolution of antiviral defense systems. *Cell Host Microbe* **19**: 142–9.
6. **Elde NC**, **Malik HS.** 2009. The evolutionary conundrum of pathogen mimicry. *Nat Rev Microbiol* **7**: 787–97.
7. **Stern A**, **Sorek R.** 2011. The phage-host arms race: shaping the evolution of microbes. *BioEssays* **33**: 43–51.
8. **Feiner R**, **Argov T**, **Rabinovich L**, **Sigal N**, et al**.** 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. *Nat Rev Microbiol* **13**: 641–50.
9. **Blanco-Melo D**, **Gifford RJ**, **Bieniasz PD**, **Bieniasz P**, et al**.** 2017. Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. *Elife* **6**: 12058–61.
10. **Paez-Espino D**, **Eloe-Fadrosh EA**, **Pavlopoulos GA**, **Thomas AD**, et al**.** 2016. Uncovering earth's virome. *Nature* **536**: 425–30.
11. **Peters DL**, **Lynch KH**, **Stothard P**, **Dennis JJ.** 2015. The isolation and characterization of two *Stenotrophomonas maltophilia* bacteriophages capable of cross-taxonomic order infectivity. *BMC Genomics* **16**: 664.
12. **Virgin HW.** 2014. The virome in mammalian physiology and disease. *Cell* **157**: 142–50.
13. **Krupovic M**, **Dutilh BE**, **Adriaenssens EM**, **Wittmann J**, et al**.** 2016. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. *Arch Virol* **161**: 1095–9.
14. **Barr JJ**, **Auro R**, **Furlan M**, **Whiteson KL**, et al**.** 2013. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci USA* **110**: 10771–6.
15. **Yafremava LS**, **Wielgos M**, **Thomas S**, **Nasir A**, et al**.** 2013. A general framework of persistence strategies for biological systems helps explain domains of life. *Front Genet* **4**: 16.
16. **Thingstad TF**, **Bratbak G.** 2016. Microbial oceanography: viral strategies at sea. *Nature* **531**: 454–5.
17. **Delwart E.** 2013. A roadmap to the human virome. *PLoS Pathog* **9**: e1003146.
18. **Touchon M**, **Bobay L-M**, **Rocha EP.** 2014. The chromosomal accommodation and domestication of mobile genetic elements. *Curr Opin Microbiol* **22**: 22–9.
19. **Brosius J.** 2003. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* **118**: 99–116.
20. **Feschotte C**, **Gilbert C.** 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* **13**: 283–96.
21. **Liu H**, **Fu Y**, **Jiang D**, **G Li**, et al**.** 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *J Virol* **84**: 11876–87.
22. **Cortez D**, **Forterre P**, **Gribaldo S.** 2009. A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. *Genome Biol* **10**: R65.
23. **Daubin V**, **Lerat E**, **Perrière G**, **Sueoka N**, et al**.** 2003. The source of laterally transferred genes in bacterial genomes. *Genome Biol* **4**: R57.
24. **Knowles B**, **Silveira CB**, **Bailey BA**, **Barott K**, et al**.** 2016. Lytic to temperate switching of viral communities. *Nature* **531**: 466–70.
25. **Conrad R**, **Vlassov AV.** 2015. The human microbiota: composition, functions, and therapeutic potential. *Med Sci Rev* **2**: 92–103.
26. **Reyes A**, **Haynes M**, **Hanson N**, **Angly FE**, et al**.** 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**: 334–8.
27. **Paul JH.** 2008. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J* **2**: 579–89.
28. **Valdivia RH**, **Heitman J**, **Dyall SD**, **Brown MT**, et al**.** 2007. Endosymbiosis: the evil within. *Curr Biol* **17**: R408–10.
29. **Partida-Martinez LP**, **Hertweck C.** 2005. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437**: 884–8.
30. **Brüssow H**, **Canchaya C**, **Hardt W-D.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* **68**: 560–602.
31. **Márquez LM**, **Redman RS**, **Rodriguez RJ**, **Roossinck MJ.** 2007. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* **315**: 513–5.
32. **Bordenstein SR**, **Bordenstein SR.** 2016. Eukaryotic association module in phage WO genomes from Wolbachia. *Nat Commun* **7**: 13155.
33. **Strand MR**, **Burke GR.** 2014. Polydnaviruses: nature's genetic engineers. *Annu Rev Virol* **1**: 333–54.
34. **Federici BA**, **Bigot Y.** 2003. Origin and evolution of polydnaviruses by symbiogenesis of insect DNA viruses in endoparasitic wasps. *J Insect Physiol* **49**: 419–32.
35. **Burke GR**, **Strand MR.** 2012. Polydnaviruses of parasitic wasps: domestication of viruses to act as gene delivery vectors. *Insects* **3**: 91–119.
36. **Zilber-Rosenberg I**, **Rosenberg E.** 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiol Rev* **32**: 723–35.
37. **Moran NA**, **Degnan PH**, **Santos SR**, **Dunbar HE**, et al**.** 2005. The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proc Natl Acad Sci USA* **102**: 16919–26.

38. **Wang X**, **Kim Y**, **Ma Q**, **Hong SH**, et al**.** 2010. Cryptic prophages help bacteria cope with adverse environments. *Nat Commun* **1**: 147.
39. **Pradeu T.** 2016. Mutualistic viruses and the heteronomy of life. *Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci* **59**: 80–8.
