INSTITUT PASTEUR

Research in Microbiology 160 (2009) 466–472

www.elsevier.com/locate/resmic

The origin of viruses

Patrick Forterre [a,b,*], David Prangishvili [b]

[a] Institut de Génétique et Microbiologie, Université Paris-Sud, CNRS UMR 8621, 91405 Orsay Cedex, France
[b] Biologie Moléculaire du Gène chez les Extrêmophiles, Département de Microbiologie, Institut Pasteur, 75724 Paris Cedex 15, France

Received 9 June 2009; accepted 15 July 2009
Available online 30 July 2009

Abstract

Viruses are parasitic organisms that live in infected cells and produce virions to disseminate their genes. Most viral proteins have no homologues in modern cells, in contradiction with the traditional view of viruses as pickpockets of cellular genes. This suggests that viral genes essentially originated in the virosphere during replication of viral genomes and/or were recruited from cellular lineages now extinct. Some specific viral proteins are present in viruses infecting members of the three domains of Life, suggesting that viruses are indeed very ancient. In particular, structural analyses of capsid proteins have revealed that at least two types of virions originated independently before the LUCA (the Last Universal Cellular Ancestor). Although several hypotheses have been recently proposed to explain the origin of viruses, the emergence of virions, as a specific mechanism for gene dissemination, remains unexplained.

Keywords: Virus; Evolution; Last universal cellular ancestor; Origin of viral genes

1. Introduction

It was recently proposed that the living world can be divided between ribosome-encoding organisms (modern cells) and capsid-encoding organisms (viruses) [61]. The origin of modern cells is straightforward, i.e. they all descend from a single ancestor, the LUCA (the last universal cellular ancestor, or the Last Universal CenAncestor) [28,39]. The LUCA was already a complex organism, since the universal protein set contains 33 ribosomal proteins [52]. This means that, in addition to the three rRNA molecules, the ribosome of the LUCA already contained AT LEAST these 33 proteins (it may have contained up to 67 proteins if ribosomal proteins specifically shared by *Archaea* and *Eukarya* were already present in the LUCA) [27]. In agreement with the assumption that the LUCA was a sophisticated organism, the modern universal optimized genetic code [72] was probably already operational in the LUCA. As all modern cells descend from the LUCA, it is theoretically possible (even if practically difficult) to draw a universal tree of Life connecting together all ribosome-encoding organisms [76,77,15]. In contrast, there is not a single informational molecule common to all viruses. In particular, structural analyses of capsid proteins indicate that different types of virions have been selected several times independently as modes of viral genes dissemination ([6] see below). Furthermore, new viruses have emerged during evolution by the mixing of different capsids and viral genomes [47,37]. Accordingly, it will never be possible to draw a universal tree of viruses analogous to the tree of the LUCA [37,57]. Understanding how modern viruses originated thus appears to be a more complex problem from the start than understanding the evolutionary history of modern cells.

2. The nature and origin of viral genes

For some time it has been assumed that viral genomes have been derived from genomes of modern cells (bacteriophages from bacterial genomes and eukaryotic viruses from eukaryotic chromosomes). According to this traditional view, viruses

* Corresponding author. Institut de Génétique et Microbiologie, Université Paris-Sud, CNRS UMR 8621, 91405 Orsay Cedex, France. Tel.: +33 1 45 68 87 91; fax: +33 1 45 68 88 84.
E-mail address: forterre@pasteur.fr (P. Forterre).

0923-2508/$ - see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2009.07.008

are essentially "gene robbers" or "pickpockets" that evolve by recruiting cellular genes (for a recent opinion paper illustrating this view, see [57]). In that paradigm, all viral genes should have a cellular origin *in fine*. As a consequence, the very existence of viral genes, i.e. genes of viral origin, is often practically denied. For instance, Moreira and Brochier-Armanet concluded from the comparison of mimivirus proteins with cellular homologues that mimivirus originated from a mixture of genomes from different cellular organisms [56]. However, those authors restricted the analysis to only 15% of the mimivirus proteome (126 genes out of 911). The origin of mimivirus proteins without cellular homologues (the vast majority) was overlooked by these authors, possibly because they do not match the current paradigm that viruses are "pickpockets" of cellular genes (see their absence in Fig. 2 of the paper by Moreira and Lopez-Garcia, 2009 [57]).

In the traditional view (viruses as robbers of cellular genes), there are only two possibilities for explaining the existence of viral proteins without homologues in modern cells: 1) they were recruited from modern cells whose genomes had not yet been sequenced, or 2) they extensively diverged from their cellular homologues, such that all traces of homology were erased at the amino acid sequence level. The first explanation seems unlikely, since the number of viral-specific proteins does not decrease with increasing numbers of sequenced cellular genomes. In contrast, new viral-specific proteins are uncovered each time a new virus is sequenced. The second explanation, hidden homology at the sequence level because of rapid viral evolution, can also be ruled out, because many viral proteins without cellular homologues, such as helicases of the SFIII family, rolling-circle Rep proteins, or else monomeric RNA polymerases, also have no structural similarity to bona fide cellular proteins [30,45]. Viruses and evolutionarily related plasmids indeed appear to be a reservoir of new protein folds that are not present in the cellular world [6,42,43,51].

