The depths of virus exaptation

Eugene V.Koonin[1] and Mart Krupovic[2]

1 National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894
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 rsistance) 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 *Lavidaviridae* [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

DNA fragments of ~4kb, as in the case of the *R. capsulatus* GTAs [41], to semi-specific packaging of larger genomic fragments of ~14kb in *Bartonella* GTAs [42]. In the latter case, GTAs preferentially package genes encoding host interaction factors, including secretion systems and putative secretion substrates such as cholera-like toxins, that are amplified from a nearby phage-derived origin of replication [43]. It has been proposed that the specific encapsidation of host-adaptation systems facilitated adaptive evolution and explosive radiation of *Bartonella*, an emerging pathogen [44]. Whereas the heads of the *R. capsulatus* GTAs are too small for encapsidation of the entire phage morphogenetic module (~14 kb), the *Bartonella* GTAs potentially could be self-transmissible [39]. A recent phylogenomic analysis of the *R. capsulatus* GTAs has shown that the GTA organization was fixed at the base of one of the α-proteobacterial branches, and further, that the GTA genes evolve much slower than the corresponding prophage genes [45]. These findings indicate that the GTAs are an exaptation that was fixed in bacterial evolution and persisted for a long evolutionary span, presumably as a dedicated device for HGT. Notably, the GTAs are beneficial only at the population level because the individual cells producing GTAs are lysed upon GTA release [46]. This "altruistic" character of the GTAs obviously mimics the defense function of the virophages discussed above.

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

host regulation [54,55]. This adaptation safeguards the complete submission of the PDV production to the host control. One may wonder whether PDVs should be considered viruses rather than fully exapted virus-derived specialized agents analogous to the GTAs or assault organelles for transfer of the wasp genes.

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 involved in bacterial cell recognition [58]. The eCIS, including R-tailocins (also called R-pyocins), *Photorhabdus* virulence cassettes (PVC) and antifeeding prophages (Afp), effectively are head-less phages that are released in the medium by bacterial cells and bind to the surface of target cells. Whereas tailocins function as deployable devices for puncturing sensitive bacterial cells, 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 mircompartments (BMCs) that compartmentalize a variety of metabolic processes, e.g., CO_2 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) polyproteins 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 wide spread of 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

1. Iranzo J, Puigbo P, Lobkovsky AE, Wolf YI, Koonin EV: **Inevitability of genetic parasites**. *Genome biology and evolution* 2016, **8**:2856-2869.
2. Koonin EV: **Viruses and mobile elements as drivers of evolutionary transitions**. *Philos Trans R Soc Lond B Biol Sci* 2016, **371**:20150442.
3. Koonin EV, Wolf YI, Katsnelson MI: **Inevitability of the emergence and persistence of genetic parasites caused by evolutionary instability of parasite-free states**. *Biology direct* 2017, **12**:31.
4. Jalasvuori M: **Vehicles, replicators, and intercellular movement of genetic information: evolutionary dissection of a bacterial cell**. *International journal of evolutionary biology* 2012, **2012**:874153.
5. Jalasvuori M, Koonin EV: **Classification of prokaryotic genetic replicators: between selfishness and altruism**. *Annals of the New York Academy of Sciences* 2015, **1341**:96-105.
6. Koonin EV, Starokadomskyy P: **Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question**. *Studies in history and philosophy of biological and biomedical sciences* 2016, **59**:125-134.
7. Iranzo J, Cuesta JA, Manrubia S, Katsnelson MI, Koonin EV: **Disentangling the effects of selection and loss bias on gene dynamics**. *Proceedings of the National Academy of Sciences of the United States of America* 2017, **114**:E5616-E5624.
8. Krupovic M, Koonin EV: **Multiple origins of viral capsid proteins from cellular ancestors**. *Proceedings of the National Academy of Sciences of the United States of America* 2017, **114**:E2401-E2410.
9. Krupovic M, Bamford DH: **Order to the viral universe**. *Journal of virology* 2010, **84**:12476-12479.
10. Iranzo J, Koonin EV, Prangishvili D, Krupovic M: **Bipartite Network Analysis of the Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless Mobile Elements**. *Journal of virology* 2016, **90**:11043-11055.
11. Koonin EV, Dolja VV: **Virus world as an evolutionary network of viruses and capsidless selfish elements**. *Microbiology and molecular biology reviews : MMBR* 2014, **78**:278-303.
12. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV: **Viruses of archaea: Structural, functional, environmental and evolutionary genomics**. *Virus research* 2018, **244**:181-193.
13. Krupovic M: **Recombination between RNA viruses and plasmids might have played a central role in the origin and evolution of small DNA viruses**. *BioEssays : news and reviews in molecular, cellular and developmental biology* 2012, **34**:867-870.
14. Koonin EV, Krupovic M: **Polintons, virophages and transpovirons: a tangled web linking viruses, transposons and immunity**. *Current opinion in virology* 2017, **25**:7-15.
15. Defraia C, Slotkin RK: **Analysis of retrotransposon activity in plants**. *Methods in molecular biology* 2014, **1112**:195-210.
16. Lopez-Flores I, Garrido-Ramos MA: **The repetitive DNA content of eukaryotic genomes**. *Genome dynamics* 2012, **7**:1-28.
17. Forterre P, Prangishvili D: **The major role of viruses in cellular evolution: facts and hypotheses**. *Current opinion in virology* 2013, **3**:558-565.

