The depths of virus exaptation
Eugene V. Koonin, Mart Krupovic

To cite this version:

HAL Id: pasteur-01977329
https://pasteur.hal.science/pasteur-01977329
Submitted on 10 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The depths of virus exaptation

Eugene V. Koonin\textsuperscript{1} and Mart Krupovic\textsuperscript{2}

\textsuperscript{1} National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894
\textsuperscript{2} Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Department of Microbiology, Institut Pasteur, 25 rue du Docteur Roux, Paris 75015, France

*For correspondence; e-mail: koonin@ncbi.nlm.nih.gov; krupovic@pasteur.fr
Abstract

Viruses are ubiquitous parasites of cellular life forms and the most abundant biological entities on earth. The relationships between viruses and their hosts involve the continuous arms race but are by no account limited to it. Growing evidence shows that, in the course of evolution, viruses and their components are repeatedly recruited (exapted) for host functions. The functions of exapted viruses typically involve either defense from other viruses or cellular competitors or transfer of nucleic acids between cells, or storage functions. Virus exaptation can reach different depths, from recruitment of a fully functional virus to exploitation of defective, partially degraded viruses, to utilization of individual virus proteins.
Introduction

Parasitic genetic elements are ubiquitous companions of cellular life forms. Theoretical argument and empirical evidence strongly suggest that emergence of parasites is inevitable in replicator systems [1-3]. Moreover, most cellular organisms are hosts to multiple types of genetic parasites that differ with respect to their degree of autonomy and impact on the host [4-6]. Some of the parasitic elements, such as plasmids and transposons, are primarily commensals that reproduce at a low cost to the host [7]. Others are virulent viruses that kill the host. The key difference between viruses and other genetic parasites is that most viruses form virions, specialized particles that encapsidate the viral genome and serve as transmission devices [8,9]. Many viruses are evolutionarily related to non-viral, particle-less genetic elements such as plasmids or transposons, and transitions between viruses and particle-less elements apparently have occurred in both directions on many occasions [10-14]. Furthermore, genomes of all types of parasitic elements integrate into host genomes, either as part of their life cycle or sporadically. In many animals and plants, integrated parasitic elements, mostly, inactivated ones, account for the majority of the genomic DNA [15,16].

Given the ubiquity of viruses in the biosphere and the billions of years of virus-host coevolution [17,18], it is obvious that the relationships between viruses and hosts cannot be limited to the proverbial arms race. Far from that, genetic material of viruses and other parasitic elements is repeatedly recruited by hosts for various functional roles. The domestication of viruses and exaptation (a concept and term introduced by Gould and Vrba to denote recruitment of a biological entity for a new function unrelated to the original one [19]) of viral genomes, individual genes or smaller fragments for host functions takes many different forms and reaches different depths, with respect to what remains of the virus genome. In this brief review, we discuss several distinct cases of virus exaptation that have been illuminated by recent discoveries.

Proviruses, virophages, and antivirus defense

Viruses with an essential stage of integration into the host genome include several families of reverse-transcribing viruses [20]. Many viruses, especially those with DNA genomes, use a “bet hedging” strategy whereby a virus switches from a lytic to a lysogenic reproduction mode, whereby the viral genome integrates into the host chromosome and is vertically inherited. Such is the lifestyle of numerous bacteriophages, particularly, those in the family Siphoviridae, as exemplified by Enterobacteria phage Lambda, the classic model of molecular biology [21,22]. Most bacterial genomes carry prophages, often, several ones. Are prophages a form of exaptation? Although prophages have been studied for decades, there is still no general answer to this question. Estimation of the fitness values for all microbial genes shows that on average, prophages are costly [7]. Nevertheless, prophages can protect the host from superinfection by other phages by a variety of mechanisms ranging from modification of the host cell surface or masking of phage receptors to active blockage of genome injection of superinfecting phages to repressor-based immunity at the level transcription [23,24]. Provirus-mediated superinfection exclusion (also referred to as superinfection immunity or superinfection resistance) is not restricted to bacteriophages but also takes place in the case of retroviral infections [25]. Interestingly, deletion of all prophages from Escherichia coli genome has led to decreased fitness of the bacterium under various environmental conditions [26].
Recently, a group of eukaryotic viruses has been discovered that mimic the prophage life style but shows clear evidence of exaptation. These are the virophages (family Lavedaviridae [27]), dsDNA viruses with small genomes that parasitize on giant viruses of the family Mimiviridae [28]. When virophage Mavirus infects the flagellate host of the giant helper virus, Cafeteria roenbergensis virus (CroV), it integrates into the cellular genome, without causing any tangible harm to the cell, and remains there “in waiting” for the infection with CroV [29,30]. Once CroV infects, the expression of the integrated Mavirus is induced, and the propagating Mavirus abrogates the giant virus reproduction. The infected cell dies nevertheless but the surrounding ones are protected and can be infected by the released Mavirus, perpetuating the protection [29]. The striking feature of the Mavirus genome that underlies this mechanism of defense is the identity of the Mavirus gene promoters to those of CroV [30]. This is a double adaptation that both allows the virophage to utilize the transcription machinery of the giant virus and to safeguard the uninfected cells in the host population. Thus, the virophage seems to represent a “transient exaptation” (Figure 1A): it is adopted by the host for a distinct function but remains a full-fledged virus.