In our opinion, the existence of viral-specific proteins requires abandoning the concept of viruses as pickpockets of cellular genes. This leaves us with the question as to whether viral genes could be of "viral" origin. In other words, is it possible for viruses to create new genes? This idea seems a heresy to some biologists because of the traditional confusion of the virus with the virion, i.e. the infectious particle composed of the replicon and the capsid. Of course, new genes cannot appear inside a virion (see [7,16] for critical remarks on this confusion). It is clear that new viral genes can only originate during replication of the viral genome in infected cells. They can be created by several mechanisms, such as gene duplication followed by divergent evolution, recombination between different genes mediated by transposable elements, insertions and deletions producing larger or smaller proteins, or else frameshift mutations producing new open-reading frames (ORFs). Many more possibilities for the creation of new proteins can be tested by variation and selection in viral genomes compared to cellular genomes, considering the huge number of viral progeny produced by each cycle of infection. Many viral genes may therefore have originated in modern or ancient infected cells during the intracellular phase of virus life.

Some specific viral proteins, such as helicases of the SFIII family or Rep proteins for initiation of rolling-circle replication, are present in apparently unrelated viruses infecting members of the three domains [30,45]. Other viral proteins, such as protein-primed DNA polymerases or viral DNA topoisomerases, have cellular homologues, but they form specific clusters in phylogenetic trees, well separated from the three cellular domains [21,22,34]. These viral proteins are probably very ancient and indeed might either have originated during viral replication in infected cells from now extinct lineages or were recruited from these cells. However, if viruses originally were derived from ancient cellular lineages by reductive evolution (see below), some of these viral-specific genes might have been also directly inherited from these lost lineages.

Koonin and colleagues have suggested that widespread viral-specific genes might have originated in an ancient viral world that preceded the cellular world [45]. However, there are many arguments supporting the idea that cells appeared very early on in the evolution of Life (for a review, see [33]). Indeed, early formation of cellular structures was probably a prerequisite for triggering Darwinian evolution of individuals with a sufficiently elaborated metabolism to produce nucleotides and energy in order to initiate formation of the RNA world, i.e. a world of cells with RNA genomes [30]. If we restrict our definition of viruses to organisms producing protein-containing virions (see below), one has to conclude that viruses originated after the first cells (but probably well before modern cells) [30–32].

Jalasvuori and Bamford [41] have recently suggested that Life started with protoviruses infecting non-living vesicles that became protocells. In this scenario, protoviruses may have been membrane-surrounded RNA. However, this corresponds to an evolutionary stage that occurred long before the invention of ribosome, and present-day viral-specific proteins cannot have originated in this protovirus world.

3. Viruses, as cellular organisms producing virions

The idea that viruses, as defined above, originated after the first cells is in line with the recent suggestion that viruses should be considered as a particular type of living cellular organism [38,36]. Lwoff noticed a long time ago that infection of bacteria by a "bacteriophage" transforms the infected cell into a viral factory [53]. In many cases, the genome of the infected cell is destroyed and the only genome remaining operational in the cellular body is the viral genome (for a recent example, see [10]). Such an infected cell is no longer a bacterium, an archaeon or a eukaryote, but the virus itself, a virion-producing organism [38,36]. The viral genome is in total control of the cellular structure that it has stolen from the host cell, which is now dead. In this view, the virus factories (or more precisely virion factory) present in infected eukaryotic cells could be considered as the transient nucleus of a novel organism.

The concept of virus as a different type of cellular organism fits well with the idea that viruses originated from ancient free-living cellular lineages that have been eliminated in the competition with ancestors or descendants of the LUCA, but have managed to survive at the expense of the LUCA's descendants, by hi-jacking their bodies [25,31,32]. In that case, the first viral proteins were by definition of cellular origin, but this origin was far in the past, well before the emergence of the LUCA (see below).

Although the cellular phase of the virus (and not the virion) can be considered as representing the true living viral organism, the production of virions remains the hallmark of viruses. The definition of viruses as capsid-encoding organisms recognizes the existence of two major mechanisms for the vertical transmission of genetic material in the biosphere, through cell division and through the dissemination of infectious virions. In this framework, infectious genetic elements that do not produce capsids, but which probably originated from capsid-encoding organisms, can be viewed as ''orphan replicons'' that have lost their capsids [61]. One of us has recently suggested defining an organism as a collection of ''molecular organs'', such as replicons, capsids, membranes or ribosomes [36]. Viruses defined as capsid-encoding organisms fit with this definition since they are formed by the association of at least two molecular organs, a replicon and a capsid, whereas plasmids or eukaryotic ''viruses'' without capsid (subvital agens [20]) only basically correspond to one ''molecular organ'', the replicon.

