Opinion
Viruses take center stage in cellular evolution
Jean-Michel Claverie

Address: Structural and Genomic Information Laboratory, CNRS-UPR2589, IBSM, Parc Scientifique de Luminy, 163 Avenue de Luminy, case 934, Marseille 13288, cedex 9, France. Email: Jean-Michel.Claverie@igs.cnrs-mrs.fr

Abstract

The origins of viruses are shrouded in mystery, but advances in genomics and the discovery of highly complex giant DNA viruses have stimulated new hypotheses that DNA viruses were involved in the emergence of the eukaryotic cell nucleus, and that they are worthy of being considered as living organisms.

The reputedly intractable problem of the origin of viruses has long been neglected. In the modern literature, 'virus evolution' has come to refer to studies more akin to population genetics, such as the worldwide scrutiny of new polymorphisms appearing daily in the H5N1 avian flu virus [1], than to the fundamental question of where viruses come from. This is now rapidly changing, as a result of the coincidence of bold new ideas (and the revival of old ones), the unexpected spectacular features of some recently isolated giant viruses [2,3], as well as the steady increase in the numbers of genomic sequences for 'regular' viruses and cellular organisms, which enhances the power of comparative genomics [4]. After being considered non-living and relegated to the wings by most biologists, viruses are now center stage: they might have been there at the origin of DNA, might have played a central role in the emergence of the eukaryotic cell, and might even have been the cause of partitioning of biological organisms into the three domains of life: Bacteria, Archaea and Eukarya. In this article, I shall briefly survey some of the recent discoveries and the new evolutionary thoughts they have prompted, before adding to the discussion with a question of my own: what if we have totally missed the true nature of (at least some) viruses?

Ancient viruses as the origin of different domains

As of April 2006, more than 1,600 viral genomes have been sequenced, approximately equally divided between RNA and DNA viruses. In view of this fundamental difference in their genetic material (and thus in their replication mechanisms, size, genetic complexity, host range and other features) it is tempting to immediately rule out the idea that viruses are monophyletic, that is, that they derive from a common ancestor. That might not be so easy to do, however. Although there are many arguments in favor of the idea that RNA and DNA viruses were generated independently - RNA viruses first, in the context of the 'RNA world' theory - their genesis might have overlapped quite significantly either before or shortly after the Last Universal Common Ancestor (LUCA, the last unique ancestor of all cellular life, reviewed in [2]), allowing a non-negligible level of genome mixing. Indeed, several proteins have homologs in both RNA and DNA viruses, the most important of all being the jelly-roll capsid protein [5], the sole protein that is found in most viruses and not found in cellular organisms [6]. Other components are shared between the two types of viruses, but these are considered to be the results of more recent lateral gene transfers; they include the chaperonin Hsp70, which is found in the giant double-stranded DNA (dsDNA) mimivirus [7] and the positive-strand RNA closteroviruses [8].

The notion that viruses might be very ancient (and even ancestral to cells, as proposed by d'Herelle, the discoverer of bacteriophages [9]) has become the starting point of increasingly daring evolutionary scenarios, modernized to take into account our present knowledge of molecular biology and genomics [10,11]. To explain the puzzling phylogenies and distribution of many DNA informational proteins (proteins involved in the replication and transcription of DNA)

between the three domains of life, it has been proposed that DNA viruses could be the origin of present-day eukaryotic replication proteins [12,13]. Other researchers postulate that a large poxvirus-like dsDNA virus might be the origin of the eukaryotic nucleus, taken in by an ancestral cell and adapted as an organelle - the notion of viral eukaryogenesis [14,15]. I personally find the general idea that a nucleus is functionally equivalent to a selfish DNA virus (that is, replicating 'its' DNA using the cellular metabolism) simple and very appealing - and even more so when one realizes that the idea can be turned on its head to envisage the nucleus of a (primitive) eukaryote (re-)turning into a large DNA virus - the notion of nuclear viriogenesis (Figure 1). Of particular interest, such a transfer of an 'infectious' nucleus is well documented in many parasitic red algae [16].

Such back-and-forth eukaryogenesis-viriogenesis could readily explain the multiplicity of present-day virus lineages, together with their diversity in size, complexity and gene complement, as well as the apparent mixture of monophyly and polyphyly (descent from more than one ancestor) exhibited by the viral world. In this context, extant complex eukaryotic DNA viruses could have originated from iterative waves of nuclear viriogenesis. But we still need some initial 'seeding' virus, the one that, for instance, invented the proto-type of the now nearly ubiquitous jelly-roll capsid protein. Reviving d'Herelle's initial 'virus first' hypothesis, Koonin

and Martin [17] paradoxically proposed that RNA viruses might have emerged even before the invention of individual cells, as selfish RNA replicons roaming prebiotic inorganic compartments. There is little chance, however, that this hypothesis could be scientifically proven anytime soon.

