



HAL
open science

The depths of virus exaptation

Eugene V. Koonin, Mart Krupovic

► **To cite this version:**

Eugene V. Koonin, Mart Krupovic. The depths of virus exaptation. *Current Opinion in Virology*, 2018, Viral evolution, 31, pp.1-8. 10.1016/j.coviro.2018.07.011 . pasteur-01977329

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.

1
2
3
4

The depths of virus exaptation

Eugene V.Koonin¹ and Mart Krupovic²

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

5
6
7

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

8
9

10 Abstract

11

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

20

21 **Introduction**

22

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

37

38 Given the ubiquity of viruses in the biosphere and the billions of years of virus-host coevolution
39 [17,18], it is obvious that the relationships between viruses and hosts cannot be limited to the
40 proverbial arms race. Far from that, genetic material of viruses and other parasitic elements is
41 repeatedly recruited by hosts for various functional roles. The domestication of viruses and
42 exaptation (a concept and term introduced by Gould and Vrba to denote recruitment of a
43 biological entity for a new function unrelated to the original one [19]) of viral genomes,
44 individual genes or smaller fragments for host functions takes many different forms and reaches
45 different depths, with respect to what remains of the virus genome. In this brief review, we
46 discuss several distinct cases of virus exaptation that have been illuminated by recent
47 discoveries.

48

49 **Provirus, virophage, and antiviral defense**

50 Viruses with an essential stage of integration into the host genome include several families of
51 reverse-transcribing viruses [20]. Many viruses, especially those with DNA genomes, use a “bet
52 hedging” strategy whereby a virus switches from a lytic to a lysogenic reproduction mode,
53 whereby the viral genome integrates into the host chromosome and is vertically inherited. Such
54 is the lifestyle of numerous bacteriophages, particularly, those in the family *Siphoviridae*, as
55 exemplified by Enterobacteria phage Lambda, the classic model of molecular biology [21,22].
56 Most bacterial genomes carry prophages, often, several ones. Are prophages a form of
57 exaptation? Although prophages have been studied for decades, there is still no general answer to
58 this question. Estimation of the fitness values for all microbial genes shows that on average,
59 prophages are costly [7]. Nevertheless, prophages can protect the host from superinfection by
60 other phages by a variety of mechanisms ranging from modification of the host cell surface or
61 masking of phage receptors to active blockage of genome injection of superinfecting phages to
62 repressor-based immunity at the level transcription [23,24]. Provirus-mediated superinfection
63 exclusion (also referred to as superinfection immunity or superinfection resistance) is not
64 restricted to bacteriophages but also takes place in the case of retroviral infections [25].
65 Interestingly, deletion of all prophages from *Escherichia coli* genome has led to decreased fitness
66 of the bacterium under various environmental conditions [26].

67
68 Recently, a group of eukaryotic viruses has been discovered that mimic the prophage life style
69 but shows clear evidence of exaptation. These are the virophages (family *Lavidaviridae* [27]),
70 dsDNA viruses with small genomes that parasitize on giant viruses of the family *Mimiviridae*
71 [28]. When virophage Mavirus infects the flagellate host of the giant helper virus, Cafeteria
72 roenbergensis virus (CroV), it integrates into the cellular genome, without causing any tangible
73 harm to the cell, and remains there “in waiting” for the infection with CroV [29,30]. Once CroV
74 infects, the expression of the integrated Mavirus is induced, and the propagating Mavirus
75 abrogates the giant virus reproduction. The infected cell dies nevertheless but the surrounding
76 ones are protected and can be infected by the released Mavirus, perpetuating the protection [29].
77 The striking feature of the Mavirus genome that underlies this mechanism of defense is the
78 identity of the Mavirus gene promoters to those of CroV [30]. This is a double adaptation that
79 both allows the virophage to utilize the transcription machinery of the giant virus and to
80 safeguard the uninfected cells in the host population. Thus, the virophage seems to represent a
81 “transient exaptation” (Figure 1A): it is adopted by the host for a distinct function but remains a
82 full-fledged virus.

83
84 Some protist genomes contain multiple insertions of intact and degraded virophage genomes
85 which appear to be traces of an active exaptation process [31]. The phenomenon of virophage
86 exaptation might be much more general. Polintons, self-synthesizing transposons that are
87 common in many diverse eukaryotes [32,33], are clearly related to virophages [30]. Although,
88 originally, polintons have not been considered viruses, it has been shown that they encode major
89 and minor capsid proteins homologous to those of adenoviruses and virophages [34,35]. Virions
90 of the polintons so far have not been observed but the conservation of all the structural elements
91 in the encoded capsid proteins implies that such particles exist. An attractive hypothesis, then, is
92 that all polintons are actually integrated virophages that have been exapted as a mechanism of
93 adaptive immunity against large virulent viruses that, in most cases, remain unknown [14,36].
94 This hypothesis is compatible with the findings indicating that at least one member of the
95 recently discovered, poorly characterized family of polinton-like viruses (PLV) is a virophage
96 that is associated with *Pheocystis globosa* virus [37,38], and multiple copies of some of the PLV
97 are integrated in algal genomes [38].

