Title
Persistent virus and addiction modules: an engine of symbiosis

Author
Villarreal, Luis P

Publication Date
2016-06-01

DOI
10.1016/j.mib.2016.03.005

Peer reviewed

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**

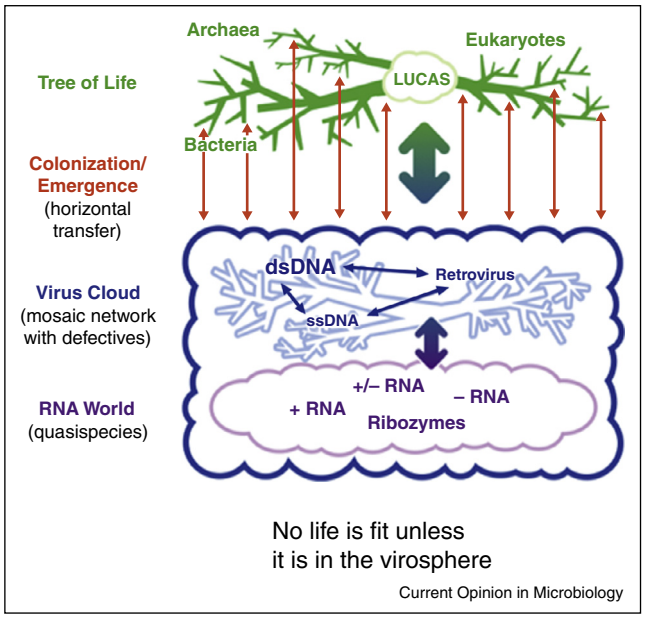
Edited by **Didier Raoult** and **Jonatas Abrahao**

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].

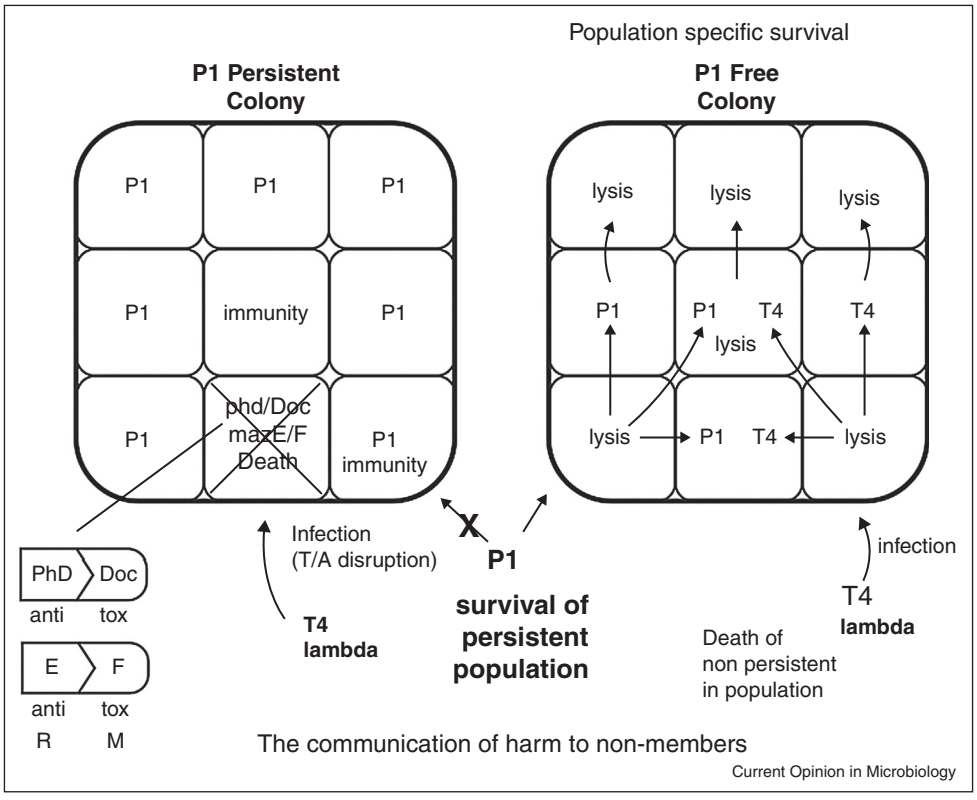
Viruses as competent editors of code

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

Figure 2



Virus addiction module. Two populations (colonies) of host cells are shown (P1 phage persistent and P1 free). The antitoxin and toxin gene sets are shown in the lower left for the P1 persistent host (PhD/DOC and restriction/modification) which induce self destruction upon lytic virus infection, preventing virus transmission. The external T4, lambda and resident P1 phage are all capable of host lysis in unprotected populations. Modified and reproduced from [25] with permission.

survival in a virus rich habitat. But this process is both inherently symbiotic and cooperative. Yet it can also be destructive and seemingly selfish as the capacity for harm (lysis) must always be retained (similar to prevent programmed cell death [26]). Since both the protective (A) and destructive (T) features must operate together to establish addiction (persistence), this state does not emerge from the fittest-type individual gene since the two functions must always be paired. If then, the new symbiotic virus–host combination should lose the viral component (via curing or reproductive loss), death of the host cell results. This is programmed cell death upon 'curing'. If we accept that virosphere survival is indeed fundamental, programmed cell death along with the loss of persisting virus can be understood via addiction module disruption. However, traditional (virus-free) concepts of evolutionary biology (e.g. neodarwinian ideas) would call this programmed cell death (especially in a unicellular, clonal organism) a spiteful behavior, which is difficult to understand [27]. Most generally, TA gene sets and virus addiction modules provide mechanisms to cooperatively link genomes that were from distinct ancestral lineages. This strategy can now allow us to reevaluate historic ideas that previously proposed to explain cooperative behaviors with kin selection and game theory.