4. The impact of viruses on the evolution of Life

The time scale of viral evolution and the diversity of viral lineages that emerged during that time would explain why viral-specific genes now seem to outnumber cellular ones in metagenomic studies [12,3], forming a huge reservoir of molecular biodiversity. The fact that viral genes have probably always outnumbered cellular ones implies that more genes have been transferred from viruses to cells than the opposite during the history of Life [38]. In complete contradiction with the view of viruses as pickpockets of cellular genes, the cells might, in fact, be the real robbers (of viral genes). Daubin and coworkers have shown that most ORFans in *Proteobacteria* are probably of viral origin [17]. This view has been disputed, based on the observation that only a very low proportion of cellular ORFans (around 3%) have homologues in viral databases [79]. However, recent identification of viral-related sequences based on dinucleotide composition and comparative genomics has indicated that the current viral database poorly represents the actual viral genosphere [80]. It was found that, on the average, about 10% of bacterial and archaeal genomes have a recent viral origin (in time course, viral genes integrated into cellular genomes are either eliminated or assimilated to cellular genes, for instance, in terms of GC content or dinucleotide composition). Strikingly, most integrated viral genes identified by in silico analysis in archaeal and bacterial genomes have no homologues in the viral database [80], explaining the puzzling observation of Yin and Fisher [79]. Interestingly, a large proportion of ORFans are located in integrated virus-like elements, strongly suggesting that the observation of Daubin and colleagues on *Proteobacteria* can be extrapolated to the whole archaeal and bacterial worlds, i.e. viruses are the major sources of ORFans in cellular genomes. Far from being simply vehicles that transfer genes from one cell to another, viruses therefore have the potential to introduce new functions in the genomes of their hosts (functions that emerged in the virosphere) with a tremendous impact on cellular evolution. This impact is not limited to the realm of *Archaea* and *Bacteria*. In fact, the proportion of viral genes in cellular genomes may be even higher in eukaryotes, whose cells can accommodate much larger genomes, reducing the selection pressure to remove foreign DNA. For instance, it has been estimated that around 45% of the human genome has a retroviral origin [8,18]. The consequences of this viral invasion have been dramatic (see below) and suggest that viruses indeed ''manipulate'' the evolution of their hosts.

The importance and antiquity of gene flux from viruses to cells (from the RNA/protein world until present) indeed implies that viruses have been major players in the evolution of Life. Not only have they always imposed high selection pressure on their hosts and manipulated the whole environment [62], but they have also been a (the) major source of novel functions in cellular lineages via the insertion of genes of viral origin into cellular genomes. According to recent hypotheses, viruses might have played a direct role in the origin of DNA [29] and DNA replication mechanisms [22,26,29,73], in the origin of the eukaryotic nucleus [9,70], cellular envelopes [41] of pathogenicity [14], alternative genetic codes [65] and formation of the three domains of Life, *Archaea*, *Bacteria* and *Eukarya* [31,41]. Although these historical hypotheses will always remain controversial, the study of more recent evolutionary events has indeed shown that viruses have the ability to create novelties and to modify the evolutionary trajectory of their hosts [13,63,74,75]. Viruses, for instance, shaped the transcription and replication systems of mitochondrial genomes [23,66]. Recently, the role of retroviruses in the evolution of novel features in vertebrates has been especially well documented (for a monography, see Ryan [64]). For instance, they have provided proteins that are pivotal for the construction of the immune system [74,75], embryo development [68], and placentation in mammals [11,55], providing immunosuppressive activity that protects the embryo from the immune system of the mother [54]. Recent data strongly suggest that endogenous retroviral proteins could also play crucial roles in brain physiology [58].

5. The origin of virions

Viruses originated from the association of a replicon, a capsid and the associated genome packaging, entry and extrusion mechanisms. The origin of replicons is a general problem, which encompasses both the origin of viruses and those of cells. The most specific aspect of the origin of viruses is therefore the problem of the origin of virions. Virions are usually formed by a nucleoprotein core which is surrounded

by capsid proteins. Some virions contain a lipid membrane that can be located either inside or outside of the capsid. The first infectious virion-like structure might have originated from nucleoprotein complexes that were transferred from one RNA cell to another by cell fusion. Such nucleoprotein structure could have appeared as a means of nucleic acid protection before being coupled to cell–cell transfer and RNA replication. It has possibly been easier to couple formation of virion-like structures to genome replication at a time when the genomes were probably smaller than in modern cells. This might have given a tremendous advantage to intracellular parasites by allowing them to multiply and disseminate their genomes independently of the genome of their hosts. It even became possible for the parasites to kill their hosts in order to benefit from its raw material to produce more virions. The parasites at the origin of viruses might have been either small infectious RNA cells, infectious RNA chromosomes, or membrane vesicles containing infectious RNA (protoviruses) [32,35,41].

Many modern virions enter their target cell by fusion of the cell membrane and the virion envelope, reminiscent of endosome formation. Escape of virions from the infected cell sometimes also occurs by a budding mechanism reminiscent of the extrusion of membrane vesicles (exosome formation). Infection of ancient cells by smaller parasitic cells (possible virus precursors) might have also involved membrane fusions, sometimes leading to the disruption of the cellular structure of the parasite. These observations and reasoning suggest that vesicle membrane trafficking and membrane reorganization might have played an important role in the origin of the first viruses.