Some protist genomes contain multiple insertions of intact and degraded virophage genomes which appear to be traces of an active exaptation process [31]. The phenomenon of virophage exaptation might be much more general. Polintons, self-synthesizing transposons that are common in many diverse eukaryotes [32,33], are clearly related to virophages [30]. Although, originally, polintons have not been considered viruses, it has been shown that they encode major and minor capsid proteins homologous to those of adenoviruses and virophages [34,35]. Virions of the polintons so far have not been observed but the conservation of all the structural elements in the encoded capsid proteins implies that such particles exist. An attractive hypothesis, then, is that all polintons are actually integrated virophages that have been exapted as a mechanism of adaptive immunity against large virulent viruses that, in most cases, remain unknown [14,36]. This hypothesis is compatible with the findings indicating that at least one member of the recently discovered, poorly characterized family of polinton-like viruses (PLV) is a virophage that is associated with Pheocystis globosa virus [37,38], and multiple copies of some of the PLV are integrated in algal genomes [38].

**Gene transfer agents, polydnaviruses and contractile injection systems: virus-derived vehicles for DNA and proteins**

The virophages discussed in the preceding section apparently were exapted by the host for an antivirus defense function but remain viruses that are competent for replication, even if only in the presence of the supporting giant virus. A deeper level of exaptation, where viruses are more highly derived, includes the Gene Transfer Agents (GTAs) (Figure 1B). The GTAs are highly derived, defective prophages that have been studied in greatest detail in the α-proteobacterium *Rhodobacter capsulatus* but subsequently have been identified in diverse bacteria and some archaea [39]. As shown by genetic methods, the genes of the GTAs are dispersed in microbial chromosomes although the cluster of genes encoding the protein subunits of the phage head and tail stays compact [40]. The GTAs have been exapted on at least five independent occasions, from different viral lineages [39]. The key feature of the GTAs is that they generally do not package the prophage genes and instead encapsidate fragments of the host bacterial DNA. However, depending on the GTA family, the encapsidated DNA varies from essentially random
Polydnaviruses (PDVs) are a group of unusual insect viruses with genomes consisting of multiple, circular segments of dsDNA [47, 48] that presents a close parallel to the prokaryotic GTAs. The PDVs are mutualistic symbionts of parasitoid wasps that are stably integrated into the wasp genomes. The PDVs appear to have evolved from insect viruses but have lost many core viral genes and instead acquired numerous genes from the hosts. Nevertheless, the PDVs retain the ability to form virus particles. Remarkably, and similarly to the GTAs, PDVs have apparently evolved on at least three independent occasions in different insect lineages from unrelated viruses because the proteins involved in the formation of the viral particles are not homologous in the three PDV groups [49]. PDVs of the genus Bracovirus have evolved from an endogenized nudivirus (genus Betanudivirus) genome ~103 million years ago [50], whereas members of the genus Ichnovirus and PDVs found in wasps of the subfamily Banchinae have respectively evolved from two other unrelated, currently unknown groups of viruses [49, 51]. The infectious cycle of the PDVs involves two hosts, the wasps and their caterpillar prey. Virus particles are produced only in the ovaries of female parasitoid wasps and serve as vehicles to deliver PDV genes into the caterpillars where they are expressed to produce virulence factors, various proteins that suppress the immune system of the caterpillars and accordingly provide for the development of the wasp eggs. Interestingly, a parallel nudivirus (from genus Alphanudivirus) domestication event has occurred in the campoplegine parasitic wasp Venturia canescens. The nudivirus has been domesticated for the same purpose as in the case of PDVs, i.e., for subversion of the host immunity. However, differently from PDVs, V. canescens produce virion-like particles lacking the DNA, which instead transport virulence proteins [52]. Recently, it has been shown that PDVs manipulate the immune system not only in the