98
99 **Gene transfer agents, polydnviruses and contractile injection systems: virus-derived**
100 **vehicles for DNA and proteins**

101 The virophages discussed in the preceding section apparently were exapted by the host for an
102 antiviral defense function but remain viruses that are competent for replication, even if only in
103 the presence of the supporting giant virus. A deeper level of exaptation, where viruses are more
104 highly derived, includes the Gene Transfer Agents (GTAs) (Figure 1B). The GTAs are highly
105 derived, defective prophages that have been studied in greatest detail in the α -proteobacterium
106 *Rhodobacter capsulatus* but subsequently have been identified in diverse bacteria and some
107 archaea [39]. As shown by genetic methods, the genes of the GTAs are dispersed in microbial
108 chromosomes although the cluster of genes encoding the protein subunits of the phage head and
109 tail stays compact [40]. The GTAs have been exapted on at least five independent occasions,
110 from different viral lineages [39]. The key feature of the GTAs is that they generally do not
111 package the prophage genes and instead encapsidate fragments of the host bacterial DNA.
112 However, depending on the GTA family, the encapsidated DNA varies from essentially random

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

129
130 Polydnnaviruses (PDVs) are a group of unusual insect viruses with genomes consisting of
131 multiple, circular segments of dsDNA [47,48] that presents a close parallel to the prokaryotic
132 GTAs. The PDVs are mutualistic symbionts of parasitoid wasps that are stably integrated into
133 the wasp genomes. The PDVs appear to have evolved from insect viruses but have lost many
134 core viral genes and instead acquired numerous genes from the hosts. Nevertheless, the PDVs
135 retain the ability to form virus particles. Remarkably, and similarly to the GTAs, PDVs have
136 apparently evolved on at least three independent occasions in different insect lineages from
137 unrelated viruses because the proteins involved in the formation of the viral particles are not
138 homologous in the three PDV groups [49]. PDVs of the genus *Bracovirus* have evolved from an
139 endogenized nudivirus (genus *Betanudivirus*) genome ~103 million years ago [50], whereas
140 members of the genus *Ichnovirus* and PDVs found in wasps of the subfamily Banchinae have
141 respectively evolved from two other unrelated, currently unknown groups of viruses [49,51]. The
142 infectious cycle of the PDVs involves two hosts, the wasps and their caterpillar prey. Virus
143 particles are produced only in the ovaries of female parasitoid wasps and serve as vehicles to
144 deliver PDV genes into the caterpillars where they are expressed to produce virulence factors,
145 various proteins that suppress the immune system of the caterpillars and accordingly provide for
146 the development of the wasp eggs. Interestingly, a parallel nudivirus (from genus
147 *Alphanudivirus*) domestication event has occurred in the campoplegine parasitic wasp *Venturia*
148 *canescens*. The nudivirus has been domesticated for the same purpose as in the case of PDVs,
149 i.e., for subversion of the host immunity. However, differently from PDVs, *V. canescens* produce
150 virion-like particles lacking the DNA, which instead transport virulence proteins [52]. Recently,
151 it has been shown that PDVs manipulate the immune system not only in the caterpillars but also
152 in the plants on which those feed, through the caterpillar saliva [53]. Thus, the PDVs, in
153 principle, analogously to polintons have been exapted by the host as a “weapon”, in this case one
154 used by a predator to subjugate the prey. In the process of co-adaptation with the host, the viral
155 enzymatic machinery has degraded substantially so that, unlike the ancestral viruses, the PDVs
156 largely depend on the host enzymes for expression and on vertical transmission in the integrated
157 state for reproduction. The genes encoding the viral structural proteins are not encapsidated into
158 the PDV virions but have been relocated into the wasp chromosome where they are under tight