Interestingly, modern cells from the three domains of Life produce membrane bond virus-like vesicles [67 and references therein]. These vesicles are often strikingly similar in appearance to spherical virions. They are also sometimes tightly bound to cellular DNA, and it has been suggested that such vesicles could be vehicles for horizontal gene transfer [68]. Recently, archaeal homologues of eukaryotic ESCRTIII proteins, which are involved in both eukaryotic vesicle trafficking and virus budding, have been found in membrane bond vesicles produced by the crenarchaeon *Sulfolobus* [19]. It is therefore tempting to speculate on a possible link between the widespread capacity of cells to produce membrane bond virus-like vesicles (that probably predated the LUCA) and the origin of virions. Jalasvuori and Bamford, for instance, suggested that protoviruses already originated at the onset of the first age of the RNA world, when a replicative RNA bound to the membrane of a primitive but large vesicle had acquired the capacity to trigger budding of small vesicles that were then able to fuse with another vesicle (infection). Thus, a better knowledge of the mechanism of production of modern vesicles could possibly shed light on the origin of viruses. It should be especially important to determine whether evolutionary relationships exist between some proteins specific for vesicle formation and capsid proteins (or more generally speaking with proteins found in virions).

Virus-like particles (not membrane bond) produced by the archaeon *Pyrococcus furiosus* [2] and the bacterium *Thermotoga maritima* [69] have been recently characterized at the structural level. These polyhedral particles are nanocompartiments formed by shell-forming proteins (encapsulin) which share homologous domains with capsid proteins of bacterial caudovirales and eukaryotic Herpesviruses (see below). They are used by cells to package enzymes and confine specific metabolic pathways. Archaeal and bacterial encapsulins are evolutionarily closely related and could be remnants of viral capsids that were recruited to form nanocompartiments possibly already before the divergence between the *Archaea* and *Bacteria* [69]. Similar but larger virus-like particles (microcompartments) are present in many bacteria. The best characterized are carboxysomes that sequester enzymes involved in carbon fixation [71,78]. Although the shell-forming proteins of carboxysomes have no structural similarity with known viral capsid proteins [44], they might also be remnants of ancient viral capsid proteins no longer present in modern viruses or present in viruses that have not yet been discovered or studied at the atomic level. Sutter et al. [69] suggest that "*viruses* (i.e. virions) could have originated from a similar cellular assembly by a switch of the specificity from encapsulating proteins to encapsulation of nucleic acids". Indeed, one can speculate that nano- or microcompartments already existed in some cells of the second age of the RNA world [30] for metabolic compartmentalization and were recruited in some early viral lineages to build more stable virions.

6. How many viral lineages?

Structural analysis of modern virions (especially of capsid proteins) has revealed that capsids evolved independently at least twice [4–6,48]. A first type of capsid is characterized by proteins with either single or double-jelly-roll folds [1,5,6,48]. These structures are present in the capsid proteins of both RNA and DNA viruses and of viruses infecting the three domains of Life, indicating that capsids based on jelly-roll proteins already had been evolved before the LUCA, probably in the second age of the RNA world [5,30,32]. A second type of protein capsid structure has been detected in bacterial head-and-tail viruses (and is probably present in all caudovirales), in herpesviruses (infecting eukaryotes) and in archaeal and bacterial encapsulin [2,4,69]. The type of capsid (HK97) corresponding to this structure was therefore probably also "invented" before the LUCA [6]. More types of capsids (and capsid proteins) will probably be defined in the next few years. Recently, Bamford and colleagues described a virus infecting halophilic *Archaea* that produces pleiomorphic virions with a membrane envelope that resembles those of viruses infecting *Mycoplasma* spp. They suggest that this type of virion could represent a third viral lineage, encompassing more than one cellular domain [59].

It will be important in the next few years, through extensive studies of the structure of capsid proteins from all known viruses, to determine how many lineages of different capsid proteins can be identified. This will tell us how many times "viruses" originated independently and whether the process has been a relatively easy one (having occurred many times) or

whether it occurred only rarely, being a highly unlikely event. It will be especially important in the future to determine the structure of structural proteins from virions with unusual morphology, such as, e.g. the unusual bottle-shaped archaeal virus of the family *Ampullaviridae* [40].

Moreira and Lopez-Garcia [57] recently questioned the antiquity of viruses by suggesting that structural similarities observed between capsid proteins from viruses infecting cells from different domains are not a convincing argument for the existence of ancestral capsid proteins that predated the LUCA, but could be explained either by convergent evolution or past transfer of viruses from one cellular domain to another, e.g. a virus infecting a cell from one of the three domains learned to infect a cell from another domain. The convergence argument has been refuted convincingly for a long time by Bamford and colleagues, based on exhaustive structural analyses of virions with double-jelly-roll fold capsid proteins (see e.g. [48]). The existence of two completely different types of capsid proteins – the jelly-roll fold and the HK97 fold – in fact argues directly against the idea that only one type of protein fold is compatible with the assemblage of a capsid structure.

