Available online at www.sciencedirect.com
SCIENCE DIRECT
Virus Research 117 (2006) 5–16
Virus Research
www.elsevier.com/locate/virusres

Review

The origin of viruses and their possible roles in major evolutionary transitions

Patrick Forterre [a,b,*]

[a] *Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris-Sud, 91405 Orsay Cedex, France*
[b] *Institut Pasteur, 25 rue du Docteur Roux, 75015 Paris, France*

Available online 14 February 2006

Abstract

Viruses infecting cells from the three domains of life, Archaea, Bacteria and Eukarya, share homologous features, suggesting that viruses originated very early in the evolution of life. The three current hypotheses for virus origin, e.g. the virus first, the escape and the reduction hypotheses are revisited in this new framework. Theoretical considerations suggest that RNA viruses may have originated in the nucleoprotein world by escape or reduction from RNA-cells, whereas DNA viruses (at least some of them) might have evolved directly from RNA viruses. The antiquity of viruses can explain why most viral proteins have no cellular homologues or only distantly related ones. Viral proteins have replaced the ancestral bacterial RNA/DNA polymerases and primase during mitochondrial evolution. It has been suggested that replacement of cellular proteins by viral ones also occurred in early evolution of the DNA replication apparatus and/or that some DNA replication proteins originated directly in the virosphere and were later on transferred to cellular organisms. According to these new hypotheses, viruses played a critical role in major evolutionary transitions, such as the invention of DNA and DNA replication mechanisms, the formation of the three domains of life, or else, the origin of the eukaryotic nucleus.

Keywords: Virus origin; RNA world; RNA/DNA transition; Mimivirus; DNA origin; DNA replication; Nucleus origin; LUCA; Universal tree of life

Contents

1. Introduction

The origin of viruses is still enigmatic and their nature controversial (in part for historical reasons, the existence of viruses challenging the cellular theory of life). It has been often stated

Abbreviations: NCLDV; Nucleo-cytoplasmic large DNA viruses

* Tel.: +33 169157489; fax: +33 169157808.
E-mail addresses: forterre@pasteur.fr, patrick.forterre@igmors.u-psud.fr.

0168-1702/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2006.01.010

that viruses are polyphyletic, i.e. that different viral lineages originated independently. In particular, RNA and DNA viruses were thought to be evolutionary unrelated. However, the overall similarity between virus structures – a protein coat enclosing a nucleoprotein filament – at least suggests a common mechanism for their appearance. Three hypotheses have been proposed to explain the emergence of viruses: (i) they are relics of pre-cellular life forms; (ii) they are derived by reduction from unicellular organisms (via parasitic-driven evolution); (iii) they originated from fragments of genetic material that escaped from the control of the cell and became parasitic (Luria and Darnell, 1967; Bandea, 1983; Forterre, 2003; Hendrix et al., 2000 and references herein).

The first hypothesis (here called *the virus-first hypothesis*) has been dismissed for a long time, since all present viruses are obligatory parasites requiring an intracellular development stage for their reproduction. The second hypothesis (here called *the reduction hypothesis*) was also usually rejected based on two arguments: (i) we don't know any intermediate form between cells and viruses; and (ii) parasites derived from cells in the three domains of life, such as *Mycoplasma* in Bacteria, *Microsporidia* in Eukaryotes or *Nanoarchaea* in Archaea, have retained their cellular characters (i.e. their own ribosomes and complete machineries for protein synthesis and ATP production). The third hypothesis (here called *the escape theory*) became popular partly by default and partly because it was a priori supported by the observation that present-day viruses can integrate cellular genes into their own genomes. In this view, plasmids and mobile elements are often considered to be viral precursors. However, the escape hypothesis has also serious drawbacks since it does not specify how a free nucleic acid could have recruited a capsid and the complex mechanisms required by viruses to deliver their nucleic acid to their host cells. Furthermore, in its traditional version, the escape hypothesis predicts that bacteriophages originated from bacterial genomes and eukaryotic viruses from eukaryotic genomes. In this case, one expects to find evolutionary affinities between viral proteins encoded by viruses from one domain and their cellular homologues in that domain. However, this is often not the case; for instance, some proteins encoded by T4 bacteriophage are more related to proteins from eukaryotes or eukaryotic viruses than to their bacterial homologues (Miller et al., 2003; Gadelle et al., 2003). Furthermore, although more than 250 cellular genomes from the three domains have now been completely sequenced, most of the viral proteins detected in viral genomes have no cellular homologues (up to 90–100% in the genomes of archaeal viruses) (Prangishvili and Garrett, 2004).

At this point, one should realize that some of the major critics against the three above hypotheses have been made in the context of the present-day biosphere (i.e. modern viruses indeed need modern cells to replicate, modern cells cannot regress to viral forms, free DNA cannot recruit proteins from modern cells to form capsids, and so on). However, things may be different if viruses originated before the formation of modern cells (*sensu* Woese, 2002): Archaea, Bacteria and Eukarya. In this case, we are less constraint by the present reality to propose new evolutionary scenarii for the origin of viruses. Of course, such speculations should be made with caution, but one cannot expect to understand the origin of modern cells and viruses by sticking to the present context. In this review, I will discuss briefly how the three hypotheses for virus origin can be revisited if one considers that viruses originated before the Last Universal Cellular Ancestor (LUCA) from which the three cellular domains diverged. I will also present recent data and new hypotheses on the involvement of viruses in the origin and early evolution of modern DNA cells.

2. Viruses as old players in life evolution

The idea that viruses are ancient was first more easily accepted for RNA viruses, in relation with the RNA world theory. Several authors have convincingly argued that present RNA viruses could be relics of the RNA world, whereas Retroviruses and/or Hepadnaviruses could be relics of the RNA/DNA transition (Wintersberger and Wintersberger, 1987; Weiner and Maizels, 1993, 1994; Makeyev and Grimes, 2004). Such vision was boosted by the discovery of tRNA-like structure linked to some viral RNA genomes (possible relic of an earlier coupling between replication and translation), and by the discovery of viral reverse transcriptase, one of the essential enzymes for the RNA to DNA transition. A priori, the idea that RNA viruses are ancient could appear at odds with their apparent predominance in eucaryotes, considering the prejudice that eucaryotes are more recent than *pro*karyotes. However, the hypothesis that RNA viruses are relics of the RNA world is supported by the fact that both single-stranded and double-stranded RNA viruses are also present in the bacterial domain (they are presently unknown in Archaea). Furthermore, double-stranded RNA viruses infecting Bacteria (*Cystoviridae*) and those infecting Eukarya (*Totiviridae* and *Reoviridae*) have a similar structure and life cycle (Bamford, 2003) and their RNA-dependent RNA polymerases are homologues (Makeyev and Grimes, 2004). Finally, RNA replicases/transcriptases of double-stranded RNA viruses are also evolutionary related to those of single-stranded RNA viruses (Makeyev and Grimes, 2004). These observations strongly support the idea that all RNA viruses are evolutionary related and both very ancient. In the traditional view of life evolution, this could simply imply that "eukaryotic" RNA viruses originated from "bacterial" RNA viruses (being therefore "only" as old as Bacteria). However, comprehensive analysis of the universal tree of life suggests that Eukarya did not originate from Bacteria, but that both evolved from a LUCA that was neither a bona fide prokaryote, nor a bona fide eukaryote, and that could even still belong to the RNA world (Woese, 1987; Leipe et al., 1999; Forterre, 2005). If this interpretation of the universal tree is correct, the finding of homologous RNA viruses in both Eukarya and Bacteria suggests that these viruses were already present at the time of LUCA, and most likely even before LUCA (i.e. probably at the epoch of the RNA-protein world).