18. Koonin EV, Dolja VV: **A virocentric perspective on the evolution of life**. *Current opinion in virology* 2013, **3**:546-557.
19. Gould SJ, Vrba ES: **Exaptation — a missing term in the science of form**. *Paleobiology* 1982, **8**:4-15.
20. Krupovic M, Blomberg J, Coffin JM, Dasgupta I, Fan H, Geering AD, Gifford R, Harrach B, Hull R, Johnson W, et al.: ***Ortervirales*: New virus order unifying five families of reverse-transcribing viruses**. *J Virol* 2018, **92**:e00515-00518.
21. Gandon S: **Why Be Temperate: Lessons from Bacteriophage lambda**. *Trends in microbiology* 2016, **24**:356-365.
22. Casjens SR, Hendrix RW: **Bacteriophage lambda: Early pioneer and still relevant**. *Virology* 2015, **479-480**:310-330.
23. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL: **Prophages mediate defense against phage infection through diverse mechanisms**. *The ISME journal* 2016, **10**:2854-2866.
24. Labrie SJ, Samson JE, Moineau S: **Bacteriophage resistance mechanisms**. *Nature reviews. Microbiology* 2010, **8**:317-327.
25. Nethe M, Berkhout B, van der Kuyl AC: **Retroviral superinfection resistance**. *Retrovirology* 2005, **2**:52.
26. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK: **Cryptic prophages help bacteria cope with adverse environments**. *Nature communications* 2010, **1**:147.
27. Krupovic M, Kuhn JH, Fischer MG: **A classification system for virophages and satellite viruses**. *Archives of virology* 2016, **161**:233-247.
28. Bekliz M, Colson P, La Scola B: **The expanding family of virophages**. *Viruses* 2016, **8**:E317.
29. Fischer MG, Hackl T: **Host genome integration and giant virus-induced reactivation of the virophage mavirus**. *Nature* 2016, **540**:288-291.
30. Fischer MG, Suttle CA: **A virophage at the origin of large DNA transposons**. *Science* 2011, **332**:231-234.
31. Blanc G, Gallot-Lavallee L, Maumus F: **Provirophages in the Bigelowiella genome bear testimony to past encounters with giant viruses**. *Proceedings of the National Academy of Sciences of the United States of America* 2015, **112**:E5318-5326.
32. 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.
33. Kapitonov VV, Jurka J: **Self-synthesizing DNA transposons in eukaryotes**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**:4540-4545.
34. Krupovic M, Bamford DH, Koonin EV: **Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses**. *Biology direct* 2014, **9**:6.
35. Krupovic M, Koonin EV: **Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution**. *Nature reviews. Microbiology* 2015, **13**:105-115.
36. Koonin EV, Krupovic M: **Virology: A parasite's parasite saves host's neighbours**. *Nature* 2016, **540**:204-205.
37. Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, Barbe V, Wommack KE, Noordeloos AA, Brussaard CP, et al.: **Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting**

eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**:10800-10805.
38. Yutin N, Shevchenko S, Kapitonov V, Krupovic M, Koonin EV: **A novel group of diverse Polinton-like viruses discovered by metagenome analysis**. *BMC biology* 2015, **13**:95.
39. Lang AS, Westbye AB, Beatty JT: **The Distribution, Evolution, and Roles of Gene Transfer Agents in Prokaryotic Genetic Exchange**. *Annual review of virology* 2017, **4**:87-104.
40. Hynes AP, Shakya M, Mercer RG, Grull MP, Bown L, Davidson F, Steffen E, Matchem H, Peach ME, Berger T, et al.: **Functional and Evolutionary Characterization of a Gene Transfer Agent's Multilocus "Genome"**. *Molecular biology and evolution* 2016, **33**:2530-2543.
41. Solioz M, Marrs B: **The gene transfer agent of Rhodopseudomonas capsulata. Purification and characterization of its nucleic acid**. *Archives of biochemistry and biophysics* 1977, **181**:300-307.
42. Barbian KD, Minnick MF: **A bacteriophage-like particle from Bartonella bacilliformis**. *Microbiology* 2000, **146 (Pt 3)**:599-609.
43. Tamarit D, Neuvonen MM, Engel P, Guy L, Andersson SGE: **Origin and Evolution of the Bartonella Gene Transfer Agent**. *Molecular biology and evolution* 2018, **35**:451-464.
44. Guy L, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, Granberg F, Naslund K, Eriksson AS, Andersson SG: **A gene transfer agent and a dynamic repertoire of secretion systems hold the keys to the explosive radiation of the emerging pathogen Bartonella**. *PLoS genetics* 2013, **9**:e1003393.
45. Shakya M, Soucy SM, Zhaxybayeva O: **Insights into origin and evolution of alpha-proteobacterial gene transfer agents**. *Virus evolution* 2017, **3**:vex036.
46. Hynes AP, Mercer RG, Watton DE, Buckley CB, Lang AS: **DNA packaging bias and differential expression of gene transfer agent genes within a population during production and release of the Rhodobacter capsulatus gene transfer agent, RcGTA**. *Molecular microbiology* 2012, **85**:314-325.
47. Drezen JM, Leobold M, Bezier A, Huguet E, Volkoff AN, Herniou EA: **Endogenous viruses of parasitic wasps: variations on a common theme**. *Current opinion in virology* 2017, **25**:41-48.
48. Strand MR, Burke GR: **Polydnaviruses: From discovery to current insights**. *Virology* 2015, **479-480**:393-402.
49. Herniou EA, Huguet E, Theze J, Bezier A, Periquet G, Drezen JM: **When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses**. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2013, **368**:20130051.
50. Theze J, Bezier A, Periquet G, Drezen JM, Herniou EA: **Paleozoic origin of insect large dsDNA viruses**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:15931-15935.
51. Lapointe R, Tanaka K, Barney WE, Whitfield JB, Banks JC, Beliveau C, Stoltz D, Webb BA, Cusson M: **Genomic and morphological features of a banchine polydnavirus: comparison with bracoviruses and ichnoviruses**. *Journal of virology* 2007, **81**:6491-6501.

