Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Persistent virus and addiction modules: an engine of symbiosis

Luis P Villarreal

The giant DNA viruses are highly prevalent and have a particular affinity for the lytic infection of unicellular eukaryotic host. The giant viruses can also be infected by inhibitory virophage which can provide lysis protection to their host. The combined protective and destructive action of such viruses can define a general model (PD) of virus-mediated host survival. Here, I present a general model for role such viruses play in the evolution of host symbiosis. By considering how virus mixtures can participate in addiction modules, I provide a functional explanation for persistence of virus derived genetic 'junk' in their host genomic habitats.

Address
Center for Virus Research, University of California, Irvine, Irvine, CA 926197, USA

Corresponding author: Villarreal, Luis P (lpvillar@uci.edu)

Current Opinion in Microbiology 2016, **31**:70–79

This review comes from a themed issue on **Megaviromes**

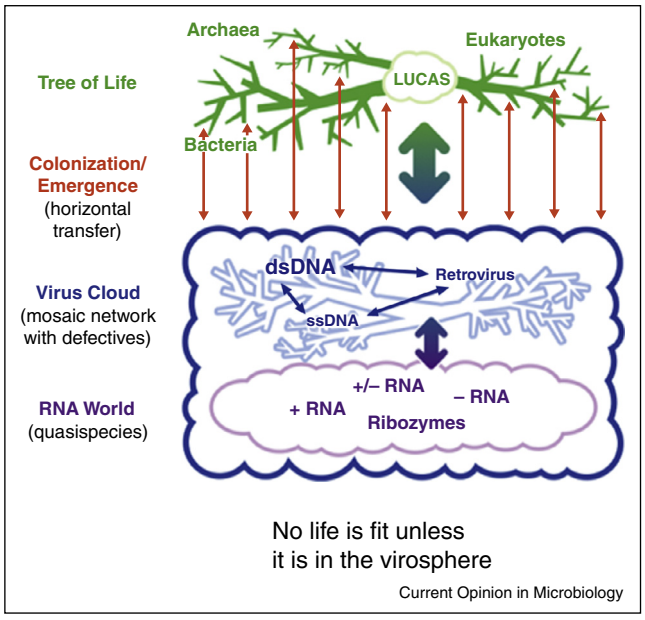
Edited by **Didier Raoult** and **Jonatas Abrahao**

http://dx.doi.org/10.1016/j.mib.2016.03.005

The historic 'virus-free' concepts of evolution

For many decades, there were arguments in the biological literature regarding the relative importance of selfish behaviors versus symbiotic behaviors for evolution [1–3]. In the 1960s, however, with the introduction of kin selection and later game theory, it appeared that essentially selfish (individual based) strategies could result in and explain the evolution of cooperative and even altruistic behaviors [4]. Yet in the 1970s the fundamental importance of symbiosis was made clear by its success at explaining the evolutionary origin of the mitochondria and chloroplast via symbiosis of two previously distinct cellular lineages [5,6]. Historically, viruses were not ever part of this discussion [7]. Indeed, viruses appeared to be the ultimate selfish agents whose capacity to kill their host resembled a predatory–prey relationship [8]. And when it was observed that virus derived genetic information had become incorporated into host genomes, this was explained by using war like metaphors resulting from 'arms races' in which following a virus 'plague sweep' the host would occasionally survive but still retain a bit of the selfish virus DNA. Thus although parasitic selfish (virus-like) information is common in the genomes of all life forms, its presence was explained as mostly defective remnants of past plague sweeps that provides no functional benefit to the host (e.g. junk). Until recently, this explanation seemed satisfactory. In the last twenty years, however, various observation-based developments have compelled us to re-evaluate this stance. Both comparative genomics and metagenomics (sequencing habitats) has made it clear that viral sequences constitute the most numerous of all genetic DNA sequences in both the various habitats that have been measured as well as within the genomes of most cellular DNA. Indeed, we can consider the microbial genomes as also composed of collections of parasitic agents that did not descend from a common ancestor [9]. Thus we have come to accept the existence of a vast 'virosphere': a vast cloud-like population of viral genomes that shows considerable exchange with other viruses and host as shown in Figure 1 [7]. The rampant occurrence of horizontal gene transfer (especially in prokaryotes) seems to have mostly resulted via the action of viruses and other related genetic parasites [10]. More recently, we have become aware that there also exists an entire putative domain of eukaryotic viruses that is much larger and more complex then previously imagined. These are the giant viruses like Mimivirus and Pandoravirus of ameobazoa, which seem to have only a lytic life cycle [11]. How these viruses might have affected host evolution is not yet clear. In addition, it is becoming increasingly clear that gene regulation in eukaryotes involves various types of non-coding regulatory RNA. Indeed, it now appears that regulatory complexity (not gene numbers or gene complexity) accounts for the much more complex multicellular biology of eukaryotes compared to prokaryotes and that the regulatory RNA that is involved in this originates mostly from parasitic (junk) DNA sequences [12,13]. In this article, I consider a different perspective to understand the virus–host relationship: the fundamental evolutionary consequence of persisting non-lytic virus infections of the host. This includes genomic, epigenomic and 'defective' virus persistence. According to this view, such virus derived information is not junk, but has provided a salvation pathway for the host lineage to survive in its virosphere. Such persistence requires an intimate virus–host molecular relationship. To understand persistence mechanisms, I consider the strategy of addiction modules (involving both destruction and protection) as a general approach to understand what binds virus to persist in host and also as a