The possible antiquity of DNA viruses was recognized more recently. I suggested in 1992 that DNA viruses probably also predated the formation of the three domains of life based on my interest in DNA replication proteins (Forterre, 1992). I was first impressed by the singularity of T4 type II DNA topoiso-

Fig. 1. Alignment of the Type IIA DNA topoisomerase in the region of the ATP binding site (Bergerat fold) (adapted from Forterre, 1992). The type II DNA topoisomerase from bacteriophage T4 branches between the eukaryotic and bacterial sequences in phylogenetic trees (see for instance Gadelle et al., 2003). This can be explained a priori either by an ancient divergence of the T4 sequence (before the last common ancestor of all bacteria), or from a rapid evolution from a bacterial sequence. Examination of the sequence alignment strongly favors the former hypothesis. In case of rapid divergence from a bacterial sequence, one would expect retention of more ancestral common signatures in the bacterial and eukaryotic sequences (lost in T4). However, there is a single position (in blue) corresponding to this situation. In contrast, T4 exhibits 2 signatures (in red) and a homologous indel (underlined) with eukaryotic sequences, something that cannot be explained by the rapid evolution hypothesis. The conservation of 13 common signatures to all Topo IIA sequences also argues again rapid evolution of the bacteriophage enzyme (at least in that fold). The presence of 5 typical bacterial signatures (in green) suggests that T4 Topo IIA is slightly more related to their bacterial homologues or that eukaryotic sequences evolved more rapidly. Only positions with the same amino-acid in the majority of bacterial or eukaryotic Topo IIA sequences (not shown here) were used to identify bacterial or eukaryotic signatures, respectively.

merase (a heterohexamer without gyrase activity) compared to bacterial DNA gyrase (a heterotetramer) and to eucaryotic type II DNA topoisomerase (a homodimer without gyrase activity). The amino-acid sequence of T4 type II DNA topoisomerase indeed turned out to exhibit a striking succession of bacterial and eucaryotic signatures, as well as unique features (Fig. 1). This protein was clearly neither of the bacterial or the eukaryotic type but seemed to represent an entire new domain by itself. Similarly, I noticed that both human adenovirus and *Bacillus subtilis* bacteriophage ϕ29, use a similar atypical protein-priming mechanism to replicate their DNA (something unknown in the cellular world) and encode a unique type of DNA polymerase that can use such template to initiate its own replication. This again suggested that these two viruses originated from a common ancestor that existed before the divergence between Eukarya and Bacteria. The validity of these two examples has been confirmed in the following decade by the accumulation of sequence data. In phylogenetic trees based on amino-acid sequence comparison, the type II DNA topoisomerases encoded by T4 and relatives indeed form a cluster of sequences clearly distinct from those of Bacteria and Eukarya (Gadelle et al., 2003). Similarly, adenovirus and ϕ29 DNA polymerases belong to a subfamily of DNA polymerases B without any cellular member (Filée et al., 2002).

The evolutionary connection between eukaryotic and bacterial viruses is now also supported by the existence of homologous features at the structural level. Hence, The architecture of the portal protein complex of herpes virus resembles those of the portal/connector proteins of bacterial viruses (Trus et al., 2004). The more convincing work was performed by Bamford and colleagues who identified, by structural analyses, a common module (the double-barrel trimer) in the capsid proteins of human adenovirus and the bacterial viruses PRD1 (Bamford, 2003). This module turned out to be also present in the capsid proteins of other icosahedral double-stranded DNA viruses with large facets, the bacterial virus Bam35 and several groups of eukaryotic viruses (*Phycodnaviridae, Iridoviridae, Asfarviridae* and *Mimiviridae*) (Benson et al., 2004). Finally, this module was detected in the putative capsid protein of the archaeal virus STV1, recently discovered in a Yellowstone hot spring (Benson et al., 2004; Rice et al., 2004). In all these viruses, the proteins containing the double-barrel trimer fold are arranged according to similar principle to form the viral capsid. All these data suggest that the capsids of these DNA viruses derived from the capsid of an ancestral virus living either before, or at the time of LUCA.

The antiquity of both RNA and DNA viruses can readily explain why, with rare exceptions, viral proteins involved in DNA/RNA transcription, replication, recombination or repair (informational proteins), are not minor variations of their cellular functional analogues. As in the case of type II DNA topoisomerases, the sequences of many viral informational proteins form specific clusters in phylogenetic trees, being as distantly related to their cellular homologues than proteins from different domains are from each others (Filée et al., 2002, 2003; Miller et al., 2003; Gadelle et al., 2003; Raoult et al., 2004; Forterre et al., 2004). In other cases, viral informational proteins have no real cellular homologues at all (except for plasmid versions or viral remnants in cellular genomes) such as rolling-circle initiator proteins (Ilyina and Koonin, 1992), monomeric DNA dependent RNA polymerases (Cermakian et al., 1997), herpes virus primase (Dracheva et al., 1995), or else RNA and DNA helicases of the SF3 superfamily (Gorbalenya et al., 1990) (see other examples in Iyer et al., 2005). These proteins only share with some cellular enzymes common domains or structural folds (Iyer et al., 2005), as expected if a common pool of structural modules present in the RNA-protein world was used repeatedly for the modular construction of all large proteins (viral and cellular). These viral-specific proteins could have been directly invented in ancient viral lineages, independently of their hosts, or they could have been recruited from ancient cellular lineages that have later on disappeared from the biosphere.

In a minority of cases, viral proteins are closely related to homologous proteins encoded by their hosts, indicating recent transfers of these proteins from cells to viruses (Moreira, 2000; Filée et al., 2002). From this observation, it has been argued by proponents of the escape hypothesis that all viral proteins have a cellular origin, and that the basal position of many of them in phylogenetic trees is an artifact of phylogenetic reconstruction, due to the rapid rate of evolution of viral sequences (long branch attraction) (Moreira, 2000). However, this explanation cannot be used for all viral proteins without cellular homologues previously described. It is logical to assume that the mechanisms

that produce these proteins unique to viruses also produce viral specific versions of enzymes with cellular homologues.

Accordingly, I think that most viral proteins that form specific clusters in phylogenetic trees (away from their cellular homologues) indeed testify for deep evolutionary divergence and for viral antiquity, and not simply for the rapid evolution of viral proteins after their acquisition from a cellular host (see legend of Fig. 1 for further explanation).

Many characteristic features of viral genomes are also typical of plasmids. Plasmids also encode a majority of genes without cellular homologues, including proteins involved in DNA replication. This is for instance the case of rolling-circle Rep proteins or else of the recently discovered DNA polymerases E (Lipps et al., 2003). Many of these proteins have homologues encoded in viral genomes (Iyer et al., 2005), indicating that plasmids and viruses are evolutionary related. If viruses are very ancient, one can suggest that plasmids originated from DNA viruses (and not the other way around) that have lost genes encoding for capsid proteins (Forterre, 1992, 2005). Plasmids probably evolved relatively late in early life evolution, i.e. after the RNA to DNA transition and the appearance of viruses with circular double-stranded DNA genomes.

3. The three hypotheses for virus origin revisited

The hypothesis that viruses were already present before the emergence of modern DNA-cells open new perspectives about their origin, and suggests to revisit the three classical hypotheses in this new context.