Lytic and defective virus in generalized virus addiction

Although some viruses (and plasmids) can often encode toxins (and TA gene sets), and many encode pore proteins involved in host lysis, such gene encoded TA strategies are not typical for a large number of viruses (including most viruses of eukaryotes). Yet persisting viruses (genomic and epigenomic, often defective) are common in all domains of life. Clearly, the P1 style (gene based TA) addiction modules cannot explain such extensive virus persistence. However, since lytic infection by virus can also provide the 'death' function needed for an addiction module, protecting a persistently infected cell from lysis need not be restricted to a protein function. In this case, persistence by the same (or partial) virus can provide the protective (immunity) function needed to complete an addiction module. Thus the combination of viral lysis and cell death as well as persistence and cell protection can together provide a virus based addiction module that compels the host to retain persistence of virus for survival

in the virosphere. In my book, I called the generalized version of an addiction module 'virus addiction' [25]. This situation is applicable to all domains of life (even in the pre-DNA RNA world [28]). Virus addiction will thus require strategies to protect the host from virus (A) along with strategies that will kill the host if sufficient virus information integrity is lost (T). This is programmed cell death via virus, not toxins. Thus, for example, we might accordingly reconsider the function of the lambda *rex* gene. This iconic and puzzling gene seems to provide an altruistic function by killing the host cell upon infection by other viruses [29]. We can instead, however, propose that it as an essential function to allow the virus–host combination to survive in its habitat that is rich with lytic virus. Thus the acquisition of persistent virus (and corresponding TA strategies) by the host cell will typically enhance virosphere survival and promote new symbiotic virus–host identity (and new immunity). In this light, any new TA set will need to be coherent (in coordination) with prior TAs sets present in the host. Thus restriction/modification genes, pore/antipore proteins, toxin/antitoxin proteins all of which are TA sets and acquired via horizontal mechanisms, must become consistent with the prior network of addiction modules. For example, successful P1 colonization of *E. coli* will require proper interaction of P1 addiction modules with the resident *E. coli* mazEF, a P1 resisting TA set, which indeed seems to be the case [30]. Thus, successful colonization with a new persisting virus requires TA network coherence and promotes network modification (via expansion and/or diminution). Editing preexisting regulatory networks becomes a dominant issue for successful virus colonization.

Can regulatory RNA mediate virus and host persistence?

An assertion that follows from the above reasoning is that most all programmed cell death systems (which appear to be altruistic), will also be involved with identity/immunity, and stress response, but such complex cellular identity will have mostly originated from persisting virus addiction systems. However, these identity programs need not be restricted to proteins based genes, but will often involve functional non-coding regulatory RNAs, especially RNAs that have stem-loop interactive features [31,32]. Indeed, the *E. coli* PAR addiction module which expresses intragenic stem-loop RNAs that can regulate translation might be the best model for understanding how non-coding RNA can regulate addiction [31]. Many related loci that also use RNA are known [33]. Also, non-coding antisense RNA is used as a component of widely distributed abortive infection systems (Abi) of bacteria that counteracts a wide array of virus [34]. In some cases, the regulatory (antitoxic) RNA can be a pseudoknot [35]. Pseudoknots are of special interest in that they interact both in cis and trans with other RNA stem-loop regions to provide context dependent functions. Clearly, the CRISPR system of prokaryotes also uses virus-derived non-coding stem-loop RNAs to target endonucleases and counteract viruses, showing clear functional (but no evolutionary) similarity to eukaryotic RNAi systems [36]. Thus, in prokaryotes, although protein based (TA's, RM's which are restriction-modification systems) antiviral systems are most common, non-coding RNA based systems are also prevalent. However, in eukaryotes there occurred a big shift in antiviral systems as well as a big shift in the nature of viruses that persist in their host. RM systems are essentially absent, for example, nor was the CRISPR system retained. Instead, in eukaryotes we see the emergence of various small (and large) non-coding RNA based systems, such as RNAi, siRNA and microRNA, all of which use stem-loop non-coding RNAs to guide their corresponding antiviral [37,38] and regulatory responses, see [39]. Later, with the emergence of jawed vertebrates, we see that the antiviral role of the RNAi system was mostly displaced by the interferon response (which also responds to dsRNA and induces apoptosis). The emergence of the interferon system is directly linked to and regulates the adaptive immune response of vertebrates, leading to another big shift in virus–host relationship. Yet, interestingly, micro RNA regulation was retained and appears to underlie control of both innate [40–42] and adaptive immunity [43–45]. From the perspective of addiction modules, the entire adaptive and innate immune system can be considered as one very complex addiction module that has tremendous capacity for cellular destruction but is kept in check by various protective mechanisms (such as clonal elimination), but likely originating from virus [46,47].