Also quite provocative is the idea that RNA viruses might be at the origin of DNA biochemistry [2,18]. According to this scenario, RNA-based viruses infecting RNA-based cells would have acquired an RNA-to-DNA modification system to resist cellular RNA-degrading enzymes (the RNA equivalent of present-day bacterial restriction and modification systems). For this to happen, RNA viruses would have had to evolve the ribonucleotide reductase enzyme, to convert diphosphate-ribonucleotides to diphosphate-deoxyribonucleotides, and thymidylate synthase, to make dTMP from dUMP, the two key pathways in DNA synthesis. Cellular RNA was then replaced by DNA in the course of evolution because of its greater stability and the capacity for repair conferred by its double-stranded structure, allowing larger, more complex genomes to out-compete the RNA-based genomes of more primitive cells [18]. Note that this scenario is nicely complementary to the viral eukaryogenesis hypothesis, the cellular RNA genes being progressively recruited within the newly acquired DNA-based 'nucleus' (see Figure 1). Interestingly, deoxyuridine is known to replace thymidine in the DNA of several bacteriophages [19].
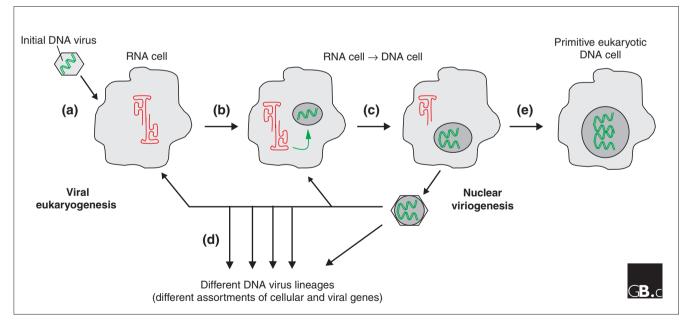

Figure 1
A possible iterative scenario for viral eukaryogenesis and nuclear viriogenesis. **(a)** A primitive DNA virus (a bacteriophage ancestor) gets trapped within an RNA cell and becomes a primitive nucleus. **(b)** Cellular genes are progressively recruited to the enlarging nucleus because of the selective advantages of DNA biochemistry. **(c)** For a while this situation remains unstable and reversible, allowing new 'pre-eukaryotic viruses' to be created. These viruses reinfect other cells at various stages of this iterative process. **(d)** This hypothetical scheme provides a mechanism for the emergence of various overlapping but not monophyletic virus lineages as well as for the rapid reassortment of genes from the viral and cellular pools before they reach their 'Darwinian threshold' [29], that is, **(e)** the evolution of a stable eukaryotic cell with a fully DNA nuclear genome.

Finally, in a paper that has already received much attention, Forterre [20] promoted (ancient) viruses to another fundamental role: to have been at the origin of the three basic cellular domains. His 'three RNA cells, three DNA viruses' hypothesis explains firstly, why there are three discrete lineages of modern cells instead of a continuum; secondly, the existence of three canonical ribosomal patterns; and thirdly, the critical differences exhibited by the, nevertheless similar, eukaryotic and archaeal replication machineries. This is readily done by postulating that DNA technology was independently transferred by three different founder DNA viruses to RNA-based ancestors of the Archaea, Bacteria, and Eukarya respectively. The reduction in rates of evolution following the transition from an RNA to a DNA genome would have stabilized the three canonical versions of translation proteins that are still recognizable today.