3.1. The virus-first hypothesis

The virus-first hypothesis has been revived in the last decade by Wolfram Zillig who suggested that viruses originated in the prebiotic word, using the primitive soup as a host (Prangishvili et al., 2001). Such hypothesis is in line with the view, advocated by several molecular biologists, that the formation of cells occurred relatively late in the evolution of life. It was common for some time to imagine the RNA world as a community of free molecules competing with each others (Gilbert, 1986). Recently, some authors have even proposed that LUCA was not a cellular entity (Kandler, 1998; Martin and Russell, 2003; Koga et al., 1998). In particular, they suggested that cellular membranes originated independently after the divergence of Archaea and Bacteria, in order to explain why archaeal lipids are so different from eukaryotic/bacterial ones (with opposite stereochemistry and different carbon chains). If all life evolution from the very beginning up to LUCA occurred in an acellular context, it is indeed possible to imagine that viruses first emerged as individual entities in a world of competing proteins and nucleic acids, either bathing in a "primitive soup" or located on a mineral platform. However, the hypothesis of a LUCA without membrane is contradicted by the existence of homologous proteins functioning at the membrane level that are encoded by all sequenced genomes from the three domains of life, strongly suggesting that these proteins (hence a membrane) were already present in LUCA (Pereto et al., 2004). In fact, the first cells probably arose well before the emergence of LUCA. For instance, it appears unlikely that a world of free molecules could have evolved to such an extent to produce a ribozyme capable to synthesize proteins (the ancestor of present-day ribosomes). Even in the early RNA world, a primitive metabolism should have produce at least precursors for RNA and lipid syntheses, as well as the energy required to perform these reactions. It is difficult to imagine the emergence of such a metabolism without Darwinian selection, and this requires the competition between well-defined individual entities (at least proto-cells). Since modern viruses contain proteins, they should have originated after the emergence of the ancestral ribosome, i.e. well after the apparition of primitive RNA-cells (in the second age of the RNA world, *sensu* Forterre, 2005). Accordingly, in my opinion, the virus first theory can be rejected for present-day viruses (even viroids would need a suitable cellular environment providing nucleotides for their emergence). In this case, one is presently left with only two possibilities: either the first RNA viruses originated from RNA cells by regressive evolution (a new version of the reduction theory), or from RNA fragments that escaped from RNA cells (a new version of the escape theory).

3.2. The escape hypothesis

The traditional hypothesis viewing viruses as elements of cell genomes that escaped from their cellular environment, becoming autonomous and infectious selfish elements, is easier to defend in the context of a pre-LUCA scenario for virus origin. In this context, one does not expect anymore any specific relationship between proteins encoded by viruses and those encoded by their hosts, since viruses now derive from genome fragments escaped from cells predating LUCA. Furthermore, one can reasonably assume that it was easier for a genome fragment to become autonomous in ancient RNA cells, since the different molecular mechanisms operating in these proto-cells were probably much simpler and less integrated than in modern DNA cells (Woese, 2002). In particular, it has been often argued that the genomes of ancestral RNA cells were fragmented (as in the case of modern double-stranded RNA viruses). These genomes could have been composed of semi-autonomous chromosomes (possibly formed by a few RNA genes) that were replicated independently and transferred randomly from cells to cells (Woese, 1987; Poole et al., 1998). Some RNA chromosomes could have encoded a coat protein that helps the proto-virus to be transferred, finally becoming infectious (Fig. 2). This process would be reminiscent of the moron theory proposed by Hendrix and co-workers for the origin of virus (Hendrix et al., 2000). Although this theory was inferred from the observation that modern DNA viruses can easily acquire more DNA by illegitimate recombination it can be easily extrapolated for the origin of RNA viruses in the RNA world.

3.3. The reduction hypothesis

The transformation of a cellular organism into a viral one could have been also much easier in a world of RNA cells,

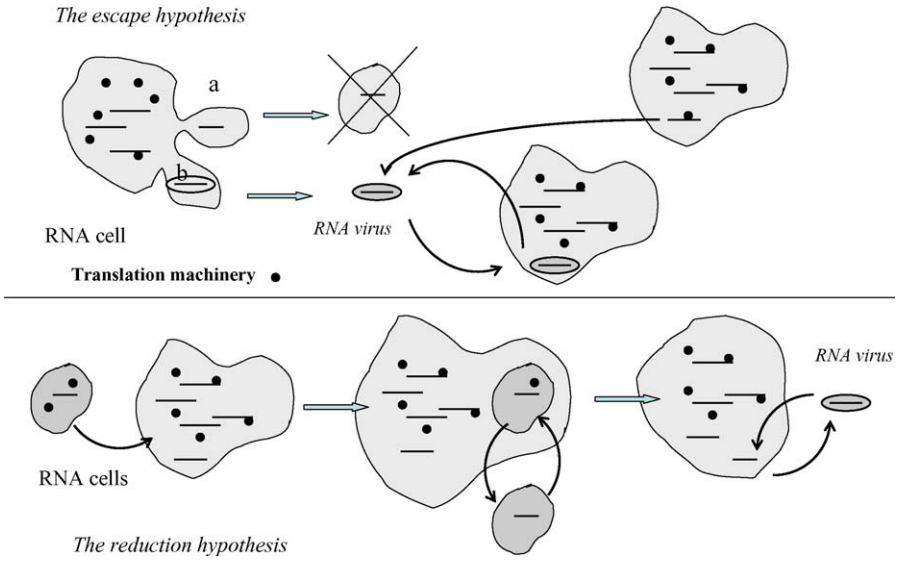

Fig. 2. Two alternative hypotheses for the origin of viruses in the second age of the RNA world (after invention of protein synthesis), black circles correspond to translation machinery (e.g. ancestral ribosomes) and lines to linear RNA chromosomes. Upper panel, the *escape hypothesis*: unequal cell division produces minicells with single chromosome (a or b) but no translation apparatus. The chromosome a will be eliminated but chromosome b will survive because it is associated with a proteins coat that allows its transfer into a new RNA cell, it becomes a virus. Lower panel, the reduction hypothesis: a small RNA cell became an endosymbiont of a larger RNA cell. It looses its translation apparatus but continue to replicate autonomously and become infectious (similar to some pathogenic bacteria in eukaryotic cells).

again because these cells were much simpler than modern ones. Just as modern parasites can loose part of their metabolic channels, an RNA-cell living as a parasitic endosymbiont in another RNA cell could have lost its own machinery for protein synthesis and for energy production, using instead those of the host (Fig. 2) (Forterre, 2005). In this model, viral capsids could have originated from the envelopes of RNA cells composed of identical proteins, resembling the S-layer of modern prokaryotes. The reduction hypothesis might have been driven by the harsh competition that most likely occurred between RNA cells all along the evolution of the RNA world. As a consequence of this competition, early life evolution probably went through several bottlenecks each time a crucial new molecular mechanism was invented. At each of these bottlenecks, the descendents of the individual endorsed with such a great selective advantage (the winners) would have eliminated all other lineages of protocells that previously coexisted with them (the losers). The only chance of survival for the loosers was to become parasites of the winners. In this hypothesis, viruses evolved by parasitic reduction from ancient lineages of RNA cells that were out-competed in the Darwinian selection process, and thus could only survive by parasiting the winner of this competition (Forterre, 1992).

4. The origin of DNA viruses

DNA viruses could have emerged either from RNA viruses or independently. It is often stated that both types of viruses indeed originated independently in agreement with the current idea that viruses are polyphyletic. However, in my opinion, the homologies that can be detected at the structural and mechanistic levels between viral RNA replicases/transcriptases, reverse transcriptases and some DNA polymerases (Hansen et al., 1997), or else between viral RNA and DNA helicases (Gorbalenya et al., 1990) strongly suggest that DNA viruses evolved from RNA viruses. This view is also in agreement with the existence of intermediate forms such as retroviruses (with a RNA genome and a RNA-DNA-RNA cycle) and hepadnaviruses (with a DNA genome and a DNA-RNA-DNA cycle). Interestingly, retroviruses and hepadnaviruses are evolutionary related, strongly supporting the idea that the transition from RNA to DNA occurred in the viral world (Miller and Robinson, 1986). The transition was probably easy for nucleic acid biosynthesis, since the specificity of RNA/DNA polymerases is not that high. For instance, it has been shown that the RNA-dependant RNA polymerase from the Brome mosaic virus can use, as templates, either RNA, DNA or RNA/DNA hybrids (Siegel et al., 1999). RNA viruses might have used cellular enzymes to transform their RNA genome into a DNA genome. However, if DNA viruses indeed originated from RNA viruses, it is tempting to suggest that the enzymes required for the RNA to DNA transition first appeared in viral genomes and to conclude that DNA is a viral invention after all (an hypothesis supported by independent argument, see below).