A large and general shift has occurred in viruses of eukaryotes regarding the type and amount of virus information that becomes 'one with' (integrates into) the host genome. With emergence of a nucleus that separates RNA synthesis and processing from translation, we see major changes in virus–host interactions, especially what type of virus nucleic acid persists as part of host DNA. In prokaryotes, DNA acquired from large dsDNA viruses (and plasmids) predominate whereas in eukaryotes, DNA derived from retroviruses, retroposons and rolling circular (rcr)-DNA viruses and transposons dominate. Also in eukaryotes, we see only modest increases in numbers of gene (ORFs) but a substantial increase in DNA involved in gene regulation. Along with this basic pattern shift, the numbers of these parasitic viruses and agents increases enormously in eukaryotes. Since many of these parasitic agents can express non-coding RNA [48], this represents a big shift in regulatory strategy. Eukaryotic regulatory RNAs can affect numerous events in RNA processing, transport, stability and translation. And viruses can be involved in, bypass or regulate all of these regulatory events. With persisting eukaryotic viruses, it is thus important to understand if regulatory non-coding RNA is also involved in virus persistence or if such RNA

can participate in addiction modules. Indeed, it has recently become clear that small regulatory RNA's are most often involved in the persistence of many, if not all eukaryotic DNA and retro viruses [49,50]. A main target of such non-coding RNAs appears to be to regulate the innate immune response, especially with respect to the control of apoptosis. The question at hand then is if the giant DNA viruses ever persist with their host. And what, if any, role does the integration of viral DNA play? Is persistence ever attained and can it involve distinct virus–host strategies? Is noncoding RNA based regulation involved in their virus–host biology? And are there additional virus–virus and virus–host relationships present in their virosphere and habitat participating in virus addiction modules?

Giant virus and the addiction module perspective

Since *Acanthamoeba polyphaga mimivirus* was discovered various other giant viruses have also been reported (*Marseilleviridae*, *Pithovirus sibericum*, *Cafeteria roenbergensis virus* and *Pandoravirus*). These giant virus families (Mimivirus, Megavirus, Pandoravirus) plus the less giant phycodnavirus have some clear similarities, besides their large size. They mostly infect unicellular eukaryotes in aquatic habitat, although some multicellular hosts are also known (such as brown algae). In addition, they mostly have lytic and non-integrating life cycle that resembles the lytic T4 like phage of bacteria. Persistent infections have not been reported. Thus from the perspective of virus mediated addiction modules, we might wonder if these giant viruses could be providing lytic half of a more extended relationship (addiction module) in which persistence in another host is preventing viral lysis, such as a lytic host being harbored in other symbiotic cells. Or can these giant viruses persist by other strategies, such as extreme chemical stability, or via the possible involvement of another inhibitory virus? Mimivirus mostly infects unicellular protist (ameobazoa) as does Pandoravirus [51,52]. Both of these hosts can be symbiotic in or with other species. The virus–host relationship is predominantly lytic, which does not seem to offer any explanations as to why these viruses need such giant genomes to infect relatively simple host. Interestingly, some amoebas (polychoas dubium) do have extremely large genomes (670 Gb), up to 100× that found in humans [53]. Yet many are intracellular parasites [54], which is usually associated with genome reduction. Further, some free living amoebas may have smaller genomes [55]. The giant viruses represent diverse morphology, genetic composition and replication programs with little in common to the extended set of large DNA viruses. Their genomes have a mosaic character with relatively little ORF similarity to the other families of DNA viruses or host genes but with clear similarity to other members of same family of virus [56] or other genes within the same genome [57]. So why are they so big? With passage in monoculture, many viral genes are lost (especially those that modify proteins) [58]. This suggests that these lost genes are needed for survival in a more complex habitat (or virosphere). Along these lines, one such 'reduced' strain was bald (lacked fibers) but did not support the sputnik virus (see below). So virus–virus interactions seem clearly relevant. This occurrence of many genes that are not needed for passage in culture is also similar to T4 phage. In the T4 case, many of the lost genes seem to interact with other viruses. One feature of the giant viruses that might require many genes is the need to create viral factories. Giant viruses undergo cytoplasmic replication via factories that accumulate and cap mRNA and assemble virus [59,60]. Thus, these factories are quite complex and provide many of the functions associated with the host nucleus. This could explain why some giant viruses appear to express up to 700 early genes. Clearly, such large viral genomes can compel major genetic reprogramming of their host cells. Still, it remains to be established why these viral genomes are so large. Being a giant virus, however, may come with some hazards. Since they reside in water habitats, they can also be food for filter feeders, such as oysters. Indeed, oysters provide excellent source for isolating Mimivirus [61].

Virus–virus and host interactions

It is interesting that within the cytoplasmic factories of Mimivirus, we can sometimes also find Sputnik virus (a virophage) which is an inhibitory viral parasite of Mimivirus [62]. Sputnik virus is an 18 kb integrating dsDNA virus that replicates as a rolling circle (rcrDNA) via a viral encoded protein primed DNA pol B. A similar parasite of phycodnavirus (Mavirus) also seems able to affect virus–host dynamics by protecting host from phycodnavirus lysis [63]. Interestingly, this Mavirus also encodes an integrase with clear similarity to those of retroviruses. Due to these viral genes, the Mavirus genome shows clear similarity to the Maverick/Politon family of 40 kb DNA transposons, which are highly prevalent in the genomes of various Eukaryotes [64]. Since the replication origins of rcr DNA viruses have clear stem-loop inverted repeat DNA structures, the presence of such elements in host genomes provides a clear capacity for the potential expression of stem-loop regulatory RNAs [65]. Clearly, if these putative stem-loop noncoding RNAs were indeed to be expressed, they would have major consequences to the ability of a host to support similar rcrDNA viruses. They might also provide parasitic (inhibitory) functions against the giant DNA viruses. Polintons are wide-spread in protists, such as single-celled entamoeba, trichomonas, but show patchy distribution in non-animal species and are absent in plants [66,67]. Unicellular amoeba (and hydra) have the basal version of the Politon DNA pol B sequence that is absent in plants. Since transposons, even if derived initially from a virus, are almost always defective for infectious virus function (or transposition), it is usually accepted that they are 'inactive' or dead. But if

they can still provide regulatory RNA, they have a major potential to regulate the virus–host dynamic state, especially in habitats rich in giant viruses and virophage. Thus, it remains possible that the giant viruses might still be providing the lytic function of a virus addiction module involving other viruses and various (symbiotic) hosts.