52. Pichon A, Bezier A, Urbach S, Aury JM, Jouan V, Ravallec M, Guy J, Cousserans F, Theze J, Gauthier J, et al.: **Recurrent DNA virus domestication leading to different parasite virulence strategies**. *Science advances* 2015, **1**:e1501150.
53. Tan CW, Peiffer M, Hoover K, Rosa C, Acevedo FE, Felton GW: **Symbiotic polydnavirus of a parasite manipulates caterpillar and plant immunity**. *Proceedings of the National Academy of Sciences of the United States of America* 2018, **115**:5199-5204.
54. Burke GR, Walden KK, Whitfield JB, Robertson HM, Strand MR: **Widespread genome reorganization of an obligate virus mutualist**. *PLoS genetics* 2014, **10**:e1004660.
55. Volkoff AN, Jouan V, Urbach S, Samain S, Bergoin M, Wincker P, Demettre E, Cousserans F, Provost B, Coulibaly F, et al.: **Analysis of virion structural components reveals vestiges of the ancestral ichnovirus genome**. *PLoS pathogens* 2010, **6**:e1000923.
56. Cascales E: **Microbiology: And Amoebophilus Invented the Machine Gun!** *Curr Biol* 2017, **27**:R1170-R1173.
57. Touchon M, Bobay LM, Rocha EP: **The chromosomal accommodation and domestication of mobile genetic elements**. *Curr Opin Microbiol* 2014, **22**:22-29.
58. Taylor NMI, van Raaij MJ, Leiman PG: **Contractile injection systems of bacteriophages and related systems**. *Molecular microbiology* 2018, **108**:6-15.
59. Scholl D: **Phage Tail-Like Bacteriocins**. *Annual review of virology* 2017, **4**:453-467.
60. Galan JE, Waksman G: **Protein-Injection Machines in Bacteria**. *Cell* 2018, **172**:1306-1318.
61. Shikuma NJ, Pilhofer M, Weiss GL, Hadfield MG, Jensen GJ, Newman DK: **Marine tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures**. *Science* 2014, **343**:529-533.
62. Koonin EV, Krupovic M: **A movable defense**. *Scientist* 2015, **29**:46-53.
63. Nichols RJ, Cassidy-Amstutz C, Chaijarasphong T, Savage DF: **Encapsulins: molecular biology of the shell**. *Crit Rev Biochem Mol Biol* 2017, **52**:583-594.
64. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M: **Bacterial microcompartments**. *Nature reviews. Microbiology* 2018, **16**:277-290.
65. Krupovic M, Koonin EV: **Cellular origin of the viral capsid-like bacterial microcompartments**. *Biology direct* 2017, **12**:25.
66. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, Yoder N, Belnap DM, Erlendsson S, Morado DR, et al.: **The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer**. *Cell* 2018, **172**:275-288 e218.
67. Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V, Thomson T: **Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons**. *Cell* 2018, **172**:262-274 e211.
68. Campillos M, Doerks T, Shah PK, Bork P: **Computational characterization of multiple Gag-like human proteins**. *Trends Genet* 2006, **22**:585-589.
69. Kaneko-Ishino T, Ishino F: **The role of genes domesticated from LTR retrotransposons and retroviruses in mammals**. *Front Microbiol* 2012, **3**:262.
70. Naville M, Warren IA, Haftek-Terreau Z, Chalopin D, Brunet F, Levin P, Galiana D, Volff JN: **Not so bad after all: retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates**. *Clin Microbiol Infect* 2016, **22**:312-323.
71. Cornelis G, Funk M, Vernochet C, Leal F, Tarazona OA, Meurice G, Heidmann O, Dupressoir A, Miralles A, Ramirez-Pinilla MP, et al.: **An endogenous retroviral**