Considering the great diversity of DNA viruses (especially in term of size), it is also possible that the first DNA viruses originated from RNA viruses and that, later on, further DNA viruses originated by reduction of primitive DNA cells. The idea that some DNA viruses originated by reduction from ancient lineages of DNA cells that existed either before or shortly after LUCA (Forterre, 1992) was recently boosted by the discovery of the mimivirus, a giant virus infecting *Entamoeba* (La Scola et al., 2003). The genome of this virus (1.2 Mb) is three times larger than the genome of the smallest cell, and four or five

times larger than the minimal genome predicted for a functional cell. The mimivirus thus can be viewed as an intermediate form between a true cell and a virus (only lacking ribosomal proteins), i.e. the previously missing link of the reduction hypothesis. Indeed, the mimivirus genome encodes several proteins that are also present in the three domains of life, and a tree based on the concatenation of these proteins place the mimivirus in-between Archaea and Eukarya, as if it was the representative of a fourth domain with no longer free-living cellular representatives (Raoult et al., 2004). The mimivirus belongs to the family of Nucleo-Cytoplasmic Large DNA Viruses (NCLDV). The proteomes of these viruses share a relatively small common core of proteins (about 30–40), mainly involved in viral structure and DNA transactions (Raoult et al., 2004). In the reduction hypothesis, the NCLDV ancestor could have been even larger than the mimivirus (closely related to a putative ancestral DNA cell) and many gene loss (together with some acquisitions) should have occurred in the different NCLDV lineages.

However, the mimivirus remains "a bona fide virus since it rely on its host for protein synthesis and energy production and has a typical capsid" (Koonin, 2005). Furthermore, the recent observation that the mimivirus capsid protein is homologous to those of the adenovirus and of some bacterial and archaeal viruses (Benson et al., 2004) supports the hypothesis of a viral origin for the mimivirus. It is indeed difficult to imagine that all these viruses originated from the same primitive DNA cell. It is more likely that NCLDV and other viruses with the double-barrel trimer module originated from a very old DNA virus that already existed before the establishment of the three cellular domains.

The position of the mimivirus in the universal tree of life has been itself strongly disputed on methodological ground by Moreira and Lopez-Garcia (2005) who argued that the mimivirus is "only" a giant pickpocket that has captured these genes from Amoeba. In support of this view, the mimivirus branch as sister group of *Entamoeba* in a refined phylogenetic tree of tyrosyl tRNA synthetase (one of the protein used in the concatenation of Raoult and co-workers). However, comparative genomic analysis suggests that, including the tyrosyl tRNA synthetase, only five of the 363 mimivirus genes with cellular homologues (out of 1262 predicted ORFs) are more related to *Entamoeba* than to other eukaryotic proteins (Ogata et al., 2005). Koonin (2005) suggested that the mimivirus genome has grown from the NCLDV core via the accretion of genes of eukaryotic and bacterial origin. In fact, most mimivirus genes are orphans and those with bacterial or eucaryotic affinities are usually very distantly related from their cellular counterparts. Many genes of the mimivirus could have been therefore not borrowed from bacterial or eukaryotic lineages but from other viruses and/or from ancient lineages of DNA or RNA cells.

5. Viruses and the puzzling phylogenomic distribution of DNA replication proteins

A major surprise that came out early from large-scale comparative genomics was the existence of two sets of non-homologous proteins for DNA replication, one in Bacteria, and the other common to Archaea and Eukarya (Leipe et al., 1999). These sets include the central components of bacterial replication forks: replicative DNA polymerase, DNA primase and replicative helicase. This puzzling situation was in fact already predicted in 1977 by Woese and Fox, based on their progenote theory (Woese and Fox, 1977). These authors had put forward that DNA replication machineries would have originated independently in Bacteria and Eukarya, as a consequence of their view that modern cells diverged from a very simple LUCA, still member of the RNA world. Once the first genomic data became available, the hypothesis of an RNA-based LUCA was endorsed by Mushegian and Koonin (1996) who posited that the DNA replication mechanism was indeed invented twice independently, once in Bacteria, and once in a common lineage leading to Archaea and Eukarya. Later on, Koonin and co-workers refined their hypothesis to explain why a few proteins involved in DNA metabolism were nevertheless present – and homologous – in the three domains. They suggested that DNA was invented before LUCA but was not yet replicated at that time, replication mechanisms having being invented only after the divergence of the archaeal and bacterial lineages (Leipe et al., 1999). In this scenario, the replicating genome of LUCA was still RNA, but it replicated with a DNA intermediate, via retro-transcription, as in the case of retroviruses.

As an alternative to an RNA-based LUCA, I proposed in 1999 that the ancestral mechanism for DNA replication originally present in LUCA was later on displaced by a second one, of viral origin, either in Bacteria or in a lineage common to Archaea and Eukarya (Forterre, 1999). This hypothesis was supported by the existence of so many viral DNA replication proteins non-homologous (but analogous) to cellular ones (as previously discussed). It was also inspired by the fact that cellular DNA replication proteins of bacterial origin were replaced by viral ones in the course of mitochondrial evolution (Filée et al., 2002, 2005, see below). Recently, two specific archaeal/eukaryal DNA replication proteins, the helicase MCM and the primase, were found to be encoded, as a fused protein, by a prophage integrated in the genome of the bacterium *Bacilllus cereus* (McGeoch and Bell, 2005). This again supports the idea that viruses could have been the vectors for the delivery in one domain of cellular DNA replication proteins from another one. The possible role of viruses in the origin of cellular DNA replication mechanisms was also hypothesized by Villarreal and De Filippis (2000), who noticed that eukaryotic DNA polymerase delta was clustered with a group of viral DNA polymerases in a phylogenetic tree of family B DNA polymerases. However, the hypothesis of a non-orthologous displacement of an ancestral cellular DNA replication mechanism by a viral one seems to be contradicted by the observation that the major components of the bacterial and eukaryotic DNA replication machineries have apparently never been displaced in modern cells by homologues encoded in cryptic proviruses or plasmids. For instance, the bacterial replicative helicase DnaB is encoded in all bacterial genomes, despite the presence of alternative viral/plasmid like helicases in many of them (Iyer et al., 2005). One can bypass the requirement for a non-orthologous displacement step in the evolution

of the DNA replication machineries, by suggesting that all types of DNA replication machineries in fact originated in the virosphere and were later on transferred to cellular organisms. I thus proposed that both the bacterial and the archaeal/eukaryal replication machineries were recruited independently from the viral world, either before of after LUCA (Forterre, 2002). In that scenario, one has to conclude that these machineries became refractory to non-orthologous displacement by further viral proteins once they were established at the origin of each cellular domain.