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

163
164 Deep exaptations of portions of bacteriophage genomes are not limited to the GTAs. The
165 contractile tails of bacteriophages of the family *Myoviridae* were exapted to function as so-called
166 Contractile Injection Systems (CIS) which include two major categories: extracellular (e)CIS and
167 Type VI Secretion Systems (T6SS) both of which play crucial roles in inter-microbial conflicts
168 [56,57]. The CIS contain all the key components of actual phage tail including the baseplate
169 involved in bacterial cell recognition [58]. The eCIS, including R-tailocins (also called R-
170 pyocins), *Photorhabdus* virulence cassettes (PVC) and antifeeding prophages (Afp), effectively
171 are head-less phages that are released in the medium by bacterial cells and bind to the surface of
172 target cells. Whereas tailocins function as deployable devices for puncturing sensitive bacterial
173 cells, which results in membrane depolarization and cell death, PVC and Afp inject the target
174 cells with various protein effectors, such as toxins. Notably, non-contractile tails of siphoviruses
175 have also been recruited to function as tailocins, named F-tailocins [59]. Overall, tailocins have
176 apparently evolved from phage tails on at least five independent occasions [59]. Similar to GTA,
177 eCIS are released after cell lysis, mediated by the phage-derived holin and endolysin genes. By
178 contrast, the T6SS are located in bacterial cytoplasm and anchored in the inner membrane. Upon
179 direct contact between bacterial cells, the T6SS inject various toxins into the cytoplasm of the
180 recipient cell. In other cases, the T6SS can target eukaryotic cells, contributing both to defense
181 and to bacterial pathogenicity [60]. Another variation on the theme is presented by the
182 metamorphosis-associated contractile assemblies (MAC), arrays comprised of ~100 contractile
183 phage tail-like structures linked by tail fibers and a dynamic hexagonal net [61]. Remarkably, a
184 marine bacterium, *Pseudoalteromonas luteoviolacea*, produces MACs to trigger metamorphosis
185 of the tubeworm *Hydroides elegans* [61]. The specific evolutionary affinity of the CIS with the
186 tails of particular groups of bacteriophages clearly defines the vector of evolution, from phages
187 to the CIS, which is the route of virus exaptation. This case of exaptation fits the “guns for hire”
188 concept [62]: a group of viral genes that encode the viral device for genome injection into
189 infected cells is exapted as a “weapon” for intercellular conflicts.

190 191 **Virus capsids for transport and storage**

192 Capsids are dedicated devices for containment of the viral nucleic acid. Potentially, capsids
193 could be used as “universal boxes” for storage and transport of various compounds (Figure 1C).
194 Many diverse bacteria and some archaea contain nanocompartments, known as encapsulins,
195 icosahedral particles that resemble bacteriophage capsids and contain multiple pores that are
196 permeable to small molecules but not to proteins [63]. The major encapsulin shell protein is
197 homologous to the HK97-like major capsid protein of tailed bacterial and archaeal viruses of the
198 order *Caudovirales* [8]. The encapsulins package cargo proteins inside the shell where these
199 proteins either catalyze certain reactions or, apparently more often, serve to store ions, e.g.
200 bacterioferritin that binds iron cations, presumably, also defending the cell against oxidative
201 stress. The evolutionary relationships between encapsulins and phage capsids is beyond doubt
202 but the direction of evolution can be questioned [8]. Nevertheless, given the wider spread of
203 viruses with HK97-like capsids compared to encapsulins, the route of virus exaptation appears
204 most likely.

205
206 The encapsulins are distinct from the much larger bacterial mircompartments (BMCs) that
207 compartmentalize a variety of metabolic processes, e.g., CO₂ fixation or catabolism of organic
208 substrates [64]. The BMCs also are icosahedral particles that resemble viral capsids in shape but
209 consist of proteins without detectable viral homologs [65]. Thus, the icosahedral shell has
210 evolved convergently on multiple occasions not only in diverse viruses but also in a cellular
211 context. It is conceivable that viruses with capsids made of BMC proteins remain to be
212 discovered.

213
214 Exaptation of virus capsids occurred in parallel in eukaryotes, in this case, involving the Gag
215 polyprotein of retroviruses. Sequences derived from retroviruses and retrotransposons abound in
216 animal and plant genomes and many cases of exaptation are known, particularly, recruitment of
217 regulatory signals but also protein-coding genes for antiviral defense (see above). Recently,
218 however, it has been demonstrated that Arc proteins that have been known to be involved in
219 neuronal plasticity in both mammalian and insect brains are Gag derivatives that form virus-like
220 particles and transfer RNA molecules between neurons [66,67]. Mammalian genomes encode
221 many additional Gag derivatives. At least 85 genes derived from the *gag* gene of LTR
222 retrotransposons, particularly, Ty3/Gypsy-like metaviruses, have been identified in the human
223 genome [68]. Many of these genes are conserved in other mammals and have been grouped into
224 three major families, the *Mart*, *PNMA* and *SCAN* families [69,70]. Whereas some members of
225 the *Mart* and *PNMA* families have retained (almost) the entire Gag protein sequence, the *SCAN*
226 domain is derived from the C-terminal domain of the capsid protein which fused with multiple
227 C2H2 zinc fingers and/or Krüppel-associated box (KRAB) domains and functions as a protein-
228 interaction module that mediates self-association or selective association with other proteins in
229 vertebrate transcription factors [69]. Proteins from all three families are implicated in various
230 forms of tissue differentiation but exact molecular mechanisms are often not known. It remains
231 to be determined whether or not any of these proteins also function through the formation of
232 virus-like particles, analogously to the case of Arc proteins.