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.
72. Pinello JF, Lai AL, Millet JK, Cassidy-Hanley D, Freed JH, Clark TG: **Structure-Function Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein HAP2**. *Curr Biol* 2017, **27**:651-660.
73. Fedry J, Liu Y, Pehau-Arnaudet G, Pei J, Li W, Tortorici MA, Traincard F, Meola A, Bricogne G, Grishin NV, et al.: **The Ancient Gamete Fusogen HAP2 Is a Eukaryotic Class II Fusion Protein**. *Cell* 2017, **168**:904-915 e910.
74. Valansi C, Moi D, Leikina E, Matveev E, Grana M, Chernomordik LV, Romero H, Aguilar PS, Podbilewicz B: **Arabidopsis HAP2/GCS1 is a gamete fusion protein homologous to somatic and viral fusogens**. *J Cell Biol* 2017, **216**:571-581.
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 evolution**. *Virology* 2015, **479-480**:26-37.
77. Feschotte C, Gilbert C: **Endogenous viruses: insights into viral evolution and impact on host biology**. *Nature reviews. Genetics* 2012, **13**:283-296.
78. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K: **Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome**. *Proceedings of the National Academy of Sciences of the United States of America* 2014, **111**:13175-13180.
79. Suzuki Y, Frangeul L, Dickson LB, Blanc H, Verdier Y, Vinh J, Lambrechts L, Saleh MC: **Uncovering the repertoire of endogenous flaviviral elements in *Aedes* mosquito genomes**. *Journal of virology* 2017, **91**:e00571-00517.
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.
81. Frank JA, Feschotte C: **Co-option of endogenous viral sequences for host cell function**. *Current opinion in virology* 2017, **25**:81-89.
82. Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, Inoue I: **Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses**. *PLoS genetics* 2017, **13**:e1006883.
83. Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-option of endogenous retroviruses**. *Science* 2016, **351**:1083-1087.
84. Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J, et al.: **Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes**. *Genome Res* 2017, **27**:1384-1394.
85. Blanco-Melo D, Gifford RJ, Bieniasz PD: **Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors**. *eLife* 2017, **6**:e22519.
86. Yap MW, Colbeck E, Ellis SA, Stoye JP: **Evolution of the retroviral restriction gene Fv1: inhibition of non-MLV retroviruses**. *PLoS pathogens* 2014, **10**:e1003968.
87. Monde K, Terasawa H, Nakano Y, Soheilian F, Nagashima K, Maeda Y, Ono A: **Molecular mechanisms by which HERV-K Gag interferes with HIV-1 Gag assembly and particle infectivity**. *Retrovirology* 2017, **14**:27.