Traditional 'cell-first' hypotheses

If, for a moment, we put aside the paradoxical virus-first hypothesis, we are left with two more traditional (cell-first) hypotheses about the origin of viruses in general. One is the 'escape hypothesis', which views viruses as originating from cells by the escape of a minimal set of cellular components necessary to constitute an infectious selfish replicating system. The other is the 'reduction hypothesis', in which viruses would have derived from a cellular organism through a progressive loss of functions until it finally became a *bona fide* virus. In real life, unfortunately, this simple dichotomy will be blurred by the accretion of genes laterally transferred between viruses (or parasitic cellular organisms) sharing identical hosts, or directly captured from the virus hosts. In that respect, bacteriophages differ markedly from most eukaryotic dsDNA viruses by exhibiting massive recombinational reassortment and accretion of genes, most probably resulting from the existence of a prophage state integrated into the host genome [21]. Yet 80% of the genes of dsDNA bacteriophages have no obvious homologs in microbial genomes, suggesting a large degree of evolutionary independence of the phage gene set [22]. A much stricter genetic isolation is exhibited by the eukaryotic nucleocytoplasmic large dsDNA viruses (NCLDV), such as the giant *Acanthamoeba polyphaga* mimivirus [7], whose 1.2 Mb genome (911 genes) exhibits little evidence of horizontal transfer [23]. This also holds true for the next-largest NCLDVs, alga-infecting phycodnaviruses (with known genome sequences in the 300-400 kb range) [24,25]. Mimivirus also exhibits a high level of genomic coherence, as shown by the homogeneity of its nucleotide composition and the strict conservation of half of its promoter sequences [26].

As more genomes of large eukaryotic viruses are sequenced, new genes keep turning up, most of them with no obvious phylogenetic affinity with known hosts or extant cellular organisms. This simple observation is definitely more favorable to the idea that these large viruses arose from the reduction of a more complex ancestral (viral) genome, than to the hypothetical accretion of numerous exogenous genes (without recognizable origin) around a primitive minimal viral genome. Recent results on coccolithovirus EhV-86 illustrate this point very nicely. Until the 407 kb genome of EhV-86 was characterized, the trademark of all previously characterized phycodnaviruses (with smaller 320 kb genomes) compared with other NCLDVs was the absence of a virus-encoded transcription machinery (a lack of DNA-directed RNA polymerase) [24]. Obviously, the presence or absence of an RNA polymerase implies major differences in virus physiology. Unexpectedly, EhV-86 was found to encode its own six-subunit transcriptional machinery [25]. Nevertheless, a phylogenetic analysis of 25 core genes common to NCLDVs firmly placed EhV-86 within the Phycodnaviridae clade [25]. In this case, the loss of the transcription apparatus by the smaller phycodnaviruses, rather than the simultaneous gain of the six subunits of an RNA polymerase by EhV-86, appears much more likely.

The reduction hypothesis received a strong boost from the discovery and genomic characterization of *A. polyphaga* mimivirus [7], the first virus to largely overlap with the world of cellular organisms, in terms of both particle size and genome complexity [2]. The finding of numerous virally encoded components of an incomplete translation apparatus strongly suggested a process of reductive evolution from an even more complex ancestor that was endowed with protein synthetic capability. Such an ancestor could either have evolved from an obligate intracellular parasitic cell (functionally similar to *Rickettsia* or *Chlamydia*), or be derived from the nucleus of a primitive eukaryote through the mechanism illustrated in Figure 1. If reduction is the scenario at the origin of mimivirus, it is most likely to apply to other NCLDVs, in particular to those exhibiting the closest phylogenetic affinity with mimivirus such as the Phycodnaviridae and Iridoviridae. Sequencing additional large genomes from representatives of these families should provide valuable insights about this postulated giant ancestor.

Changing the viewpoint on viruses

At first sight, bacterial obligate intracellular parasites such as *Rickettsia* and *Chlamydia* have little functional resemblance to mimivirus despite a comparable genomic complexity. On one side, the bacteria are metabolically active, stealing ATP and biochemical precursors from their hosts to transcribe their genomes, translate their proteins, replicate their DNA, and divide. On the other side, one sees a large but metabolically silent viral particle, not deserving to be described as living by most biologists. This traditional view might, however, be a case of 'when the finger points to the stars, the fool looks at the finger'. Rather than comparing a parasitic cell to the virus particle, I believe we should compare it to the 'virus factory' [27]. Not much is yet known specifically on mimivirus factories, but upon infection, all

complex eukaryotic viruses such as iridoviruses, poxviruses, and asfarviruses give rise to complex intracellular structures that transcribe the viral genome, translate transcripts into proteins, and replicate the viral DNA, before packaging it into sophisticated vehicles designed to reproduce the virus factory upon the infection of another host cell (Figure 2). The virus factory is enclosed by a membrane (often derived from the rough endoplasmic reticulum) to exclude cellular organelles, but contains ribosomes and cytoskeletal elements. In the meantime, the virus factory recruits the mitochondria at its periphery, from which it obtains ATP [27]. At this stage, the overall functional resemblance between an intracellular parasitic bacterium and a large eukaryotic virus is quite striking. From this point of view, the genomic complexity of NCLDVs is no longer paradoxical, as it is commensurate with the complexity of the cell-like virus factory, but not with the particle used to reproduce it. Interpreting the virion particle as 'the virus', is very much like looking at a
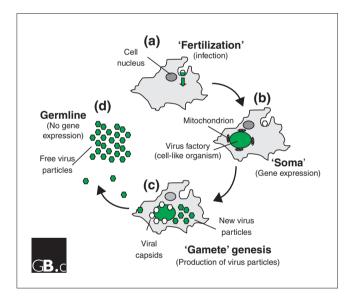