6. Viruses and the origin of DNA

It is usually considered that RNA was "logically" replaced by DNA in the course of evolution for two reasons: (i) it is more stable, thanks to the removal of the reactive oxygen in position 2′ of the ribose; and (ii) modification in the genetic message produced by deamination of cytosine into uracil (a common spontaneous chemical reaction) can be recognized and repaired in DNA, but not in RNA. As a consequence of this stability and more faithful replication, the substitution of RNA by DNA as cellular genetic material allowed genome size to increase, DNA cells with enlarging genomes becoming more complex and finally out-competing their RNA-based ancestors (Lazcano et al., 1988). However, this argumentation cannot fully explain the origin of DNA, since the evolution of a DNA repair system to remove uracil from DNA, and, more generally, the advantages to have a large genome might have been major factors in the evolution process only after populations of DNA cells were already established. Instead, if the replacement of RNA by DNA first occurred in a virus, modification of its RNA genome into a DNA one would have immediately produced a benefit for the virus (a prerequisite for Darwinian selection). We know that some modern viruses have indeed chemically altered their DNA genomes to become resistant to the nucleases of their hosts (for instance via methylation, hydroxymethylation or even more complex chemical modifications, for review see Warren, 1980). As a first step, the emergence or recruitment of ancestral ribonucleotide reductase activities in viruses would have modified its RNA genome into a uracile-containing DNA (U-DNA) genome. This intermediate step in the RNA to DNA transition is inferred from the synthesis of dTMP from dUMP in modern cells (Fig. 3). Some bacterial viruses that have a U-DNA genome could be relics of this first transition (Takahashi and Marmur, 1963). Viruses that have conserved an RNA genome have evolved alternative mechanisms to protect their genetic material against RNA-degrading or modifying enzymes. Some of them keep their RNA genomes into their capsids all along the infection process, whereas others encode proteins that inhibit cellular RNA degradation or modification mechanisms (e.g. demethylases). In a second step, the emergence of thymidylate synthase activity in some U-DNA virus lineages would have produced viruses with the modern form of thymidine-containing DNA (T-DNA) (Fig. 3). This second step would have occurred at least twice independently in order to explain the existence of two non-homologous thymidylate synthase, ThyA and ThyX (Myllykallio et al., 2002). In agreement with the idea that ribonucleotide reductase and thymidylate synthase were invented in the virosphere, it is noteworthy that many DNA viruses encode their own ribonucleotide reductase or thymidylate synthase. Furthermore, these proteins are usually only distantly related to those encoded by their hosts in phylogenetic trees, as in the case of the other typical viral specific proteins previously discussed (Myllykallio et al., 2002; Filée et al., 2003; Miller et al., 2003; Forterre et al., 2004).

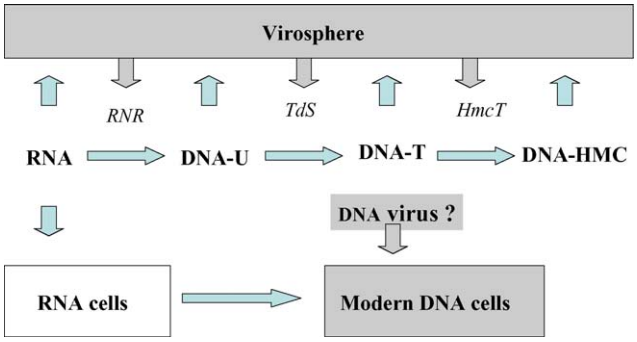

Fig. 3. Schematic representation of genome evolution from RNA to modified DNA genomes. All types of genomes are present in the virosphere but only DNA-T genomes in modern cells. The latter could have originated via the transfer of DNA from a double-stranded DNA virus to an RNA cell (question mark). RNR; ribonucleotide-reductase, TdS: thymidylate-synthase, HmcT: hydroxymethylcytosine-transferase.

The hypothesis of a viral origin for DNA has the advantage to readily explain the diversity of DNA replication proteins and mechanisms observed in modern viruses. In this hypothesis, after emergence of the first DNA virus, DNA replication mechanisms probably evolved in several steps (as initially proposed by Wintersberger and Wintersberger, 1987) and several times independently in different lineages of DNA viruses (Forterre et al., 2004). In a first step, the simplest scenario assumes that a single-stranded RNA virus with a double-stranded RNA replication intermediate became a single-stranded DNA virus that replicated its genome via a DNA/RNA intermediate. At this stage, a single enzyme could have both transcribed DNA into RNA and retro-transcribed RNA into DNA. Later on, the diversification of the initial DNA viral lineage and the invention of DNA-dependent DNA polymerases would have given birth to DNA viruses replicating their DNA without any RNA intermediate, with production of double-stranded DNA. The first double-stranded DNA genomes were probably replicated one strand after the other, much alike the genomes of modern double-stranded RNA viruses and some DNA viruses. This mechanism only requires a DNA polymerase with strand-displacement activity. It might have been later on improved by the introduction of helicases and processivity factors to help the polymerase. The next logical step is to initiate replication of the second strand before completion of the first one. Such partial coupling requires the invention of a primase to initiate the replication of the second strand (becoming the lagging strand). At the end of this evolutionary process, the rapid replication of very large genomes became possible by coupling the replication of the lagging and leading strands at the replication fork, as in all modern cells and in many DNA viruses. Such coupling requires the tight coordi-

nation of the helicase, primase and DNA polymerase activities in a single "replication factory".

In this sequential model for the evolution of DNA replication mechanisms, new proteins were probably recruited at each step by modifying the specificity of proteins already present in the RNA world (DNA polymerases and primases from RNA polymerases, DNA helicase from RNA helicase, DNA binding proteins from RNA binding protein and so on) and/or by the different associations of various protein modules pre-existing in the RNA world. This idea is supported by the mechanistic and structural homologies between some of these enzymes (already discussed) and by the presence of RNA binding modules in many DNA informational proteins. For instance, several of these proteins contain various derived versions of the RNA Recognition Motif, the RRM fold (Iyer et al., 2005). Since the sequential evolution of DNA replication mechanisms by recruitment and patchwork construction probably occurred independently in several viral lineages, this model readily explains the existence of many DNA informational proteins that are not homologous but nevertheless perform the same function (Forterre et al., 2004).

If DNA and DNA replication mechanisms indeed originated in the ancient virosphere, one has to explain when and how they were later on transferred into the cellular world (Fig. 3). There are several possibilities; one can imagine that this transfer occurred at an early step in the evolution of the DNA replication mechanism (for instance at the stage of DNA/RNA hybrid genomes) and that later stages of this evolution occurred in parallel in both viruses and cells. Alternatively, the evolution of DNA replication mechanisms could have occurred entirely in the viral world, and transfer of DNA from viruses to modern cells only occurred at the end of the process, when the fully symmetric mechanism for DNA replication present in modern cells became available (Forterre et al., 2004).

Different mechanisms can be also imagined for the transfer itself. Cells could have learnt from viruses how to make DNA by capturing selectively viral genes encoding ribonucleotide reductases, thymidylate synthases and replication proteins (much like bacterial cells later on probably recruited viral restriction-modification systems to fight viral infections). Another possibility is that some DNA viruses got the ability to take complete control of the RNA cell that they originally infected, transforming it from within into a DNA cell (Forterre, 2005). Indeed, by analogy with the existence of modern DNA virus living in a carrier state inside DNA cells, many ancient RNA cells probably contained resident viral DNA genomes beside their own cellular RNA genomes. Some DNA viruses living in a carrier state inside a RNA cell could have evolved into "plasmid" after loosing the genes encoding their capsid proteins and/or proteins involved in the infectious process. If the virus or the cell also encoded a reverse transcriptase, the DNA plasmid (of viral origin) could have progressively integrated by retro-transcription more and more cellular genes since they could be replicated more efficiently and faithfully than the RNA chromosome. The end of this process would have been the complete elimination of the ancient cellular RNA genome and the formation of a modern cellular DNA chromosome.

7. Viruses and the origin of the three modern cellular domains

As mentioned previously, the hypothesis that DNA was transferred from viruses to cells, together with a complete machinery for its replication, was originally proposed to explain the existence of two sets of non-homologous DNA replication proteins, one in Bacteria, the other in Archaea and Eukarya. It was thus necessary to imagine at least two independent transfers to take into account this observation (Fig. 4A, B). Recently, I have suggested that three independent transfers of DNA from viruses to RNA cells actually occurred, leading to the formation of the three domains, Archaea, Bacteria and Eukarya (Fig. 4C) (Forterre, in press). This "*three viruses three domains*" hypothesis can explain why eukaryal and archaeal DNA replication machineries, although similar, exhibit critical differences. Hence, major eukaryal DNA topoisomerases (Topo IB and Topo IIA) are phylogenetically unrelated to archaeal ones (Topo IA and Topo IIB) (Gadelle et al., 2003; Krogh and Shuman, 2002). In addition, Archaea contain a DNA polymerase of the D family that has no eukaryotic homologues (Cann et al., 1998). Furthermore, archaeal and eukaryal DNA polymerases of the B family, although homologous, are not specifically related, but both cluster with different viral groups in a DNA polymerases B phylogeny (Filée et al., 2002). There are many other differences between DNA recombination, repair and recombination mechanisms between Archaea and Eucaryotes that are better explained by a partially independent origin of these two domains than by the transformation of Archaea into Eukarya. Finally, the existence of three viruses at the origin of the three domains can explain that, despite a LUCA with an RNA genome, some proteins involved in DNA metabolism are homologous in all cellular organisms (such as RecA-like recombinases or the DNA polymerase processivity factors PCNA) (Leipe et al., 1999). One has simply to postulate that the three DNA viruses at the origin of all modern cells shared this small set of homologous DNA informational proteins. In agreement with this hypothesis, many viruses (e.g. bacteriophage T4) encoded virus-specific versions of RecA or PCNA.