233
234 **Viral fusogens for cell fusion**
235 Not only the Gag but also proteins of the retrovirus envelope (Env) have been exapted for host
236 functions (Figure 1C). Syncitins, the placental receptors, are Env derivatives. The original
237 function of the Env, i.e., to promote fusion between the viral envelope and cellular membrane
238 during virus entry, has been repurposed for the fusion of trophoblast cells during placentation.
239 Remarkably, *env* sequences have been convergently domesticated from different retroviruses
240 infecting germ cells in at least 10 different mammalian lineages [70] as well as in viviparous
241 *Mabuya* lizards [71]. The pan-eukaryotic fusogen HAP2 is homologous to Class II viral fusion
242 proteins [72-74]. In this case, the ultimate provenance of this key protein remains unclear [75]
243 but, given the ubiquity of HAP2 in eukaryotes, cellular origin appears more likely, with viruses
244 acquiring the fusogenic capacity at a later stage. The biological connotation, though, is the same,
245 namely, shared functions of cellular and viral proteins, promoting a two-way gene flow. Indeed,
246 some of the key viral proteins, most notably those involved in virion formation, have been
247 exapted from functionally diverse cellular proteins on multiple independent occasions [8].

248
249 **Endogenous Viral Elements: Sporadic exaptation**
250

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

283

284 **Concluding remarks**

285 Given the ubiquity of viruses in the biosphere and the wide spread of integration of virus
286 genomes into those of the hosts, exaptation of viruses and individual viral genes for host
287 functions is bound to be common, and indeed, an increasing number of cases are being
288 identified. In this brief article, we only considered selected examples that were chosen in such a
289 manner as to illustrate a notable trend: the varying “depths” of the derivation of viruses in the
290 course of exaptation (Figure 2). A virus can be repurposed for a defense function virtually
291 without losing its identity as is the case for Mavirus and, possibly, other polinton-like viruses;
292 else, this type of exaptation can involve notable degradation of the virus as is the case with
293 PDVs. The next step is the complete loss of the viral replication module whereas the structural-
294 morphogenetic module remains intact as illustrated by the GTAs or at least partially retained as
295 in the CIS. Finally, an exapted virus can be reduced to a single gene or even part of a gene that,
296 however, retains certain key capacities such as capsid formation or fusogenic activity.

297
298 In functional terms, exaptation exploits the central properties of viruses. As aggressive parasites,
299 viruses abrogate host functions, particularly, those involved in defense, and this capacity can be
300 repurposed for host protection as it occurs in the case of the virophages. Also, viruses are devices
301 for storage and transfer of nucleic acids, and of that capacity, hosts appear to make ample use for
302 various forms of intercellular communication.

303
304 The most remarkable cases of virus exaptation including the virophages, encapsulins, CIS and
305 Arc particles have been discovered or characterized only recently. There is little doubt that the
306 true depth and breadth of virus exaptation remain to be uncovered.