Figure 1



Diagrammatic representation of the relationship of the Tree of Life (green dendrogram) to the Virosphere (blue cloud). The blue dendrogram within the viral cloud represents species specific persisting viruses.
Reproduced from [7] with permission.

general strategy to bind any two lineages of life and promote symbiosis [14].

Viruses as competent editors of code

The DNA genome has been considered a linear language or code. Its evolution is accepted to occur mostly via genetic errors that generate diversity for natural selection to operate on for the natural selection of individuals. However, if DNA is indeed an authentic code, it will need to also address the concepts of language theory. In particular, it has been argued that editing a 'real' language or code cannot emerge via errors and must involve populations of 'editors' (competent users of code) [15]. This abstract concept does not initially appears to make sense in the context of modern molecular genetics as the needed population of editors seems not to exist. However, viruses could provide the populations of such competent editors. In prokaryotes, the results of comparative genomic seem most consistent with heavy virus involvement in host evolution, mainly involving horizontal gene transfer [10]. In addition, gene regulation usually seems to involve regulatory networks. By definition, networks are reticulated and usually involve complex positive and negative interactions between network participants (members). Such complex regulatory networks are especially applicable to eukaryotes. However, networks originating from error-based variation in individuals poses numerous problems as networks are fundamentally reticulated and do not adhere to tree based (graphic) analysis. Since viruses can operate as diffuse populations, they might also promote the establishment and editing of networks en mass. This is especially evident with RNA (and retro) viruses of eukaryotes. RNA (and retro) viruses in particular operate via quasispecies which are coherent RNA populations [16] able to also colonize host DNA as provirus. Since viruses are inherently competent in all forms of host genetic and epigenetic code, it has been suggested that they are the main editors of DNA code [17]. In terms of the human genome, there are 330 000 solo LTRs (retroviral long terminal repeats) that mostly have originated from full retrovirus integration [18]. Thus during our evolution, about 3.3 Gb of DNA bps (equal to our entire genome) was once retrovirus sequence that underwent editing (recombination based deletion) to generate these solo LTRs. And similar LTRs are now providing complex gene network regulation (specifying multicellular identity), such as in the placenta [19,20] and for primate p53 regulation [21]. Why would virus be involved (become symbiotic) in this way? The prevailing view has been that viruses have simply provided a diverse source of new genes (and regulators) to be 'exapted' by the host for host evolution. However, if we instead consider what might compel a virus to establish a symbiotic state with host and install a new persistent regulation of itself and its host we can propose an alternative view. The virus has 'addicted' the host to its presence and created a new virus–host entity that is more successful in the virosphere.

History of addiction module

The existence of addiction modules was first reported in the early 1990s by Yarmalinsky and colleagues at NIH [22,23]. As they sought to understand how the P1 virus can stably infect its host as an episome and why host cell death occurs when the P1 plasmid is lost, they discovered a strategy in which P1 encodes stable toxins as well as a less stable but matching antitoxins. Thus a counter acting toxin/antitoxin (TA) gene pair promotes the retention of P1 for host survival. Fundamentally, this strategy can allow the stable linkage of two previously distinct lineages of life (virus and host). The P1 TA strategy is thus used as an exemplar to account for how two (or more) genetic lineages can be merged into one. Thus, the infected *Escherichia coli* and the P1 phage (an epigenomic plasmid) now act as (have become) one entity and will also work together to oppose other genetic parasites, such as T4 and lambda, lytic virus prevalent in the virosphere [24]. As survival in the virosphere is proposed to be a basic necessity of all life, the stable persistence of virus itself can provide a 'virus addiction' by resisting other similar and sometimes different lytic viruses. This is a generalized 'virus addiction' module mediated by toxic (T) virus lysis and counter-acted (A) by virus persistence [25]. This 'virus addiction' is shown diagrammatically in Figure 2 for two populations of *E. coli* (P1 persistent and P1 free). Virus addiction thus provides a basic mechanism for