Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
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-