307
308
309 **Acknowledgements**

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

313 **References**

- 314 1. Iranzo J, Puigbo P, Lobkovsky AE, Wolf YI, Koonin EV: **Inevitability of genetic parasites.**
315 *Genome biology and evolution* 2016, **8**:2856-2869.
- 316 2. Koonin EV: **Viruses and mobile elements as drivers of evolutionary transitions.** *Philos*
317 *Trans R Soc Lond B Biol Sci* 2016, **371**:20150442.
- 318 3. Koonin EV, Wolf YI, Katsnelson MI: **Inevitability of the emergence and persistence of**
319 **genetic parasites caused by evolutionary instability of parasite-free states.** *Biology*
320 *direct* 2017, **12**:31.
- 321 4. Jalasvuori M: **Vehicles, replicators, and intercellular movement of genetic information:**
322 **evolutionary dissection of a bacterial cell.** *International journal of evolutionary biology*
323 2012, **2012**:874153.
- 324 5. Jalasvuori M, Koonin EV: **Classification of prokaryotic genetic replicators: between**
325 **selfishness and altruism.** *Annals of the New York Academy of Sciences* 2015, **1341**:96-
326 105.
- 327 6. Koonin EV, Starokadomskyy P: **Are viruses alive? The replicator paradigm sheds decisive**
328 **light on an old but misguided question.** *Studies in history and philosophy of biological*
329 *and biomedical sciences* 2016, **59**:125-134.
- 330 7. Iranzo J, Cuesta JA, Manrubia S, Katsnelson MI, Koonin EV: **Disentangling the effects of**
331 **selection and loss bias on gene dynamics.** *Proceedings of the National Academy of*
332 *Sciences of the United States of America* 2017, **114**:E5616-E5624.
- 333 8. Krupovic M, Koonin EV: **Multiple origins of viral capsid proteins from cellular ancestors.**
334 *Proceedings of the National Academy of Sciences of the United States of America* 2017,
335 **114**:E2401-E2410.
- 336 9. Krupovic M, Bamford DH: **Order to the viral universe.** *Journal of virology* 2010, **84**:12476-
337 12479.
- 338 10. Iranzo J, Koonin EV, Prangishvili D, Krupovic M: **Bipartite Network Analysis of the**
339 **Archaeal Virosphere: Evolutionary Connections between Viruses and Capsidless**
340 **Mobile Elements.** *Journal of virology* 2016, **90**:11043-11055.
- 341 11. Koonin EV, Dolja VV: **Virus world as an evolutionary network of viruses and capsidless**
342 **selfish elements.** *Microbiology and molecular biology reviews : MMBR* 2014, **78**:278-
343 303.
- 344 12. Krupovic M, Cvirkaite-Krupovic V, Iranzo J, Prangishvili D, Koonin EV: **Viruses of**
345 **archaea: Structural, functional, environmental and evolutionary genomics.** *Virus*
346 *research* 2018, **244**:181-193.
- 347 13. Krupovic M: **Recombination between RNA viruses and plasmids might have played a**
348 **central role in the origin and evolution of small DNA viruses.** *BioEssays : news and*
349 *reviews in molecular, cellular and developmental biology* 2012, **34**:867-870.
- 350 14. Koonin EV, Krupovic M: **Polintons, virophages and transpovirons: a tangled web linking**
351 **viruses, transposons and immunity.** *Current opinion in virology* 2017, **25**:7-15.
- 352 15. Defraia C, Slotkin RK: **Analysis of retrotransposon activity in plants.** *Methods in*
353 *molecular biology* 2014, **1112**:195-210.
- 354 16. Lopez-Flores I, Garrido-Ramos MA: **The repetitive DNA content of eukaryotic genomes.**
355 *Genome dynamics* 2012, **7**:1-28.
- 356 17. Forterre P, Prangishvili D: **The major role of viruses in cellular evolution: facts and**
357 **hypotheses.** *Current opinion in virology* 2013, **3**:558-565.

- 358 18. Koonin EV, Dolja VV: **A virocentric perspective on the evolution of life.** *Current opinion*
359 *in virology* 2013, **3**:546-557.
- 360 19. Gould SJ, Vrba ES: **Exaptation — a missing term in the science of form.** *Paleobiology*
361 1982, **8**:4-15.
- 362 20. Krupovic M, Blomberg J, Coffin JM, Dasgupta I, Fan H, Geering AD, Gifford R, Harrach B,
363 Hull R, Johnson W, et al.: **Ortervirales: New virus order unifying five families of**
364 **reverse-transcribing viruses.** *J Virol* 2018, **92**:e00515-00518.
- 365 21. Gandon S: **Why Be Temperate: Lessons from Bacteriophage lambda.** *Trends in*
366 *microbiology* 2016, **24**:356-365.
- 367 22. Casjens SR, Hendrix RW: **Bacteriophage lambda: Early pioneer and still relevant.**
368 *Virology* 2015, **479-480**:310-330.
- 369 23. Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell
370 **KL: Prophages mediate defense against phage infection through diverse**
371 **mechanisms.** *The ISME journal* 2016, **10**:2854-2866.
- 372 24. Labrie SJ, Samson JE, Moineau S: **Bacteriophage resistance mechanisms.** *Nature reviews.*
373 *Microbiology* 2010, **8**:317-327.
- 374 25. Nethe M, Berkhout B, van der Kuyl AC: **Retroviral superinfection resistance.**
375 *Retrovirology* 2005, **2**:52.
- 376 26. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK: **Cryptic prophages**
377 **help bacteria cope with adverse environments.** *Nature communications* 2010, **1**:147.
- 378 27. Krupovic M, Kuhn JH, Fischer MG: **A classification system for virophages and satellite**
379 **viruses.** *Archives of virology* 2016, **161**:233-247.
- 380 28. Bekliz M, Colson P, La Scola B: **The expanding family of virophages.** *Viruses* 2016,
381 **8**:E317.
- 382 29. Fischer MG, Hackl T: **Host genome integration and giant virus-induced reactivation of**
383 **the virophage mavirus.** *Nature* 2016, **540**:288-291.
- 384 30. Fischer MG, Suttle CA: **A virophage at the origin of large DNA transposons.** *Science*
385 2011, **332**:231-234.
- 386 31. Blanc G, Gallot-Lavallee L, Maumus F: **Provirophages in the Bigelowiella genome bear**
387 **testimony to past encounters with giant viruses.** *Proceedings of the National Academy*
388 *of Sciences of the United States of America* 2015, **112**:E5318-5326.
- 389 32. Pritham EJ, Putliwala T, Feschotte C: **Mavericks, a novel class of giant transposable**
390 **elements widespread in eukaryotes and related to DNA viruses.** *Gene* 2007, **390**:3-17.
- 391 33. Kapitonov VV, Jurka J: **Self-synthesizing DNA transposons in eukaryotes.** *Proceedings of*
392 *the National Academy of Sciences of the United States of America* 2006, **103**:4540-4545.
- 393 34. Krupovic M, Bamford DH, Koonin EV: **Conservation of major and minor jelly-roll**
394 **capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide**
395 **viruses.** *Biology direct* 2014, **9**:6.
- 396 35. Krupovic M, Koonin EV: **Polintons: a hotbed of eukaryotic virus, transposon and**
397 **plasmid evolution.** *Nature reviews. Microbiology* 2015, **13**:105-115.
- 398 36. Koonin EV, Krupovic M: **Virology: A parasite's parasite saves host's neighbours.** *Nature*
399 2016, **540**:204-205.
- 400 37. Santini S, Jeudy S, Bartoli J, Poirot O, Lescot M, Abergel C, Barbe V, Wommack KE,
401 Noordeloos AA, Brussaard CP, et al.: **Genome of Phaeocystis globosa virus PgV-16T**
402 **highlights the common ancestry of the largest known DNA viruses infecting**