The possibility that a virus shifts its specificity for a cellular host from one domain to another also appears unlikely. In fact, preliminary data suggest that at least some viruses have co-evolved with their hosts from the time of the LUCA. The double-jelly-roll fold story is a good example. Indeed, whereas only structural similarities can be detected between double-jelly-roll fold capsid proteins from viruses infecting hosts from different domains, it is possible to detect sequence similarities between double-jelly-roll fold capsid proteins from viruses infecting different phyla of the same domain [49]. A challenge in upcoming years will be to obtain much more structural and phylogenetic data to test the co-evolution hypothesis in a rigorous way.

While one cannot draw a universal tree of Life for viruses since there is no universal viral protein, it has been suggested that we immerse the universal tree of cellular Life into a viral ocean [5]. In our opinion, this metaphore is somehow misleading if viruses indeed mainly co-evolved with their hosts, and transfer of viruses from one domain to another never or only rarely occurred. To take into account both long-range virus-host co-evolution and the existence of evolutionary relationships between viral lineages, we recently suggested a model in which three different parts of the ancient viral world were selected and split up by the formation of the three cellular domains [60]. This model tries to explain why viruses infecting members of different domains are clearly different, whilst some of them exhibit homologous capsid proteins.

7. The universal viral network of Life

It has been suggested that viral lineages which are based on virion structure should be considered true viral lineages, with the capsid being the hallmark of viruses [5]. This notion causes a problem, considering that different viruses with evolutionarily related virions can have non-homologous replication machineries and vice-versa [50,46,47]. Indeed, as previously discussed, viruses as organisms can be viewed as assemblages of several gene clusters or cassettes encoding ‘‘molecular organs’’ [36]. Schematically, one can distinguish three types of cassettes in viral genomes: one grouping genes encoding proteins involved in virion architecture (the capsid cassette), another grouping genes encoding proteins involved in genome replication (the replicon cassette), and the third grouping genes involved in the manipulation of host functions (i.e. in the transformation of the infected cell into a virion factory). The capsid and replicon cassettes are usually quite stable, forming a core of conserved genes in viral families. Their genes also often exhibit conserved synteny in their organization along the chromosome (this has been nicely analyzed in the case of the bacteriovirus T4 group, see [24]). In each type of cassettes, proteins apparently co-evolved because of functional and/or physical constraints due to their tight interactions (this is probably especially true in the case of the capsid cassette). A given association of a replicon and capsid cassettes forms a specific viral family. Such association can be very stable (especially in the case of viruses with large genomes) and both types of cassettes can co-evolve together for long evolutionary periods [24]. However, new viruses can apparently appear from time to time by exchange of clusters originating from different virus lineages. This has been nicely demonstrated by the work of Krupovic and Bamford, who have shown that bacterial corticoviruses used at least four different replication cassettes (including several of plasmid origin) to replicate their genomes [47]. New types of viruses have thus originated in the ‘‘corticovirus’’ lineage by exchanging their replication cassettes with replication cassettes of plasmid origin. In fact, plasmid replication cassettes probably originated from ancient viruses that have lost their capsid cassette and/or from genomes of ancient cellular lineages that have now disappeared as free-living entities but manage to parasitize cells from successful lineages. The focus on the capsid (the virus ‘‘self’’) is in line with the definition of viruses as organisms producing virions [3,61]. However, it will also be very important to reconstruct the history of the replicons in order to gain a comprehensive view of virus origin and evolution. Instead of a single tree, as in the case of the descendents of the LUCA, the evolution of viral lineages will be best represented by a network of connected capsid and replicon lineages [37]. In this framework, one should distinguish the ‘‘primordial’’ origin of viruses, i.e. the first invention of virions in ancient cellular lineages, from the more recent origin of new viruses by recombination of pre-existing viruses and/or plasmids.

8. Conclusion: viruses testify of our ancient history

Finally, is it relevant to discuss the origin of viruses in the context of the origin of Life? Although viruses (defined as capsid-encoding organisms) clearly originated at a relatively late stage of the evolution of Life (the second age of the RNA world, after the ‘‘invention’’ of modern proteins [30]), they are an intrinsic part of Life as we know it and, accordingly, the origin of life encompasses the origin of viruses. Furthermore,

from their very ancient origin, modern viruses still appear to carry molecular relics, both in term of proteins and functions of cellular lineages that have otherwise disappeared [25]. We have mentioned several times in this essay the existence of extinct cellular lineages that thrived on our planet either before or after the LUCA. It is sometimes argued that it is not parsimonious to refer to such extinct lineages and that one should explain all aspects of ancient cellular evolution based on the characteristics of extant cells (this reasoning explains the numerous speculations in which eukaryotes emerged via the association of a bacterium and an archaeon). Accordingly, in the case that we had no fossils of any extinct *Homo* species, some scientists could then argue that it is not parsimonious to explain the origin of *Homo sapiens* by its evolution from *Homo* species that have now disappeared! Evolution is not necessarily parsimonious and the very existence of viral-specific proteins, such as reverse transcriptase or else protein-primed polymerase, can be viewed as a window into a time when the diversity of cellular lineages and molecular mechanisms was much greater than today. One should be grateful for viruses to remind us of some lost roads of our common history and to enrich our vision of our origins.