539
540
541

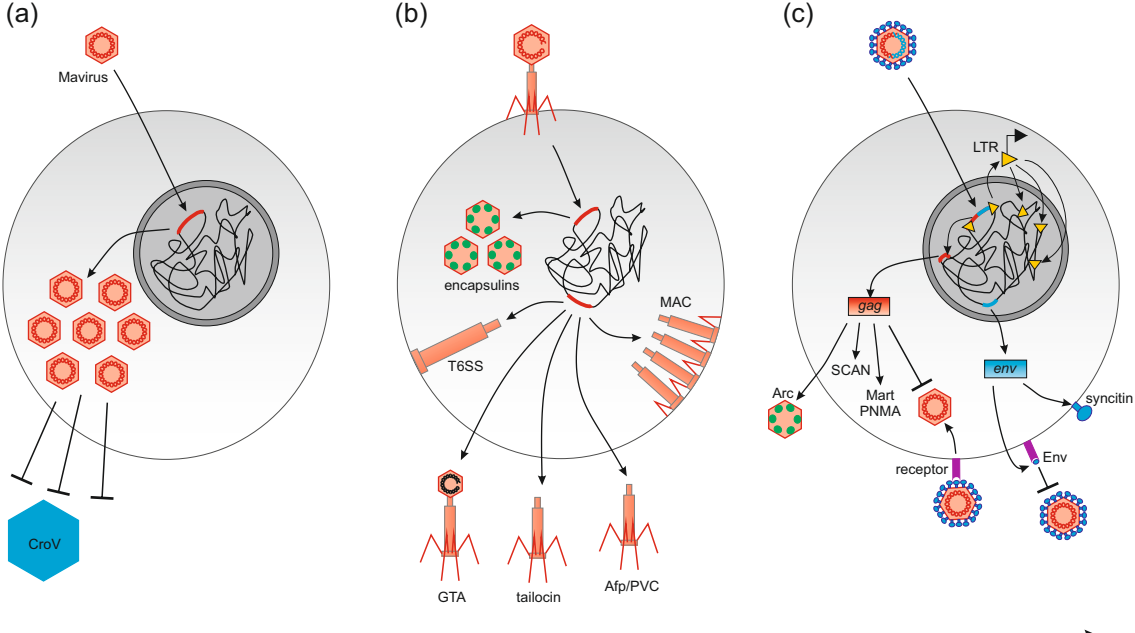
Figure legends

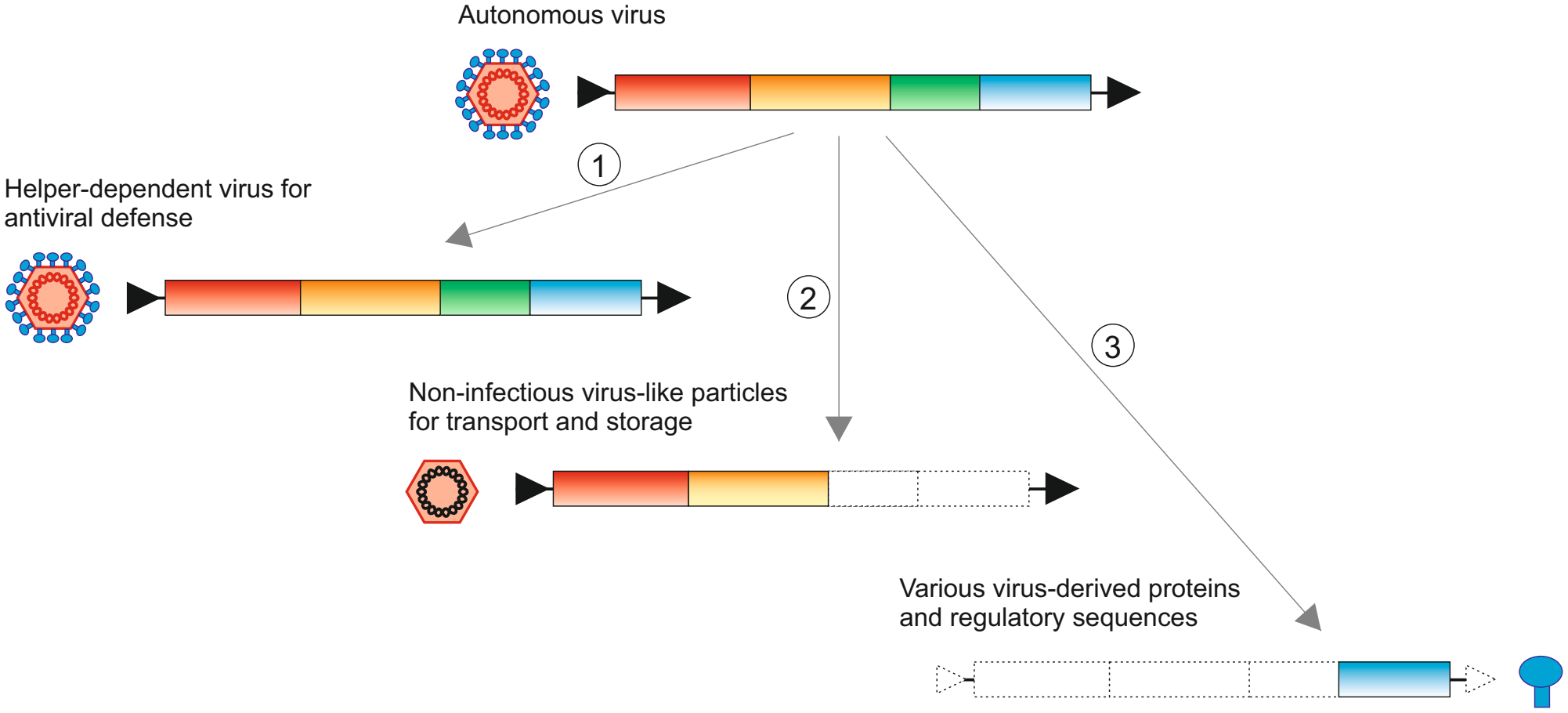
Figure 1. Distinct routes of virus exaptation:
(a) competent virus serves as a defense systems;
(b) Defective virus particles as vehicles for DNA and protein transfer and storage;
(c) Exaptation of single viral proteins or regulatory sequences for antiviral defense and other functions.

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


(a)
Mavirus
CroV
(b)
encapsulins
T6SS
MAC
GTA
tailocin
Afp/PVC
(c)
LTR
gag
SCAN
Mart
PNMA
Arc
receptor
env
syncitin
Env
Depth of virus exaptation


Autonomous virus
1
Helper-dependent virus for
antiviral defense
2
Non-infectious virus-like particles
for transport and storage
3
Various virus-derived proteins
and regulatory sequences