- 403 **eukaryotes.** *Proceedings of the National Academy of Sciences of the United States of*
404 *America* 2013, **110**:10800-10805.
- 405 38. Yutin N, Shevchenko S, Kapitonov V, Krupovic M, Koonin EV: **A novel group of diverse**
406 **Polinton-like viruses discovered by metagenome analysis.** *BMC biology* 2015, **13**:95.
- 407 39. Lang AS, Westbye AB, Beatty JT: **The Distribution, Evolution, and Roles of Gene**
408 **Transfer Agents in Prokaryotic Genetic Exchange.** *Annual review of virology* 2017,
409 **4**:87-104.
- 410 40. Hynes AP, Shakya M, Mercer RG, Grull MP, Bown L, Davidson F, Steffen E, Matchem H,
411 Peach ME, Berger T, et al.: **Functional and Evolutionary Characterization of a Gene**
412 **Transfer Agent's Multilocus "Genome".** *Molecular biology and evolution* 2016,
413 **33**:2530-2543.
- 414 41. Solioz M, Marrs B: **The gene transfer agent of *Rhodopseudomonas capsulata*.**
415 **Purification and characterization of its nucleic acid.** *Archives of biochemistry and*
416 *biophysics* 1977, **181**:300-307.
- 417 42. Barbian KD, Minnick MF: **A bacteriophage-like particle from *Bartonella bacilliformis*.**
418 *Microbiology* 2000, **146 (Pt 3)**:599-609.
- 419 43. Tamarit D, Neuvonen MM, Engel P, Guy L, Andersson SGE: **Origin and Evolution of the**
420 ***Bartonella* Gene Transfer Agent.** *Molecular biology and evolution* 2018, **35**:451-464.
- 421 44. Guy L, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, Granberg F, Naslund K,
422 Eriksson AS, Andersson SG: **A gene transfer agent and a dynamic repertoire of**
423 **secretion systems hold the keys to the explosive radiation of the emerging pathogen**
424 ***Bartonella*.** *PLoS genetics* 2013, **9**:e1003393.
- 425 45. Shakya M, Soucy SM, Zhaxybayeva O: **Insights into origin and evolution of alpha-**
426 **proteobacterial gene transfer agents.** *Virus evolution* 2017, **3**:vex036.
- 427 46. Hynes AP, Mercer RG, Watton DE, Buckley CB, Lang AS: **DNA packaging bias and**
428 **differential expression of gene transfer agent genes within a population during**
429 **production and release of the *Rhodobacter capsulatus* gene transfer agent, RcGTA.**
430 *Molecular microbiology* 2012, **85**:314-325.
- 431 47. Drezen JM, Leobold M, Bezier A, Huguet E, Volkoff AN, Herniou EA: **Endogenous viruses**
432 **of parasitic wasps: variations on a common theme.** *Current opinion in virology* 2017,
433 **25**:41-48.
- 434 48. Strand MR, Burke GR: **Polydnviruses: From discovery to current insights.** *Virology*
435 2015, **479-480**:393-402.
- 436 49. Herniou EA, Huguet E, Theze J, Bezier A, Periquet G, Drezen JM: **When parasitic wasps**
437 **hijacked viruses: genomic and functional evolution of polydnviruses.** *Philosophical*
438 *transactions of the Royal Society of London. Series B, Biological sciences* 2013,
439 **368**:20130051.
- 440 50. Theze J, Bezier A, Periquet G, Drezen JM, Herniou EA: **Paleozoic origin of insect large**
441 **dsDNA viruses.** *Proceedings of the National Academy of Sciences of the United States of*
442 *America* 2011, **108**:15931-15935.
- 443 51. Lapointe R, Tanaka K, Barney WE, Whitfield JB, Banks JC, Beliveau C, Stoltz D, Webb
444 BA, Cusson M: **Genomic and morphological features of a banchine polydnvirus:**
445 **comparison with bracoviruses and ichnoviruses.** *Journal of virology* 2007, **81**:6491-
446 6501.