Integrated virophage (and related transposons)

Giant viruses typically infect unicellular host, but in such host there is no giant viral DNA integration or episomal DNA persistence (or small RNA regulation) described so far. The relationship appears to be strictly lytic. And even if virophage is present and could modify the Mimivirus infection outcome by preventing Mimivirus production, the virophage-giant virus infected host still dies. With virophage, however, broad based DNA integration is much more possible and likely. Here we might expect the existence of populations of host with silent integrated virophage. In a sense, such integrated hosts are immune to the corresponding giant virus since if they become infected, they simply produce more virophage and die (like programmed cell death) and do not produce more giant virus and also do not transmit virophage to other members of the virophage persisting host population. This situation is clearly very similar to how the P1 infected *E. coli* (described above, Figure 2) will induce programmed cell death upon infection with other lytic viruses, such as T4. Here, we can propose that giant viruses are providing the lytic component of virus addiction. Interestingly, giant viruses have related versions of intienes (protein splicing) found in both phycodnavirus [68] and Mimivirus DNA pol ORF [69]. These self splicing elements are functionally similar to T4 phage homing endonuclease also found in the DNA pol ORF which uses stem-loop RNA cleavage endonuclease sites for exclusion of other phage [70].

Thus it seems most plausible that giant virus mediated addiction and survival is occurring via the persistent infection with virophage. The host persistently colonized by virophage survives the ubiquitous presence of lytic giant virus. Given this scenario, it is worth considering how integrated relatives of their rcrDNA viruses (Politons), which are observed in large numbers and diverse protist [71], might also affect the outcome of lytic infection. We cannot currently answer this question as experimental evaluation is lacking. However, we can hypothesize major consequences to virus–host dynamics due to the presence of such defective and potentially interfering viral code.

Integration of giant virus DNA

Although most giant viruses do not integrate into host DNA, there are some very interesting exceptions to this. For example, Phycodnaviruses that infect multicellular brown algae do integrate their DNA. Ectocarpus siliculosus shows the world wide occurrence of the endogenized virus; EsV-1 [72]. Feldmania brown algal species, such as *Emiliania Huxley*, are similar. With EsV-1, the DNA integrates as one copy and establishes a lifelong persistent infection in its host. In addition, there are distinct subgenomic loci (fragments) of EsV-1 related viral DNA integrated at various sites in the host [72]. These loci appear able to express viral encoded endonucleases and transposase. The Feldmania virus is closely associated with host reproductive biology. The virus infects free swimming wall-less gametes and establishes a genome endogenization. The virus becomes latent in the vegetative cells of the host but is expressed in reproductive tissue. Although gametes show viral pathology, no pathology is seen in latently infected algae. The genomes of these viruses, however, have some interesting distinctions with those of lytic Phycodnaviruses and giant viruses. They are much less gene-dense and have a significant amount of inverted repeated and non-coding sequences. In addition, no introns are found. This is also similar to the Pithovirus genome which also has a large fraction of noncoding repeated sequence with a 150 bp palindrom that shows clear similarity to mobile elements [73]. The functional significant of these repeat sequences is not yet clear. It remains unknown if they encode ncRNA or participate in persistence. However, it remains clear that multicellular brown algae have a distinct and persistent relationship with their giant DNA viruses. And it seems most likely (but not yet established) that the presence of virus related genetic code in these species is of relevance to this persistent relationship.

Reproduction and programming a cryptic life style

Unicellular green algae and amoeba are often free swimming (or symbiotic in other species) and do not differentiate cell types. Most reproduce by cell fission. Multicellular brown algae do differentiate in cell types but also reproduce sexually by producing free swimming gametes. But more complex eukaryotes (especially plants) also differentiate cell types and can produce cysts, spores or seeds for sexual reproduction. Amoeba can also produce cysts. Cysts, spores and seeds are all examples that represent a very complex changes in genetic programs that result in the production of a stable, cryptic (inactive) cell that can tolerate environmental stress but preserve the capacity to generate new life. The networks needed for such complex programs must first be acquired for such reproductive biology to emerge. For example, cyst-trophozoite differentiation of *Acanthamoeba castellanii* involves massive turnover and remodeling of cellular components so as to produce a cyst, a cryptobiotic cell, resistant to desiccation, heat, freezing, and chemical treatments. Complex regulation must certainly be involved. Such cysts are also resistant to Mimivirus infection. Virus resistance could therefore provide additional selection for their emergence of cryptic life programs.

Thus it is very interesting that when Mimivirus infects this amoeba, it lytically produces a giant virus that is extremely stable, but is cryptic in that is inert but it has preserved the capacity for subsequent virus life. This clearly resembles a viral cyst and identifies a viral life strategy in which persistence in the environment may be crucial. Interestingly, Mimivirus not only superimposes a new genetic program for producing this stable 'cyst-like' virus, it also prevent the cyst production of its host amoeba [74]. In a sense, the virus has displaced the host program with a new viral programming for a cryptic state of viral life. Neither virions or the cellular cysts, spores or seeds can be considered as living states, but all are complex latent or cryptic states that preserve the capacity for life. The persistence of this capacity for life is key to survival of the organism. That viruses are able to program such states onto the host can also suggest a viral origin for such a cellular life strategy.