caterpillars but also in the plants on which those feed, through the caterpillar saliva [53]. Thus, the PDVs, in principle, analogously to polintons have been exapted by the host as a “weapon”, in this case one used by a predator to subjugate the prey. In the process of co-adaptation with the host, the viral enzymatic machinery has degraded substantially so that, unlike the ancestral viruses, the PDVs largely depend on the host enzymes for expression and on vertical transmission in the integrated state for reproduction. The genes encoding the viral structural proteins are not encapsidated into the PDV virions but have been relocated into the wasp chromosome where they are under tight regulation...
Deep exaptations of portions of bacteriophage genomes are not limited to the GTAs. The contractile tails of bacteriophages of the family Myoviridae were exapted to function as so-called Contractile Injection Systems (CIS) which include two major categories: extracellular (e)CIS and Type VI Secretion Systems (T6SS) both of which play crucial roles in inter-microbial conflicts [56,57]. The CIS contain all the key components of actual phage tail including the baseplate, which results in membrane depolarization and cell death, PVC and Afp inject the target cells with various protein effectors, such as toxins. Notably, non-contractile tails of siphoviruses have also been recruited to function as tailocins, named F-tailocins [59]. Overall, tailocins have apparently evolved from phage tails on at least five independent occasions [59]. Similar to GTA, eCIS are released after cell lysis, mediated by the phage-derived holin and endolysin genes. By contrast, the T6SS are located in bacterial cytoplasm and anchored in the inner membrane. Upon direct contact between bacterial cells, the T6SS inject various toxins into the cytoplasm of the recipient cell. In other cases, the T6SS can target eukaryotic cells, contributing both to defense and to bacterial pathogenicity [60]. Another variation on the theme is presented by the metamorphosis-associated contractile assemblies (MAC), arrays comprised of ~100 contractile phage tail-like structures linked by tail fibers and a dynamic hexagonal net [61]. Remarkably, a marine bacterium, Pseudoalteromonas luteoviolacea, produces MACs to trigger metamorphosis of the tubeworm Hydroides elegans [61]. The specific evolutionary affinity of the CIS with the tails of particular groups of bacteriophages clearly defines the vector of evolution, from phages to the CIS, which is the route of virus exaptation. This case of exaptation fits the “guns for hire” concept [62]: a group of viral genes that encode the viral device for genome injection into infected cells is exapted as a “weapon” for intercellular conflicts.

**Virus capsids for transport and storage**

Capsids are dedicated devices for containment of the viral nucleic acid. Potentially, capsids could be used as “universal boxes” for storage and transport of various compounds (Figure 1C). Many diverse bacteria and some archaea contain nanocompartments, known as encapsulins, icosahedral particles that resemble bacteriophage capsids and contain multiple pores that are permeable to small molecules but not to proteins [63]. The major encapsulin shell protein is homologous to the HK97-like major capsid protein of tailed bacterial and archaeal viruses of the order Caudovirales [8]. The encapsulins package cargo proteins inside the shell where these proteins either catalyze certain reactions or, apparently more often, serve to store ions, e.g. bacterioferritin that binds iron cations, presumably, also defending the cell against oxidative stress. The evolutionary relationships between encapsulins and phage capsids is beyond doubt but the direction of evolution can be questioned [8]. Nevertheless, given the wider spread of viruses with HK97-like capsids compared to encapsulins, the route of virus exaptation appears most likely.
The encapsulins are distinct from the much larger bacterial microcompartments (BMCs) that compartmentalize a variety of metabolic processes, e.g., CO₂ fixation or catabolism of organic substrates [64]. The BMCs also are icosahedral particles that resemble viral capsids in shape but consist of proteins without detectable viral homologs [65]. Thus, the icosahedral shell has evolved convergently on multiple occasions not only in diverse viruses but also in a cellular context. It is conceivable that viruses with capsids made of BMC proteins remain to be discovered.