- 447 52. Pichon A, Bezier A, Urbach S, Aury JM, Jouan V, Ravallec M, Guy J, Cousserans F, Theze
448 J, Gauthier J, et al.: **Recurrent DNA virus domestication leading to different parasite**
449 **virulence strategies**. *Science advances* 2015, **1**:e1501150.
- 450 53. Tan CW, Peiffer M, Hoover K, Rosa C, Acevedo FE, Felton GW: **Symbiotic polydnavirus**
451 **of a parasite manipulates caterpillar and plant immunity**. *Proceedings of the National*
452 *Academy of Sciences of the United States of America* 2018, **115**:5199-5204.
- 453 54. Burke GR, Walden KK, Whitfield JB, Robertson HM, Strand MR: **Widespread genome**
454 **reorganization of an obligate virus mutualist**. *PLoS genetics* 2014, **10**:e1004660.
- 455 55. Volkoff AN, Jouan V, Urbach S, Samain S, Bergoin M, Wincker P, Demetree E, Cousserans
456 F, Provost B, Coulibaly F, et al.: **Analysis of virion structural components reveals**
457 **vestiges of the ancestral ichtnovirus genome**. *PLoS pathogens* 2010, **6**:e1000923.
- 458 56. Cascales E: **Microbiology: And Amoebophilus Invented the Machine Gun!** *Curr Biol*
459 2017, **27**:R1170-R1173.
- 460 57. Touchon M, Bobay LM, Rocha EP: **The chromosomal accommodation and domestication**
461 **of mobile genetic elements**. *Curr Opin Microbiol* 2014, **22**:22-29.
- 462 58. Taylor NMI, van Raaij MJ, Leiman PG: **Contractile injection systems of bacteriophages**
463 **and related systems**. *Molecular microbiology* 2018, **108**:6-15.
- 464 59. Scholl D: **Phage Tail-Like Bacteriocins**. *Annual review of virology* 2017, **4**:453-467.
- 465 60. Galan JE, Waksman G: **Protein-Injection Machines in Bacteria**. *Cell* 2018, **172**:1306-
466 1318.
- 467 61. Shikuma NJ, Pilhofer M, Weiss GL, Hadfield MG, Jensen GJ, Newman DK: **Marine**
468 **tubeworm metamorphosis induced by arrays of bacterial phage tail-like structures**.
469 *Science* 2014, **343**:529-533.
- 470 62. Koonin EV, Krupovic M: **A movable defense**. *Scientist* 2015, **29**:46-53.
- 471 63. Nichols RJ, Cassidy-Amstutz C, Chaijarasphong T, Savage DF: **Encapsulins: molecular**
472 **biology of the shell**. *Crit Rev Biochem Mol Biol* 2017, **52**:583-594.
- 473 64. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M: **Bacterial microcompartments**.
474 *Nature reviews. Microbiology* 2018, **16**:277-290.
- 475 65. Krupovic M, Koonin EV: **Cellular origin of the viral capsid-like bacterial**
476 **microcompartments**. *Biology direct* 2017, **12**:25.
- 477 66. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, Yoder N,
478 Belnap DM, Erlendsson S, Morado DR, et al.: **The Neuronal Gene Arc Encodes a**
479 **Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA**
480 **Transfer**. *Cell* 2018, **172**:275-288 e218.
- 481 67. Ashley J, Cordy B, Lucia D, Fradkin LG, Budnik V, Thomson T: **Retrovirus-like Gag**
482 **Protein Arc1 Binds RNA and Traffics across Synaptic Boutons**. *Cell* 2018, **172**:262-
483 274 e211.
- 484 68. Campillos M, Doerks T, Shah PK, Bork P: **Computational characterization of multiple**
485 **Gag-like human proteins**. *Trends Genet* 2006, **22**:585-589.
- 486 69. Kaneko-Ishino T, Ishino F: **The role of genes domesticated from LTR retrotransposons**
487 **and retroviruses in mammals**. *Front Microbiol* 2012, **3**:262.
- 488 70. Naville M, Warren IA, Haftek-Terreau Z, Chalopin D, Brunet F, Levin P, Galiana D, Volff
489 JN: **Not so bad after all: retroviruses and long terminal repeat retrotransposons as a**
490 **source of new genes in vertebrates**. *Clin Microbiol Infect* 2016, **22**:312-323.
- 491 71. Cornelis G, Funk M, Vernochet C, Leal F, Tarazona OA, Meurice G, Heidmann O,
492 Dupressoir A, Miralles A, Ramirez-Pinilla MP, et al.: **An endogenous retroviral**