Integrated virus DNA, addictive symbiosis and plant evolution

Free living amoeba are an especially common host for giant viruses. As noted above, some amoeba have extremely large genomes. Although the composition of amoeba genomes are poorly studied, it seems certain that such large genomes cannot simply code for genes (ORFs). They must have a large amount of repeated and likely parasitic DNA. It is known that politon transposons (related to polB and integrase of Mavirus) are present in the genomes of some amoeba [71]. And it is interesting that the polB sequence of this politon is phylogenetically basal to that version found in aquatic animals. In contrast, no multicellular green plants or land animals have retained these politons. As argued above, we can expect large effects on virus–host relationship due to the presence of virus related sequences. Thus it is very interesting that cnidarians and lower plants have acquired some Mimivirus-like sequences into their genomes [75]. Both a moss species and hydra (cnidarian) have acquires at least 23 core genes from a Mimi-like virus. Yet none of these species are known to support lytic giant viruses. Moss (a bryophyte) are early haploid spore forming avascular plants, that emerged prior to flowering (seed forming) plants. Bryophytes and lycophytes (early spore forming vascular plant) have up to 16 ORFs from genes of large DNA viruses, which is a subset of the core replication virus specific genes [76]. The consequence to the virus–host dynamic of these integrated viral sequences is currently unknown. However, I can predict that such virus derived DNA will counteract related lytic virus and alter host susceptibility (via virus addiction). Such cells may be immune and now be capable of supporting some type of persistent presence of the giant viruses. How this integration occurred is also interesting, since the giant viruses are non-integrating. Possibly a retrovirus (like the ancient chromovirus) could have helped by providing integration functions as they are found as ERVs in the genomes of early eukaryotes. It is most interesting that with the emergence of early multicellular green plants, we see another big shift in virus–host relationship. In higher plants, absent are the giant DNA viruses of unicellular eukaryotes or the large dsDNA viruses of bacteria or green or brown algae. Indeed, few viruses have been reported for these simple nonseed producing multicellular plants like moss. Although some RNA viruses (dsRNA and +RNA) are known, these are nowhere near the large diversity of +RNA viruses found in seed producing and flowering plants. Nor are ds DNA viruses found in higher plants. However, pararetroviral derived DNA is highly prevalent in higher plant genomes. Small non-coding RNA's seem crucial for viral control in eukaryotes (including plants). However their activity in moss-like plants is not well studied. In Cnidarians (like hydra), however, small regulatory RNAs resemble both miRNA and siRNA but can regulate their target RNAs by RNA cleavage (a TA like function) [77]. It is unknown if these RNAs might regulate persistent virus infections.

Giant virus as mediators of multicellular symbiosis?

Symbiotic relationships seem highly prevalent in lower eukaryotes (in both plant-like and animal-like lineages). Indeed, the origin of unicellular green algae is thought to have occurred via symbiosis with photosynthetic cyanobacteria. These organisms also often support prevalent large lytic DNA viruses. I have argued that virus addiction modules provide a strategy that can compel a symbiotic state between two previously distinct lineages of organisms (virus and host). A resident (often defective) virus can protect the host against prevalent lytic viruses in the virosphere. But this strategy can also be extended to include multiple cellular host as well as multiple viruses. Consider, for example, a host cell like hydra. With integrated sequences from both giant viruses and/or virophage, it can have genetic resistance to these lytic viruses. However, if it is colonized internally by another cell type (for example a symbiotic photosynthetic unicellular green algae) the hydra can provide a protected hiding place, such as chlorella species protection against chlorella virus. The hydra then benefits from the photosynthetic activity of the chlorella cell and the chlorella cell has found a safe haven from chlorella virus. This would be an example of a multi species virus addiction module. This state can provide exemplar of how lytic virus can play a key role in symbiosis by promoting interspecies virus addiction.

Amoebas can symbiotically harbor various other prokaryotic and eukaryotic organisms. What is the virus story behind these relationships? Does the extremely large amoeba genome with lots of parasitic DNA affect this capacity to support symbiosis? We do not know the answer to such questions. We do know, that amoeba can protect Mimivirus from UV inactivation and desiccation [78]. And preclusion between Mimivirus and symbiotic bacteria has