Acknowledgments

We would like to thank Simonetta Gribaldo for stimulating and critical discussions on the nature of viruses.

References

[1] Abrescia, N.G., Grimes, J.M., Kivelä, H.M., Assenberg, R., Sutton, G.C., Butcher, S.J., Bamford, J.K., Bamford, D.H., Stuart, D.I. (2008) Insights into virus evolution and membrane biogenesis from the structure of the marine lipid-containing bacteriophage PM2. Mol. Cell 31, 749–761.

[2] Akita, F., Chong, K.T., Tanaka, H., Yamashita, E., Miyazaki, N., Nakaishi, Y., Suzuki, M., Namba, K., Ono, Y., Tsukihara, T., Nakagawa, A. (2007) The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. J. Mol. Biol. 368, 1469–1483.

[3] Angly, F.E., Felts, B., Breitbart, M., Salamon, P., Edwards, R.A., Carlson, C., Chan, A.M., Haynes, M., Kelley, S., Liu, H., Mahaffy, J.M., Mueller, J.E., Nulton, J., Olson, R., Parsons, R., Rayhawk, S., Suttle, C.A., Rohwer, F. (2006) The marine viromes of four oceanic regions. PLoS Biol. 4(11), e368.

[4] Baker, M.L., Jiang, W., Rixon, F.J., et al. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. J. Virol. 79, 14967–14970.

[5] Bamford, D.H. (2003) Do viruses form lineages across different domains of life? Res. Microbiol. 154, 231–236.

[6] Bamford, D.H., Grimes, J.M., Stuart, D.I. (2005) What does structure tell us about virus evolution? Curr. Opin. Struct. Biol. 15, 655–663.

[7] Bandea, C. (1983). J. Theor. Biol. 105, 591–602.

[8] Bannert, N., Kurth, R. (2004) Retroelements and the human genome: new perspectives on an old relation. Proc. Natl. Acad. Sci. U.S.A. 101(Suppl. 2), 14572–14579.

[9] Bell, P.J. (2001) Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus? J. Mol. Evol. 53, 251–256.

[10] Bize, A., Karlsson, E.A., Ekefjärd, K., Quax, T.E.F., Pina, M., Prevost, M.-C., Forterre, P., Tenaillon, O., Bernander, R., Prangishvili, D. (2009) A unique virus release mechanism in the Archaea. Proc. Natl. Acad. Sci. U.S.A 10, 11306–11311.

[11] Blond, J.-L., Lavillette, D., Cheynet, V., et al. (2000) An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. J. Virol. 74, 3321–3329.

[12] Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J.M., Segall, A.M., Mead, D., Azam, F., Rohwer, F. (2002) Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. U.S.A. 99, 14250–14255.

[13] Brosius, J. (2003) The contribution of RNAs and retroposition to evolutionary novelties. Genetica 118, 99–116.

[14] Brüssow, H. (2007) Bacteria between protists and phages: from antipredation strategies to the evolution of pathogenicity. Mol. Microbiol. 36, 583–589.

[15] Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283–1287.

[16] Claverie, J.M. (2006) Viruses take center stage in cellular evolution. Genome Biol. 7, 110.

[17] Daubin, V., Lerat, E., Perrière, G. (2003) The source of laterally transferred genes in bacterial genomes. Genome Biol. 4, R57.

[18] De Parseval, N., Heidmann, T. (2005) Human endogenous retroviruses: from infectious elements to human genes. Cytogenet. Genome Res. 110, 318–332.

[19] Ellen, A.F., Albers, S.V., Huibers, W., Pitcher, A., Hobel, C.F., Schwarz, H., Folea, M., Schouten, S., Boekema, E.J., Poolman, B., Driessen, A.J. (2009) Proteomic analysis of secreted membrane vesicles of archaeal Sulfolobus species reveals the presence of endosome sorting complex components. Extremophiles 13, 67–79.

[20] Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Bail, L.A. (2005) Virus taxonomy, eight report of the international committee on taxonomy of viruses. Amsterdam: Elsevier.

[21] Filée, J., Forterre, P., Sen-Lin, T., Laurent, J. (2002) Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. J. Mol. Evol. 54, 763–773.

[22] Filée, J., Forterre, P., Laurent, J. (2003) The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. Res. Microbiol. 154, 237–243.

[23] Filée, J., Forterre, P. (2005) Viral proteins functioning in organelles: a cryptic origin? Trends Microbiol. 13, 510–513.

[24] Filée, J., Bapteste, E., Susko, E., Krisch, H.M. (2006) A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol. Biol. Evol. 23, 1688–1696.

[25] Forterre, P. (1992) New hypotheses about the origins of viruses, prokaryotes and eukaryotes. In J.K. Thanh Vân, J.C. Mounolou, J. Shneider, & C. Mc Kay (Eds.), Frontiers of Life, Edition Frontières (pp. 221–234), Gif-sur-Yvette, France.