Figure 2
What is a virus? The life cycle of a complex dsDNA virus (for example NCLDVs) is shown. **(a)** A virus particle infects the cell and releases its DNA into the cytoplasm. **(b)** The viral DNA replicates and capsid proteins are synthesized within a 'virus factory' in the cytoplasm to which are recruited cellular ribosomes and the protein-synthesis machinery, as well as mitochondria to provide ATP. **(c)** New infectious viral particles are produced (while the nucleus is fading) and **(d)** released from the cell to begin another round of infection and replication. I propose that the true nature of complex eukaryotic dsDNA viruses is found in the transient virus factory they produce at each generation, rather than in the reproductive virus particle with which they have been equated. The virus factory is proposed to represent the result of the progressive reductive evolution of an obligate parasitic cellular organism, committed to the viral evolutionary pathway by the loss of a functional translation machinery. For a viral organism, the virus factory exhibits all the properties of the soma, in which genes are expressed, while the particle state corresponds to the germline (*sensu* August Weismann [28]) which remains unchanged. If we follow this line of thought, one might think of infection as being analogous with fertilization and the production of new virus particles as being akin to the formation of gametes.

spermatozoid and calling it a human: a 3,000 Mb genome would seem like overkill for such a unicellular organism (the similar thought arises when considering the size of plant genomes when looking at metabolically inert pollen grains).

Conceptually, the analogy between a virus life cycle and the reproductive cycle of a nondividing organism can be extended further. *Sensu* August Weismann, the virus particle possesses all the property of the *Germen* (the germline, the continuous immortal lineage responsible for carrying one generation to the next), whereas the transient virus factory exhibits all the property of the *Soma*, the body or somatic cells [28]. Also, according to Weismann, such a partition implies the phenomenon of aging: once the opportunity to pass germplasm on has passed (that is, once viral particles have been produced), there is no need to maintain the integrity of the somaplasm. In this interpretation, the virus factory now becomes the ultimate illustration of a disposable soma, vanishing immediately after viral particles have been produced. Nevertheless, I believe that the virus factory should be considered the actual virus organism when referring to a virus. Incidentally, in this interpretation the living nature of viruses is undisputable, on the same footing as intracellular bacterial parasites. Focusing on the structure of the virus factory rather than on the morphology of the virus particle might help us reach a better understanding of the evolutionary history of viruses.

A serious difficulty in the reductive hypothesis for the origin of viruses (when considered as particles) is to propose reasonable mechanisms by which a cell, even a highly parasitic cell, might switch from a cellular dividing mode to a host-supported particle-replication mode all at once. Focusing on viruses as cell-like factories rather than particles makes it much easier to conceive a gradual transition. I would like to propose the following scenario. The event committing a parasitic cell towards the reductive viral evolution pathway would be the loss of an essential component of its translation apparatus (for example, a ribosomal protein): the presence or absence of an encoded protein synthesis system clearly remains the last unambiguous genomic divide between the viral and the cellular worlds. In order to survive, the now translation-defective cell would have had to adopt new strategies to gain access to the ribosomes of its hosts. At the same time, this translation-defective cell could now dispense with the rest of its ribosome-encoding genes. Such an intermediate protoviral cell could survive in its original host while improving the design of a *bona fide* virus factory. Finally, a gamete-like genome-packaging process could emerge, following the acquisition of a capsid protein gene from an ancestral RNA virus. Such an event would allow the reduced cellular genome to be reproduced in many more copies, at the same time relieving the burden of maintaining the viability of the infected host cells. The soma-like virus factory could then become the transient organism we observe today.

In summary, the past few years have seen a spectacular renaissance of the field of viral evolution, prompted equally by the publication of increasing bold theories on the origin of life, the realization that viruses are the dominant life form on Earth, an exponential increase of genomic data, and the serendipitous discovery of few giant viruses. Viruses have come a long way from being unwanted inhabitants of the Tree of Life, to being given a central role in all major evolutionary transitions [6]. The challenge is now to unify the many evolutionary scenarios that have been proposed, using hard facts and experimental data, without getting sidetracked by the many spectacular but anecdotal features that individual virus families have incorporated during their long and probably chaotic history.