been observed with passage of the host amoeba [79]. However, with the addition of Sputnik virus as a parasite of Mimivirus, the symbiotic bacteria is retained. Clearly there are complex interactions between symbiotic cellular organisms, lytic virus and viral parasites. But we can further add to this mix additional cellular symbiosis and virus. For example, what cell can protect the amoeba host from lytic Mimivirus? A Hydra? As presented above, Hydra may be inherently resistant to these viruses and thus provide a protective habitat for symbiosis of algae and amoebas. Symbiosis is prevalent in cnidarians (e.g. coral). But when we add virus addiction as providing the 'glue' that compels interspecies symbiosis, we see a Russian doll like situation involving multiple, nested host species. This is a multi-species network that is enforced and maintained via virus. Amoeba and chlorella can also be symbiotic in hydra. Thus the phycodnaviruses and giant lytic virus would be essential to preserve and enforce these multicellular symbiotic states. This, therefore, may be an example of an extended viral addiction network. And the overall fitness of the cellular and viral participants is dependent not simply on individual (selfish) fitness, but how persistence of the virus–host addiction module is maintained. And for this, both the capacity for lysis and resistance to lysis (PD) must be maintained under appropriate conditions. This can provide an extended network of addiction strategies involving mixed virus–host lineages with enhanced complexity via virus TA selective pressure. Could the giant viruses have thus promoted emergence of symbiosis and help explain the origin of complex, mosaic multicellular organisms? And could retroviruses similarly promoted the emergence of complex innate immunity [80••] and placental [81•] networks in vertebrates? I suggest the answer is yes and if this is indeed the case, we might think '*Ex Virus Omnia*' (from virus everything).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Margulis L, Nealson KH: **Symbiosis as the source of evolutionary innovation**. *Endocytobiosis Cell Res* 1989, **6**: 235-239.
2. Wilson DS: **What is wrong with absolute individual fitness?** *Trends Ecol Evol* 2004, **19**:245-248.
3. Dawkins R: *The selfish gene [Internet]*. Oxford University Press; 2006.
4. Smith JM, Wynneedwards VC: **Group selection + kin selection**. *Nature* 1964, **201**:1145-1147.
5. Margulis L: **Origins of species — acquired genomes and individuality**. *Biosystems* 1993, **31**:121-125.
6. Sagan D, Margulis L: **Evolution means cooperation, not just competition**. *Scientist* 1986, **1**:10.
7. Villarreal LP, Witzany G: **Viruses are essential agents within the roots and stem of the tree of life**. *J Theor Biol* 2010, **262**:698-710.
8. Nowak MA, May RM: *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford University Press; 2000.
9. Merhej V, Raoult D: **Rhizome of life, catastrophes, sequence exchanges, gene creations, and giant viruses: how microbial genomics challenges Darwin [Internet]**. *Front Cell Infect Microbiol* 2012:2.
10. Koonin EV: *The Logic of Chance: The Nature and Origin of Biological Evolution*. FT Press; 2011.
11. Raoult D, Forterre P: **Redefining viruses: lessons from Mimivirus**. *Nat Rev Microbiol* 2008, **6**:315-319.
12. Mattick JS: **The central role of RNA in the genetic programming of complex organisms**. *Acad Bras Cienc* 2010, **82**:933-939.
13. Mattick JS: **Rocking the foundations of molecular genetics**. *Proc Natl Acad Sci U S A* 2012, **109**:16400-16401.
14. Villarreal LP: **The addiction module as a social force**. In *Viruses: Essential Agents of Life*. Edited by Witzany G. Dordrecht: Springer Science + Business Media; 2012.
15. Witzany G: *Biocommunication and Natural Genome Editing*. Springer; 2011.
16. Domingo E, Sheldon J, Perales C: **Viral quasispecies evolution**. *Microbiol Mol Biol Rev* 2012, **76**:159-216.
17. Witzany G: **Natural genome-editing competences of viruses**. *Acta Biotheor* 2006, **54**:235-253.
18. Moelling K: **What contemporary viruses tell us about evolution: a personal view**. *Arch Virol* 2013, **158**:1833-1848.
19. Chuong EB, Rumi MAK, Soares MJ, Baker JC: **Endogenous retroviruses function as species-specific enhancer elements in the placenta [Internet]**. *Nat Genet* 2013 http://dx.doi.org/10.1038/ng.2553.
20. Emera D, Wagner GP: **Transposable element recruitments in the mammalian placenta: impacts and mechanisms**. *Brief Funct Genomics* 2012, **11**:267-276.
21. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D: **Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53**. *Proc Natl Acad Sci U S A* 2007, **104**:18613-18618.
22. Lehnherr H, Maguin E, Jafri S, Yarmolinsky MB: **Plasmid addiction genes of bacteriophage P1: doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained**. *J Mol Biol* 1993, **233**:414-428.
23. Lehnherr H, Yarmolinsky MB: **Addiction protein Phd of plasmid prophage P1 is a substrate of the ClpXP serine protease of *Escherichia coli***. *Proc Natl Acad Sci U S A* 1995, **92**:3274-3277.
24. Cooper TF, Heinemann JA: **Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids**. *Proc Natl Acad Sci U S A* 2000, **97**:12643-12648.
25. Villarreal LP: *Viruses and the Evolution of Life [Internet]*. ASM Press; 2005.
26. Engelberg-Kulka H, Glaser G: **Addiction modules and programmed cell death and antideath in bacterial cultures**. *Annu Rev Microbiol* 1999, **53**:43-70.
27. Hamilton WD: **Selfish and spiteful behaviour in an evolutionary model**. *Nature* 1970, **228**:1218-1220.
28. Villarreal LP: **Force for ancient and recent life: viral and stem-loop RNA consortia promote life**. *Ann N Y Acad Sci* 2014 http://dx.doi.org/10.1111/nyas.12565.
29. Hayes S, Bull HJ, Tulloch J: **The Rex phenotype of altruistic cell death following infection of a λ lysogen by T4rII mutants is suppressed by plasmids expressing OOP RNA**. *Gene* 1997, **189**:35-42.
30. Hazan R, Sat B, Reches M, Engelberg-Kulka H: **Postsegregational killing mediated by the P1 phage "addiction module" phd-doc requires the *Escherichia coli* programmed cell death SystemmazEF**. *J Bacteriol* 2001, **183**:2046-2050.