Interestingly, the three viruses three domains hypothesis can take into account essential features of the universal tree of life that are not readily explained in classical models of early cellular evolution (Forterre, in press). For instance, it explains why there are three discrete lineages of modern cells and not an evolutionary continuum of cellular organisms, since it involves (and specify) three different "major qualitative evolutionary changes" or "dramatic evolutionary events" at the origin of each domain (Forterre and Philippe, 1999; Woese, 2000), i.e. the fusion of RNA cells with DNA viruses. Since RNA genomes were a priori replicated and repaired less faithfully than DNA genomes, this new hypothesis also specifies the difference in evolutionary mode responsible for the drastic reduction in the rate of protein and rRNA evolution that should have occurred after the formation of the three domains (Woese and Fox, 1977). This reduction in evolutionary rate is required to explain why ribosomal proteins and rRNA diverged more profoundly in the "short" time period that took place between LUCA and the three last com-

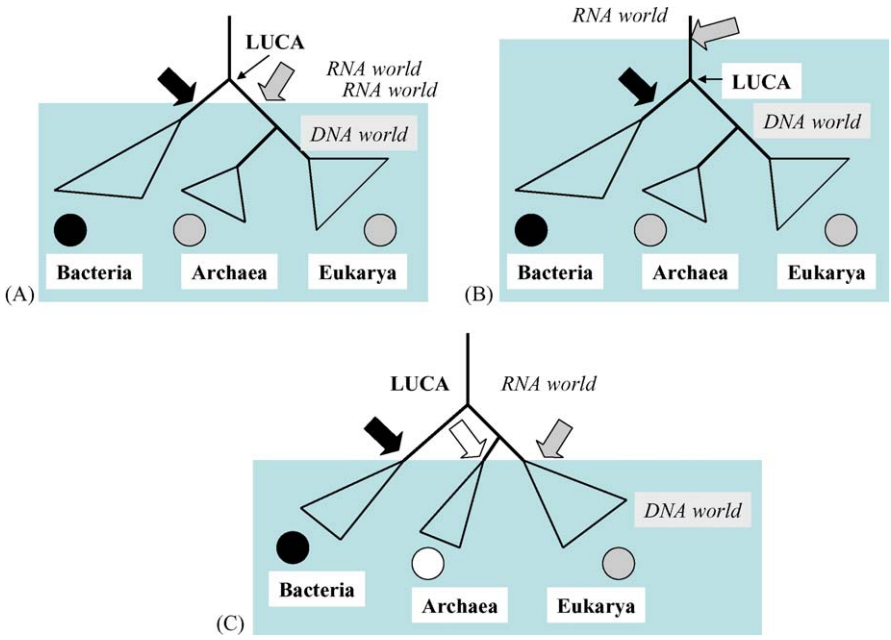

Fig. 4. Several models for the possible transfer of DNA from viruses to cells, black circle symbolize the bacterial DNA replication apparatus, whereas grey circles symbolize the archaeal/eukaryal DNA replication apparatus. (A) Two independent transfers after a LUCA with an RNA genome; (B) one transfer before LUCA followed by non-orthologous displacement by another DNA replication system in the bacterial branch. In that scenario, LUCA has a DNA genome. (C) Three independent transfers after LUCA, leading to the formation of the three cellular domains (differences between the archaeal and eukaryal DNA replication machineries are illustrated by different color, white and grey circles, respectively).

mon ancestors of each domain, than in the time period (much longer) that has taken place between the formation of the three domains and the present time (Woese and Fox, 1977; Woese, 2000). Finally, the three viruses three domains hypothesis can also explain why members of different domains are infected by specific groups of viruses. If ancestors of all present viruses were already thriving at the time of LUCA, member of a given viral families (for instance *Fuselloviridae*) should be found to infect cells from the three cellular domains. This is apparently not the case, since the presence of fuselloviruses appears to be restricted to Archaea. However, in the three viruses three domains scenario, following an initial diversification of viral lineages along with divergence of RNA cell lineages, only viruses that were able to infect RNA cells at the origin of each domain could have survived the massive elimination of RNA cells (and RNA/DNA viruses) that occurred after the emergence of the three lineages of DNA cells. This would have selected from each domain a subpopulation of different viral families that were pre-adapted to infect the newly formed DNA cells.

8. Viruses and the origin of the eukaryotic nucleus

Recent theories suggesting a viral origin for the eukaryotic nucleus are another example of the comeback made by viruses in the mind of evolutionists during the last five years. The origin of the eukaryotic nucleus is one of the great mysteries of early life evolution (Pennisi, 2004). It has been advocated by several authors that the nucleus originated from an ancient archaeon that was engulfed by a bacterial-like cell. However, this hypothesis does not explain the existence of many eukaryotic specific proteins that have no homologous counterparts in Archaea or Bacteria (Hartman and Fedorov, 2002), as well as nuclear features that are unknown in Archaea (spliceosome, nuclear pores, mRNA capping and so on). Furthermore, endosymbiosis of an archaeon into a bacterium would have produced a nucleus with a double-membrane, one of bacterial and one of archaeal origin. In contrast, the nuclear membrane is a single membrane folded onto itself that originates by recruitment of membrane vesicles from the endoplasmic reticulum. Finally, a major question mark in all theories for the origin of the nucleus (autogeneous or endosymbiotic) is the origin of nuclear pores. It is difficult to imagine why nuclear pores could have appeared in the absence of a nuclear envelope but it's not possible to imagine the appearance of a nuclear envelope without pre-existing nuclear pore.

In 2001, two authors independently proposed a completely new hypothesis to explain the origin of the nucleus, i.e. the viral eukaryogenesis hypothesis (Bell, 2001; Takemura, 2001). They suggested that the nucleus originated from a large double-stranded DNA viruses that infected a prokaryotic-like organism. This hypothesis has the great advantage to solve the nuclear pore problem, since viruses have evolved a number of complex molecular devices to translocate RNA or DNA through various types of membranes (including cellular membranes). Both Bell and Takemura suggested that the virus at the origin of the nucleus was related to poxviruses. Several features of the poxviruses cell cycle are indeed reminiscent of the eukaryotic nucleus biology and some poxviruses enzymes (DNA polymerase, mRNA capping enzymes) are evolutionary related to their eukaryotic counterparts, however forming viral specific clusters in phylogenetic trees. Interestingly, early transcription of poxviruses occurs inside a viral core and the RNA should be transported to the cytoplasm through some kind of pores (Broyles, 2003). These pores

have been recently visualized in vaccinia virus by cryo-electron tomography (Cyrklaff et al., 2005). After its release from the viral core (which is dependent of products from early transcription), the viral DNA is replicated in «mininuclei» that are formed by the recruitment of the endoplasmic reticulum, a process reminiscent of the formation of the nuclear membrane (Tolonen et al., 2001). As in the case of the nucleus, DNA replication only occurs once these mininuclei are fully assembled, and the mininuclei are disassembled only after completion of DNA replication (both processes being regulated by protein phosphorylation).