References

1. Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, Spiro DJ, Sitz J, Koo H, Bolotov P, *et al.*: **Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution.** *Nature* 2005, **437:**1162-1166.
2. Claverie JM, Ogata H, Audic S, Abergel C, Suhre K, Fournier PE: **Mimivirus and the emerging concept of "giant" virus.** *Virus Res* 2006, **117:**133-144.
3. **GiantVirus.org** [www.giantvirus.org]
4. Koonin EV, Dolja VV: **Evolution of complexity in the viral world: the dawn of a new vision.** *Virus Res* 2006, **117:**1-4.
5. Bamford DH, Grimes JM, Stuart DI: **What does structure tell us about virus evolution?** *Curr Opin Struct Biol* 2005, **15:**655-663.
6. Forterre P: **The origin of viruses and their possible roles in major evolutionary transitions.** *Virus Res* 2006, **117:**5-16
7. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie JM: **The 1.2-megabase genome sequence of Mimivirus.** *Science* 2004, **306:**1344-1350.
8. Dolja VV, Kreuze JF, Valkonen JP: **Comparative and functional genomics of closteroviruses.** *Virus Res* 2006, **117:**38-51.
9. D'Herelle F: *The Bacteriophage; Its Role in Immunity.* Baltimore: Williams and Wilkins; 1922.
10. Bamford DH: **Do virus form lineages across different domain of life?** *Res Microbiol* 2003, **154:**231-236.
11. Forterre P: **The great virus comeback: from an evolutionary perspective.** *Res Microbiol* 2003, **154:**223-225.
12. Villarreal LP, DeFilippis VR: **A hypothesis for DNA viruses as the origin of eukaryotic replication proteins.** *J Virol* 2000, **74:**7079-7084.
13. Forterre P: **Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins.** *Mol Microbiol* 1999, **33:**457-465.
14. Takemura M: **Poxviruses and the origin of the eukaryotic nucleus.** *J Mol Evol* 2001, **52:**419-425.
15. Bell PJ: **Viral eukaryogenesis: was the ancestor of the nucleus a complex DNA virus?** *J Mol Evol* 2001, **53:**251-256.
16. Goff LJ, Coleman AW: **Fate of parasite and host organelle DNA during cellular transformation of red algae by their parasites.** *Plant Cell* 1995, **7:**1899-1911.
17. Koonin EV, Martin W: **On the origin of genomes and cells within inorganic compartments.** *Trends Genet* 2005, **21:**647-654.
18. Forterre P: **The origin of DNA genomes and DNA replication proteins.** *Curr Opin Microbiol* 2002, **5:**525-532.
19. Takahashi I, Marmur J: **Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for *Bacillus subtilis*.** *Nature* 1963, **197:**794-795.
20. Forterre P: **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 USA* 2006, **103:**3669-3674.
21. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S: **The origin and ongoing evolution of viruses.** *Trends Microbiol* 2000, **8:**504-508.
22. Liu J, Glazko G, Mushegian A: **Protein repertoire of double-stranded bacteriophages.** *Virus Res* 2006, **117:**68-80.
23. Ogata H, Abergel C, Raoult D, Claverie JM: **Response to Comment on "The 1.2-megabase genome sequence of Mimivirus".** *Science* 2005, **308:**1114.
24. Dunigan DD, Fitzgerald LA, Van Etten JL: **Phycodnaviruses: A peek at genetic diversity.** *Virus Res* 2006, **117:**119-132.
25. Wilson WH, Schroeder DC, Allen MJ, Holden MT, Parkhill J, Barrell BG, Churcher C, Hamlin N, Mungall K, Norbertczak H, *et al.*: **Complete genome sequence and lytic phase transcription profile of a Coccolithovirus.** *Science* 2005, **309:**1090-1092.
26. Suhre K, Audic S, Claverie JM: **Mimivirus gene promoters exhibit an unprecedented conservation among all eukaryotes.** *Proc Natl Acad Sci USA* 2005, **102:**14689-14693.
27. Novoa RR, Calderita G, Arranz R, Fontana J, Granzow H, Risco C: **Virus factories: associations of cell organelles for viral replication and morphogenesis.** *Biol Cell* 2005, **97:**147-172.
28. Weismann A: *Essays upon Heredity and Kindred Biological Problems.* Oxford: Clarendon Press, 1889.
29. Woese CR: **On the evolution of cells.** *Proc Natl Acad Sci USA* 2002, **99:**8742-8747.