31. Shokeen S, Patel S, Greenfield TJ, Brinkman C, Weaver KE: **Translational regulation by an intramolecular stem-loop is required for intermolecular RNA regulation of the par addiction module**. *J Bacteriol* 2008, **190**:6076-6083.

32. Shokeen S, Greenfield TJ, Ehli EA, Rasmussen J, Perrault BE, Weaver KE: **An intramolecular upstream helix ensures the stability of a toxin-encoding RNA in *Enterococcus faecalis***. *J Bacteriol* 2009, **191**:1528-1536.

33. Weaver KE, Reddy SG, Brinkman CL, Patel S, Bayles KW, Endres JL: **Identification and characterization of a family of toxin–antitoxin systems related to the *Enterococcus faecalis* plasmid pAD1 par addiction module**. *Microbiology* 2009, **155**:2930-2940.

34. Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, Salmond GP: **The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair**. *Proc Natl Acad Sci U S A* 2009, **106**:894-899.

35. Blower TR, Pei XY, Short FL, Fineran PC, Humphreys DP, Luisi BF, Salmond GP: **A processed noncoding RNA regulates an altruistic bacterial antiviral system**. *Nat Struct Mol Biol* 2011, **18**:185-190.

36. Marraffini LA, Sontheimer EJ: **CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea**. *Nat Rev Genet* 2010, **11**:181-190.

37. Lu LF, Liston A: **MicroRNA in the immune system, microRNA as an immune system**. *Immunology* 2009, **127**:291-298.

38. Yeung ML, Benkirane M, Jeang KT: **Small non-coding RNAs, mammalian cells, and viruses: regulatory interactions?** *Retrovirology* 2007, **4**:74.

39. Morris KV, Mattick JS: **The rise of regulatory RNA**. *Nat Rev Genet* 2014, **15**:423-437.

40. O'Connell RM, Rao DS, Baltimore D: **microRNA regulation of inflammatory responses**. *Annu Rev Immunol* 2012, **30**:295-312.

41. David M: **Interferons and microRNAs**. *J Interferon Cytokine Res* 2010, **30**:825-828.

42. Sedger LM: **microRNA control of interferons and interferon induced anti-viral activity**. *Mol Immunol* 2013, **56**:781-793.

43. Xiao C, Rajewsky K: **MicroRNA control in the immune system: basic principles**. *Cell* 2009, **136**:26-36.

44. Faltejskova P, Slaby O, Hezova R, Michalek J. Role of microRNAs in the immune system. Cas Lek Cesk [date unknown];149:10–5.

45. Zhu S, Pan W, Qian Y: **MicroRNA in immunity and autoimmunity**. *J Mol Med* 2013, **91**:1039-1050.

46. Villarreal LP: **The source of self: genetic parasites and the origin of adaptive immunity**. *Ann N Y Acad Sci* 2009, **1178**:194-232.

47. Villarreal LP: **Viruses in host evolution: virus-mediated self idenity**. In *Self and Nonself*. Edited by Lopez-Larrea C. Landes Bioschience and Springer + Business Media; 2012:185-217.

48. Mattick JS, Taft RJ, Faulkner GJ: **A global view of genomic information — moving beyond the gene and the master regulator**. *Trends Genet* 2010, **26**:21-28.

49. Grundhoff A, Sullivan CS. Virus-encoded microRNAs. Virology [date unknown];411:325–43.

50. Cox JE, Sullivan CS: **Balance and stealth: the role of noncoding RNAs in the regulation of virus gene expression**. *Annu Rev Virol* 2014, **1**:89-109.

51. Philippe N, Legendre M, Doutre G, Couté Y, Poirot O, Lescot M, Arslan D, Seltzer V, Bertaux L, Bruley C *et al.*: **Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes**. *Science* 2013, **341**:281-286.

52. Reteno DG, Benamar S, Khalil JB, Andreani J, Armstrong N, Klose T, Rossmann M, Colson P, Raoult D, Scola BL: **Faustovirus, an asfarvirus-related new lineage of giant viruses infecting amoebae**. *J Virol* 2015 http://dx.doi.org/10.1128/JVI.00115-15.

53. Parfrey LW, Lahr DJG, Katz LA: **The dynamic nature of eukaryotic genomes**. *Mol Biol Evol* 2008, **25**:787-794.

54. Moliner C, Fournier PE, Raoult D: **Genome analysis of microorganisms living in amoebae reveals a melting pot of evolution**. *FEMS Microbiol Rev* 2010, **34**:281-294.

55. Marzzoco A, Colli W: **Characterization of the genome of the small free-living amoeba *Acanthamoeba castellanii***. *Biochim Biophys Acta BBA — Nucleic Acids Protein Synth* 1975, **395**: 525-534.

56. Yoosuf N, Yutin N, Colson P, Shabalina SA, Pagnier I, Robert C, Azza S, Klose T, Wong J, Rossmann MG *et al.*: **Related giant viruses in distant locations and different habitats: *Acanthamoeba polyphaga moumouvirus* represents a third lineage of the mimiviridae that is close to the megavirus lineage**. *Genome Biol Evol* 2012, **4**:1324-1330.

57. Suhre K: **Gene and genome duplication in *Acanthamoeba polyphaga Mimivirus***. *J Virol* 2005, **79**:14095-14101.

58. Boyer M, Azza S, Barrassi L, Klose T, Campocasso A, Pagnier I, Fournous G, Borg A, Robert C, Zhang X *et al.*: **Mimivirus shows dramatic genome reduction after intraamoebal culture**. *Proc Natl Acad Sci U S A* 2011, **108**:10296-10301.