[26] Forterre, P. (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. Mol. Microbiol. 33, 457–465.

[27] Forterre, P., Philippe, H. (1999) Where is the root of the universal tree of life? Bioessays 10, 871–879.

[28] Forterre, P. (2001) Genomics and early cellular evolution. The origin of the DNA world. C.R. Acad. Sci. III 324, 1067–1076.

[29] Forterre, P. (2002) The origin of DNA genomes and DNA replication. Curr. Opin. Microbiol. 5, 525–532.

[30] Forterre, P. (2005) The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie 87, 793–803.

[31] Forterre, P. (2006) Three RNA cells for ribosomal lineages and three DNA viruses to replicate their genomes: a hypothesis for the origin of cellular domain. Proc. Natl. Acad. Sci. U.S.A. 103, 3669–3674.

[32] Forterre, P. (2006) The origin of viruses and their possible roles in major evolutionary transitions. Virus Res. 117, 5–16.

[33] Forterre, P., Gribaldo, S. (2007) The origin of modern terrestrial life. HFSP J. 1, 156–168.

[34] Forterre, P., Gribaldo, S., Gadelle, D., Serre, M.C. (2007) Origin and evolution of DNA topoisomerases. Biochimie 89, 427–446.

[35] Forterre, P. (2008) Origin of viruses. In B.W.J. Mahy, & M.H.V. Van Regenmortel (Eds.), (3rd ed.)Encyclopedia of virology vol. 5 (pp. 472–479) Oxford: Elsevier.