Exaptation of virus capsids occurred in parallel in eukaryotes, in this case, involving the Gag polyprotein of retroviruses. Sequences derived from retroviruses and retrotransposons abound in animal and plant genomes and many cases of exaptation are known, particularly, recruitment of regulatory signals but also protein-coding genes for antiviral defense (see above). Recently, however, it has been demonstrated that Arc proteins that have been known to be involved in neuronal plasticity in both mammalian and insect brains are Gag derivatives that form virus-like particles and transfer RNA molecules between neurons [66,67]. Mammalian genomes encode many additional Gag derivatives. At least 85 genes derived from the gag gene of LTR retrotransposons, particularly, Ty3/Gypsy-like metaviruses, have been identified in the human genome [68]. Many of these genes are conserved in other mammals and have been grouped into three major families, the Mart, PNMA and SCAN families [69,70]. Whereas some members of the Mart and PNMA families have retained (almost) the entire Gag protein sequence, the SCAN domain is derived from the C-terminal domain of the capsid protein which fused with multiple C2H2 zinc fingers and/or Krüppel-associated box (KRAB) domains and functions as a protein-interaction module that mediates self-association or selective association with other proteins in vertebrate transcription factors [69]. Proteins from all three families are implicated in various forms of tissue differentiation but exact molecular mechanisms are often not known. It remains to be determined whether or not any of these proteins also function through the formation of virus-like particles, analogously to the case of Arc proteins.

Viral fusogens for cell fusion

Not only the Gag but also proteins of the retrovirus envelope (Env) have been exapted for host functions (Figure 1C). Syncitins, the placental receptors, are Env derivatives. The original function of the Env, i.e., to promote fusion between the viral envelope and cellular membrane during virus entry, has been repurposed for the fusion of trophoblast cells during placentation. Remarkably, env sequences have been convergently domesticated from different retroviruses infecting germ cells in at least 10 different mammalian lineages [70] as well as in viviparous Mabuya lizards [71]. The pan-eukaryotic fusogen HAP2 is homologous to Class II viral fusion proteins [72-74]. In this case, the ultimate provenance of this key protein remains unclear [75] but, given the ubiquity of HAP2 in eukaryotes, cellular origin appears more likely, with viruses acquiring the fusogenic capacity at a later stage. The biological connotation, though, is the same, namely, shared functions of cellular and viral proteins, promoting a two-way gene flow. Indeed, some of the key viral proteins, most notably those involved in virion formation, have been exapted from functionally diverse cellular proteins on multiple independent occasions [8].

Endogenous Viral Elements: Sporadic exaptation
Recent genome analyses have increasingly revealed integration into host eukaryotic genomes of complete or partial copies of diverse viruses that are not known to include an integration step in their reproduction cycle [76,77]. Most of these Endogenous Viral Elements (EVEs) accumulate multiple mutations and effectively are decaying remnants of viral genomes. However, there are notable exceptions whereby EVEs preserve certain viral genes, and a few have been shown to inhibit infection by closely related viruses. For example, nucleoproteins produced by endogenous bornaviruses (negative-sense RNA genomes) in ground squirrels inhibit bornavirus reproduction, apparently, by down-regulating the viral RNA-dependent RNA polymerase activity [78]. Other EVEs, such as endogenous flavivirus (positive-sense RNA genomes)-derived elements in mosquitoes, serve as templates for production of small interfering RNAs that might inhibit superinfection [79]. The multitude of consequences of EVE exaptation has been extensively studied in the case of endogenous retroviruses (ERVs), a variety of EVEs derived from retroviruses. Whereas EVEs are generated extremely rarely, endogenization of retroviruses is prevalent, even if sporadic, particularly in mammalian genomes [80,81]. Dynamic proliferation of ERVs and long terminal repeats (LTRs), which flank the retroviral genomes and contain regulatory sequences, has resulted in massive rewiring of various cellular regulatory circuits. For instance, it has been recently demonstrated that ERVs/LTRs tend to be located near the genes involved in innate immune responses such as cytokine-mediated signaling [82] and have shaped the evolution of transcriptional networks underlying the interferon response, a major branch of innate immunity, on multiple independent occasions in diverse mammalian lineages [83]. LTRs also play a role in recycling pseudogenes and de novo formation of new protein-coding genes by providing not only the promoter regions but also the first exon [84]. In the case of the D6Ertd527e gene expressed in mouse oocytes, an LTR provided a promoter and the 5’ exon with a functional start codon while the bulk of the protein-coding sequence evolved through a CAG repeat expansion [84]. Furthermore, ERV-derived envelope (Env) and group-specific antigen (Gag) polypeptides have been shown to restrict exogenous retroviruses at different stages of infection, from inhibition of viral entry by competing with exogenous Env for cell surface receptors [85] to virion disassembly prior to integration into the host genome [86], to virions assembly and release [87]. On the whole, the formation of EVEs is a widespread, nearly ubiquitous phenomenon in eukaryotes that appears to be a by-product of virus reproduction. From this large pool, EVEs or their components seem to have been recurrently exapted for various forms of antiviral defense.