- 493 **envelope syncytin and its cognate receptor identified in the viviparous placental**
494 **Mabuya lizard.** *Proceedings of the National Academy of Sciences of the United States of*
495 *America* 2017, **114**:E10991-E11000.
- 496 72. Pinello JF, Lai AL, Millet JK, Cassidy-Hanley D, Freed JH, Clark TG: **Structure-Function**
497 **Studies Link Class II Viral Fusogens with the Ancestral Gamete Fusion Protein**
498 **HAP2.** *Curr Biol* 2017, **27**:651-660.
- 499 73. Fedry J, Liu Y, Pehau-Arnaudet G, Pei J, Li W, Tortorici MA, Traincard F, Meola A,
500 Bricogne G, Grishin NV, et al.: **The Ancient Gamete Fusogen HAP2 Is a Eukaryotic**
501 **Class II Fusion Protein.** *Cell* 2017, **168**:904-915 e910.
- 502 74. Valansi C, Moi D, Leikina E, Matveev E, Grana M, Chernomordik LV, Romero H, Aguilar
503 PS, Podbilewicz B: **Arabidopsis HAP2/GCS1 is a gamete fusion protein homologous**
504 **to somatic and viral fusogens.** *J Cell Biol* 2017, **216**:571-581.
- 505 75. Hernández JM, Podbilewicz B: **The hallmarks of cell-cell fusion.** *Development* 2017,
506 **144**:4481-4495.
- 507 76. Aiewsakun P, Katzourakis A: **Endogenous viruses: Connecting recent and ancient viral**
508 **evolution.** *Virology* 2015, **479-480**:26-37.
- 509 77. Feschotte C, Gilbert C: **Endogenous viruses: insights into viral evolution and impact on**
510 **host biology.** *Nature reviews. Genetics* 2012, **13**:283-296.
- 511 78. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K: **Inhibition of Borna disease**
512 **virus replication by an endogenous bornavirus-like element in the ground squirrel**
513 **genome.** *Proceedings of the National Academy of Sciences of the United States of*
514 *America* 2014, **111**:13175-13180.
- 515 79. Suzuki Y, Frangeul L, Dickson LB, Blanc H, Verdier Y, Vinh J, Lambrechts L, Saleh MC:
516 **Uncovering the repertoire of endogenous flaviviral elements in Aedes mosquito**
517 **genomes.** *Journal of virology* 2017, **91**:e00571-00517.
- 518 80. Greenwood AD, Ishida Y, O'Brien SP, Roca AL, Eiden MV: **Transmission, evolution, and**
519 **endogenization: Lessons learned from recent retroviral invasions.** *Microbiol Mol Biol*
520 *Rev* 2018, **82**:e00044-00017.
- 521 81. Frank JA, Feschotte C: **Co-option of endogenous viral sequences for host cell function.**
522 *Current opinion in virology* 2017, **25**:81-89.
- 523 82. Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, Inoue I: **Systematic**
524 **identification and characterization of regulatory elements derived from human**
525 **endogenous retroviruses.** *PLoS genetics* 2017, **13**:e1006883.
- 526 83. Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-**
527 **option of endogenous retroviruses.** *Science* 2016, **351**:1083-1087.
- 528 84. Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H,
529 Cernohorska M, Urbanova J, et al.: **Long terminal repeats power evolution of genes**
530 **and gene expression programs in mammalian oocytes and zygotes.** *Genome Res*
531 2017, **27**:1384-1394.
- 532 85. Blanco-Melo D, Gifford RJ, Bieniasz PD: **Co-option of an endogenous retrovirus**
533 **envelope for host defense in hominid ancestors.** *eLife* 2017, **6**:e22519.
- 534 86. Yap MW, Colbeck E, Ellis SA, Stoye JP: **Evolution of the retroviral restriction gene Fv1:**
535 **inhibition of non-MLV retroviruses.** *PLoS pathogens* 2014, **10**:e1003968.
- 536 87. Monde K, Terasawa H, Nakano Y, Soheilian F, Nagashima K, Maeda Y, Ono A: **Molecular**
537 **mechanisms by which HERV-K Gag interferes with HIV-1 Gag assembly and**
538 **particle infectivity.** *Retrovirology* 2017, **14**:27.

539
540
541

542 **Figure legends**

543 **Figure 1.** Distinct routes of virus exaptation:

544 (a) competent virus serves as a defense systems;

545 (b) Defective virus particles as vehicles for DNA and protein transfer and storage;

546 (c) Exaptation of single viral proteins or regulatory sequences for antiviral defense and other
547 functions.

548

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

550

551 **Highlights**

- 552 • Viruses and their components are repeatedly exapted for diverse host functions
- 553 • Virus components are often exapted to function in antiviral defense
- 554 • Defective viruses are employed for gene transfer and nutrient storage
- 555 • Retroviruses and LTR retrotransposons are a rich source of new cellular functions