59. Mutsafi Y, Zauberman N, Sabanay I, Minsky A: **Vaccinia-like cytoplasmic replication of the giant Mimivirus**. *Proc Natl Acad Sci U S A* 2010, **107**:5978-5982.

60. Benarroch D, Qiu ZR, Schwer B, Shuman S: **Characterization of a mimivirus RNA cap guanine-N2 methyltransferase**. *Rna-Publ Rna Soc* 2009, **15**:666-674.

61. Andrade KR, Boratto PPVM, Rodrigues FP, Silva LCF, Dornas FP, Pilotto MR, Scola BL, Almeida GMF, Kroon EG, Abrahão JS: **Oysters as hot spots for mimivirus isolation**. *Arch Virol* 2014, **160**:477-482.

62. La Scola B, Desnues C, Pagnier I, Robert C, Barrassi L, Fournous G, Merchat M, Suzan-Monti M, Forterre P, Koonin E *et al.*: **The virophage as a unique parasite of the giant mimivirus**. *Nature* 2008, **455** 100-U65.

63. Yau S, Lauro FM, DeMaere MZ, Brown MV, Thomas T, Raftery MJ, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Gibson JA *et al.*: **Virophage control of antarctic algal host–virus dynamics**. *Proc Natl Acad Sci U S A* 2011, **108**:6163-6168.

64. Fischer MG, Suttle CA: **A virophage at the origin of large DNA transposons**. *Science* 2011, **332**:231-234.

65. Platt RN, Vandewege MW, Kern C, Schmidt CJ, Hoffmann FG, Ray DA: **Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats**. *Mol Biol Evol* 2014 http://dx.doi.org/10.1093/molbev/msu112.

66. Pritham EJ, Putliwala T, Feschotte C: **Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses**. *Gene* 2007, **390**:3-17.

67. Haapa-Paananen S, Wahlberg N, Savilahti H: **Phylogenetic analysis of Maverick/Polinton giant transposons across organisms**. *Mol Phylogenet Evol* 2014, **78**:271-274.

68. Nagasaki K, Takao Y, Shirai Y, Mizumoto H, Tomaru Y: **Molecular ecology of microalgal viruses**. *Uirusu* 2005, **55**:127-132.

69. Ogata H, Raoult D, Claverie JM: **A new example of viral intein in Mimivirus**. *Virol J* 2005, **2**:8.

70. Brok-Volchanskaya VS, Kadyrov FA, Sivogrivov DE, Kolosov PM, Sokolov AS, Shlyapnikov MG, Kryukov VM, Granovsky IE: **Phage T4 SegB protein is a homing endonuclease required for the preferred inheritance of T4 tRNA gene region occurring in co-infection with a related phage**. *Nucleic Acids Res* 2008, **36**: 2094-2105.

71. Katzourakis A, Aswad A: **The origins of giant viruses, virophages and their relatives in host genomes**. *BMC Biol* 2014, **12**:51.

72. Delaroque N, Boland W: **The genome of the brown alga *Ectocarpus siliculosus* contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses**. *BMC Evol Biol* 2008, **8**:110.

73. Legendre M, Bartoli J, Shmakova L, Jeudy S, Labadie K, Adrait A, Lescot M, Poirot O, Bertaux L, Bruley C *et al.*: **Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology**. *Proc Natl Acad Sci U S A* 2014 http://dx.doi.org/10.1073/pnas.1320670111.

74. Boratto P, Albarnaz JD, Almeida GM, de F, Botelho L, Fontes ACL, Costa AO, Santos DA, Bonjardim CA, Scola BL, Kroon EG *et al.*: ***Acanthamoeba polyphaga mimivirus*** **prevents amoebal encystment-mediating serine proteinase expression and circumvents cell encystment**. *J Virol* 2014 http://dx.doi.org/10.1128/JVI.03177-14.

75. Filée J: **Multiple occurrences of giant virus core genes acquired by eukaryotic genomes: the visible part of the iceberg?** *Virology* 2014, **466–467**:53-59.

76. Maumus F, Epert A, Nogué F, Blanc G: **Plant genomes enclose footprints of past infections by giant virus relatives [Internet]**. *Nat Commun* 2014:5.

77. Moran Y, Fredman D, Praher D, Li XZ, Wee LM, Rentzsch F, Zamore PD, Technau U, Seitz H: **Cnidarian microRNAs frequently regulate targets by cleavage**. *Genome Res* 2014, **24**:651-663.

78. Boratto PVM, Dornas FP, Andrade KR, Rodrigues R, Peixoto F, Silva LCF, Scola BL, Costa AO, Almeida GMF, de Kroon EG *et al.*: **Amoebas as mimivirus bunkers: increased resistance to UV light, heat and chemical biocides when viruses are carried by amoeba hosts**. *Arch Virol* 2013, **159**:1039-1043.

79. Slimani M, Pagnier I, Raoult D, Scola BL: **Amoebae as battlefields for bacteria, giant viruses, and virophages**. *J Virol* 2013, **87**:4783-4785.

80. •• Villarreal LP: **Viruses and the placenta: the essential virus first view**. *APMIS* 2016, **124**:20-30.
An overview of how endogenous viruses contributed to the prigin of placental biology in various mammals.

81. • Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-option of endogenous retroviruses**. *Science* 2016, **351**:1083-1087.
A very recent report showing how endogenous viruses participated in the origin of the regulation of the antiviral interferon network.