40. **Nasir A**, **Forterre P**, **Kim KM**, **Caetano-Anollés G.** 2014. The distribution and impact of viral lineages in domains of life. *Front Microbiol* **5**: 194.
41. **Nasir A**, **Sun FJ**, **Kim KM**, **Caetano-Anolles G.** 2015. Untangling the origin of viruses and their impact on cellular evolution. *Ann NY Acad Sci* **1341**: 61–74.
42. **Boussau B**, **Blanquart S**, **Necsulea A**, **Lartillot N** et al**.** 2008. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**: 942–5.
43. **Groussin M**, **Gouy M.** 2011. Adaptation to environmental temperature is a major determinant of molecular evolutionary rates in archaea. *Mol Biol Evol* **28**: 2661–74.
44. **Kim KM**, **Nasir A**, **Hwang K**, **Caetano-Anollés G.** 2014. A tree of cellular life inferred from a genomic census of molecular functions. *J Mol Evol* **79**: 240–62.
45. **Forterre P.** 2013. The common ancestor of archaea and eukarya was not an archaeon. *Archaea* **2013**: 372396.
46. **Prangishvili D.** 2013. The wonderful world of archaeal viruses. *Annu Rev Microbiol* **67**: 565–85.
47. **Forterre P.** 2011. A new fusion hypothesis for the origin of Eukarya: better than previous ones, but probably also wrong. *Res Microbiol* **162**: 77–91.
48. **Koonin EV**, **Dolja VV**, **Krupovic M.** 2015. Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* **479**: 2–5.
49. **Woese C.** 1998. The universal ancestor. *Proc Natl Acad Sci USA* **95**: 6854–9.
50. **Spang A**, **Saw JH**, **Jørgensen SL**, **Zaremba-Niedzwiedzka K**, et al**.** 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**: 173–9.
51. **Rivera MC**, **Lake JA.** 2004. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**: 152–5.
52. **Caetano-Anollés G**, **Nasir A**, **Zhou K**, **Caetano-Anollés D**, et al**.** 2014. Archaea: the first domain of diversified life. *Archaea* **2014**: 590214.
53. **Forterre P.** 2005. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. *Biochimie* **87**: 793–803.
54. **Fox NK**, **Brenner SE**, **Chandonia JM.** 2014. SCOPe: structural classification of proteins-extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res* **42**: D304–9.
55. **Nasir A**, **Caetano-Anollés G.** 2015. A phylogenomic data-driven exploration of viral origins and evolution. *Sci Adv* **1**: e1500527.
56. **Rath D**, **Amlinger L**, **Rath A**, **Lundgren M.** 2015. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* **117**: 119–28.
57. **Nilsson AS.** 2014. Phage therapy-constraints and possibilities. *Ups J Med Sci* **119**: 192–8.
58. **Erez Z**, **Steinberger-Levy I**, **Shamir M**, **Doron S**, et al**.** 2017. Communication between viruses guides lysis-lysogeny decisions. *Nature* **541**: 488–93.
59. **Oppenheim AB**, **Kobiler O**, **Stavans J**, **Court DL**, et al**.** 2005. Switches in bacteriophage lambda development. *Annu Rev Genet* **39**: 409–29.
60. **Abrescia NGA**, **Bamford DH**, **Grimes JM**, **Stuart DI.** 2012. Structure unifies the viral universe. *Annu Rev Biochem* **81**: 795–822.
61. **Krupovic M**, **Ravantti JJ**, **Bamford DH.** 2009. Geminiviruses: a tale of a plasmid becoming a virus. *BMC Evol Biol* **9**: 112.
62. **Abrescia NGA**, **Grimes JM**, **Fry EE**, **Ravantti JJ**, et al**.** 2010. What does it take to make a virus: the concept of the viral "self". In Stockley, PG. Twarock, R. ed; *Emerging Topics in Physical Virology*. London: Imperial College Press. p. 35–58.
63. **Nasir A**, **Caetano-Anollés G.** 2017. Identification of capsid/coat related protein folds and their utility for virus classification. *Front Microbiol* **8**: 380.
64. **Chen Y**, **Yao H**, **Thompson EJ**, **Tannir NM**, et al**.** 2013. VirusSeq: software to identify viruses and their integration sites using next-generation sequencing of human cancer tissue. *Bioinformatics* **29**: 266–7.
65. **Wang Q**, **Jia P**, **Zhao Z**, **Lace M**, et al**.** 2013. VirusFinder: software for efficient and accurate detection of viruses and their integration sites in host genomes through next generation sequencing data. *PLoS ONE* **8**: e64465.
66. **Ode H**, **Nakashima M**, **Kitamura S**, **Sugiura W**, et al**.** 2012. Molecular dynamics simulation in virus research. *Front Microbiol* **3**: 258.
67. **Sheftel H**, **Shoval O**, **Mayo A**, **Alon U.** 2013. The geometry of the Pareto front in biological phenotype space. *Ecol Evol* **3**: 1471–83.
68. **Garoff H**, **Hewson R**, **Opstelten DJ.** 1998. Virus maturation by budding. *Microbiol Mol Biol Rev* **62**: 1171–90.
69. **Snyder JC**, **Brumfield SK**, **Kerchner KM**, **Quax TEF**, et al**.** 2013. Insights into a viral lytic pathway from an archaeal virus-host system. *J Virol* **87**: 2186–92.