[36] Forterre, P. Defining life: the virus viewpoint. Origin of life and Evol. Biosphere, in press.
[37] Forterre, P. (2009) Evolution, viral. In M. Schaechter (Ed.), Encyclopedia of microbiology, (3rd ed.) (pp. 370–389) Oxford: Elsevier.
[38] Forterre, P., Prangishvili, D. The great billion-year war between ribosome- and capsid-encoding organisms (cells and viruses) as the major source of evolutionary novelties. Proc. N.Y. Acad. Sci., in press.
[39] Glansdorff, N., Xu, Y., Labedan, B. (2008) The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. Biol. Direct 3, 29.
[40] Häring, M., Rachel, R., Peng, X., Garrett, R.A., Prangishvili, D. (2005) Viral diversity in hot springs of Pozzuoli, Italy, and characterization of a unique archaeal virus, acidianus bottle-shaped virus, from a new family, the Ampullaviridae. J. Virol. 79(15), 9904–9911.
[41] Jalasvuori, M., Bamford, J.K.H. (2008) Structural co-evolution of viruses and cells in the primordial world. Orig. Life Evol. Biosph. 38, 165–181.
[42] Keller, J., Leulliot, N., Collinet, B., Campanacci, V., Cambillau, C., Pranghisvilli, D., van Tilbeurgh, H. (2009) Crystal structure of AFV1-102, a protein from the acidianus filamentous virus 1. Protein Sci. 18, 845–849.
[43] Keller, J., Leulliot, N., Soler, N., Collinet, B., Vincentelli, R., Forterre, P., Van Tilbeurgh, H. (2009) A protein encoded by a new family of mobile elements from euryarchaea exhibits three domains with novel folds. Protein Sci. 18(4), 850–855.
[44] Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M., Yeates, T.O. (2005) Protein structures forming the shell of primitive bacterial organelles. Science 309, 936–938.
[45] Koonin, E.V., Senkevich, T.G., Dolja, V.V. (2006) The ancient virus world and evolution of cells. Biol. Direct 9, 1–29.
[46] Koonin, E.V., Wolf, Y.I., Nagasaki, K., Dolja, V.V. (2009) The complexity of the virus world. Nat. Rev. Microbiol. [Epub ahead of print].
[47] Krupovic, M., Bamford, D.H. (2007) Putative prophages related to lytic tailess marine dsDNA phage PM2 are widespead in the genome of aquatic bacteria. BMC Genomics 8, 236.
[48] Krupovic, M., Bamford, D.H. (2008) Virus evolution: how far does the double beta-barrel viral lineage extend? Nat. Rev. Microbiol. 6, 941–948.
[49] Krupovic, M., Bamford, D.H. (2008) Archaeal proviruses TKV4 and MVV extend the PRD1-adenovirus lineage to the phylum Euryarchaeota. Virology 375, 292–300.
[50] Krupovic, M., Bamford, D.H. (2009) Does the evolution of viral polymerases reflect the origin and evolution of viruses? Nat. Rev. Microbiol. 7, 250.
[51] Lawrence, C.M., Menon, S., Eilers, B.J., Bothner, B., Khayat, R., Douglas, T., Young, M.J. (2009) Structural and functional studies of archaeal viruses. J. Biol. Chem. 284, 12599–12603.
[52] Lecompte, O., Ripp, R., Thierry, J.C., Moras, D., Poch, O. (2002) Comparative analysis of ribosomal proteins in complete genomes: an example of reductive evolution at the domain scale. Nucleic Acids Res. 30, 5382–5390.
[53] Lwoff, A. (1967) Principles of classification and nomenclature of viruses. Nature 215, 13–14.
[54] Manganey, M., Renard, M., Schlecht-Louf, G., Bouallaga, I., Heidmann, O., Letzelter, C., Richaud, A., Ducos, B., Heidmann, T. (2007) Placental syncytins: genetic dysjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. Proc. Natl. Acad. Sci. U.S.A. 104, 20534–20539.
[55] Mi, S., Lee, X., Li, X.P., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., Keith Jr., J.C., McCoy, J.M. (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403, 785–789.
[56] Moreira, D., Brochier-Armanet, C. (2008) Giant viruses, giant chimeras: the multiple evolutionary histories of Mimivirus genes. BMC Evol. Biol. 8, 12.
[57] Moreira, D., López-García, P. (2009) Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol 4, 306–311.
[58] Perron, H., Lazarini, F., Ruprecht, K., Pechoux-Longin, C., Seilhean, D., Sazdovitch, W., Creange, A., Battail-Poirot, N., Sibaii, G., Santoro, L., Jolivet, M., Darlix, J.L., Rieckmann, P., Arzberger, T., Hauw, J.J., Lassmann, H. (2005) Human endogenous retrovirus (HERV)-W ENV and GAG proteins: physiological expression in human brain and pathophysiological modulation in multiple sclerosis lesions. J. Neurovirol. 11, 23–33.
[59] Pietilä, M.K., Roine, E., Paulin, L., Kalkkinen, N., Bamford, D.H. (2009) An ssDNA virus infecting Archaea: a new lineage of viruses with a membrane envelope. Mol. Microbiol. 72, 307–319.
[60] Prangishvili, D., Forterre, P., Garrett, R.A. (2006) Viruses of the Archaea: a unifying view. Nat. Rev. Microbiol. 4, 837–848.
[61] Raoult, D., Forterre, P. (2008) Redefining viruses: lessons from Mimivirus. Nat. Rev. Microbiol. 6, 315–319.
[62] Rohwer, F., Thurber, R.V. (2009) Viruses manipulate the marine environment. Nature 459, 207–212.
[63] Ryan, R.F. (2007) Viruses as symbionts. Symbiosis 44, 11–21.
[64] Ryan, R.F. (2009) Virolution. Collins.
[65] Shackelton, L.A., Holmes, E.C. (2008) The role of alternating genetic codes in viral evolution and emergence. J. Theor. Biol. 254, 128–134.
[66] Shutt, T.E., Gray, M.W. (2006) Bacteriophage origins of mitochondrial replication and transcription proteins. Trends Genet. 22, 90–95.
[67] Soler, N., Marguet, E., Verbavatz, J.M., Forterre, P. (2008) Virus-like vesicles and extracellular DNA produced by hyperthermophilic Archaea of the order Thermococcales. Res. Microbiol. 159, 390–399.
[68] Spadafora, C. (2008) A reverse transcriptase-dependent mechanism plays central roles in fundamental biological processes. Syst. Biol. Reprod. Med. 54, 11–21.
[69] Sutter, M., Boehringer, D., Gutmann, S., Günther, S., Prangishvili, D., Loessner, M.J., Stetter, K.O., Weber-Ban, E., Ban, N. (2008) Structural basis of enzyme encapsulation into a bacterial nanocompartment. Nat. Struct. Mol. Biol. 15, 939–947.
[70] Takemura, M. (2001) Poxviruses and the origin of the eukaryotic nucleus. J. Mol. Evol. 52, 419–425.
[71] Tanaka, S., Kerfeld, C.A., Sawaya, M.R., Cai, F., Heinhorst, S., Cannon, G.C., Yeates, T.O. (2008) Atomic-level models of the bacterial carboxysome shell. Science 319, 1083–1086.
[72] Vetsigian, K., Woese, C., Goldenfeld, N. (2006) Collective evolution and the genetic code. Proc. Natl. Acad. Sci. U.S.A 103, 10696–10701.
[73] Villarreal, L.P., De Filippis, V.R. (2000) A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J. Virol. 74, 7079–7084.
[74] Villarreal, L.P. (2005) Viruses and the evolution of life. Washington: ASM Press.
[75] Villarreal, L. The source of self: genetic parasites and the origin of adaptive immunity. Annals of the New York Academy of Sciences, in press.
[76] Woese, C.R., Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. U.S.A. 74, 5088–5090.
[77] Woese, C.R., Kandler, O., Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. U.S.A. 87, 4576–4579.
[78] Yeates, T.O., Kerfeld, C.A., Heinhorst, S., Cannon, G.C., Shively, J.M. (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nat. Rev. Microbiol. 6, 681–691.
[79] Yin, Y., Fischer, D. (2006) On the origin of microbial ORFans: quantifying the strength of the evidence for viral lateral transfer. BMC Evol. Biol. 6, 63.
[80] Cortez, D., Forterre, P., Gribaldo, S. (2009) A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. Genome Biol. 10(6), R65.