The classical view is that poxviruses and other NCLDV viruses have adapted their molecular biology and cell cycle to those of their eukaryotic hosts. However, as noticed by Villarreal (2005), many other viruses successfully infect eukaryotic cells by using completely different strategies. One can therefore argue that all characteristic features of poxviruses were of viral origin, and were later on transferred to the host. The number of genes encoded by poxviruses could appear to be relatively low as starting material for eukaryotic nucleus. This criticism was eliminated after the discovery of the mimivirus whose genome is only twice smaller than the complete genome of the Eukarya *Encephalitozoon cruzi*. A giant NCLDV virus, such as the mimivirus could be a good starting point for the eukaryogenesis hypothesis (Raoult et al., 2004).

Both Bell and Takemura suggested that the host of the ancestral virus at the origin of the nucleus was an archaeon, in order to explain the similarity between the translation apparatus of Archaea and Eukarya. However, this creates a problem: why was the archaeal membrane replaced later on by the "bacterial-like" membrane of modern eucaryotes? If the viral eukaryogenesis is correct, the host of the virus at the origin of the nucleus was more likely an ancestral proto-eukaryotic cell with a bacterial-like membranes but an archaeal-like translation apparatus. Such ancient cell has been named the urkaryote (Woese, 1987) or the chronocyte (Hartman and Fedorov, 2002). It is also possible to couple the viral eukaryogenesis hypothesis with the three viruses three domains hypothesis, by supposing that the nucleus originated at the same time as the eukaryotic domain, i.e. from one (or several) large DNA virus that infected an urkaryote (Forterre, in press). In any case, one should be aware that some bacteria also contain a nucleus with fascinating structural similarities (not yet prove to be homologies) with the eukaryotic one (Fuerst, 2005). This raises difficult questions (with presently no satisfactory answer) for all present theories on the origin of the eukaryotic nucleus, including the viral eukaryogenesis hypothesis.

9. The role of viruses in the evolution of mitochondria and chloroplasts

It is obvious now for everybody that viruses have played an important role in recent cellular evolution by transferring some of their genes to cellular organisms. However, it is usually considered that modifications introduced in cellular lineages by viruses are restricted to "superficial" traits which have a direct selective advantage for the cell in a specific environment (e.g. pathogenecity islands) but do not drastically change its genomic make-up. It is thus difficult for many biologists to take seriously new hypotheses that postulate a major role for viruses in the origin and evolution of fundamental cellular features, such as the nature and the organisation of their genome. However, the great impact that viruses can have on the genetic systems is well illustrated by the evolution of mitochondria. Comparative genomics and molecular phylogenetic data have shown that mitochondria originated from a free-living α-proteobacterium (Gray and Lang, 1998). However, to only one exception (*Reclinomonas elongata*), the RNA polymerase that transcribed the mitochondrial genome is not homologous to bacterial RNA polymerases (as would have been expected) but to monomeric RNA polymerases encoded by bacterioviruses T3/T7 and relatives. Phylogenetic analyses have recently revealed that the helicase and DNA polymerases involved in mitochondrial DNA replication are also specifically related to the same group of viruses (Filée et al., 2002, 2003). Interestingly, homologues of these proteins are encoded by cryptic proviruses in the genomes of several proteobacteria (Filée and Forterre, 2005). This suggests that the α-proteobacterium at the origin of mitochondria contained a cryptic provirus related to T3/T7, and that the RNA polymerase, DNA helicase and DNA polymerase encoded by this provirus replaced the original DNA transcription and replication proteins of the ancestral α-proteobacterium during its transformation into mitochondria.

Amazingly, the ancestral bacterial type RNA polymerase (of cyanobacterial origin) and a T3/T7-like type RNA polymerase are both used in the transcription of chloroplast genomes (Gray and Lang, 1998). The viral-like enzyme of the chloroplast originated from a duplication of the gene encoding the mitochondrial RNA polymerase. Plants thus contain two viral-like RNA polymerases, one targeted to the mitochondrion, the other to the chloroplast. This indicates the existence of a strong selection pressure that has pushed for the replacement of cellular enzymes by viral ones in mitochondria and chloroplasts. In both organelles, this replacement has been associated with profound modifications in the mechanism of DNA replication and chromosome structure. This demonstrates that viruses can be the source of new "cellular" DNA informational proteins and can definitely play a critical role in shaping the cellular genome.

10. Conclusion

For a long time, viruses were not included in the universal tree of life, since they had no ribosomal RNA. As seen in this review, if the viruses are brought back into the tree (in fact if the tree and its root are immersed in a viral ocean, as suggested by Bamford (2003)), one can propose radically new hypotheses to solve problems encountered in deciphering ancient relationships between the three domains and the history of DNA replication mechanisms.

A reappraisal of the role of viruses in early cellular evolution is therefore under way. The realisation that viruses are probably very ancient allows to better understand their extraordinary diversity, explaining why most viral proteins inferred from genome sequencing have no cellular homologues. The continuous rain of viral genes into cellular genomes could thus be partly responsible for the existence of an irreducible fraction of

orphan genes in cellular genomes (Daubin and Ochman, 2004). There seems to exist in the biosphere an illimited reservoir of viral proteins that has provided opportunity at different steps of the evolutionary process to introduce new functions into cellular organisms.

The idea that virus origin predated the emergence of LUCA also suggests that modern viruses have inherited from ancient RNA and DNA cells molecular mechanisms that have disappeared from modern DNA cells (an illuminating example could be the mechanism for protein-primed DNA replication). This would explain why the molecular biology of the viral world for transcription and replication is more diverse than those of the cellular world. Many yet unknown molecular mechanisms (and their associated proteins) thus probably remain to be discovered in the virosphere. If these considerations are correct, this means that the exploration of the viral diversity will be for sure one of the major challenges of biology in this new century.

Acknowledgment

I would like to thank Eugene Koonin for inviting me to write this review.

References

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

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

Benson, S.D., Bamford, J.K., Bamford, D.H., Burnett, R.M., 2004. Does common architecture reveal a viral lineage spanning all three domains of life? Mol. Cell 16, 673–685.

Bandea, C.I., 1983. A new theory on the origin and the nature of viruses. J. Theor. Biol. 105, 591–602.

Broyles, S.S., 2003. Vaccinia virus transcription. J. Gen. Virol. 84, 2293–2303.

Cann, I.K., Komori, K., Toh, H., Kanai, S., Ishino, Y., 1998. A heterodimeric DNA polymerase: evidence that members of Euryarchaeota possess a distinct DNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95, 14250–14255.

Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Gray, M.W., Cedergren, R., 1997. On the evolution of the single-subunit RNA polymerases. J. Mol. Evol. 45, 671–681.

Cyrklaff, M., Risco, C., Fernandez, J.J., Jimenez, M.V., Esteban, M., Baumeister, W., Carrascosa, J.L., 2005. Cryo-electron tomography of vaccinia virus. Proc. Natl. Acad. Sci. U.S.A. 102, 2772–2777.

Daubin, V., Ochman, H., 2004. Start-up entities in the origin of new genes. Curr. Opin. Genet. Dev. 14, 616–619.

Dracheva, S., Koonin, E.V., Crute, J.J., 1995. Identification of the primase active site of the herpes simplex virus type 1 helicase-primase. J. Biol. Chem. 270, 14148–14153.

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

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 (4), 237–243.

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.

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

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.

Forterre, P., 2002. The origin of DNA genomes and DNA replication proteins. Curr. Opin. Microbiol. 5, 525–532.

Forterre, P., 2003. The great virus comeback-from an evolutionary perspective. Res. Microbiol. 154, 223–225.

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.

Forterre, P., in press. 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.

Forterre, P., Philippe, H., 1999. Where is the root of the universal tree of life? Bioessays 21, 871–879.

Forterre, P., Filée, J., Myllykallio, H., 2004. Origin and evolution of DNA and DNA replication machineries. In: Ribas de Pouplana, L. (Ed.), The Genetic Code and the Origin of Life. Landes Bioscience, pp. 145–168.

Fuerst, J.A., 2005. Intracellular compartimentation in Planctomycetales. Annu. Rev. Microbiol. 59, 299–328.