**Concluding remarks**

Given the ubiquity of viruses in the biosphere and the widespread integration of virus genomes into those of the hosts, exaptation of viruses and individual viral genes for host functions is bound to be common, and indeed, an increasing number of cases are being identified. In this brief article, we only considered selected examples that were chosen in such a manner as to illustrate a notable trend: the varying “depths” of the derivation of viruses in the course of exaptation (Figure 2). A virus can be repurposed for a defense function virtually without losing its identity as is the case for Mavirus and, possibly, other polinton-like viruses; else, this type of exaptation can involve notable degradation of the virus as is the case with PDVs. The next step is the complete loss of the viral replication module whereas the structural-morphogenetic module remains intact as illustrated by the GTAs or at least partially retained as in the CIS. Finally, an exapted virus can be reduced to a single gene or even part of a gene that, however, retains certain key capacities such as capsid formation or fusogenic activity.
In functional terms, exaptation exploits the central properties of viruses. As aggressive parasites, viruses abrogate host functions, particularly, those involved in defense, and this capacity can be repurposed for host protection as it occurs in the case of the virophages. Also, viruses are devices for storage and transfer of nucleic acids, and of that capacity, hosts appear to make ample use for various forms of intercellular communication.

The most remarkable cases of virus exaptation including the virophages, encapsulins, CIS and Arc particles have been discovered or characterized only recently. There is little doubt that the true depth and breadth of virus exaptation remain to be uncovered.

Acknowledgements
EVK is supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine). MK was supported by l’Agence Nationale de la Recherche (France) project ENVIRA.
References


envelope syncytin and its cognate receptor identified in the viviparous placental
Mabuya lizard. *Proceedings of the National Academy of Sciences of the United States of
America* 2017, **114**:E10991-E11000.

Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein

Bricogne G, Grishin NV, et al.: The Ancient Gamete Fusogen HAP2/GCS1 is a gamete fusion protein homologous

75. Hernández JM, Podbilewicz B: The hallmarks of cell-cell fusion. *Development* 2017,
**144**:4481-4495.

76. Aiewsakun P, Katzourakis A: Endogenous viruses: Connecting recent and ancient viral

80. Greenwood AD, Ishida Y, O'Brien SP, Roca AL, Eiden MV: Transmission, evolution, and
endogenization: Lessons learned from recent retroviral invasions. *Microbiol Mol Biol
Rev* 2018, **82**:e00044-00017.
Figure legends

**Figure 1.** Distinct routes of virus exaptation:
(a) competent virus serves as a defense systems;
(b) Defective virus particles as vehicles for DNA and protein transfer and storage;
(c) Exaptation of single viral proteins or regulatory sequences for antiviral defense and other functions.

**Figure 2.** The different depths of virus exaptation, a generalized scheme.
Highlights

- Viruses and their components are repeatedly exapted for diverse host functions
- Virus components are often exapted to function in antiviral defense
- Defective viruses are employed for gene transfer and nutrient storage
- Retroviruses and LTR retrotransposons are a rich source of new cellular functions
Env receptor
SCAN Arc LTR T6SS MAC encapsulins GTA tailocin Afp/PVC

(a) Mavirus
(b) Env
(c) Depth of virus exaptation

Depth of virus exaptation
Autonomous virus

Helper-dependent virus for antiviral defense

Non-infectious virus-like particles for transport and storage

Various virus-derived proteins and regulatory sequences