Gadelle, D., Filée, J., Buhler, C., Forterre, P., 2003. Phylogenomics of type II DNA topoisomerases. Bioessays 25, 232–242.

Gilbert, W., 1986. The RNA world. Nature 319, 618.

Gorbalenya, A.E., Koonin, E.V., Wolf, Y.I., 1990. A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett. 262, 145–148.

Gray, M.W., Lang, B.F., 1998. Transcription in chloroplasts and mitochondria: a tale of two polymerases. Trends Microbiol. 6, 1–3.

Hansen, J.L., Long, A.M., Schultz, S.C., 1997. Structure of the RNA-dependent RNA polymerase of poliovirus. Structure 5, 1109–1122.

Hartman, H., Fedorov, A., 2002. The origin of the eukaryotic cell: a genomic investigation. Proc. Natl. Acad. Sci. U.S.A. 99, 1420–1425.

Hendrix, R.W., Lawrence, J.G., Hatfull, G.F., Casjens, S., 2000. The origins and ongoing evolution of viruses. Trends Microbiol. 8, 504–508.

Ilyina, T.V., Koonin, E.V., 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. 20, 3279–3285.

Iyer, L.M., Koonin, E.V., Leipe, D.D., Aravind, L., 2005. Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. 33, 3875–3896.

Kandler, O., 1998. The early diversification of life and the origin of the three domains, a proposal. In: Wiegel, J., Adams, M.W.W. (Eds.), Thermophiles: the Keys to Molecular Evolution and the Origin of Life. Taylor and Francis, pp. 19–28.

Koga, Y., Kyuragi, T., Nishihara, M., Sone, N., 1998. Archaeal and bacterial cells arise independently from noncellular precursors? A hypothesis stating that the advent of membrane phospholipid with enantiomericglycerophosphate backbones caused the separation of the two lines of descent. J. Mol. Evol. 46, 54–63.

Koonin, E.V., 2005. Virology: Gulliver among the Lilliputians. Curr. Biol. 15, R167–R169.

Krogh, B.O., Shuman, S., 2002. A poxvirus-like type IB topoisomerase family in bacteria. Proc. Natl. Acad. Sci. U.S.A. 99, 1853–1858.

Lazcano, A., Guerrero, R., Margulis, L., Oro, J., 1988. The evolutionary transition from RNA to DNA in early cells. J. Mol. Evol., 283–290.

La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.M., Raoult, D., 2003. A giant virus in amoebae. Science 299, 2033.

Leipe, D.D., Aravind, L., Koonin, E.V., 1999. Did DNA replication evolve twice independently? Nucleic Acids Res. 27, 3389–3401.

Lipps, G., Rother, S., Hart, C., Krauss, G., 2003. A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. EMBO J. 22, 2516–2525.

Luria, S., Darnell, 1967. General Virology. John Wiley and Sons, New York.

McGeoch, A.T., Bell, S.D., 2005. Eukaryotic/Archaeal primase and MCM proteins encoded in a bacteriophage genome. Cell 120, 167–168.

Makeyev, E.V., Grimes, J.M., 2004. NA-dependent RNA polymerases of dsRNA bacteriophages. Virus Res. 101, 45–55.

Martin, W., Russell, M.J., 2003. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. Philos. Trans. R. Soc. Lond. Biol. Sci. 358, 59–83.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 86–156.

Miller, R.H., Robinson, W.S., 1986. Common evolutionary origin of hepatitis B virus and retroviruses. Proc. Natl. Acad. Sci. U.S.A. 83, 2531–2535.

Moreira, D., 2000. Multiple independent horizontal transfers of informational genes from bacteria to plasmids and phages: implications for the origin of bacterial replication machinery. Mol. Microbiol. 35, 1–5.

Moreira, D., Lopez-Garcia, P., 2005. Comment on "The 1.2-megabase genome sequence of Mimivirus". Science 308, 1114 (author reply 1114).

Mushegian, A.R., Koonin, E.V., 1996. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc. Natl. Acad. Sci. U.S.A. 93, 10268–10273.

Myllykallio, H., Lipowski, G., Leduc, D., Filée, J., Forterre, P., Liebl, U., 2002. An alternative Flavin-dependent mechanism for Thymidylate synthesis. Science 297, 105–107.

Ogata, H., Abergel, C., Raoult, D., Claverie, J.M., 2005. Response to comment on "The 1.2-megabase genome sequence of Mimivirus". Science 308, 1114.

Pennisi, E., 2004. The birth of the nucleus. Science 305, 766–768.

Pereto, J., Lopez-Garcia, P., Moreira, D., 2004. Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem. Sci. 29, 469–477.

Poole, A.M., Jeffares, D.C., Penny, D., 1998. The path from the RNA world. J. Mol. Evol. 46, 1–17.

Prangishvili, D., Garrett, R.A., 2004. Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. Biochem. Soc. Trans. 32, 204–208.

Prangishvili, D., Stedman, K., Zillig, W., 2001. Viruses of the extremely thermophilic archaeon Sulfolobus. Trends Microbiol. 9, 39–43.

Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.M., 2004. The 1.2-megabase genome sequence of Mimivirus. Science 306, 1344–1350.

Rice, G., Tang, L., Stedman, K., Roberto, F., Spuhler, J., Gillitzer, E., Johnson, J.E., Douglas, T., Young, M., 2004. The structure of a thermophilic archaeal virus shows a double-stranded DNA viral capsid type that spans all domains of life. Proc. Natl. Acad. Sci. U.S.A. 101, 7716–7720.

Siegel, R.W., Bellon, L., Beigelman, L., Kao, C.C., 1999. Use of DNA, RNA, and chimeric templates by a viral RNA-dependent RNA polymerase: evolutionary implications for the transition from the RNA to the DNA world. J. Virol. 73, 6424–6429.

Takahashi, I., Marmur, J., 1963. Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for *Bacillus subtilis*. Nature 197, 794–795.

Takemura, M., 2001. Poxviruses and the origin of the eukaryotic nucleus. J. Mol. Evol. 52, 419–425.

Tolonen, N., Doglio, L., Schleich, S., Krijnse Locker, J., 2001. Vaccinia virus DNA replication occurs in endoplasmic reticulum-enclosed cytoplasmic mini-nuclei. Mol. Biol. Cell 12, 2031–2046.

Trus, B.L., Cheng, N., Newcomb, W.W., Homa, F.L., Brown, J.C., Steven, A.C., 2004. Structure and polymorphism of the UL6 portal protein of herpes simplex virus type 1. J. Virol. 78, 12668–12671.

Villarreal, L.P., De Filippis, A., 2000. Hypothesis for DNA viruses as the origin of eukaryotic replication proteins. J. Virol. 74, 7079–7084.

Villarreal, L.P., 2005. Viruses and the Evolution of Life. ASM Press, Washington.

Warren, R.A.J., 1980. Modified bases in bacteriophage DNAs. Ann. Rev. Microbiol. 34, 137–158.

Wintersberger, U., Wintersberger, E., 1987. RNA makes DNA: a speculative view of the evolution of the DNA replication mechanism. Trends Genet. 3, 198–202.

Weiner, A.M., Maizels, N., 1993. The genomic taghypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication. In: Gesteland, R.F., Atkins, J.F. (Eds.), The RNA World, second ed. Cold Spring Harbor Press, NY, pp. 577–602.

Weiner, A.M., Maizels, N., 1994. Molecular evolution: unlocking the secrets of retroviral evolution. Curr. Biol. 4, 560–563.

Woese, C.R., 1987. Bacterial evolution. Ann. Rev. Microbiol. 270, 221–271.

Woese, C.R., 2000. Interpreting the universal phylogenetic tree. Proc. Natl. Acad. Sci. U.S.A. 97, 8392–8396.

Woese, C.R., 2002. On the evolution of cells. Proc. Natl. Acad. Sci. U.S.A. 99, 8742–8747.

Woese, C.R., Fox, G.E., 1977. The concept of cellular evolution. J. Mol. Evol. 10, 1–6.